The four pillars of deflationism and their logic

Carlo Nicolai

Slides available at https://carlonicolai.github.io

Four Pillars

1. The meaning of ‘is true’ is fixed by the Tarski-biconditionals

Four Pillars

1. The meaning of ‘is true’ is fixed by the Tarski-biconditionals
2. The truth predicate enables one to express infinite conjunctions and disjunctions

Four Pillars

1. The meaning of ‘is true’ is fixed by the Tarski-biconditionals

2. The truth predicate enables one to express infinite conjunctions and disjunctions

3. The truth predicate is fundamentally a device to perform sentential quantification over pronominal variables

Four Pillars

1. The meaning of ‘is true’ is fixed by the Tarski-biconditionals
2. The truth predicate enables one to express infinite conjunctions and disjunctions
3. The truth predicate is fundamentally a device to perform sentential quantification over pronominal variables
4. Truth does not contribute substantially to philosophical and scientific explanations

TRUTH AND SUBSTANTIALITY

The minimalist theory entails that a truth predicate should be conservative over a given theory that is stated without the truth predicate. (Horsten 1995, p. 183)

The minimalist theory entails that a truth predicate should be conservative over a given theory that is stated without the truth predicate. (Horsten 1995, p. 183)

I submit that, in one form or another, conservativeness is essential to deflationism […] How thin can the notion of arithmetic truth be, if by invoking it, we can learn more about natural numbers? (Shapiro 1998, pp. 497-99)

The minimalist theory entails that a truth predicate should be conservative over a given theory that is stated without the truth predicate. (Horsten 1995, p. 183)

I submit that, in one form or another, conservativeness is essential to deflationism [. . .] How thin can the notion of arithmetic truth be, if by invoking it, we can learn more about natural numbers? (Shapiro 1998, pp. 497-99)

. . . if truth is not substantial – as deflationists claim – then the theory of truth should be conservative. Roughly: non-substantiality ≡ conservativeness. (Ketland 1999, p. 79)

CONSERVATIVENESS

- T is *proof-theoretically or deductively conservative* over S iff for all $\varphi \in \mathscr{L}_S$, if $T \vdash \varphi$, then $S \vdash \varphi$;

CONSERVATIVENESS

- T is *proof-theoretically or deductively conservative* over S iff for all $\varphi \in \mathscr{L}_S$, if $T \vdash \varphi$, then $S \vdash \varphi$;
- T is *semantically conservative* over S iff in any model $\mathscr{M} \models S$, the interpretation of any additional notion of $\mathscr{L}_T \supseteq \mathscr{L}_S$ can be provided already in $\mathscr{M}$.

CONSERVATIVENESS

- T is *proof-theoretically or deductively conservative* over S iff for all $\varphi \in \mathscr{L}_S$, if $T \vdash \varphi$, then $S \vdash \varphi$;
- T is *semantically conservative* over S iff in any model $\mathscr{M} \models S$, the interpretation of any additional notion of $\mathscr{L}_T \supseteq \mathscr{L}_S$ can be provided already in $\mathscr{M}$.

Semantic conservativeness implies deductive conservativeness. The converse does not hold.

- The requirement of *semantic conservativeness*, to make sense, needs to postulate *other* notions of truth besides the notion of *truth simpliciter*, e.g. the **notion of truth in $\mathbb{N}$**. (Cieśliński 2015, 2017)

- The requirement of *semantic conservativeness*, to make sense, needs to postulate *other* notions of truth besides the notion of *truth simpliciter*, e.g. the **notion of truth in** $\mathbb{N}$. (Cieśliński 2015, 2017)
- It is not clear whether the requirement of *deductive conservativeness* – nonetheless closer to the deflationist's spirit – has not yet been convincingly linked to the way in which deflationists traditionally understood the notion of *explanatory power*. (Halbach 2001, Field, Cieśliński 2015)

- The requirement of *semantic conservativeness*, to make sense, needs to postulate *other* notions of truth besides the notion of *truth simpliciter*, e.g. the **notion of truth in $\mathbb{N}$**. (Cieśliński 2015, 2017)
- It is not clear whether the requirement of *deductive conservativeness* – nonetheless closer to the deflationist's spirit – has not yet been convincingly linked to the way in which deflationists traditionally understood the notion of *explanatory power*. (Halbach 2001, Field, Cieśliński 2015)
- Virtually none of the natural theories of truth defended in the literature are *semantically conservative*. Very few of them are *syntactically conservative*.

Truth Simpliciter

- Let's work over PA and call our theory of truth *Th*. And ask ourselves what happens if we want to say that it is true that some finite $B \subset \mathsf{PA}$ is consistent. We require out theory of truth *Th* to prove
$$\mathsf{Tr}\ulcorner \mathsf{Con}(B) \urcorner$$

Truth Simpliciter

- Let's work over PA and call our theory of truth Th. And ask ourselves what happens if we want to say that it is true that some finite $B \subset \mathsf{PA}$ is consistent. We require out theory of truth Th to prove
$$\mathsf{Tr}\ulcorner \mathsf{Con}(B)\urcorner$$
Of course we mean that (i) Tr **expresses** truth simpliciter, but that's guaranteed by our axioms, and that (ii) $\mathsf{Con}(B)$ **expresses** the consistency of B. This is essentially because it is so in $\mathbb{N}$.

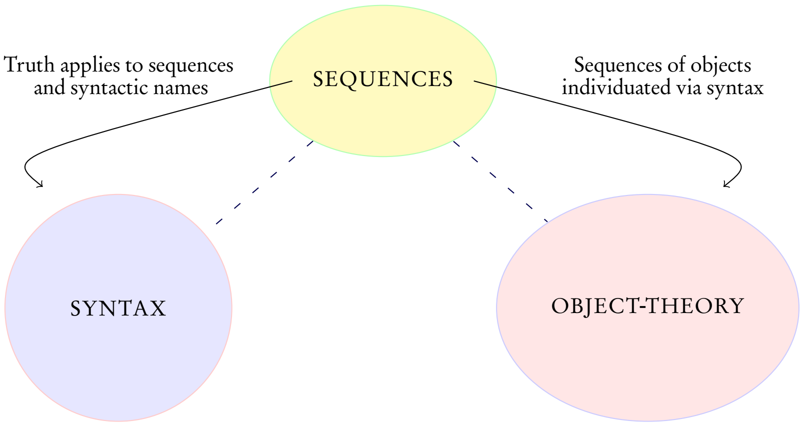
Truth Simpliciter

- Let's work over PA and call our theory of truth *Th*. And ask ourselves what happens if we want to say that it is true that some finite $B \subset \mathsf{PA}$ is consistent. We require out theory of truth *Th* to prove

$$\mathsf{Tr}\ulcorner \mathsf{Con}(B)\urcorner$$

  Of course we mean that (i) Tr **expresses** truth simpliciter, but that's guaranteed by our axioms, and that (ii) $\mathsf{Con}(B)$ **expresses** the consistency of B. This is essentially because it is so in $\mathbb{N}$.

- A reasonable reaction is that the development of the syntax of $\mathscr{L}_{\mathsf{PA}} \cup \{\mathsf{Tr}\}$ in PA is 'isomorphic' to our usual understanding of syntax. But this is vague. If this is so, and we want to eliminate truth-in-$\mathbb{N}$, we should do better.



SEQUENCES
Truth applies to sequences and syntactic names
Sequences of objects individuated via syntax
SYNTAX
OBJECT-THEORY

Nicolai 2015, Nicolai 2016

THE THREE-SORTED THEORY $\mathsf{T}[O]$

- Let O be *any* RE-theory interpreting Robinson's arithmetic Q.

Nicolai 2015, Nicolai 2016

THE THREE-SORTED THEORY $\mathsf{T}[O]$

- Let O be *any* RE-theory interpreting Robinson's arithmetic Q.
- Consider a *fixed, natural, minimal* theory of concatenation – in particular, a theory Σ^b_1-NIA from Ferreira 1990 mutually interpretable with S^1_2 and Q.

Nicolai 2015, Nicolai 2016

THE THREE-SORTED THEORY $\mathsf{T}[O]$

- Let O be *any* RE-theory interpreting Robinson's arithmetic Q.
- Consider a *fixed, natural, minimal* theory of concatenation – in particular, a theory Σ_1^b-NIA from Ferreira 1990 mutually interpretable with S_2^1 and Q.
- Link the two theories via a theory of sequences (empty sequence, prolongation).

Nicolai 2015, Nicolai 2016

THE THREE-SORTED THEORY $\mathsf{T}[O]$

- Let O be *any* RE-theory interpreting Robinson's arithmetic Q.
- Consider a *fixed, natural, minimal* theory of concatenation – in particular, a theory Σ^b_1-NIA from Ferreira 1990 mutually interpretable with S^1_2 and Q.
- Link the two theories via a theory of sequences (empty sequence, prolongation).
- Add *compositional*, *typed* truth axioms.

$\mathsf{T}[O]$ is trivially semantically conservative over O.

Nicolai 2015, Nicolai 2016

THEOREM

Let $\mathsf{Con}(O)$ be given via an interpretation of Σ^b_1-NIA in O.

- For O *finitely* axiomatized, verifiably in Σ^b_1-NIA or S^1_2, $\mathsf{T}[O]$ is **mutually interpretable** with $O + \mathsf{Con}(O)$.
- For O *schematically* axiomatized, verifiably in Σ^b_1-NIA or S^1_2, $\mathsf{T}[O]$ + 'all axioms of O are true' is **mutually interpretable** with $O + \mathsf{Con}(O)$.
- In fact, in both cases, $\mathsf{Con}(O)$ is the **unique** Π_1-sentence – modulo EA-provable equivalence, enjoying this property.

THESIS 1

Truth simpliciter corresponds *at least* to our belief in the consistency of O, for *any* reasonable O. This is a genuinely *metalinguistic* claim, and I propose that the right form of *explanation* that a truth theory can give is in fact primarily metalinguistic.

THE MEANING OF ‘TRUE’

By calling the sentence [‘snow is white’] true, we call snow white. The truth predicate is a device of disquotation. (Quine, Philosophy of Logic, *p. 12)*

I [. . .] simply say that in the purely disquotational sense of “true”, the claim that u is true (where u is an utterance I understand) is cognitively equivalent for me to u itself (as I understand it). [. . .] If I understand “Snow is white”, and if I also understand a notion of disquotational truth as explained above, then I will understand “ ‘Snow is white’ is true”, since it will be equivalent to “Snow is white” (Field, The deflationary view of meaning and content, *p. 251)*

I take that ‘calling’ and ‘cognitively equivalence’ entail a strong form of conceptual necessity: Tr- biconditionals are **constitutive** of the meaning of truth.

Modalized Disquotationalism MD

For simplicity, consider a negation-free language with Tr and F ($\mathscr{L}_{\mathsf{Tr},\mathsf{F}}$) and a modality $\Box$ ($\mathscr{L}^{\Box}_{\mathsf{Tr},\mathsf{F}}$) – also a predicate!

I take that 'calling' and 'cognitively equivalence' entail a strong form of conceptual necessity: Tr- biconditionals are **constitutive** of the meaning of truth.

MODALIZED DISQUOTATIONALISM MD

For simplicity, consider a negation-free language with Tr and F ($\mathscr{L}_{\mathsf{Tr},\mathsf{F}}$) and a modality $\Box$ ($\mathscr{L}^{\Box}_{\mathsf{Tr},\mathsf{F}}$) – also a predicate!

- '$\Box$ is closed under logic'
  $(\forall \varphi, \psi : \mathscr{L}^{\Box}_{\mathsf{Tr},\mathsf{F}})\big((\mathsf{Pr}_{\varnothing}(\varphi) \to \Box\varphi) \land (\Box(\varphi \to \psi) \land \Box\varphi \to \Box\psi)\big)$

I take that 'calling' and 'cognitively equivalence' entail a strong form of conceptual necessity: Tr - biconditionals are **constitutive** of the meaning of truth.

MODALIZED DISQUOTATIONALISM MD

For simplicity, consider a negation-free language with Tr and F ($\mathscr{L}_{\mathsf{Tr},\mathsf{F}}$) and a modality $\Box$ ($\mathscr{L}^{\Box}_{\mathsf{Tr},\mathsf{F}}$) – also a predicate!

- '$\Box$ is closed under logic'
  $(\forall \varphi, \psi : \mathscr{L}^{\Box}_{\mathsf{Tr},\mathsf{F}})\big((\mathsf{Pr}_{\varnothing}(\varphi) \rightarrow \Box\varphi) \wedge (\Box(\varphi \rightarrow \psi) \wedge \Box\varphi \rightarrow \Box\psi)\big)$
- 'uniform disquotation (for positive formulas) is analytic'
  $\Box\big(\forall x(\mathsf{Tr}\ulcorner \varphi(\dot{x})\urcorner \leftrightarrow \varphi(x)) \wedge \forall x(\mathsf{F}\ulcorner \varphi^{D}(\dot{x})\urcorner \leftrightarrow \varphi(x))\big)$

I take that 'calling' and 'cognitively equivalence' entail a strong form of conceptual necessity: Tr - biconditionals are **constitutive** of the meaning of truth.

MODALIZED DISQUOTATIONALISM MD

For simplicity, consider a negation-free language with Tr and F ($\mathscr{L}_{\mathsf{Tr},\mathsf{F}}$) and a modality $\square$ ($\mathscr{L}^{\square}_{\mathsf{Tr},\mathsf{F}}$) – also a predicate!

- '$\square$ is closed under logic'
  $(\forall \varphi, \psi : \mathscr{L}^{\square}_{\mathsf{Tr},\mathsf{F}})\big((\mathsf{Pr}_{\varnothing}(\varphi) \to \square\varphi) \wedge (\square(\varphi \to \psi) \wedge \square\varphi \to \square\psi)\big)$
- 'uniform disquotation (for positive formulas) is analytic'
  $\square\big(\forall x(\mathsf{Tr}\ulcorner \varphi(\dot{x})\urcorner \leftrightarrow \varphi(x)) \wedge \forall x(\mathsf{F}\ulcorner \varphi^{D}(\dot{x})\urcorner \leftrightarrow \varphi(x))\big)$
- Modal axioms (T) and (Nec):
  - for $A(v) \in \mathscr{L}_{\mathsf{Tr},\mathsf{F}}$: $\forall x(\square A(x) \to A(x))$;
  - for $A \in \mathscr{L}^{\square}_{\mathsf{Tr},\mathsf{F}}$, $\dfrac{A}{\square A}$

LEMMA

Over a suitable B, modalized disquotationalism MD includes the axioms of the compositional theory KF (Kripke-Feferman) and it is proof-theoretically reducible to it.

LEMMA

Over a suitable B, modalized disquotationalism MD includes the axioms of the compositional theory KF (Kripke-Feferman) and it is proof-theoretically reducible to it.

It is also natural, however, to read the modal axiom (T) $\forall x(\Box A(x) \to A(x))$ via a truth predicate in the metatheory. But we have a truth predicate in the **object-theory**:

LEMMA

Over a suitable B, modalized disquotationalism MD includes the axioms of the compositional theory KF (Kripke-Feferman) and it is proof-theoretically reducible to it.

It is also natural, however, to read the modal axiom (T) $\forall x(\Box A(x) \to A(x))$ via a truth predicate in the metatheory. But we have a truth predicate in the **object-theory**:

OBSERVATION

The truth predicate of MD in which reflection is replaced by the schema, for $\sigma \in \mathscr{L}_{\mathsf{Tr},\mathsf{F}}$:

$$(\mathsf{GRP}^-) \qquad \Box \ulcorner \sigma \urcorner \to \mathsf{Tr} \ulcorner \sigma \urcorner$$

is **inconsistent**.

Everything works if one moves to a setting in which the truth predicate is **transparent** (Nicolai 2018).

THESIS 2

Careful reconstruction of the assumptions behind the modal status of disquotation strongly suggests shifting to a sub-classical logic. It is not clear whether this is compatible with deflationism.

THE TRUTH THAT CANNOT BE QUANTIFIED

We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. We need it to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent. (Quine, Philosophy of Logic, *p. 12)*

It is not surprising that we should have use for a predicate P with the property that" '__' is P " and " " are always interdeducible. For we frequently find ourselves in a position to assert each sentence in a certain infinite set z (e.g., when all the members of z share a common form); lacking the means to formulate infinite conjunctions, we find it convenient to have a single sentence which is warranted precisely when each member of z is warranted. A predicate P with the property described allows us to construct such a sentence: $(x)(x \in z \rightarrow P(x))$. *Truth is thus a notion that we might reasonably want to have on hand, for expressing semantic ascent and descent, infinite conjunction and disjunction. And given that we want such a notion, it is not difficult to explain how it is that we have been able to invent one. (Leeds,* Theories of Reference and Truth, *p. 121)*

From Gupta's *A Critique of Deflationism*:

(1) _____ & snow is white

(2) (Sky is blue & snow is white) &
(Salzburg is blue & snow is white) & …

(3) ('Sky is blue' is true & snow is white) &
('Salzburg is blue' is true & snow is white) & …

(4) for all σ: σ is true and snow is white

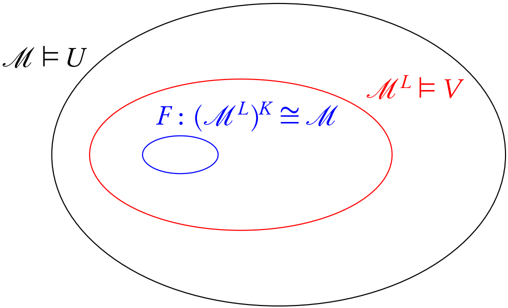
Gupta's main point is that (4) can **express** (2) **via (1)** iff (3) and (2) are **equivalent** in the strong sense of **sameness of sense**.

Synonymy and theoretical equivalence

Visser (2006): *Categories of theories and interpretations*. Let U, V be extensions of EA (Kalmar's Elementary Arithmetic).

RETRACT

U is a **retract** of V iff there are $K\colon U \to V$ and $L\colon V \to U$ and there is a U-definable isomorphism F between $L \circ K$ and id_U.



Synonymy and theoretical equivalence

Visser (2006): *Categories of theories and interpretations*. Let U, V be extensions of EA (Kalmar's Elementary Arithmetic).

RETRACT

U is a **retract** of V iff there are $K\colon U \to V$ and $L\colon V \to U$ and there is a U-definable isomorphism F between $L \circ K$ and id_U.

BI-INTERPRETABILITY

U and V are **biinterpretable** if there are $K\colon U \to V$ and $L\colon V \to U$ s.t. the following both hold:

1. U is a retract of V, witnessed by $F, L \circ K, \mathsf{id}_U$.
2. V is a retract of U, witnessed by $G, K \circ L, \mathsf{id}_V$.

Synonymy and theoretical equivalence

Visser (2006): *Categories of theories and interpretations*. Let U, V be extensions of EA (Kalmar's Elementary Arithmetic).

RETRACT

U is a **retract** of V iff there are $K\colon U \to V$ and $L\colon V \to U$ and there is a U-definable isomorphism F between $L \circ K$ and id_U.

BI-INTERPRETABILITY

U and V are **biinterpretable** if there are $K\colon U \to V$ and $L\colon V \to U$ s.t. the following both hold:

1. U is a retract of V, witnessed by $F, L \circ K, \mathsf{id}_U$.
2. V is a retract of U, witnessed by $G, K \circ L, \mathsf{id}_V$.

Under natural circumstances (identity-preserving interpretations), bi-interpretability entails definitional equivalence.

Let's fix a finite support theory B, and suppose a set S of $\mathscr{L}_{\mathsf{B}}$-sentences is defined by an $\mathscr{L}_B$-formula $F(v)$.

My best way to study the **sameness of sense** between affirming *instances of S* and the claim 'all instances of S are true' is framed in terms of the bi-interpretability of them over B.

Let's fix a finite support theory B, and suppose a set S of $\mathscr{L}_{\mathsf{B}}$-sentences is defined by an $\mathscr{L}_B$-formula $F(v)$.

My best way to study the **sameness of sense** between affirming *instances of S* and the claim 'all instances of S are true' is framed in terms of the bi-interpretability of them over B.

PROPOSITION (Nicolai 2017)

Let B^{Tr} be a *finite* extension of B proving all typed Tr-sentences fpr $\mathscr{L}_B$. Then the theories

- $\mathsf{B} + \{F(\ulcorner \psi \urcorner) \to \psi \mid \psi \in S\}$
- $\mathsf{B}^{\mathsf{Tr}} + \forall x(F(x) \to \mathsf{Tr}\, x)$

can **never be bi-interpretable**, let alone synonymous (definitionally equivalent).

THESIS 3

For truth to serve its expressive role, infinite conjunctions (let alone disjunctions) and corresponding truth-theoretic claims should be equivalent in a precise formal sense. And this is simply not possible.

It might be argued, however, that there is too much emphasis on the *disquotation thesis* and that the focus should really be on the *quantificational nature* of truth, **no matter what truth theoretic device is employed** to fully disclose it.

But in what sense can truth be a quantificational device?

Let us start with (relational version of) the language of second-order arithmetic $\mathscr{L}_2$ and consider PA extended on the one hand with **arithmetical, parameter-free predicative comprehension** (ACA$^{\mathsf{pf}}$):

$$(\mathsf{CA}^{\mathsf{pf}}) \quad \exists X \forall u(u \in X \leftrightarrow \varphi(u)) \quad \text{in } \varphi \text{ no second-order variables },$$

on the other **typed uniform disquotation** (UTB):

$$(\mathsf{UTB}) \qquad \forall x(\mathsf{Tr}\ulcorner A(\dot{x})\urcorner \leftrightarrow A(x)) \qquad \text{for all } A(v) \in \mathscr{L}_{\mathbb{N}}.$$

Induction *is extended* to truth and second-order variables.

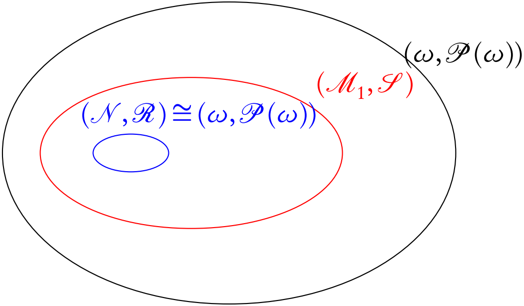
PROPOSITION (Nicolai 2017)

$\mathsf{ACA}^{\mathsf{pf}}$ and UTB are **not** bi-interpretable.

PROPOSITION (Nicolai 2017)

$\mathsf{ACA}^{\mathsf{pf}}$ and UTB are **not** bi-interpretable.

That UTB is a retract of $\mathsf{ACA}^{\mathsf{pf}}$ follows from the standard interpretation. The converse, however, cannot hold:



The conceptual point is that the truth predicate is a ‘bad quantifier’, disallowing the richness of interpretations of the ‘properties’ involved: to restore the bi-interpretability, as in Nicolai (2017), one needs to say

> ‘all properties are definable by a first-order formula of the language of arithmetic’

It seems to me that this is not what we mean by a quantifier, at least in general.

Four Pillars

1. The meaning of ‘is true’ is fixed by the Tarski-biconditionals
   (and we need a nonclassical logic for this)

Four Pillars

1. The meaning of ‘is true’ is fixed by the Tarski-biconditionals **(and we need a nonclassical logic for this)**

2. The truth predicate enables one to express infinite conjunctions and disjunctions **(and we cannot do this if we mean by express a form of synonymy)**

Four Pillars

1. The meaning of 'is true' is fixed by the Tarski-biconditionals **(and we need a nonclassical logic for this)**

2. The truth predicate enables one to express infinite conjunctions and disjunctions **(and we cannot do this if we mean by express a form of synonymy)**

3. The truth predicate is fundamentally a device to perform sentential quantification over pronominal variables **(and a rather restrictive device)**

Four Pillars

1. The meaning of ‘is true’ is fixed by the Tarski-biconditionals **(and we need a nonclassical logic for this)**

2. The truth predicate enables one to express infinite conjunctions and disjunctions **(and we cannot do this if we mean by express a form of synonymy)**

3. The truth predicate is fundamentally a device to perform sentential quantification over pronominal variables **(and a rather restrictive device)**

4. Truth does not contribute substantially to philosophical and scientific explanations **(but it does so meta-theoretically)**