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 ...

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

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:

$$Tx \leftrightarrow_{def} \Phi(x)$$

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

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

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.

[1] Or proposition, statement, utterance, etc.

[2] Note that this is not an explicit definition!

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:

$$Tx \leftrightarrow_{def} \Phi(x)$$

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) $$T\ulcorner\Phi\urcorner \leftrightarrow \Phi$$

[1] Or proposition, statement, utterance, etc.

[2] Note that this is not an explicit definition!

Shattered by his own theorem

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

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

$$\Downarrow$$

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

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

Shattered by his own theorem

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

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

$$\Downarrow$$

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

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

Should we reject:

- The reasoning that led us to a contradiction? The reasoning that takes us from a contradiction to triviality?

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.

Shattered by his own theorem

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

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

$$\Downarrow$$

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

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

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)?

Shattered by his own theorem

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

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

$$\Downarrow$$

$$T\ulcorner\lambda\urcorner \leftrightarrow \lambda$$

$$T\ulcorner\lambda\urcorner \leftrightarrow \neg T\ulcorner\lambda\urcorner$$

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.

Shattered by his own theorem

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

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

$$\Downarrow$$

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

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

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)?

Shattered by his own theorem

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

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

$$\Downarrow$$

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

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

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 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".

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.

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.

Object languages and their metalanguages

Object languages must contain finitely many primitive symbols.

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 Φ 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 $\exists$.

Tarskian Truth Definitions

Example 1: A toy case

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

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:

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

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 : $1 + 1 = 2$

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:

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: $\mathrm{T}\ulcorner p \urcorner$,

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: $\mathrm{T}\ulcorner p \urcorner$, $\neg \mathrm{T}\ulcorner q \urcorner$,

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

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

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

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 Φ is a sentence, $\neg\Phi$ is a sentence.
  - If Φ and Ψ are sentences, $(\Phi \wedge \Psi)$ and $(\Phi \vee \Psi)$ are also sentences.

Example 2: Complicating things a bit

Let the object language contain:

- Atomic sentences: p, q, r
- Molecular sentences:
  - If Φ is a sentence, $\neg\Phi$ is a sentence.
  - If Φ and Ψ 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:

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

Example 2: Complicating things a bit

Let the object language contain:

- Atomic sentences: p, q, r
- Molecular sentences:
  - If Φ is a sentence, $\neg\Phi$ is a sentence.
  - If Φ and Ψ 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:

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

With help of the syntactic predicates:

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

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

Example 2: Complicating things a bit cont'd

What is needed is a recursive definition:

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

With help of the syntactic predicates:

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

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

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

Thus:

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

Example 3: Complicating things a bit more

Let the object language now contain:

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

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 Φ is a formula, $\neg\Phi$ is a formula.
  - If Φ and Ψ are formulae, $(\Phi \wedge \Psi)$ and $(\Phi \vee \Psi)$ are also formulae.
  - If Φ 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:

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

Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

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

Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

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

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.

Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

- Uni($\ulcorner \forall x\Phi \urcorner, \ulcorner \Phi \urcorner, \ulcorner x \urcorner$)
- Exi($\ulcorner \exists x\Phi \urcorner, \ulcorner \Phi \urcorner, \ulcorner x \urcorner$)
- 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
- Px : x is a philosopher

Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

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

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
- $\mathrm{P}x$: x is a philosopher

We have that: $\mathrm{T}\ulcorner \mathrm{Pa} \urcorner$, $\neg \mathrm{T}\ulcorner \mathrm{Pb} \urcorner$,

Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

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

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
- Px : x is a philosopher

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

Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

- Uni($\ulcorner \forall x \Phi \urcorner, \ulcorner \Phi \urcorner, \ulcorner x \urcorner$)
- Exi($\ulcorner \exists x \Phi \urcorner, \ulcorner \Phi \urcorner, \ulcorner x \urcorner$)
- 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
- Px : x is a philosopher

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

Example 3: Complicating things a bit more cont'd

With help of the additional syntactic predicates and function:

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

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
- $\text{P}x$: x is a philosopher

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

Thus:

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

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 T_0 for T in the metalanguage, $\mathscr{L}_1$.

No truth definition for the metalanguage?

Call the object language from the previous example $\mathscr{L}_0$ and the result of replacing T_0 for T in the metalanguage, $\mathscr{L}_1$.

Let $\mathscr{L}_2$ extend $\mathscr{L}_1$ with a new monadic predicate symbol T_1, names $\ulcorner\Phi\urcorner$ for sentences of $\mathscr{L}_1$, and syntactic predicates and functions for expressions of $\mathscr{L}_1$.

No truth definition for the metalanguage?

Call the object language from the previous example $\mathscr{L}_0$ and the result of replacing T_0 for T in the metalanguage, $\mathscr{L}_1$.

Let $\mathscr{L}_2$ extend $\mathscr{L}_1$ with a new monadic predicate symbol 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 T_1 as a truth predicate for $\mathscr{L}_1$ in $\mathscr{L}_2$:

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

No truth definition for the metalanguage? cont'd

With help of the additional syntactic predicates and function:

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

No truth definition for the metalanguage? cont'd

With help of the additional syntactic predicates and function:

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

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

No truth definition for the metalanguage? cont'd

With help of the additional syntactic predicates and function:

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

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

No truth definition for the metalanguage? cont'd

With help of the additional syntactic predicates and function:

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

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

No truth definition for the metalanguage? cont'd

With help of the additional syntactic predicates and function:

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

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

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

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

No truth definition for the metalanguage? cont'd

With help of the additional syntactic predicates and function:

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

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

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

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

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 T_n, names $\ulcorner \Phi \urcorner$ for sentences, and syntactic predicates and functions for expressions, of $\mathscr{L}_n$.

The Tarskian Hierarchy

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

The Tarskian Hierarchy

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

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

The Tarskian Hierarchy cont'd

We have that: $T_n(\ulcorner Pa \urcorner)$,

The Tarskian Hierarchy cont'd

We have that: $T_n(\ulcorner Pa \urcorner)$, $T_n(\ulcorner \ldots T_1(\ulcorner T_0(\ulcorner Pa \urcorner)\urcorner) \ldots \urcorner)$

The Tarskian Hierarchy cont'd

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

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

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

The Tarskian Hierarchy cont'd

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

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

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

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

$$\text{(T-schema)} \qquad \mathsf{T}\ulcorner\Phi\urcorner \leftrightarrow \Phi$$

on pain of triviality. In particular, not the one for the Liar sentence, λ.

Beyond Tarski

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

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

on pain of triviality. In particular, not the one for the Liar sentence, λ.

Tarski prescribes that we only allow for instances not containing T, but this seems too restrictive.

Beyond Tarski

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

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

on pain of triviality. In particular, not the one for the Liar sentence, λ.

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.

Beyond Tarski

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

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

on pain of triviality. In particular, not the one for the Liar sentence, λ.

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.

Beyond Tarski

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

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

on pain of triviality. In particular, not the one for the Liar sentence, λ.

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.

Beyond Tarski

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

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

on pain of triviality. In particular, not the one for the Liar sentence, λ.

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 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 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?

[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?

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

[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?

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]

[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?

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 λ_1 and λ_2 are jointly inconsistent, but consistent on their own.

[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?

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 λ_1 and λ_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?

$$\forall x \forall y (\mathrm{Neg}(x, y) \to (\mathrm{T}x \leftrightarrow \neg \mathrm{T}y))$$

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

What truth principles? cont'd

Compositional principles?

$$\forall x \forall y (\mathrm{Neg}(x, y) \rightarrow (\mathrm{T}x \leftrightarrow \neg \mathrm{T}y))$$

$$\forall x \forall y \forall z (\mathrm{Con}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \mathrm{T}y \wedge \mathrm{T}z))$$

Metarules?

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

What truth principles? cont'd

Compositional principles?

$$\forall x \forall y(\mathrm{Neg}(x,y) \rightarrow (\mathrm{T}x \leftrightarrow \neg \mathrm{T}y))$$

$$\forall x \forall y \forall z(\mathrm{Con}(x,y,z) \rightarrow (\mathrm{T}x \leftrightarrow \mathrm{T}y \wedge \mathrm{T}z))$$

Metarules?

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

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

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.

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 ϵ of $\mathscr{L}_{\mathsf{T}}$ with a natural number n, and we say n is the gödelnumber of ϵ.

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 ϵ of $\mathscr{L}_{\mathsf{T}}$ with a natural number n, and we say n is the gödelnumber of ϵ.

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 ϵ coded by n. To indicate this, we often write $\ulcorner\epsilon\urcorner$ instead of $\overline{n}$.

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 ϵ of $\mathscr{L}_{\mathsf{T}}$ with a natural number n, and we say n is the gödelnumber of ϵ.

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 ϵ 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)	$\forall x(Sx \neq 0)$
(PA2)	$\forall x \forall y(Sx = Sy \rightarrow x = y)$
(PA3)	$\forall x(x + 0 = x)$
(PA4)	$\forall x \forall y(x + Sy = S(x + y))$
(PA5)	$\forall x(x \times 0 = 0)$
(PA6)	$\forall x \forall y(x \times Sy = x \times y + x)$
(Induction)	$\Phi(0) \wedge \forall x(\Phi(x) \rightarrow \Phi(Sx)) \rightarrow \forall x \Phi(x)$

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.

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)	$\forall x \forall y(\mathsf{S}x = \mathsf{S}y \rightarrow x = y)$
(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)	$\Phi(0) \wedge \forall x(\Phi(x) \rightarrow \Phi(\mathsf{S}x)) \rightarrow \forall x \Phi(x)$

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.

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)	$\forall x \forall y(\mathsf{S}x = \mathsf{S}y \rightarrow x = y)$
(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)	$\Phi(0) \wedge \forall x(\Phi(x) \rightarrow \Phi(\mathsf{S}x)) \rightarrow \forall x \Phi(x)$

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

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

Let PAT be PA formulated in $\mathscr{L}_{\mathrm{T}}$ with an instance of induction for each formula $\Phi(x)$ of $\mathscr{L}_{\mathrm{T}}$.

The base theory

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

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

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

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

Let PAT be PA formulated in $\mathscr{L}_{\mathrm{T}}$ with an instance of induction for each formula $\Phi(x)$ of $\mathscr{L}_{\mathrm{T}}$. 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 $\mathscr{L}_{\text{PA}}$.

[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 1: Tarski Biconditionals

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

Shortcomings:

[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 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.

[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 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.

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

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

(CT2) $\forall x \forall y(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Neg}(x, y) \rightarrow (\mathrm{T}x \leftrightarrow \neg \mathrm{T}y))$

(CT3) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Con}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \mathrm{T}y \wedge \mathrm{T}z))$

(CT4) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Dis}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \mathrm{T}y \vee \mathrm{T}z))$

(CT5) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Uni}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \forall w \mathrm{Tsub}(x, \dot{w}, z)))$

(CT6) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Exi}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \exists w \mathrm{Tsub}(x, \dot{w}, z)))$

With help of the additional syntactic predicate and function:

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

System 2: Compositional Truth

CT extends PAT with the following axioms:

(CT1) $\forall x \forall y \forall z(\mathrm{Ide}(x, y, z) \rightarrow (\mathrm{T}z \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)))$

(CT2) $\forall x \forall y(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Neg}(x, y) \rightarrow (\mathrm{T}x \leftrightarrow \neg \mathrm{T}y))$

(CT3) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Con}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \mathrm{T}y \wedge \mathrm{T}z))$

(CT4) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Dis}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \mathrm{T}y \vee \mathrm{T}z))$

(CT5) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Uni}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \forall w \mathrm{Tsub}(x, \dot{w}, z)))$

(CT6) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \wedge \mathrm{Exi}(x, y, z) \rightarrow (\mathrm{T}x \leftrightarrow \exists w \mathrm{Tsub}(x, \dot{w}, z)))$

With help of the additional syntactic predicate and function:

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

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

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

System 2: Compositional Truth

CT extends PAT with the following axioms:

(CT1) $\forall x \forall y \forall z(\mathrm{Ide}(x, y, z) \to (\mathrm{T}z \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y)))$

(CT2) $\forall x \forall y(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \land \mathrm{Neg}(x, y) \to (\mathrm{T}x \leftrightarrow \neg \mathrm{T}y))$

(CT3) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \land \mathrm{Con}(x, y, z) \to (\mathrm{T}x \leftrightarrow \mathrm{T}y \land \mathrm{T}z))$

(CT4) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \land \mathrm{Dis}(x, y, z) \to (\mathrm{T}x \leftrightarrow \mathrm{T}y \lor \mathrm{T}z))$

(CT5) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \land \mathrm{Uni}(x, y, z) \to (\mathrm{T}x \leftrightarrow \forall w \mathrm{Tsub}(x, \dot{w}, z)))$

(CT6) $\forall x \forall y \forall z(\mathrm{Sent}_{\mathscr{L}_{\mathrm{PA}}}(x) \land \mathrm{Exi}(x, y, z) \to (\mathrm{T}x \leftrightarrow \exists w \mathrm{Tsub}(x, \dot{w}, z)))$

With help of the additional syntactic predicate and function:

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

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

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:

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: $\mathrm{T}\ulcorner 0 = 0 \urcorner$ and $\neg\neg\mathrm{T}\ulcorner 0 = 0 \urcorner$ are T-positive, but $\neg\mathrm{T}\ulcorner 0 = 0 \urcorner$ and λ aren't.

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: $T\ulcorner 0 = 0 \urcorner$ and $\neg\neg T\ulcorner 0 = 0 \urcorner$ are T-positive, but $\neg T\ulcorner 0 = 0 \urcorner$ and λ aren't.

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

(Logical truth) $$0 = 0$$

(Positive T-schema) $$T\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$$

(Positive T-schema) $$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner \leftrightarrow T\ulcorner 0 = 0 \urcorner$$

$$\Downarrow$$

$$T\ulcorner 0 = 0 \urcorner$$

$$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner$$

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: $T\ulcorner 0 = 0 \urcorner$ and $\neg\neg T\ulcorner 0 = 0 \urcorner$ are T-positive, but $\neg T\ulcorner 0 = 0 \urcorner$ and λ aren't.

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

(Logical truth) $$0 = 0$$

(Positive T-schema) $$T\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$$

(Positive T-schema) $$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner \leftrightarrow T\ulcorner 0 = 0 \urcorner$$

$$\Downarrow$$

$$T\ulcorner 0 = 0 \urcorner$$

$$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner$$

Shortcomings:

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: $T\ulcorner 0 = 0 \urcorner$ and $\neg\neg T\ulcorner 0 = 0\urcorner$ are T-positive, but $\neg T\ulcorner 0 = 0\urcorner$ and λ aren't.

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

(Logical truth) $0 = 0$

(Positive T-schema) $T\ulcorner 0 = 0\urcorner \leftrightarrow 0 = 0$

(Positive T-schema) $T\ulcorner T\ulcorner 0 = 0\urcorner\urcorner \leftrightarrow T\ulcorner 0 = 0\urcorner$

$$\Downarrow$$

$$T\ulcorner 0 = 0\urcorner$$

$$T\ulcorner T\ulcorner 0 = 0\urcorner\urcorner$$

Shortcomings:

- It doesn't entail compositional principles.

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: $T\ulcorner 0 = 0 \urcorner$ and $\neg\neg T\ulcorner 0 = 0 \urcorner$ are T-positive, but $\neg T\ulcorner 0 = 0 \urcorner$ and λ aren't.

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

(Logical truth) $$0 = 0$$

(Positive T-schema) $$T\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$$

(Positive T-schema) $$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner \leftrightarrow T\ulcorner 0 = 0 \urcorner$$

$$\Downarrow$$

$$T\ulcorner 0 = 0 \urcorner$$

$$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner$$

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

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

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

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

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

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

(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:

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

System 4: Friedman-Sheard cont'd

FS is fully compositional, very natural and also untyped:

(Logical truth) $$0 = 0$$

(FS1) $$T\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$$

$$\Downarrow$$

$$T\ulcorner 0 = 0 \urcorner$$

(NEC) $$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner$$

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

System 4: Friedman-Sheard cont'd

FS is fully compositional, very natural and also untyped:

(Logical truth) $0 = 0$

(FS1) $T\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$

$$\Downarrow$$

$$T\ulcorner 0 = 0 \urcorner$$

(NEC) $T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner$

Shortcomings:

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

System 4: Friedman-Sheard cont'd

FS is fully compositional, very natural and also untyped:

(Logical truth) $0 = 0$

(FS1) $T\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$

$$\Downarrow$$

$$T\ulcorner 0 = 0 \urcorner$$

(NEC) $T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner$

Shortcomings:

- It is ω-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]

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

System 4: Friedman-Sheard cont'd

FS is fully compositional, very natural and also untyped:

(Logical truth)	$0 = 0$
(FS1)	$T\ulcorner 0 = 0 \urcorner \leftrightarrow 0 = 0$
	$\Downarrow$
	$T\ulcorner 0 = 0 \urcorner$
(NEC)	$T\ulcorner T\ulcorner 0 = 0 \urcorner \urcorner$

Shortcomings:

- It is ω-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 ω-paradox

McGee's ω-paradox

Consider the following provable equivalence:

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

McGee's ω-paradox

Consider the following provable equivalence:

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

μ, 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 . . . .

McGee's ω-paradox

Consider the following provable equivalence:

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

μ, 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 ω-inconsistency:

McGee's ω-paradox

Consider the following provable equivalence:

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

μ, 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 ω-inconsistency:

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

McGee's ω-paradox

Consider the following provable equivalence:

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

μ, 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 ω-inconsistency:

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

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