# Part II: Typed Truth and Tarski's Hierarchy

Lavinia Picollo
Assistant Professor

Munich Center for Mathematical Philosophy - LMU Munich
Lavinia.Picollo@lrz.uni-muenchen.de
laviniapicollo.com

## Tarski's Program

## Tarski's dream

Tarski's goal was to give an explicit definition of the truth predicate, i.e. a definition of the form:

A sentence[1] $x$ is true iff $x$ is ...

Formally:

$$\mathrm{T}x \leftrightarrow_{def} \Phi(x)$$

where $\Phi(x)$ doesn't contain T but simpler, already understood notions.

Previous attempts had (arguably) failed the last requirement, e.g. the correspondence theory of truth:

A sentence $x$ is true iff $x$ corresponds with reality/to a fact.

as it failed to clarify what correspondence amounts to.

Tarski simplified the debate by identifying an adequacy condition every definition of truth for a language must entail for each of its sentences:[2]

(T-schema) $\mathrm{T}\ulcorner\Phi\urcorner \leftrightarrow \Phi$

[1] Or proposition, statement, utterance, etc.
[2] Note that this is not an explicit definition!



## Shattered by his own theorem

(T-schema) $\mathrm{T}\ulcorner\Phi\urcorner \leftrightarrow \Phi$

(Liar equivalence) $\lambda \leftrightarrow \neg\mathrm{T}\ulcorner\lambda\urcorner$

$$\Downarrow$$

$$\mathrm{T}\ulcorner\lambda\urcorner \leftrightarrow \lambda$$

$$\mathrm{T}\ulcorner\lambda\urcorner \leftrightarrow \neg\mathrm{T}\ulcorner\lambda\urcorner$$

Should we reject:

- The reasoning that led us to a contradiction? The reasoning that takes us from a contradiction to triviality? No, Tarski wants to remain classical. We will explore this route in Part III.
- The existence of the liar sentence (i.e. a premise)? No, as Gödel's work and Kripke's Jack argument show.
- The T-schema (i.e. another premise)? Yes!

## Tarski's pessimism

Tarski's moral: No language can contain its own truth predicate, on pain of triviality.

Tarski's strategy, Typing: Define truth for a particular formal interpreted language, the "object language", in a 'richer' formal interpreted language, the "metalanguage".

All instances of the T-schema for sentences of the object language should follow from this definition. It should also follow that only sentences of this language can be true.



## Tarskian Truth Definitions

## Object languages and their metalanguages

Object languages must contain finitely many primitive symbols.

The corresponding metalanguages must contain:

- (translations of) the object language primitive symbols;
- names $\ulcorner \Phi \urcorner$ for each sentence $\Phi$ of the object language;
- syntactic vocabulary, to talk about expressions of the object language (in most cases);
- a predicate T to express truth for sentences of the object language;
- individual variables $x, y, z, \ldots$;
- $=, \neg, \wedge, \vee, (\rightarrow, \leftrightarrow), \forall$ and $\exists$.



## Example 1: A toy case

Let the object language consist only of the sentences: $p$, $q$, $r$

The following is an explicit definition of truth for the object language in an adequate metalanguage:

$$\mathrm{T}x \leftrightarrow_{def} (x = \ulcorner p \urcorner \wedge p) \vee (x = \ulcorner q \urcorner \wedge q) \vee (x = \ulcorner r \urcorner \wedge r)$$

- $p$ : Snow is white.
- $q$ : The moon is made of green cheese.
- $r$ : $1 + 1 = 2$

We have that: $\mathrm{T}\ulcorner p \urcorner$, $\neg \mathrm{T}\ulcorner q \urcorner$, $\mathrm{T}\ulcorner r \urcorner$

Thus, only sentences of the object language can be true and, for each of them

$$\mathrm{T}\ulcorner \Phi \urcorner \leftrightarrow \Phi$$

This definition might not give us the essence or intension of truth but it gets its extension right.

## Example 2: Complicating things a bit

Let the object language contain:

- Atomic sentences: $p$, $q$, $r$
- Molecular sentences:
  - If $\Phi$ is a sentence, $\neg\Phi$ is a sentence.
  - If $\Phi$ and $\Psi$ are sentences, $(\Phi \wedge \Psi)$ and $(\Phi \vee \Psi)$ are also sentences.

The object language now contains infinitely many sentences. So the following is a bad idea, because we will never finish writing the definition down:

$$\begin{aligned}\mathsf{T}x \leftrightarrow_{def}\ & (x = \ulcorner p \urcorner \wedge p) \vee (x = \ulcorner q \urcorner \wedge q) \vee (x = \ulcorner r \urcorner \wedge r) \vee \\ & (x = \ulcorner \neg p \urcorner \wedge \neg p) \vee (x = \ulcorner \neg q \urcorner \wedge \neg q) \vee (x = \ulcorner \neg r \urcorner \wedge \neg r) \vee \\ & (x = \ulcorner \neg p \urcorner \wedge \neg\neg p) \vee (x = \ulcorner \neg\neg q \urcorner \wedge \neg\neg q) \vee (x = \ulcorner \neg r \urcorner \wedge \neg\neg r) \ \vee \ldots\end{aligned}$$

(Note: we could include individual constants, predicates, function symbols, and quantifiers to the metalanguage, but we don't, to keep things simple.)



## Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

$$\begin{aligned}\mathsf{T}x \leftrightarrow_{def}\ & (x = \ulcorner p \urcorner \wedge p) \vee (x = \ulcorner q \urcorner \wedge q) \vee (x = \ulcorner r \urcorner \wedge r) \vee \\ & \exists y(\mathsf{Neg}(x,y) \wedge \neg \mathsf{T}y) \vee \\ & \exists y \exists z(\mathsf{Con}(x,y,z) \wedge \mathsf{T}y \wedge \mathsf{T}z) \vee \\ & \exists y \exists z(\mathsf{Dis}(x,y,z) \wedge (\mathsf{T}y \vee \mathsf{T}z))\end{aligned}$$

With help of the syntactic predicates:

- $\mathsf{Neg}(\ulcorner \neg\Phi \urcorner, \ulcorner \Phi \urcorner)$
- $\mathsf{Con}(\ulcorner \Phi \wedge \Psi \urcorner, \ulcorner \Phi \urcorner, \ulcorner \Psi \urcorner)$
- $\mathsf{Dis}(\ulcorner \Phi \vee \Psi \urcorner, \ulcorner \Phi \urcorner, \ulcorner \Psi \urcorner)$

Recursive definitions can be turned into explicit ones if enough resources are available.

We have that: $\mathsf{T}\ulcorner p \urcorner$, $\neg\mathsf{T}\ulcorner q \urcorner$, $\mathsf{T}\ulcorner r \urcorner$, $\mathsf{T}\ulcorner \neg q \urcorner$, $\mathsf{T}\ulcorner \neg\neg p \urcorner$, $\mathsf{T}\ulcorner p \wedge r \urcorner$, $\neg\mathsf{T}\ulcorner p \wedge q \urcorner$, $\mathsf{T}\ulcorner p \vee q \urcorner$

Thus:

$$\mathsf{T}\ulcorner \Phi \urcorner \leftrightarrow \Phi$$



## Example 3: Complicating things a bit more

Let the object language now contain:

- Individual constants: a, b
- Predicate letter: P
- Individual variables: $x, y, z, \ldots$
- Atomic formulae: if $t$ is an individual constant or variable, $Pt$ is a formula.
- Molecular formulae:
  - If $\Phi$ is a formula, $\neg\Phi$ is a formula.
  - If $\Phi$ and $\Psi$ are formulae, $(\Phi \wedge \Psi)$ and $(\Phi \vee \Psi)$ are also formulae.
  - If $\Phi$ is a formula and $v$ is a variable, $\forall v\Phi$ and $\exists v\Phi$ are formulae.

We can extend our recursive definition as follows:

$$\begin{aligned}\mathsf{T}x \leftrightarrow_{def}\ & (x = \ulcorner \mathsf{Pa} \urcorner \wedge \mathsf{Pa}) \vee (x = \ulcorner \mathsf{Pb} \urcorner \wedge \mathsf{Pb}) \vee \\ & \exists y(\mathsf{Neg}(x,y) \wedge \neg \mathsf{T}y) \vee \\ & \exists y \exists z(\mathsf{Con}(x,y,z) \wedge \mathsf{T}y \wedge \mathsf{T}z) \vee \\ & \exists y \exists z(\mathsf{Dis}(x,y,z) \wedge (\mathsf{T}y \vee \mathsf{T}z)) \vee \\ & \exists y \exists z(\mathsf{Uni}(x,y,z) \wedge \mathsf{Tsub}(y, \ulcorner \mathsf{a} \urcorner, z) \wedge \mathsf{Tsub}(y, \ulcorner \mathsf{b} \urcorner, z)) \vee \\ & \exists y \exists z(\mathsf{Exi}(x,y,z) \wedge (\mathsf{Tsub}(y, \ulcorner \mathsf{a} \urcorner, z) \vee \mathsf{Tsub}(y, \ulcorner \mathsf{b} \urcorner, z)))\end{aligned}$$



## Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

- $\mathsf{Uni}(\ulcorner \forall x\Phi \urcorner, \ulcorner \Phi \urcorner, \ulcorner x \urcorner)$
- $\mathsf{Exi}(\ulcorner \exists x\Phi \urcorner, \ulcorner \Phi \urcorner, \ulcorner x \urcorner)$
- $\mathsf{Sub}(\ulcorner \Phi \urcorner, \ulcorner t \urcorner, \ulcorner x \urcorner) = \ulcorner \Phi[t/x] \urcorner$

We assume our first-order object languages contain names for each object the language is about. If this is not the case, definitions are slightly more complicated but still possible.

- a: Aristotle
- b: Beyoncé
- $\mathsf{P}x$ : $x$ is a philosopher

We have that: $\mathsf{T}\ulcorner \mathsf{Pa} \urcorner$, $\neg\mathsf{T}\ulcorner \mathsf{Pb} \urcorner$, $\mathsf{T}\ulcorner \exists x\mathsf{P}x \urcorner$, $\mathsf{T}\ulcorner \neg\forall x\mathsf{P}x \urcorner$

Thus:

$$\mathsf{T}\ulcorner \Phi \urcorner \leftrightarrow \Phi$$

## Tarski's Hierarchy

## No truth definition for the metalanguage?

Call the object language from the previous example $\mathscr{L}_0$ and the result of replacing $\mathsf{T}_0$ for $\mathsf{T}$ in the metalanguage, $\mathscr{L}_1$.

Let $\mathscr{L}_2$ extend $\mathscr{L}_1$ with a new monadic predicate symbol $\mathsf{T}_1$, names $\ulcorner\Phi\urcorner$ for sentences of $\mathscr{L}_1$, and syntactic predicates and functions for expressions of $\mathscr{L}_1$.

The following defines $\mathsf{T}_1$ as a truth predicate for $\mathscr{L}_1$ in $\mathscr{L}_2$:

$$\begin{aligned}\mathsf{T}_1(x) \leftrightarrow_{def}\ & (x = \ulcorner\mathsf{Pa}\urcorner \wedge \mathsf{Pa}) \vee (x = \ulcorner\mathsf{Pb}\urcorner \wedge \mathsf{Pb}) \vee \\ & \exists y(\mathsf{Sent}_{\mathscr{L}_0}(y) \wedge \mathsf{Tru}_0(x,y) \wedge \mathsf{T}_0(y)) \vee \\ & \exists y(\mathsf{Neg}(x,y) \wedge \neg\mathsf{T}_1(y)) \vee \\ & \exists y\exists z(\mathsf{Con}(x,y,z) \wedge \mathsf{T}_1(y) \wedge \mathsf{T}_1(z)) \vee \\ & \exists y\exists z(\mathsf{Dis}(x,y,z) \wedge (\mathsf{T}_1(y) \vee \mathsf{T}_1(z))) \vee \\ & \exists y\exists z(\mathsf{Uni}(x,y,z) \wedge \mathsf{T}_1(\mathsf{sub}(y,\ulcorner\mathsf{a}\urcorner,z)) \wedge \mathsf{T}_1(\mathsf{sub}(y,\ulcorner\mathsf{b}\urcorner,z)) \wedge \\ & \qquad \forall w(\mathsf{Sent}_{\mathscr{L}_0}(w) \rightarrow \mathsf{T}_1(\mathsf{sub}(y,\dot{w},z)))) \vee \\ & \exists y\exists z(\mathsf{Exi}(x,y,z) \wedge (\mathsf{T}_1(\mathsf{sub}(y,\ulcorner\mathsf{a}\urcorner,z)) \vee \mathsf{T}_1(\mathsf{sub}(y,\ulcorner\mathsf{b}\urcorner,z)) \vee \\ & \qquad \exists w(\mathsf{Sent}_{\mathscr{L}_0}(w) \wedge \mathsf{T}_1(\mathsf{sub}(y,\dot{w},z)))))\end{aligned}$$



## No truth definition for the metalanguage? cont'd

With help of the additional syntactic predicates and function:

- $\mathsf{Sent}_{\mathscr{L}_0}(x)$ : $x$ is a sentence of $\mathscr{L}_0$
- $\mathsf{Tru}_0(\ulcorner\mathsf{T}_0(\ulcorner\Phi\urcorner)\urcorner, \ulcorner\Phi\urcorner)$
- $\ulcorner\dot{\Phi}\urcorner = \ulcorner\ulcorner\Phi\urcorner\urcorner$

We have that: $\mathsf{T}_1(\ulcorner\mathsf{Pa}\urcorner)$, $\mathsf{T}_1(\ulcorner\mathsf{T}_0(\ulcorner\mathsf{Pa}\urcorner)\urcorner)$, $\mathsf{T}_1(\ulcorner\mathsf{T}_0(\ulcorner\neg\forall x\mathsf{P}x\urcorner)\urcorner)$

Thus, for each sentence $\Phi$ of $\mathscr{L}_0$:

$$\mathsf{T}_1(\ulcorner\Phi\urcorner) \leftrightarrow \Phi$$

Note that everything that is true$_0$ is also true$_1$.



## The Tarskian Hierarchy

For each natural number $n$, let $\mathscr{L}_{n+1}$ extend $\mathscr{L}_n$ with a new monadic predicate symbol $\mathsf{T}_n$, names $\ulcorner\Phi\urcorner$ for sentences, and syntactic predicates and functions for expressions, of $\mathscr{L}_n$. The following defines $\mathsf{T}_n$ as a truth predicate for $\mathscr{L}_n$ in $\mathscr{L}_{n+1}$:

$$\begin{aligned}\mathsf{T}_n(x) \leftrightarrow_{def}\ & (x = \ulcorner\mathsf{Pa}\urcorner \wedge \mathsf{Pa}) \vee (x = \ulcorner\mathsf{Pb}\urcorner \wedge \mathsf{Pb}) \vee \\ & \exists y(\mathsf{Sent}_{\mathscr{L}_0}(y) \wedge \mathsf{Tru}_0(x,y) \wedge \mathsf{T}_0(y)) \vee \\ & \exists y(\mathsf{Sent}_{\mathscr{L}_1}(y) \wedge \mathsf{Tru}_1(x,y) \wedge \mathsf{T}_1(y)) \vee \\ & \qquad \ldots \\ & \exists y(\mathsf{Sent}_{\mathscr{L}_n}(y) \wedge \mathsf{Tru}_n(x,y) \wedge \mathsf{T}_n(y)) \vee \\ & \exists y(\mathsf{Neg}(x,y) \wedge \neg\mathsf{T}_n(y)) \vee \\ & \exists y\exists z(\mathsf{Con}(x,y,z) \wedge \mathsf{T}_n(y) \wedge \mathsf{T}_n(z)) \vee \\ & \exists y\exists z(\mathsf{Dis}(x,y,z) \wedge (\mathsf{T}_n(y) \vee \mathsf{T}_n(z))) \vee \\ & \exists y\exists z(\mathsf{Uni}(x,y,z) \wedge \mathsf{T}_n(\mathsf{sub}(y,\ulcorner\mathsf{a}\urcorner,z)) \wedge \mathsf{T}_n(\mathsf{sub}(y,\ulcorner\mathsf{b}\urcorner,z)) \wedge \\ & \qquad \forall w(\mathsf{Sent}_{\mathscr{L}_n}(w) \rightarrow \mathsf{T}_n(\mathsf{sub}(y,\dot{w},z)))) \vee \\ & \exists y\exists z(\mathsf{Exi}(x,y,z) \wedge (\mathsf{T}_n(\mathsf{sub}(y,\ulcorner\mathsf{a}\urcorner,z)) \vee \mathsf{T}_n(\mathsf{sub}(y,\ulcorner\mathsf{b}\urcorner,z)) \vee \\ & \qquad \exists w(\mathsf{Sent}_{\mathscr{L}_n}(w) \wedge \mathsf{T}_n(\mathsf{sub}(y,\dot{w},z)))))\end{aligned}$$

## The Tarskian Hierarchy cont'd

We have that: $\mathrm{T}_n(\ulcorner \mathrm{Pa} \urcorner)$, $\mathrm{T}_n(\ulcorner \ldots \mathrm{T}_1(\ulcorner \mathrm{T}_0(\ulcorner \mathrm{Pa} \urcorner)\urcorner) \ldots \urcorner)$

Thus, for each sentence $\Phi$ of $\mathscr{L}_n$:

$$\mathrm{T}_n(\ulcorner \Phi \urcorner) \leftrightarrow \Phi$$

The hierarchy is cumulative.

# Part IV: The Axiomatic Approach

Lavinia Picollo
Assistant Professor

Munich Center for Mathematical Philosophy - LMU Munich
Lavinia.Picollo@lrz.uni-muenchen.de
laviniapicollo.com

# The project

## Beyond Tarski

If we want keep classical reasoning and the possibility of self-reference, we cannot have all instances of the

(T-schema) $\quad T\ulcorner\Phi\urcorner \leftrightarrow \Phi$

on pain of triviality. In particular, not the one for the Liar sentence, $\lambda$.

Tarski prescribes that we only allow for instances not containing T, but this seems too restrictive. For instance, $T\ulcorner 0 = 0\urcorner$ seems as innocuous as $0 = 0$ itself.

If we drop the object language/metalanguage distinction, truth might no longer be definable.

Then so be it. Definitions are evaluated according to the principles they entail (e.g. instances of the T-schema). What if, instead of introducing a truth predicate to the language by definition, we added sound truth principles (axioms or rules) to our favorite theories?



## The building blocks

In an axiomatic theory of truth:

- The language of the theory contains a predicate symbol for truth, names for its own expressions, and predicates and function symbols for its own syntactic notions. It may contain other non-semantic expressions.
- The base theory, formulated in this language and which we extend with adequate truth principles, contains no truth-specific axioms or rules. Ideally, it can prove syntactic facts about itself, it contains a syntax theory for its own language.

## What truth principles?

The T-schema restricted in a way such as to exclude paradoxical instances?

Problem: The most obvious restriction, to leave aside the instances that lead to contradiction (e.g. the Liar and Curry sentences) is not feasible, by McGee's theorem: there are infinitely many sets of maximally consistent collections of instances of the T-schema, none of which is axiomatizable.[1] Consider the following 2-liar cycle:

$$\lambda_1 \leftrightarrow \neg \mathsf{T}\ulcorner \lambda_2 \urcorner$$
$$\lambda_2 \leftrightarrow \mathsf{T}\ulcorner \lambda_1 \urcorner$$

The corresponding instances of the T-schema for $\lambda_1$ and $\lambda_2$ are jointly inconsistent, but consistent on their own. We should adopt more refined restrictions.

[1] Roughly, there is no way to describe these sets so we can know what instances of the T-schema belong to them and which don't.



## What truth principles? cont'd

Compositional principles?

$$\forall x \forall y (\mathsf{Neg}(x, y) \to (\mathsf{T}x \leftrightarrow \neg \mathsf{T}y))$$
$$\forall x \forall y \forall z (\mathsf{Con}(x, y, z) \to (\mathsf{T}x \leftrightarrow \mathsf{T}y \wedge \mathsf{T}z))$$

Metarules?

$$(\mathrm{NEC}) \ \frac{\vdash \Phi}{\vdash \mathsf{T}\ulcorner \Phi \urcorner} \qquad (\mathrm{CONEC}) \ \frac{\vdash \mathsf{T}\ulcorner \Phi \urcorner}{\vdash \Phi}$$

Note that NEC only allow us to derive $\mathsf{T}\ulcorner \Phi \urcorner$ from $\Phi$ if we have proved (and not merely assumed) $\Phi$, and similarly for CONEC.



# Famous Axiomatic Systems

## The language of truth

Our base language will be $\mathscr{L}_{\mathsf{PA}}$, the language of first-order Peano arithmetic. It contains logical symbols $=, \neg, \wedge, \vee, \forall$, and $\exists$, individual variables $x, y, z, \ldots$, an individual constant 0, a one-place function symbol S (for the successor function), and two two-place function symbols $+$ and $\times$.

Let $\mathscr{L}_{\mathsf{T}}$ extend $\mathscr{L}_{\mathsf{PA}}$ with a monadic predicate, T.

Following Gödel, we code each expression $\epsilon$ of $\mathscr{L}_{\mathsf{T}}$ with a natural number $n$, and we say $n$ is the gödelnumber of $\epsilon$.

For ever number $n$, $\overline{n}$ $(= \overbrace{\mathsf{S} \ldots \mathsf{S}}^{n \text{ times}} 0)$ is a name for $n$ in $\mathscr{L}_{\mathsf{PA}}$. Via the coding, $\overline{n}$ can also serve as a name for the expression $\epsilon$ coded by $n$. To indicate this, we often write $\ulcorner \epsilon \urcorner$ instead of $\overline{n}$.

Thus, via the coding, $\mathscr{L}_{\mathsf{PA}}$ talks about the expressions of $\mathscr{L}_{\mathsf{T}}$ and expresses many of its syntactic properties and functions (they are just numerical!).

## The base theory

Our base theory will be first-order Peano arithmetic, PA. It consists of the following axioms:

| | |
|---|---|
| (PA1) | $\forall x(\mathsf{S}x \neq 0)$ |
| (PA2) | $\forall x \forall y(\mathsf{S}x = \mathsf{S}y \to x = y)$ |
| (PA3) | $\forall x(x + 0 = x)$ |
| (PA4) | $\forall x \forall y(x + \mathsf{S}y = \mathsf{S}(x + y))$ |
| (PA5) | $\forall x(x \times 0 = 0)$ |
| (PA6) | $\forall x \forall y(x \times \mathsf{S}y = x \times y + x)$ |
| (Induction) | $\Phi(0) \wedge \forall x(\Phi(x) \to \Phi(\mathsf{S}x)) \to \forall x \Phi(x)$ |

Induction is not a single axiom but a schema. For each formula $\Phi(x)$ of $\mathscr{L}_{\mathsf{PA}}$ the corresponding instance of induction is an axiom of PA.

PA can prove many syntactic facts about the expressions of $\mathscr{L}_{\mathsf{T}}$ (i.e. about numbers). It can serve both as our favorite theory and as a syntax theory.

Let PAT be PA formulated in $\mathscr{L}_{\mathsf{T}}$ with an instance of induction for each formula $\Phi(x)$ of $\mathscr{L}_{\mathsf{T}}$. To obtain an axiomatic theory of truth we just need to add truth-specific principles to PAT.



## System 1: Tarski Biconditionals

TB extends PAT with all instances of the T-schema for sentences of $\mathscr{L}_{\mathsf{PA}}$.

Shortcomings:

- TB doesn't overcome the Tarskian restrictions, it's too weak.
- Tarski objected to disquotational systems[2] because the instances of the T-schema for a class of sentences (closed under logical operators) entail the instances of compositional principles, e.g.

$$\mathsf{T}\ulcorner \Phi \wedge \Psi \urcorner \leftrightarrow \mathsf{T}\ulcorner \Phi \urcorner \wedge \mathsf{T}\ulcorner \Psi \urcorner$$

but not the compositional principles themselves, i.e.

$$\forall x \forall y \forall z(\mathsf{Con}(x, y, z) \to (\mathsf{T}x \leftrightarrow \mathsf{T}y \wedge \mathsf{T}z))$$

[2] That is, systems whose axioms consist of instances of the T-schema, known as a principle of disquotation, for 'removing' corner quotes, $\ulcorner . \urcorner$



## System 2: Compositional Truth

CT extends PAT with the following axioms:

| | |
|---|---|
| (CT1) | $\forall x \forall y \forall z(\mathsf{Ide}(x, y, z) \to (\mathsf{T}z \leftrightarrow \mathsf{val}(x) = \mathsf{val}(y)))$ |
| (CT2) | $\forall x \forall y(\mathsf{Sent}_{\mathscr{L}_{\mathsf{PA}}}(x) \wedge \mathsf{Neg}(x, y) \to (\mathsf{T}x \leftrightarrow \neg \mathsf{T}y))$ |
| (CT3) | $\forall x \forall y \forall z(\mathsf{Sent}_{\mathscr{L}_{\mathsf{PA}}}(x) \wedge \mathsf{Con}(x, y, z) \to (\mathsf{T}x \leftrightarrow \mathsf{T}y \wedge \mathsf{T}z))$ |
| (CT4) | $\forall x \forall y \forall z(\mathsf{Sent}_{\mathscr{L}_{\mathsf{PA}}}(x) \wedge \mathsf{Dis}(x, y, z) \to (\mathsf{T}x \leftrightarrow \mathsf{T}y \vee \mathsf{T}z))$ |
| (CT5) | $\forall x \forall y \forall z(\mathsf{Sent}_{\mathscr{L}_{\mathsf{PA}}}(x) \wedge \mathsf{Uni}(x, y, z) \to (\mathsf{T}x \leftrightarrow \forall w \mathsf{T}\mathsf{sub}(x, \dot{w}, z)))$ |
| (CT6) | $\forall x \forall y \forall z(\mathsf{Sent}_{\mathscr{L}_{\mathsf{PA}}}(x) \wedge \mathsf{Exi}(x, y, z) \to (\mathsf{T}x \leftrightarrow \exists w \mathsf{T}\mathsf{sub}(x, \dot{w}, z)))$ |

With help of the additional syntactic predicate and function:

- $\mathsf{Ide}(\ulcorner s \urcorner, \ulcorner t \urcorner, \ulcorner s = t \urcorner)$
- $\mathsf{val}(\ulcorner t \urcorner) = t$

CT is compositional and contains TB, it entails all instances of the T-schema for sentences of $\mathscr{L}_{\mathsf{PA}}$.

Shortcoming: it doesn't overcome the Tarskian restrictions, it's again too weak.

This theory can be iterated just like Tarskian definitions. The resulting systems are known as systems of Ramified Truth.



## System 3: Positive Tarski Biconditionals

PTB extends PAT with instances of the T-schema for T-positive sentences, i.e. sentences in which T occurs only in the scope of an even number of negation symbols: $\mathsf{T}\ulcorner 0 = 0 \urcorner$ and $\neg\neg \mathsf{T}\ulcorner 0 = 0 \urcorner$ are T-positive, but $\neg \mathsf{T}\ulcorner 0 = 0 \urcorner$ and $\lambda$ aren't.

This theory is untyped, as the truth predicate applies to sentences containing the truth predicate:

| | |
|---|---|
| (Logical truth) | $0 = 0$ |
| (Positive T-schema) | $\mathsf{T}\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$ |
| (Positive T-schema) | $\mathsf{T}\ulcorner \mathsf{T}\ulcorner 0 = 0 \urcorner \urcorner \leftrightarrow \mathsf{T}\ulcorner 0 = 0 \urcorner$ |

$$\Downarrow$$

$$\mathsf{T}\ulcorner 0 = 0 \urcorner$$

$$\mathsf{T}\ulcorner \mathsf{T}\ulcorner 0 = 0 \urcorner \urcorner$$

Shortcomings:

- It doesn't entail compositional principles.
- It is said to be ad hoc, that the restriction to T-positive sentences seems philosophically unmotivated.

## System 4: Friedman-Sheard

FS extends PAT with the following axioms and metarules:

(FS1) $\forall x \forall y \forall z(\mathrm{Ide}(x, y, z) \to (\mathrm{T}z \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)))$

(FS2) $\forall x \forall y(\mathrm{Sent}_{\mathscr{L}_\mathrm{T}}(x) \wedge \mathrm{Neg}(x, y) \to (\mathrm{T}x \leftrightarrow \neg \mathrm{T}y))$

(FS3) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_\mathrm{T}}(x) \wedge \mathrm{Con}(x, y, z) \to (\mathrm{T}x \leftrightarrow \mathrm{T}y \wedge \mathrm{T}z))$

(FS4) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_\mathrm{T}}(x) \wedge \mathrm{Dis}(x, y, z) \to (\mathrm{T}x \leftrightarrow \mathrm{T}y \vee \mathrm{T}z))$

(FS5) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_\mathrm{T}}(x) \wedge \mathrm{Uni}(x, y, z) \to (\mathrm{T}x \leftrightarrow \forall w \mathrm{Tsub}(x, \dot{w}, z)))$

(FS6) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_\mathrm{T}}(x) \wedge \mathrm{Exi}(x, y, z) \to (\mathrm{T}x \leftrightarrow \exists w \mathrm{Tsub}(x, \dot{w}, z)))$

(NEC) $\dfrac{\vdash \Phi}{\vdash \mathrm{T}\ulcorner \Phi \urcorner}$ (CONEC) $\dfrac{\vdash \mathrm{T}\ulcorner \Phi \urcorner}{\vdash \Phi}$



## System 4: Friedman-Sheard cont'd

FS is fully compositional, very natural and also untyped:

(Logical truth) $0 = 0$

(FS1) $\mathrm{T}\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$

$\Downarrow$

$\mathrm{T}\ulcorner 0 = 0 \urcorner$

(NEC) $\mathrm{T}\ulcorner \mathrm{T}\ulcorner 0 = 0 \urcorner \urcorner$

Shortcomings:

- It is $\omega$-inconsistent: there is a formula $\Phi(x)$ such that $\neg \forall x \Phi(x)$ is a theorem, but also $\Phi(\overline{n})$ for every $n$.[3] FS is unsound.

[3] This doesn't mean that FS is inconsistent!



## McGee's $\omega$-paradox

Consider the following provable equivalence:

(McGee equivalence) $\mu \leftrightarrow \neg \forall x \mathrm{T}^x \ulcorner \mu \urcorner$

$\mu$, known as "McGee's sentence", says of itself that it's not true or it's not true that it's true or it's not true that it's true or . . . .

In FS McGee's sentence entails an $\omega$-inconsistency:

If $\neg \mathrm{T}\ulcorner \mu \urcorner$, we have that $\neg \mathrm{T}\ulcorner \neg \forall x \mathrm{T}^x \ulcorner \mu \urcorner \urcorner$. By FS2, this implies that $\neg\neg \mathrm{T}\ulcorner \forall x \mathrm{T}^x \ulcorner \mu \urcorner \urcorner$, i.e. $\mathrm{T}\ulcorner \forall x \mathrm{T}^x \ulcorner \mu \urcorner \urcorner$ and, by FS5, we have that $\forall x \mathrm{T}\ulcorner \mathrm{T}^x \ulcorner \mu \urcorner \urcorner$ or, what is the same, $\forall x \mathrm{T}^{x+1} \ulcorner \mu \urcorner$. Instantiating $x$ in 0, we have $\mathrm{T}\ulcorner \mu \urcorner$. Thus, $\neg \mathrm{T}\ulcorner \mu \urcorner \to \mathrm{T}\ulcorner \mu \urcorner$, which means we can prove $\mathrm{T}\ulcorner \mu \urcorner$, that is, $\mathrm{T}^0 \ulcorner \mu \urcorner$. By successive applications of NEC, we obtain $\mathrm{T}^1 \ulcorner \mu \urcorner$, $\mathrm{T}^2 \ulcorner \mu \urcorner$, and so on.

But, at the same time, $\mathrm{T}\ulcorner \mu \urcorner$ implies $\mathrm{T}\ulcorner \neg \forall x \mathrm{T}^x \ulcorner \mu \urcorner \urcorner$. By FS2, we have that $\neg \mathrm{T}\ulcorner \forall x \mathrm{T}^x \ulcorner \mu \urcorner \urcorner$ and, by FS5, that $\neg \forall x \mathrm{T}\ulcorner \mathrm{T}^x \ulcorner \mu \urcorner \urcorner$, i.e. $\neg \forall x \mathrm{T}^{x+1} \ulcorner \mu \urcorner$. This entails $\neg \forall x \mathrm{T}^x \ulcorner \mu \urcorner$.