Truth and Paradox

Leon Horsten
University of Bristol

Summer School on
Set Theory and Higher-Order Logic
London, 1–6 August 2011

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

Structure of the Tutorial

1. Introduction: Truth and the Liar
2. Definition, Models, Axioms
3. Typed Disquotational Theories
4. Typed Compositional Theories
5. Type-free Disquotational Theories
6. Type-free Compositional Theories
7. The Revision Theory of Truth
8. Selected References

Tarski-biconditionals

The *disquotational intuition*: supposing ϕ is tantamount to supposing the truth of ϕ, and conversely.

A plausible axiom scheme for truth:

$$\text{‘}\phi\text{’ is true if and only if } \phi.$$

- This scheme has infinitely many instances
- Instances of this scheme are called *Tarski-biconditionals*.

(after the logician Alfred Tarski)

Paradoxical sentences

The *liar sentence* L:

This sentence is not true.

The *truth teller*:

This sentence is true.

- L will be our paradigmatic test case
- but we should be aware that it is not the only paradoxical sentence

Semantic paradox

The argument of the *liar paradox*:

Proof.
It is an instance of the Tarski biconditional scheme that L is true if and only if L. But L if and only if L is not true —for this is what L says of itself. So L is true if and only if L is not true: a short truth table calculation convinces us that we have lapsed into inconsistency. □

How credible is this argument?

The framework

For simplicity, we want a toy model for English.

`Language:` The language of truth ($\mathcal{L}_T$): the language of arithmetic plus a truth predicate T.

`Background theory`: First-order Peano arithmetic (PA^T). [The truth predicate T is allowed in the induction scheme.]

The arithmetical background

Why arithmetic?

Coding

- *PA* can serve as a theory of syntax (via coding);
- *PA* allows the formation of self-referential sentences (via coding).

`Note:` We will be sloppy about notation.

Defining truth

Tarski showed us how to give a *definition* of truth for a formal language in purely logical and mathematical terms.

`material adequacy condition:` a definition of truth for a language $\mathcal{L}$ should imply all the Tarski-biconditionals for sentences of $\mathcal{L}$.

$\Rightarrow$ *Metawissenschaft*

Defining first-order arithmetical truth in second-order arithmetic

There is a first-order formula val^+ that defines the atomic arithmetical truths.
Consider the condition $\Psi(X, \phi)$:

$$\phi \text{ atomic closed} \rightarrow [X(\phi) \leftrightarrow val^+(\phi)] \wedge$$
$$\exists \psi, \lambda : \phi = \psi \wedge \lambda \rightarrow [X(\phi) \leftrightarrow X(\psi) \wedge X(\lambda)] \wedge \ldots$$

Definition

$true(\phi) \equiv \forall X : \forall \lambda \Psi(X, \lambda) \rightarrow X(\phi)$.

The second-order arithmetical predicate $true(x)$ satisfies Tarski's material adequacy condition.

Models for truth

A theory of truth for a language $\mathcal{L}$ should describe a class of intended *models* for $\mathcal{L}$.

[This thesis was popular in the 1970s and 1980s.]

Universality

We want a theory of truth for *our* language:

We want a truth theory for English in English.

$$\text{Metalanguage} = \text{Objectlanguage}$$

Definitions and universality

`Problem for the definitional approach:` A truth definition for $\mathcal{L}$ (*objectlanguage*) can only be given in a language (*metalanguage*) $\mathcal{L}'$ that is essentially richer than $\mathcal{L}$:

Theorem (Undefinability theorem I)

There is no first-order arithmetical formula that defines first-order arithmetical truth.

But perhaps a theory of truth for English as it now is can only be given in future English ...

Models and universality

A class of models for a language $\mathcal{L}$ is described in an essentially richer metalanguage (*incompleteness theorem*).

`Additional problem for the model-theoretic approach:` the domain of a model is a set

The domain of discourse of English does not form a set.

But perhaps the techniques can be adapted to proper class domains . . .

Axioms for Truth

Thesis

A theory of truth for a language $\mathcal{L}$ should posit rules of inference and / or axioms for $\mathcal{L}$.

The problems of the definitional and the model-theoretic approach are not applicable.

$\Rightarrow$ there is no immediately apparent obstacle to our universal ambitions

Definition

Let $\mathcal{L}_T$ be $\mathcal{L}_{PA}$ plus a primitive truth predicate T.

`Aim:` Formulate a truth theory for $\mathcal{L}_T$ in $\mathcal{L}_T$.

The role of models and definitions

Models can have a strong heuristic force: they give pictures.

We can give truth definitions for fragments of our language.

Soundness and completeness

Desiderata for axiomatic truth theories:

- soundness
- truth theoretic completeness

More specific desiderata will be discussed at the end of the lectures.

The naive theory of truth

Recall the disquotational intuition.

$\Rightarrow$ Take all the Tarski-biconditionals as your theory of truth.

The formal theory NT:

1. PA^T
2. $T(\phi) \leftrightarrow \phi$ for all $\phi \in \mathcal{L}_T$

Gödel's diagonal lemma

Theorem (Gödel)

For each formula $\phi(x) \in \mathcal{L}_T$, there is a sentence $\lambda \in \mathcal{L}_T$ such that PA^T proves

$$\lambda \leftrightarrow \phi(\lambda).$$

This sentence λ is self-referential.

Tarski's undefinability theorem

Theorem (Undefinability theorem 2)

No consistent extension S of PA^T proves $T(\phi) \leftrightarrow \phi$ for all $\phi \in \mathcal{L}_T$.

Proof.

Use the diagonal lemma to produce a (liar) sentence λ:

$$PA^T \vdash \lambda \leftrightarrow \neg T(\lambda).$$

If the theory S in question indeed proves $T(\phi) \leftrightarrow \phi$ for all $\phi \in \mathcal{L}_T$, then in particular S proves $T(\lambda) \leftrightarrow \lambda$. Putting these equivalences together, we obtain a contradiction in S:

$$S \vdash T(\lambda) \leftrightarrow \neg T(\lambda).$$

□

`Consequence:` NT is inconsistent.

The typed disquotational theory

`Tarski's diagnosis:` the root of the disease lies in allowing the Tarski-biconditionals to regulate the truth-conditions of sentences that themselves contain the truth predicate (such as L).

$\Rightarrow$ **typed truth theories**

The disquotational theory DT:

DT1 PA^T;

DT2 $T(\phi) \leftrightarrow \phi$ for all $\phi \in \mathcal{L}_{PA}$.

Tarski's strictures are respected.

Soundness

Proposition

DT has a nice model.

Proof.

Consider the model

$$\mathfrak{M} =: \langle \mathbb{N}, \{\phi \mid \phi \in \mathcal{L}_{PA} \wedge \mathbb{N} \models \phi\} \rangle,$$

i.e., the model in which as the extension of the truth predicate we take all arithmetical truths. An induction on the length of proofs in DT verifies that $\mathfrak{M} \models DT$. □

`Consequence:` *DT* is arithmetically sound.

The Tarskian hierarchy 1

Conversation:

A: It is true that $0 = 0$.

B: What you have just said is true.

Proposition

$DT \nvdash T(T(0 = 0))$

The Tarskian hierarchy 2

The theory DT_1:

1. PA^{T,T_1};
2. $T(\phi) \leftrightarrow \phi$ for all $\phi \in \mathcal{L}_{PA}$;
3. $T_1(\phi) \leftrightarrow \phi$ for all $\phi \in \mathcal{L}_T$.

Theorem

DT_1 has a nice model.

Proposition

$DT_1 \vdash T_1(T(0 = 0))$

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

The Tarskian hierarchy 3

Extended conversation:

A: It is true that $0 = 0$.
B: What you have just said is true.
C: Yes, B, that is very true.

$\Rightarrow$ the Tarskian hierarchy DT, DT_1, DT_2,...

$\Rightarrow$ the notion of truth is irrevocably fragmented

Contextualist theories

(Burge, Gaifman, Barwise & Perry, Glanzberg, . . .)

Thesis

Truth is a uniform but indexical concept.

> *Sentence L is not* $true_0$.
> *Sentence L is* $true_1$.

Because of the indexical shift in extension of the truth predicate between context 0 and context 1, this is not a contradiction.

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

The strengthened liar paradox

(S) *Sentence S is not true in any context.*

- this is called a *strengthened liar sentence*
- it is not hard to figure out that there is no context in which S can be coherently evaluated.

The weakness of disquotationalism

The *compositional intuition*: truth distributes over the logical connectives.

Proposition

For all $\phi \in \mathcal{L}_{PA}$*:* $DT \vdash T(\phi) \vee T(\neg\phi)$

Proof.

Already propositional logic alone proves $\phi \vee \neg\phi$*. Two restricted Tarski-biconditionals are* $T(\phi) \leftrightarrow \phi$ *and* $T(\neg\phi) \leftrightarrow \neg\phi$*. Combining these facts yields the result.* $\square$

Proposition (Tarski)

$DT \nvdash \forall\phi \in \mathcal{L}_{PA} : T(\phi) \vee T(\neg\phi)$

The typed compositional theory

(Davidson...)

The compositional theory TC:

TC1 PA^T;

TC2 $\forall$ atomic $\phi \in \mathcal{L}_{PA} : T(\phi) \leftrightarrow val^+(\phi)$;

TC3 $\forall \phi \in \mathcal{L}_{PA} : T(\neg\phi) \leftrightarrow \neg T(\phi)$;

TC4 $\forall \phi, \psi \in \mathcal{L}_{PA} : T(\phi \wedge \psi) \leftrightarrow (T(\phi) \wedge T(\psi))$;

TC5 $\forall \phi(x) \in \mathcal{L}_{PA} : T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$.

Theorem

TC has nice models, and $DT \subseteq TC$.

Substitution

For a moment we have to be sticklers for notation...

A sentence of the form $\exists x T(\phi(x))$ is really expressed along the following lines:

> *"There is a number x such that when the standard numeral for x is substituted for the variable x in $\phi(x)$, a true sentence results."*

Thus a *substitution function* (expressible in the language of arithmetic) appears in formulae such as TC5, but also in all the other compositional axioms.

True and true of

What if we do not have standard names for all elements of the language of discourse ($\mathbb{R}$,. . .)?

$\Rightarrow$ we work with a satisfaction relation ('true of') and define the truth predicate in terms of it.

The axiom TC5 then becomes (roughly):

$$\forall x, \forall \phi(y) : Sat(x, \forall y \phi(y)) \leftrightarrow \forall z Sat(z, \phi(y)))$$

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

Deflationism and conservativeness

Deflationism: The concept of truth does not play a substantial role in philosophical, mathematical, scientific debates.

Definition

A theory of truth S is *arithmetically conservative* over PA if for every sentence $\phi \in \mathcal{L}_{PA}$, if $S \vdash \phi$, then already $PA \vdash \phi$.

Proposition

DT is arithmetically conservative over PA.

The non-conservativeness of truth 1

Theorem

$TC \vdash \forall \phi \in \mathcal{L}_{PA} : Bew_{PA}(\phi) \rightarrow T(\phi)$, where $Bew_{PA}(...)$ is an arithmetical predicate that expresses provability in Peano arithmetic in a natural way.

Proof.

This is proved by an induction, inside TC, on the length of proofs. [Here we need that T is allowed in the induction scheme.]

For the basis case, we have to prove that all the axioms of true. [For the subcase of mathematical induction we again need that T is allowed in the induction scheme.]

$\square$

The non-conservativeness of truth 2

Corollary

$TC \vdash \neg Bew_{PA}(0 = 1)$.

Proof.

This follows from the previous theorem and the fact that $DT \subseteq TC$ by instantiating $0 = 1$ for ϕ. □

So by Gödel's second incompleteness theorem, TC is not conservative over PA.

The non-conservativeness of truth 3

More precise information:

Definition

The second-order system *ACA* contains full second-order induction but only those instances of Comprehension

$$\exists X \forall y : X(y) \leftrightarrow \phi(y)$$

where $\phi(y)$ contains no bound second-order quantifiers (and does not contain X free).

Theorem

The first-order arithmetical strength of TC is exactly that of the second-order system ACA.

Back to deflationism

Thesis (deflationism)

The concept of truth is not substantial but provides extra conceptual power.

$\Rightarrow$ An axiomatic theory of truth must be conservative but noninterpretable.

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

A deflationist truth theory 1

Fischer proposed a minimally adequate truth theory PT^-

Definition

$tot(\phi(x)) \equiv \forall y : T(\phi(y) \vee T(\neg\phi(y))$

A deflationist truth theory 2

PT-1 PA^T without the induction axiom;

PT-2 $\forall\phi(x) \in \mathcal{L}_T : [tot(\phi(x)) \wedge T(\phi(0)) \wedge \forall y(T(\phi(y)) \rightarrow T(\phi(y+1))] \rightarrow \forall x T(\phi(x))$;

PT-3 $\forall$ atomic $\phi \in \mathcal{L}_{PA} : T(\phi) \leftrightarrow val^+(\phi)$;

PT-4 $\forall$ atomic $\phi \in \mathcal{L}_{PA} : T(\neg\phi) \leftrightarrow \neg val^+(\phi)$;

PT-5 $\forall\phi,\psi \in \mathcal{L}_{PA} : T(\phi \wedge \psi) \leftrightarrow (T(\phi) \wedge T(\psi))$;

PT-6 $\forall\phi,\psi \in \mathcal{L}_{PA} : T(\neg(\phi \wedge \psi)) \leftrightarrow (T(\neg\phi) \vee T(\neg\psi))$;

PT-7 $\forall\phi(x) \in \mathcal{L}_{PA} : T(\forall x\phi(x)) \leftrightarrow \forall x T(\phi(x))$;

PT-8 $\forall\phi(x) \in \mathcal{L}_{PA} : T(\neg\forall x\phi(x)) \leftrightarrow \exists x T(\neg\phi(x))$;

PT-9 $\forall\phi \in \mathcal{L}_{PA} : T(\neg\neg\phi) \leftrightarrow T(\phi)$.

A deflationist truth theory 3

`Observation:` PT^- is still highly compositional, but not as compositional as TC.

Theorem

PT^- is conservative over PA but is not interpretable in PA.

PT^- seems sufficient for capturing the technical use of the concept of truth that is made in mathematics (such as model theory).

Type-free truth

- Truth theories which are formulated in languages that contain truth predicates of different levels and which prove iterated truth ascriptions only if the hierarchy constraints are satisfied, are called *typed theories of truth*.
- There also exist truth systems which contain a single truth predicate but which do validate sentences of the form $T(T(0 = 0))$. These systems are called *type-free theories of truth* (*reflexive*, *semantically closed* theories).

The strength of type-free disquotation

Theorem (McGee)

Any theory extending PA can be reaxiomatized by the axioms of PA and a set of Tarski-biconditionals.

Proof.

Consider an axiom ψ of S. Using the diagonal lemma, we can find a sentence such that

$$\lambda \leftrightarrow (T(\lambda) \leftrightarrow \psi)$$

is provable in *PA*. This equivalence is logically equivalent to:

$$\psi \leftrightarrow (T(\lambda) \leftrightarrow \lambda).$$

So ψ is *PA*-provably equivalent to the Tarski-biconditional $T(\lambda) \leftrightarrow \lambda$. $\square$

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

Positive disquotation

(Halbach)

The theory $PUTB$ is the theory given by PA^T and the set of all sentences

$$\forall x\,(T(\phi(x)) \leftrightarrow \phi(x)),$$

where in the formula $\phi(x)$, T must not occur in the scope of an odd number of negation symbols in the formula ϕ.

Theorem

$PUTB \vdash T(T(0 = 0))$

But: $PUTB \nvdash \forall \phi \in L_{PA} : T(\neg\phi) \leftrightarrow \neg T(\phi)$

Going partial

(Kripke)

Structure of the argument of the liar paradox:

1. $L \vee \neg L$
2. $L \Rightarrow \bot$
3. $\neg L \Rightarrow \bot$

So, $\bot$

`Moral:` Do not assert the law of excluded third.

Partial models

- The aim is build a model for the language $\mathcal{L}_T$ in stages
- The arithmetical vocabulary is interpreted throughout as in the standard model $\mathbb{N}$.
- The truth predicate T will be the only partially interpreted symbol: it will receive, at each ordinal stage, an *extension* $\mathcal{E}$ and an *anti-extension* $\mathcal{A}$.

$$\mathfrak{M} = (\mathcal{E}, \mathcal{A})$$

`Note:`

1. $\mathcal{E} \cap \mathcal{A} = \emptyset$
2. $\mathcal{E} \cup \mathcal{A}$ does not typically exhaust the domain, for otherwise T would be a total predicate.

Strong Kleene valuation

- For any atomic formula $Fx_1...x_n$:
  1. $\mathfrak{M} \models_{sk} Fk_1...k_n$ if the n-tuple $(k_1, ..., k_n)$ belongs to the extension of F;
  2. $\mathfrak{M} \models_{sk} \neg Fk_1...k_n$ if the n-tuple $(k_1, ..., k_n)$ belongs to the anti-extension of F.

- For any formulae ϕ, ψ :
  1. $\mathfrak{M} \models_{sk} \phi \wedge \psi$ if and only if $\mathfrak{M} \models_{sk} \phi$ and $\mathfrak{M} \models_{sk} \psi$;
  2. $\mathfrak{M} \models_{sk} \neg(\phi \wedge \psi)$ if and only if either $\mathfrak{M} \models_{sk} \neg\phi$ or $\mathfrak{M} \models_{sk} \neg\psi$ (or both);
  3. $\mathfrak{M} \models_{sk} \forall x\phi$ if and only if for all n, $\mathfrak{M} \models_{sk} \phi(n/x)$;
  4. $\mathfrak{M} \models_{sk} \neg\forall x\phi$ if and only if for at least one n, $\mathfrak{M} \models_{sk} \neg\phi(n/x)$;
  5. $\mathfrak{M} \models_{sk} \neg\neg\phi$ if and only if $\mathfrak{M} \models_{sk} \phi$.

A chain of partial models 1

$$\mathfrak{M}_0 = (\mathcal{E}_0, \mathcal{A}_0) =: (\emptyset, \emptyset)\,;$$

$$\mathcal{E}_{\alpha+1} =: \{\phi \in \mathcal{L}_{\mathsf{T}} \mid \mathfrak{M}_\alpha \models_{sk} \phi\} \text{ and}$$
$$\mathcal{A}_{\alpha+1} =: \{\phi \in \mathcal{L}_{\mathsf{T}} \mid \mathfrak{M}_\alpha \models_{sk} \neg\phi\}\,;$$

For λ limit ordinal, we set:

$$\mathcal{E}_\lambda =: \bigcup_{\kappa<\lambda} \mathcal{E}_\kappa,$$
$$\mathcal{A}_\lambda =: \bigcup_{\kappa<\lambda} \mathcal{A}_\kappa.$$

A chain of partial models 2

This inductive definition gives rise to a transfinite chain of partial models:

$$(\mathcal{E}_0, \mathcal{A}_0), (\mathcal{E}_1, \mathcal{A}_1), \ldots, (\mathcal{E}_\omega, \mathcal{A}_\omega), \ldots$$

Theorem (monotonicity)

For any two partial models $(\mathcal{E}_a, \mathcal{A}_a), (\mathcal{E}_b, \mathcal{A}_b)$, if $\mathcal{E}_a \subseteq \mathcal{E}_b$ and $\mathcal{A}_a \subseteq \mathcal{A}_b$, then

$$\{\phi \mid (\mathcal{E}_a, \mathcal{A}_a) \models_{sk} \phi\} \subseteq \{\phi \mid (\mathcal{E}_b, \mathcal{A}_b) \models_{sk} \phi\}.$$

The minimal fixed point model

A consequence of this is that:

Corollary

For all α, β with $\alpha < \beta$, we have:

$$\{\phi \mid (\mathcal{E}_\alpha, \mathcal{A}_\alpha) \models_{sk} \phi\} \subseteq \{\phi \mid (\mathcal{E}_\beta, \mathcal{A}_\beta) \models_{sk} \phi\}.$$

Proposition

For some ordinal ρ, $\mathcal{E}_\rho = \mathcal{E}_{\rho+1}$ and $\mathcal{A}_\rho = \mathcal{A}_{\rho+1}$.

This ordinal ρ is called the (Strong Kleene) *minimal fixed point*, and $(\mathcal{E}_\rho, \mathcal{A}_\rho)$ is called the *minimal fixed point model*.

Theorem (Kripke)

The minimal Strong Kleene fixed point is a Π^1_1-complete set.

Almost having it all

The minimal fixed point almost makes the unrestricted Tarski-biconditionals true:

Theorem

For all sentences ϕ in $\mathcal{L}_T$:

$$\mathfrak{M}_\rho \models_{sk} \phi \Leftrightarrow \mathfrak{M}_\rho \models_{sk} T(\phi).$$

Proof.

First, suppose $\mathfrak{M}_\rho \models_{sk} \phi$. Then by the definition of the sequence of partial models, $\mathfrak{M}_{\rho+1} \models_{sk} T(\phi)$. But since $\mathfrak{M}_\rho$ is a fixed point, we have $\mathfrak{M}_{\rho+1} = \mathfrak{M}_\rho$. So $\mathfrak{M}_\rho \models_{sk} T(\phi)$. Second, suppose $\mathfrak{M}_\rho \models_{sk} T(\phi)$. Then there must be an ordinal $\alpha < \rho$ such that $\mathfrak{M}_\alpha \models_{sk} \phi$. And therefore, by monotonicity, $\mathfrak{M}_\rho \models_{sk} \phi$. $\square$

Paradoxical sentences

Definition

Closing off a partial model means letting its anti-extension spill over into all of the complement of its extension.

Theorem

$\mathfrak{M}_\rho \nvDash_{sk} T(L)$ *and* $\mathfrak{M}_\rho \nvDash_{sk} \neg T(L)$.

Proof.

By closing off the model $\mathfrak{M}_\rho$ *we obtain a classical model* $\mathfrak{M}_\rho^c$. *Suppose that* $\mathfrak{M}_\rho \models_{sk} T(L)$, *and thereby* $\mathfrak{M}_\rho \models_{sk} L$. *Then by monotonicity, also* $\mathfrak{M}_\rho^c \models_{sk} T(L)$ *and* $\mathfrak{M}_\rho^c \models_{sk} L$. *But since* $\mathfrak{M}_\rho^c$ *is just a classical model, the diagonal lemma holds in it. So we have*

$$\mathfrak{M}_\rho^c \models_{sk} L \leftrightarrow \neg T(L).$$

But putting these three facts together gives us a contradiction. So we deny our supposition and conclude that $\mathfrak{M}_\rho \nvDash_{sk} T(L)$. □

Other valuation schemes

So far we have been using the Strong Kleene valuation scheme...

- Weak Kleene: "if a component is gappy, then the whole is gappy." (Feferman)
- Paraconsistent: "Truth gluts instead of truth gaps " (Priest)
- ...

Formally, it does not make a lot of difference ...

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

Supervaluation

In the supervaluation approach, a formula $\phi \in \mathcal{L}_T$ is regarded as true in a partial model $\mathfrak{M} = (\mathcal{E}, \mathcal{A})$ if and only if ϕ is true in all total (or classical) models $\mathfrak{M}_c = \langle \mathbb{N}, \mathcal{C} \rangle$ for which the interpretation $\mathcal{C}$ of the truth predicate is such that $\mathcal{E} \subseteq \mathcal{C}$ and $\mathcal{A} \subseteq \mathbb{N} \backslash \mathcal{C}$.

Similarly, we say that a formula $\phi \in \mathcal{L}_T$ is regarded as false in a partial model $\mathfrak{M} = (\mathcal{E}, \mathcal{A})$ if and only if ϕ is false in all total (or classical) models $\mathfrak{M}_c = \langle \mathbb{N}, \mathcal{C} \rangle$ for which the interpretation $\mathcal{C}$ of the truth predicate is such that $\mathcal{E} \subseteq \mathcal{C}$ and $\mathcal{A} \subseteq \mathbb{N} \backslash \mathcal{C}$.

Properties of supervaluation fixed points

The laws of classical logic are supervaluation-true in the minimal fixed point of the supervaluation scheme.

The supervaluation fixed point is not compositional.

Theorem

The supervaluation fixed point is a complete Π^1_1 *set.*

Axiomatising Strong Kleene Fixed Points

The theory KF:

KF1 $\forall$ atomic $\phi \in \mathcal{L}_{PA} : T(\phi) \leftrightarrow val^{+}(\phi)$;

KF2 $\forall$ atomic $\phi \in \mathcal{L}_{PA} : T(\neg\phi) \leftrightarrow val^{-}(\phi)$;

KF3 $\forall \phi \in \mathcal{L}_T : T(\neg\neg\phi) \leftrightarrow T(\phi)$;

KF4 $\forall \phi, \psi \in \mathcal{L}_T : T(\phi \wedge \psi) \leftrightarrow (T(\phi) \wedge T(\psi))$;

KF5 $\forall \phi, \psi \in \mathcal{L}_T : T(\neg(\phi \wedge \psi)) \leftrightarrow (T(\neg\phi) \vee T(\neg\psi))$;

KF6 $\forall \phi(x) \in \mathcal{L}_T : T(\forall x \phi(x)) \leftrightarrow \forall y T(\phi(y))$;

KF7 $\forall \phi(x) \in \mathcal{L}_T : T(\neg\forall x \phi(x)) \leftrightarrow \exists y T(\neg\phi(y))$;

KF8 $\forall \phi \in \mathcal{L}_T : T(T(\phi)) \leftrightarrow T(\phi)$;

KF9 $\forall \phi \in \mathcal{L}_T : T(\neg T(\phi)) \leftrightarrow T(\neg\phi)$;

KF10 $\forall \phi \in \mathcal{L}_T : \neg(T(\phi) \wedge T(\neg\phi))$.

Inner logic and outer logic

Theorem

KF has nice models.

Proof.

The closed off minimal fixed point model.

□

Definition

The inner logic of KF is the collection of sentences $\phi \in \mathcal{L}_T$ such that $KF \vdash T(\phi)$.

Proposition

$KF \vdash L \wedge \neg T(L)$, *where L is the liar sentence.*

So the inner logic of KF does not coincide with the external logic of KF.

Ramified Analysis

Let us add new second-order variables $X^1, Y^1, \ldots$ to the language of second-order arithmetic. Add a new comprehension principle

$$\exists X^1 \forall y : X^1(y) \leftrightarrow \phi(y)$$

where $\phi(y)$ contains no bound second-order quantifiers *of the new kind* (and does not contain X^1 free).

The resulting system is called ACA_1.

The strength of *KF*

We can go on, and construct $ACA_n, \ldots, ACA_{<\omega}$...

Definition

ϵ_0 is the least upper bound of $\{\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots\}$.

Theorem (Feferman)

The first order arithmetical theorems of KF are exactly those of ACA_{ϵ_0}.

Truth and Classes

Definition

$x \in y \equiv y \in \mathcal{L}_{\mathcal{T}} \wedge T(y(x))$

Then *KF* can be seen as a theory of definable or *predicative* classes (Feferman).

$\Rightarrow$ *KF* as a way of ascending from *PA* to predicative analysis.

A way out?

What is the source of the divergence between inner and outer logic?

Motivation of Kripke's theory = *partial*

Logic of *KF* = *classical*

$\Rightarrow$ Mismatch

`Solution:` Formalise Kripke's theory in partial logic!

Logic and mathematics

Restricted conditionalisation:

$$\frac{T(\phi) \vee T(\neg\phi) \quad \begin{array}{c} \phi\,(Hyp) \\ \vdots \\ \psi\,(Hyp) \end{array}}{\phi \rightarrow \psi}$$

Rule of induction:

$$\frac{\phi(0) \quad \begin{array}{c} \phi(x)\,(Hyp) \\ \vdots \\ \phi(s(x))\,(Hyp) \end{array}}{\forall x \phi(x)}$$

Here $\phi(x)$ ranges over all formulae of $\mathcal{L}_T$.

The system *PKF*

PKF1

$$\frac{val^{+}(t_1 = t_2)}{T(t_1 = t_2)} \qquad \frac{T(t_1 = t_2)}{val^{+}(t_1 = t_2)}$$

PKF2

$$\frac{T(\phi) \wedge T(\psi)}{T(\phi \wedge \psi)} \qquad \frac{T(\phi \wedge \psi)}{T(\phi) \wedge T(\psi)}$$

PKF3

$$\frac{T(\phi) \vee T(\psi)}{T(\phi \vee \psi)} \qquad \frac{T(\phi \vee \psi)}{T(\phi) \vee T(\psi)}$$

PKF4

$$\frac{\forall x T(\phi(x))}{T(\forall x \phi(x))} \qquad \frac{T(\forall x \phi(x))}{\forall x T(\phi(x))}$$

The system *PKF* Ctd

PKF5

$$\frac{\exists x T(\phi(x))}{T(\exists x \phi(x))} \quad \frac{T(\exists x \phi(x))}{\exists x T(\phi(x))}$$

PKF6

$$\frac{T(\phi)}{T(T(\phi))} \quad \frac{T(T(\phi))}{T(\phi)}$$

PKF7

$$\frac{\neg T(\phi)}{T(\neg \phi)} \quad \frac{T(\neg \phi)}{\neg T(\phi)}$$

Properties of *PKF*

Theorem

Inner logic of PKF = Outer logic of PKF

Theorem

PKF holds in all fixed point models of the Strong Kleene construction.

Theorem

The first-order arithmetical theorems of PKF are those of ACA_{ω^ω}.

A chain of classical models

$$\mathfrak{M}_0 =: \langle \mathbb{N}, \emptyset \rangle$$

$$\mathfrak{M}_{\alpha+1} =: \langle \mathbb{N}, \{\phi \in \mathcal{L}_T \mid \mathfrak{M}_\alpha \models \phi\} \rangle.$$

For λ a limit ordinal:

$$\mathfrak{M}_\lambda =: \langle \mathbb{N}, \{\phi \in \mathcal{L}_T \mid \exists \beta \forall \gamma : (\gamma \geq \beta \wedge \gamma < \lambda) \Rightarrow \mathfrak{M}_\gamma \models \phi\} \rangle.$$

Stable truth

`Stable truth:`
A sentence $\phi \in \mathcal{L}_T$ is said to be stably true if at some ordinal stage α, ϕ enters in the extension of the truth predicate of $\mathfrak{M}_\alpha$ and stays in the extension of the truth predicate in all later models.

`Stable falsehood:`
A sentence $\phi \in \mathcal{L}_T$ is said to be stably false if at some ordinal stage α, ϕ is outside the extension of the truth predicate of $\mathfrak{M}_\alpha$ and stays out forever thereafter.

Nearly stable truth

`Nearly stable truth:`
A sentence $\phi \in \mathcal{L}_T$ is said to be nearly stably true if for every stage α after some stage β, there is a natural number n such that for all natural numbers $m \geq n$, ϕ is in the extension of the truth predicate of $\mathfrak{M}_{\alpha+m}$.

`Nearly stable falsehood:`
A sentence $\phi \in \mathcal{L}_T$ is said to be nearly stably false if for every stage α after some stage β, there is a natural number n such that for all natural numbers $m \geq n$, ϕ is outside the extension of the truth predicate of $\mathfrak{M}_{\alpha+m}$.

Properties of the (nearly) stable truths

The liar sentence is neither (nearly) stably true nor (nearly) stably false.

Proposition

The chain of revision models is eventually periodic.

Theorem

The (nearly) stable truths form a set that is more complicated than a complete Π^1_1 *set.*

The Friedman-Sheard theory 1

The theory FS:

FS1 PA^T;

FS2 $\forall$ atomic $\phi \in \mathcal{L}_{PA} : T(\phi) \leftrightarrow val^+(\phi)$;

FS3 $\forall \phi \in \mathcal{L}_T : T(\neg\phi) \leftrightarrow \neg T(\phi)$;

FS4 $\forall \phi, \psi \in \mathcal{L}_T : T(\phi \wedge \psi) \leftrightarrow T(\phi) \wedge T(\psi)$;

FS5 $\forall \phi(x) \in \mathcal{L}_T : T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$.

NEC From a proof of ϕ, infer $T(\phi)$;

CNEC From a proof of $T(\phi)$, infer ϕ.

The Friedman-Sheard theory 2

Proposition

Inner logic of FS = outer logic of FS

Proposition

FS is not stably true.

Proof.

FS3 is not stably true: both the liar sentence and its negation are false at all limit ordinals □

Proposition

FS is nearly stably true.

The strength of *FS*

FS only proves finite truth-iterations.

Theorem

The first-order arithmetical consequences of FS are exactly those of ACA_ω.

Corollary

FS is arithmetically sound.

Truth and the Liar

Definitions, Models, Axioms

Typed Disquotational Theories

Typed Compositional Theories

Type-free Disquotational Theories

Type-free compositional theories

The Revision Theory

Selected References

Omega-inconsistency

Even though FS is consistent and indeed even arithmetically sound, it is in some sense "almost inconsistent":

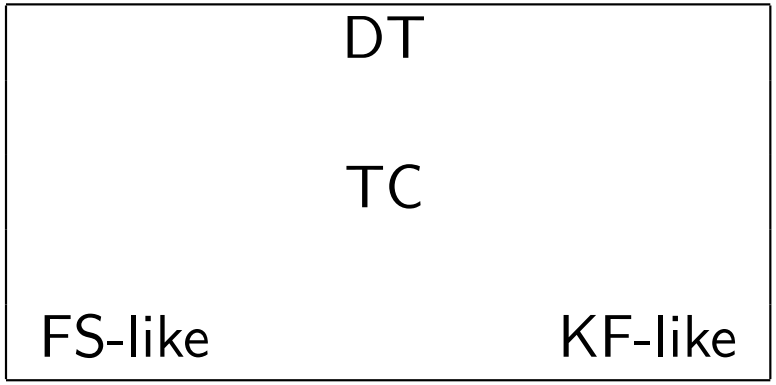
Definition

An arithmetical theory T is ω-inconsistent if for some formula $\phi(x)$, the theory T proves $\exists x \phi(x)$ while at the same time for every $n \in \mathbb{N}$, T proves $\neg\phi(n)$

Theorem (McGee)

For some formula $\phi(x) \in \mathcal{L}_T$: $FS \vdash \exists x \phi(x)$ and $FS \vdash \neg\phi(n)$ for all $n \in \mathbb{N}$.

A fork in the road



Desiderata for axiomatic truth theories

- Coherence
- Tarski-biconditionals
- Compositionality
- Sustaining ordinary reasoning
- Strength
- Capturing a picture

Fulfillment of these norms is a matter of degree.

References

- Cantini, A. *Logical Frameworks for Truth and Abstraction.* North-Holland, 1996.
- Feferman, S. *Reflecting on Incompleteness.* Journal of Symbolic Logic **56**(1991), p. 1–49.
- Gupta, A. & Belnap, N. *The Revision Theory of Truth.* MIT Press, 1993.
- Halbach, V. *Axiomatic Theories of Truth.* Cambridge University Press, 2011.
- Horsten L. *The Tarskian Turn. Deflationism and axiomatic truth.* MIT Press, 2011.
- McGee, V. *Truth, Vagueness and Paradox. An essay on the logic of truth.* Hackett, 1991.
- Visser, A. *Semantics and the Liar Paradox.* In: D. Gabbay et al (eds) *Handbook of Philosophical Logic. Volume 4.* Reidel, 1984, p. 617–706.