Erkenntnis
https://doi.org/10.1007/s10670-022-00516-y

ORIGINAL RESEARCH

Cognitive Projects and the Trustworthiness of Positive Truth

Matteo Zicchetti[1]

Received: 6 February 2021 / Accepted: 5 January 2022

Abstract
The aim of this paper is twofold: first, I provide a cluster of theories of truth in classical logic that is (internally) consistent with global reflection principles: the theories of positive truth (and falsity). After that, I analyse the *epistemic value* of such theories. I do so employing the framework of cognitive projects introduced by Wright (Proc Aristot Soc 78:167–245, 2004), and employed—in the context of theories of truth—by Fischer et al. (Noûs 2019. https://doi.org/10.1111/nous.12292). In particular, I will argue that theories of positive truth are *trustworthy*, analogously to the theories of full disquotational truth. Moreover, I argue that, for a given cognitive project, *if* the acceptance of trustworthy theories is taken to be an *epistemic norm* of cognitive project, then one has good reasons to accept theories of positive truth over other rival theories of truth in classical logic. On the other hand, the latter theories are deemed epistemically unacceptable.

1 Outline

In their recent paper Fischer et al. (2019), Fischer, Horsten and Nicolai provided a novel epistemological investigation of axiomatic theories of truth. They investigate the epistemic value of theories of full disquotational truth in the context of our epistemic practices; they show that theories of full disquotational truth are to be prefered as theories to be employed in our cognitive projects, in contrast to—what they call—theories of scientific truth in classical logic. Theories of full disquotational truth are to be prefered because they are *trustworthy*, in constrast to many theories of scientific truth in classical logic. Moreover, Fischer et al. support their philosophical analysis and epistemological claim with their technical results provided in Fischer et al. (2017): the trustworthiness of theories (of truth) is spelled out as their consistency and internal consistency with global reflection principles, and it is proved that theories of full

✉ Matteo Zicchetti
matteo.zicchetti@bristol.ac.uk

[1] Department of Philosophy, Bristol, University of Bristol, Bristol, UK

Published online: 25 February 2022

disquotational truth are (internally) consistent if closed under global reflection principles. The aim of this paper is twofold: first, I provide a cluster of theories of truth in classical logic that is (internally) consistent with global reflection principles: the theories of positive truth (and falsity). After that, I analyse the *epistemic value* of such theories. I do so employing the framework of cognitive projects introduced by Wright in (2004), and employed—in the context of theories of truth—by Fischer et al. (2019). In particular, I will argue that theories of positive truth are *trustworthy*, analogously to the theories of full disquotational truth. Moreover, I argue that, for a given cognitive project, *if* the acceptance of trustworthy theories is taken to be an *epistemic norm* of cognitive project, then one has good reasons to accept theories of positive truth over other rival theories of truth in classical logic. On the other hand, the latter theories are deemed epistemically unacceptable.

The paper has the following structure: in Sect. 2, I introduce the relevant notions and conventions about arithmetic, theories of truth, and reflection principles. After that, in Sect. 2.2 I briefly present and discuss the theories of truth in non-classical logic presented by Fischer et al. in Fischer et al. (2017), and their result (Proposition 1) together with their observation that for some natural theories of truth—axiomatised in classical logic—**S**, **S** is internally inconsistent if closed under *global reflection*. I pursue my first aim in Sect. 3: I introduce and investigate the theories of positive truth and falsity, and show that these are consistent and internally consistent if closed under global reflection. Section 4 is devoted to the epistemological investigation of theories of positive truth: after presenting the relevant context of cognitive projects, the so-called authenticity-conditions, and epistemic norms, I introduce the distinction between *full disquotational* and *scientific* truth, and the notion of *trustworthiness*. After that, I argue for my claim that theories of positive are trustworthy. Moreover, I will argue that, if to accept trustworthy theories is an epistemic norm, then one has good reasons to accept theories of positive truth over well-known rival classical theories, which turn out to be epistemically unacceptable. In the remainder of Sect. 4, I will discuss a few worries and problems for the proponent of theories of positive truth, and suggest some possible ways to respond to these worries.[1]

2 Introduction: Notation and Conventions

In this section, I present the notation and conventions adopted in the paper.[2] Here, we focus on **PA** and on theories of truth (and falsity) extending **PA**. We assume $=, \neg, \wedge, \forall$ as primitive logical symbols (and take $\vee, \rightarrow, \leftrightarrow, \exists$ as standardly defined). We call the base language of arithmetic $\mathcal{L}_0$. Terms of $\mathcal{L}_0$ are build in the usual way from variables, the constant 0, and by the application of successor, $+$ and $\times$. For a truth theory over **PA**, its language is called $\mathcal{L}_{\mathsf{T}}$, and expands the arithmetical vocabulary with the addition of a unary truth predicate T. Similarly, a theory of truth and falsity over **PA** is formulated in a language $\mathcal{L}_{\mathsf{TF}}$ expanding the arithmetical vocabulary with additional unary truth and falsity predicates T and F. Here we focus on theories of *type-free* truth

[1] However, a thorough investigation of the possible responses would exceed the scope of the paper.

[2] These will be standard. See Halbach (2014) for more details.

(and falsity), that is, theories where the truth (resp. falsity) predicate can also apply to (codes of) formulas of the language containing both the truth and the falsity predicates. Otherwise, we say that the theory of truth is *typed*.

Since coding works perfectly fine in **PA**, I use common conventions. For an expression e, we say that $\#e$ is the Gödelnumber of e, and $\ulcorner e \urcorner$ is the code of e, i.e., the term in the language $\mathcal{L}_0$ representing $\#e$.

For the language $\mathcal{L}_0$, we have the usual formulas representing syntactic properties: we use $ter_0(x)$ for the set of (Gödelnumbers of) terms of $\mathcal{L}_0$, $ct_0(x)$ for the set of (Gödelnumbers of) closed terms of $\mathcal{L}_0$, $var_0(x)$ for the set of (Gödelnumbers of) variables, $form_0{}^n(x)$ for the set of (Gödelnumbers of) formulas with at most n free distinct variables, $sent_0(x)$ for the set of (Gödelnumbers of) sentences of $\mathcal{L}_0$, where a sentences is a formula with at most 0 free distinct variables. We represent syntactic properties of the language $\mathcal{L}_{\mathsf{T}}$ (resp. $\mathcal{L}_{\mathsf{TF}}$) similarly. So for instance, we use $sent_{\mathsf{TF}}(x)$ for the set of (Gödelnumbers of) sentences of $\mathcal{L}_{\mathsf{TF}}$. Moreover, we use $val(x)$ to represent the evaluation function VAL, which for each Gödelnumber $\#t$ of a closed term t, it returns $t^{\mathbb{N}}$, that is, the value of t (in the standard model). We use variables $s, t, ...$ as ranging over closed terms, and we use the formulation $\forall s...$ as short for $\forall x(ct(x) \rightarrow ...)$. Moreover, we use the following conventions: $\varphi(\dot{x})$ as a shorthand for $sub(\varphi(v), v, num(x))$, informally standing for the result of substituting, in the formula $\varphi(v)$ the x^{th}-numeral for the free variable v. We use $\ulcorner \varphi \dot{x} \urcorner$ as a shorthand for $sub(\ulcorner \varphi v \urcorner, \ulcorner v \urcorner, num(x))$, informally standing for the result of substituting, in the code of the formula $\varphi(v)$, the code of the free variable v with the x^{th}-numeral. Moreover, we employ the dot notation for the representation of the respective syntactic functions, such as $\dot{\neg}$, $\dot{\wedge}$ and $\dot{\forall}$.

We have the usual Σ_1-formula $prov_{PA}(x)$, expressing formal provability in **PA**, reading informally as 'x is provable in **PA**' and is short for '$\exists x(prf_{PA}(x, y))$', where $prf_{PA}(x, y)$ is a Δ_0 formula expressing informally that x is a proof in **PA** of y. When reasoning about provability in **PA**, we always talk about standard provability, that it, we employ a standardly defined provability predicate for which the well-known derivability conditions hold. Moreover, we adopt the following understanding of consistency and internal consistency: for any theory of truth and falsity **S** extending **PA**, we say that **S** is consistent just in case there is no sentence φ such that $\mathbf{S} \vdash \varphi \wedge \neg\varphi$. Moreover, we say that **S** is internally consistent just in case there is no φ such that $\mathbf{S} \vdash \mathsf{T}\ulcorner \varphi \wedge \neg\varphi \urcorner$.[3]

2.1 Reflection Principles

For some first-order theory of truth (and falsity) **S** (containing **PA**) formulated in the expansion $\mathcal{L}_{\mathsf{TF}}$ of the arithmetical language $\mathcal{L}_0$, a proof-theoretic reflection principle for **S** is a 'soundness statement' for **S**, i.e., a statement expressing that everything provable in **S** is true.[4]

[3] This formulation of internal consistency is called for instance 'T-consistency' by Friedman and Sheard (1988). One can formulate the notion of F-consistency analogously. However, in the paper we only focus on T-consistency.

[4] Reflection principles have been extensively investigated. See Turing (1939) and Feferman (1962, 1964) for the first investigations of reflection principles in the context of first-order arithmetical theories. For some

We have the following reflection principles formulated in the language of **S**

$$prov_S\ulcorner 0 = 1\urcorner \rightarrow 0 = 1 \tag{CON_S}$$

$$prov_S\ulcorner \varphi\urcorner \rightarrow \varphi \tag{$LRFN_S$}$$

$$prov_S\ulcorner \varphi\dot{x}\urcorner \rightarrow \varphi(x) \tag{$URFN_S$}$$

CON_S is equivalent to $\neg prov_S\ulcorner 0 = 1\urcorner$ and is called *consistency statement* for **S**. $LRFN_S$ and $URFN_S$ are called *local* and *uniform* reflection principles for **S**.

Of course, since we focus on theories of truth **S** formulated in the language $\mathcal{L}_{\mathsf{TF}}$, we can formulate reflection principles more explicitly, employing the truth predicate T:[5]

$$\forall x(sent_{\mathsf{TF}}(x) \wedge prov_S(x) \rightarrow \mathsf{T}(x)) \tag{GRP_S}$$

Since here I focus on theories of type-free truth (and falsity), I formulate GRP_S unrestrictedly, so that it does not only express that sentences of $\mathcal{L}_0$—provable in **S**—are true, but that sentences containing both occurrences of the truth and falsity predicate—provable in **S**—are true.[6] Unless explicitly specified, when talking about global reflection or using the expression GRP_S, I am going to intend the unrestricted version. Finally, for some theory **S** and some reflection principle R, we use the notation **R[S]** to denote the theory resulting from adding the reflection principle R to **S**.

2.2 Reflection Over Non-classical Truth Theories

The aim of this section is to briefly present the result by Fischer et al. (2017) that theories of full of disquotational and compositional type-free truth formulated in non-classical logic are consistent and internally consistent if closed under global reflection. They reason about the theory called $\mathbf{UTS_0}$, which is a an extension of Elementary arithmetic **EA**, where **EA** is the same as **PA**, with the only distinction that in **EA** the induction schema is formulated for Δ_0 statements.[7]

The theory $\mathbf{UTS_0}$ is formulated in a double-sided sequent calculus over the logic called *Basic De Morgan logic*. In a nutshell, a sequent is an expression of the form $\Gamma \Rightarrow \Delta$, where Γ, Δ are finite sets of formulas. Informally, the formulas preceding the sequent arrow '⇒' are treated as assumptions, and the formulas in the succedent

Footnote 4 continued
more recent philosophical discussion of these principles see for instance Franzen (2004), Cieśliński (2017) and Horsten and Zicchetti (2021).

[5] This principle was originally formulated by Kreisel and Lévy (1968).

[6] Clearly, one could also formulate GRP_S for a typed truth predicate. One example of such formulation would be GRP_S only for provable sentences of $\mathcal{L}_0$.

[7] The fact that they reason about **EA** won't be relevant for my investigation.

are disjunctively joined to form a single conclusion. Basic De Morgan logic is a subsystem of classical logic; it is obtained from classical logic by weakening the usual clauses for negation.[8] The axiom schemata for the truth predicate of $\mathbf{UTS_0}$ are the following unrestricted disquotational principles:

$$\varphi(x) \Rightarrow \mathsf{T}\ulcorner \varphi\dot{x} \urcorner \tag{T1}$$

$$\mathsf{T}\ulcorner \varphi\dot{x} \urcorner \Rightarrow \varphi(x) \tag{T2}$$

Subsequently, they formulate, for any given theory of truth **S** containing $\mathbf{UTS_0}$, uniform and global reflection principles in the following way (the formulation is adapted to the weaker logic):

$$\frac{\Rightarrow prov_S\ulcorner \varphi\dot{x} \urcorner}{\Rightarrow \varphi(x)} \tag{$WRFN_S$}$$

$$\frac{\Rightarrow sent_T(x) \wedge prov_S(x)}{\Rightarrow \mathsf{T}(x)} \tag{$WGRP_S$}$$

Fischer et al. show that any such theory **S** containing $\mathbf{UTS_0}$ is consistent and internally consistent if closed under $WGRP_S$. They do so using the fact that **S** is consistent and internally consistent if closed under $WRFN_S$, together with the following proposition:

Proposition 1 (Fischer et al. 2017, Proposition 1) *Let a theory* **S** *contain* $\mathbf{UTS_0}$. *Then* **WRFN[S]** *and* **WGRP[S]** *are identical theories.*

It is well-known that this desirable connection between uniform and global reflection is lost for many theories of type-free truth axiomatised over classical logic. Fischer et al. correctly claim that there are many natural theories of type-free truth in classical logic, where this connection stated in Proposition 1, is lost:

> There is an intuitive connection between uniform and global reflection: both are intended to express the soundness of the base theory. It turns out, however, that this connection is lost in the classical axiomatizations of Kripke's fixed point construction considered by Horsten and Leigh (2017). For **S** an axiomatization of Kripke's fixed point construction in classical logic, in fact, the result of adding GRP_S to it determines a severe restriction of the class of acceptable models: all consistent fixed points are excluded, i.e., if $(\mathbb{N}, S)$ models **GRP[S]** with S a fixed point, then S is inconsistent. In contrast, **RFN[S]** can have models of the form $(\mathbb{N}, S)$ for S a consistent fixed point (in fact all consistent fixed points). (Fischer et al. 2017, p. 2638)

[8] This logic is presented in detail for instance by Fischer et al. (2017) (Section 2.1, Table 1.).

I agree with Fischer et al. that this connection between uniform and global reflection is natural and desirable. Of course, I also agree with them that many axiomatisations of Kripke's fixed point construction in classical logic are internally inconsistent with global reflection, although consistent with uniform reflection; as Fischer et al. show, the argument for the internal inconsistency of such theories **S** is straightforward: for a liar sentence λ, **S** proves—by classical logic—the following statement: $(\lambda \wedge \neg \mathsf{T}\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge \mathsf{T}\ulcorner\lambda\urcorner)$, and from this, it is straightforward to prove in **GRP[S]** that the liar sentence is both true and untrue.[9]

Their claim is certainly true of the theory of truth **KF**.[10] In **KF**, the connection between uniform and global reflection breaks: **KF** is internally inconsistent with GRP_{KF}, and nevertheless consistent and internally consistent with RFN_{KF}. However, the question of whether this intimate connection between uniform and global reflection is preserved for the theories of type-free positive truth and falsity investigated by Horsten and Leigh (2017) and by Leigh (2016) is still open.[11]

In the next section, I focus precisely on investigating this question. In particular, I investigate whether the theories of positive truth and falsity considered by Horsten and Leigh (2017) and by Leigh (2016) are consistent and internally consistent with global reflection. If these theories were to be consistent and internally consistent with global reflection, they would provide an example of axiomatisations of Kripke's fixed point construction over classical logic, for which the deep connection between uniform and global reflection remains intact.

3 Reflection Over Classical Positive Truth and Falsity

The aim of this section is to prove that the theories of positive truth and falsity introduced in Leigh (2016) and Horsten and Leigh (2017) are consistent and internally consistent if closed under global reflection. In Sect. 2.1, I introduce the theories and their models, and in Sect. 2.2, I argue for my claim that these theories are consistent and internally consistent with global reflection. I do so in two steps. First, I show that these theories are consistent and internally consistent if closed under the rules of *Necessitation* and *Conecessitation* (Propositions 5 and 6). After that, I show that standard models of these theories—closed under Necessitation and Conecessitation—are models of global reflection (Theorem 1).

[9] Although this is folklore, Fischer et al. show this in (Fischer et al. 2017, Footnote 13, p. 2638).

[10] This theory has been presented for instance in Feferman (1991) and in Cantini (1989) and has been extensively studied.

[11] It is important to point out that the quote mentioned is slightly misleading; Fischer et al. *prima facie* claim that the theories investigated by Horsten and Leigh are internally inconsistent with global reflection. However, in their proof of the internal inconsistency in their (Fischer et al. 2017, Footnote 13, p. 2638), they employ truth-theoretic principles, which *are not available* in the theories of positive truth and falsity investigated by Horsten and Leigh. So, from a charitable reading of their quote, they cannot possibly mean that the theories of positive truth and falsity are internally inconsistent with global reflection. Nevertheless, this leaves the question of whether these theories are internally consistent with global reflection open. I am grateful to an anonymous referee of this journal to point this out to me.

3.1 Theories of Typefree Positive Truth and Falsity

The theory of positive truth and falsity biconditionals **TFB** extends **PA**, and expands the language $\mathcal{L}_0$ of **PA** to the language $\mathcal{L}_{\mathsf{TF}}$ with fresh truth and falsity predicates T and F. For any given φ in $\mathcal{L}_{\mathsf{TF}}$ we denote the dual of φ by $\overline{\varphi}$. The duals are introduced by recursion:[12]

$$\begin{aligned} \overline{\varphi} &= \neg\varphi (\text{for } \varphi \text{ in } \mathcal{L}_0 \text{ and atomic}) \quad \overline{\neg\varphi} = \varphi \\ \overline{\varphi \wedge \psi} &= \overline{\varphi} \vee \overline{\psi} \quad \overline{\varphi \vee \psi} = \overline{\varphi} \wedge \overline{\psi} \\ \overline{\forall x \varphi} &= \exists x \overline{\varphi} \quad \overline{\exists x \varphi} = \forall x \overline{\varphi} \\ \overline{\mathsf{T}s} &= \mathsf{F}s \text{ and } \overline{\mathsf{F}s} = \mathsf{T}s \end{aligned}$$

The language $\mathcal{L}^+_{\mathsf{TF}}$ is the *strictly positive* sub-language of $\mathcal{L}_{\mathsf{TF}}$. For any φ, we say that φ is in $\mathcal{L}^+_{\mathsf{TF}}$ to say that φ is strictly positive, i.e., that any occurrence of the truth and falsity predicates T and F in φ are under the scope of no negation symbols. We say that φ is negative otherwise. We denote with $\mathrm{sent}^+_{\mathsf{TF}}$ the set of (Gödelnumbers of) strictly positive sentences in the language $\mathcal{L}_{\mathsf{TF}}$.

The theory **TFB** extends **PA** with the following axiom schemata:

$$\mathsf{T}\ulcorner\varphi\urcorner \leftrightarrow \varphi \tag{TFB1}$$

$$\mathsf{F}\ulcorner\varphi\urcorner \leftrightarrow \overline{\varphi}, \tag{TFB2}$$

for all sentences φ in $\mathcal{L}^+_{\mathsf{TF}}$. That is, we restrict the T-biconditionals and F-biconditionals to strictly positive sentences. **TFB** is a theory of *local* positive disquotational truth and falsity, because its biconditionals are formulated for sentences and not for open formulas. On the other hand, the theory of *uniform* positive disquotational truth and falsity, **UTFB**, extends **PA** with axiom schemata similar to the one of **TFB**, although formulated for open formulas:

$$\mathsf{T}\ulcorner\varphi\dot{x}\urcorner \leftrightarrow \varphi(x) \tag{UTFB3}$$

$$\mathsf{F}\ulcorner\varphi\dot{x}\urcorner \leftrightarrow \overline{\varphi(x)}, \tag{UTFB4}$$

for open formulas $\varphi(x)$ in $\mathcal{L}^+_{\mathsf{TF}}$.

$\mathbf{KF_{pos}}$ is the theory of positive compositional truth and falsity and extends **PA** with the following axioms:

$$\forall s \forall t ((\mathsf{T}(s \underset{\cdot}{=} t) \leftrightarrow val(s) = val(t)) \wedge (\mathsf{T} \underset{\cdot}{\neg} (s \underset{\cdot}{=} t) \leftrightarrow \neg(val(s) = val(t)))) \tag{KF1}$$

[12] This is Leigh's definition in Leigh (2016), p. 576.

$$\forall s \forall t ((\mathsf{F}(s \underset{\cdot}{=} t) \leftrightarrow \neg(val(s) = val(t))) \land (\ F \underset{\cdot}{\neg} (s \underset{\cdot}{=} t) \leftrightarrow val(s) = val(t))) \tag{KF2}$$

$$\forall x \forall y (sent^{+}_{\mathsf{TF}}(x \underset{\cdot}{\land} y) \rightarrow (\mathsf{T}(x \underset{\cdot}{\land} y) \leftrightarrow \mathsf{T}x \land \mathsf{T}y)) \tag{KF3}$$

$$\forall x \forall y (sent^{+}_{\mathsf{TF}}(x \underset{\cdot}{\lor} y) \rightarrow (\mathsf{T}(x \underset{\cdot}{\lor} y) \leftrightarrow \mathsf{T}x \lor \mathsf{T}y)) \tag{KF4}$$

$$\forall x \forall y (sent^{+}_{\mathsf{TF}}(x \underset{\cdot}{\land} y) \rightarrow (\mathsf{F}(x \underset{\cdot}{\land} y) \leftrightarrow \mathsf{F}x \,\overline{\land}\, \mathsf{F}y)) \tag{KF5}$$

$$\forall x \forall y (sent^{+}_{\mathsf{TF}}(x \underset{\cdot}{\lor} y) \rightarrow (\mathsf{F}(x \underset{\cdot}{\lor} y) \leftrightarrow \mathsf{F}x \,\overline{\lor}\, \mathsf{F}y)) \tag{KF6}$$

$$\forall x \forall y (form(x) \land var(y) \land (sent^{+}_{\mathsf{TF}}(\underset{\cdot}{\forall} yx)) \rightarrow (\mathsf{T} \underset{\cdot}{\forall} yx \leftrightarrow \forall z \mathsf{T}(x\dot{z}))) \tag{KF7}$$

$$\forall x \forall y (form(x) \land var(y) \land (sent^{+}_{\mathsf{TF}}(\underset{\cdot}{\forall} yx)) \rightarrow (\mathsf{F} \underset{\cdot}{\forall} yx \leftrightarrow \overline{\forall} z \mathsf{F}(x\dot{z}))) \tag{KF8}$$

$$\forall x \forall y (form(x) \land var(y) \land (sent^{+}_{\mathsf{TF}}(\underset{\cdot}{\exists} yx)) \rightarrow (\mathsf{T} \underset{\cdot}{\exists} yx \leftrightarrow \exists z \mathsf{T}(x\dot{z}))) \tag{KF9}$$

$$\forall x \forall y (form(x) \land var(y) \land (sent^{+}_{\mathsf{TF}}(\underset{\cdot}{\exists} yx)) \rightarrow (\mathsf{F} \underset{\cdot}{\exists} yx \leftrightarrow \overline{\exists} z \mathsf{F}(x\dot{z}))) \tag{KF10}$$

$$\forall x (\mathsf{T}\ulcorner \mathsf{T}\dot{x} \urcorner \leftrightarrow \mathsf{T}(x) \land \mathsf{T}\ulcorner \mathsf{F}\dot{x} \urcorner \leftrightarrow \mathsf{F}(x)) \tag{KF11}$$

$$\forall x (\mathsf{F}\ulcorner \mathsf{T}\dot{x} \urcorner \leftrightarrow \mathsf{F}(x) \land \mathsf{F}\ulcorner \mathsf{F}\dot{x} \urcorner \leftrightarrow \mathsf{T}(x)), \tag{KF12}$$

where $\overline{\land} = \lor$, $\overline{\lor} = \land$, $\overline{\forall} = \exists$ and $\overline{\exists} = \forall$.

For the aim of this paper, we focus on standard models of these theories, i.e., models of these theories expanding the class of standard models of arithmetic, $\mathbb{N}$. We call models of theories of positive truth simply $\mathcal{L}^{+}_{\mathsf{TF}}$-structures. An $\mathcal{L}^{+}_{\mathsf{TF}}$-structure $\mathfrak{M} = (\mathbb{N}, S_1, S_2)$ is an expansion of $\mathbb{N}$ with a set S_1, interpreted as the extension of the truth predicate, and a set S_2, interpreted as the extension of the falsity predicate. We want to obtain the extensions of the truth and falsity predicates by starting from two sets S_1 and S_2 with the iteration of a positive inductive operation on the pair (S_1, S_2), denoted by $\Gamma(S_1, S_2) = [\Gamma^{+}(S_1, S_2), \Gamma^{-}(S_1, S_2)]$, such that

$$\Gamma^{+}(S_1, S_2) = \{\#\varphi | \varphi \in \text{sent}^{+}_{\mathsf{TF}} \text{ and } (\mathbb{N}, S_1, S_2) \models \varphi\} \cup \{\#\varphi | \varphi \notin \text{sent}^{+}_{\mathsf{TF}} \text{ and } \varphi \in S_1\}$$

$$\Gamma^-(S_1, S_2) = \{\#\varphi | \varphi \in \mathrm{sent}^+_{\mathsf{TF}} \text{ and } (\mathbb{N}, S_1, S_2) \models \overline{\varphi}\} \cup \{\#\varphi | \varphi \notin \mathrm{sent}^+_{\mathsf{TF}} \text{ and } \varphi \in S_2\}.$$

If one starts with the first expansion of $\mathbb{N}$ being $(\mathbb{N}, \emptyset, \emptyset)$, i.e., with S_1 and S_2 being empty, then $\Gamma^+(S_1, S_2)$ and $\Gamma^-(S_1, S_2)$] are the following:

$$\Gamma^+(S_1, S_2) = \{\#\varphi | \varphi \in \mathrm{sent}^+_{\mathsf{TF}} \text{ and } (\mathbb{N}, S_1, S_2) \models \varphi\}$$
$$\Gamma^-(S_1, S_2) = \{\#\varphi | \varphi \in \mathrm{sent}^+_{\mathsf{TF}} \text{ and } (\mathbb{N}, S_1, S_2) \models \overline{\varphi}\}.$$

In order for S_1 and S_2 to be possible candidates for our extensions of the truth and falsity predicates, we need first of all the operation $\Gamma(S_1, S_2)$ to reach fixed points, that is, we want it to reach a point where $\Gamma(S_1, S_2) = (S_1, S_2)$. In order to show that Γ reaches fixed points, it is sufficient to show that Γ is monotone, i.e., we need to show that, if the pair (S_1', S_2') extends the pair (S_1, S_2), then $\Gamma(S_1', S_2')$ extends $\Gamma(S_1, S_2)$. We need to show that if $(S_1, S_2) \leq (S_1', S_2')$, then $\Gamma(S_1, S_2) \leq \Gamma(S_1', S_2')$.[13]

Proposition 2 Γ *is monotone.*

Monotonicity follows simply by the fact that Γ is a positive inductive operation. For any positive statement φ, one can easily show that by the definition of $\leq$ and the definition of Γ, if the code of φ is in $\Gamma^+(S_1, S_2)$, then the code of φ is $\Gamma^+(S_1', S_2')$.[14] One treats positive statements in $\Gamma^-(S_1, S_2)$ analogously. Negative statements are trivially taken care of by the definition of Γ.

For our investigation, we still need to show that fixed points of Γ are precisely the models of positive truth and falsity. First of all, I show that the fixed points of Γ are precisely the models of the theories **TFB** and **UTFB**.

Proposition 3 *Assume that* $S_1,\ S_2 \subseteq \omega$. *Then the* $\mathcal{L}^+_{\mathsf{TF}}$*-structure* $(\mathbb{N},\ S_1,\ S_2)$ *is a model of* **TFB** (*and also of* **UTFB**) *if and only if* $\Gamma(S_1, S_2) = (S_1, S_2)$.[15]

Proof (Sketch) For the left-to-right direction we assume that $(\mathbb{N},\ S_1,\ S_2) \models$ **TFB**. To show that $\#\varphi \in (S_1, S_2)$ if and only if $\#\varphi \in \Gamma(S_1, S_2)$ we have two cases to take care of: (i) φ is positive; (ii) φ is not positive.

(i) If φ is positive, we have the following equivalences:

1. $\#\varphi \in S_1$
   if anf only if $(\mathbb{N}, S_1, S_2) \models T\ulcorner\varphi\urcorner$
   if and only if $(\mathbb{N}, S_1, S_2) \models \varphi$ (by the assumption that $(\mathbb{N}, S_1, S_2) \models$ **TFB**)
   if and only if $\#\varphi \in \Gamma^+(S_1, S_2)$ (by the definition of Γ^+ and by the fact that φ is positive and that $(\mathbb{N}, S_1, S_2) \models \varphi$).
2. $\#\varphi \in S_2$
   if and only if $(\mathbb{N}, S_1, S_2) \models \mathsf{T}\ulcorner\varphi\urcorner$
   if and only if $(\mathbb{N}, S_1, S_2) \models \overline{\varphi}$ (by the assumption that $(\mathbb{N}, S_1, S_2) \models$ **TFB**)
   if and only if $\#\varphi \in \Gamma^-(S_1, S_2)$ (by the definition of Γ^- and by the fact that φ is a positive and that $(\mathbb{N}, S_1, S_2) \models \overline{\varphi}$).

[13] For clarity, for any to sets A, B we understand $(A, B) \leq (A', B')$ as $A \subseteq A'$ and $B \subseteq B'$.

[14] This is essentially Halbach's proof in (Halbach 2014, Lemma 19.13.).

[15] Here I am following Halbach's proof (Halbach 2014, Theorem 19.15) and adapting it to the context with duals and the falsity predicate.

(ii) If φ is not positive, then the claim that $\#\varphi \in (S_1, S_2)$ if and only if $\#\varphi \in \Gamma(S_1, S_2)$ follows trivially from the definition of Γ.

For the right-to-left direction we assume that (S_1, S_2) is a fixed point of Γ and reason about some arbitrary φ in $\mathcal{L}^{+}_{\mathsf{TF}}$. We have the following equivalences:

3. $(\mathbb{N}, S_1, S_2) \models \mathsf{T}\ulcorner\varphi\urcorner$
   if and only if $\#\varphi \in S_1$
   if and only if $\#\varphi \in \Gamma^{+}(S_1, S_2)$ (by the assumption that $\Gamma(S_1, S_2) = (S_1, S_2)$)
   if and only if $(\mathbb{N}, S_1, S_2) \models \varphi$ (by the definition of Γ^{+})
4. $(\mathbb{N}, S_1, S_2) \models \mathsf{F}\ulcorner\varphi\urcorner$
   if and only if $\#\varphi \in S_2$
   if and only if $\#\varphi \in \Gamma^{-}(S_1, S_2)$ (by the assumption that $\Gamma(S_1, S_2) = (S_1, S_2)$)
   if and only if $(\mathbb{N}, S_1, S_2) \models \overline{\varphi}$ (by the definition of Γ^{-}.)

Therefore, we conclude that the structures $(\mathbb{N}, S_1, S_2)$ verify the local disquotation axioms of **TFB** for all positive sentences. Moreover, as these structures $(\mathbb{N}, S_1, S_2)$ are standard, they also satisfy the axioms schemata of **UTFB**

$$\forall s_1 ... \forall s_n (\mathsf{T}\ulcorner \varphi s_1, ..., s_n \urcorner \leftrightarrow (\varphi(s_1, ..., s_n))$$
$$\forall s_1 ... \forall s_n (\mathsf{F}\ulcorner \varphi s_1, ..., s_n \urcorner \leftrightarrow \overline{(\varphi(s_1, ..., s_n)})$$

for all positive positive formulas $\varphi(x_1, ..., x_n)$. □

Now, we also need to show that these $\mathcal{L}^{+}_{\mathsf{TF}}$-structures of **TFB** and **UTFB** are also models of $\mathbf{KF_{pos}}$.

Proposition 4 [16] *Assume that* $S_1,\ S_2 \subseteq \omega$. *Then the following are equivalent:*

1. $(\mathbb{N},\ S_1,\ S_2)$ *is a model of* **UTFB**
2. $\Gamma(S_1, S_2) = (S_1, S_2)$
3. $(\mathbb{N},\ S_1,\ S_2)$ *is a model of* $\mathbf{KF_{pos}}$

Proof (Sketch) The equivalence between 1. and 2. is simply Proposition 3. We also have that 3. implies 1. by the fact that **UTFB** is a sub-theory of $\mathbf{KF_{pos}}$.[17] So we only need to show that 2. implies 3. To show this, the idea is to follow the strategy adopted for Proposition 3. We assume that $\Gamma(S_1, S_2) = (S_1, S_2)$ and reason about some positive φ. The axioms KF1, KF2, KF11, KF12 are instances of the biconditionals of **UTFB**, so we don't need to consider them. For the axioms KF3–KF6, we argue informally that these axioms are, when considered schematically, instances of the biconditionals of **UTFB** and therefore each instance of these axioms is also satisfied by the $\mathcal{L}^{+}_{\mathsf{TF}}$-structures $(\mathbb{N}, S_1, S_2)$ by the equivalence with 1. The quantified versions of the axioms is satisfied by induction on the complexities of φ, ψ. For the axioms KF7–KF10, we take KF7 as an example: for some positive $\underset{\cdot}{\forall} v\varphi$ we assume that $(\mathbb{N}, S_1, S_2) \models \mathsf{T}\ulcorner \underset{\cdot}{\forall} v\varphi\urcorner$ and see that we have the following equivalences: $(\mathbb{N},\ S_1,\ S_2) \models \mathsf{T}\ulcorner \underset{\cdot}{\forall} v\varphi\urcorner$, if and

[16] This is proposition 1.2 in (Leigh 2016, proposition 1.2).

[17] This is for instance a lemma in (Leigh 2016, lemma 5.2). A similar result has been shown in (Cantini 1989, lemma 3.2 (ii)), for the versions of disquotational and compositional truth without the falsity predicate.

only if $(\mathbb{N}, S_1, S_2) \models \forall x\varphi(x)$ (because $\underset{\cdot}{\forall} v\varphi$ is a positive sentence and we have the biconditionals of **UTFB**), if and only if for all n, $(\mathbb{N}, S_1, S_2) \models \varphi(n)$, if and only if $(\mathbb{N}, S_1, S_2) \models \mathsf{T}\ulcorner\varphi n\urcorner$. One reasons analogously about the axioms KF8–KF10. □

In the next section, I will reason about the theory $\mathbf{KF_{pos}}$ and show that (1) $\mathbf{KF_{pos}}$ is consistent and internally consistent if closed under the rules of *Necessitation* and *Conecessitation* for the truth predicate; (2) $\mathbf{KF^*_{pos}}$, i.e., the theory resulting by closing $\mathbf{KF_{pos}}$ under *Necessitation* and *Conecessitation* for the truth predicate, is consistent and internally consistent if closed under global reflection.

3.2 Positive Truth and Falsity is (Internally) Consistent with Global Reflection

The main aim of this section is to show that $\mathbf{KF_{pos}}$—and therefore also **TFB** and **UTFB**—is consistent and internally consistent, if closed under global reflection. In order to do so, I first extend $\mathbf{KF_{pos}}$ to the theory $\mathbf{KF^*_{pos}}$, by closing $\mathbf{KF_{pos}}$ under the following rules for the truth predicate T:[18]

$$\frac{\varphi}{\mathsf{T}\ulcorner\varphi\urcorner}\,\text{NEC};\quad \frac{\mathsf{T}\ulcorner\varphi\urcorner}{\varphi}\,\text{CONEC}$$

One might naturally ask why I do not investigate analogous rules for the falsity predicate F. I am not doing so, because my aim is to show that $\mathbf{KF_{pos}}$ is consistent and internally consistent with global reflection, and the closure under NEC and CONEC is here only technically useful. To investigate rules for the falsity predicate is not necessary for this purpose.

Informally, my aim is to show that $\mathbf{KF^*_{pos}}$ has standard models, and that $\mathbf{KF^*_{pos}}$ remains internally consistent.[19] More precisely, I prove the following propositions:

Proposition 5 *There are standard models of* $\mathbf{KF^{nec}_{pos}}$, *where* $\mathbf{KF^{nec}_{pos}}$ *is* $\mathbf{KF_{pos}}$ *together with* NEC.

Proposition 6 *Any application of* CONEC *in* $\mathbf{KF^*_{pos}}$ *is admissible in the theory. In other words,* $\mathbf{KF^*_{pos}}$ *proves the same theorems as* $\mathbf{KF^{nec}_{pos}}$.

Proof of Proposition 5 We want to construct a standard model $\mathfrak{M}^*$ of $\mathbf{KF^{nec}_{pos}}$. The model $\mathfrak{M}^*$ is supposed to be the model that we get by closing $(\mathbb{N}, S_1, S_2)$ under Γ, starting with $S_2 = \emptyset$ and with S_1 be the following set A of (codes of) non-positive statements. We define A as the set of codes of non-positive sentences provable in $\mathbf{KF^{nec}_{pos}}$:

$$A := \{\#\varphi | \varphi \notin sent^{+}_{\mathsf{TF}} \text{ and } \mathbf{KF^{nec}_{pos}} \vdash \varphi\}$$

[18] These rules are not allowed in proofs from premises. These rules should be then understood as closure conditions on theories. Looking at NEC for instance, one understands the rule in the following manner: if $\mathbf{KF^*_{pos}}$ proves φ, then it also proves $\mathsf{T}\ulcorner\varphi\urcorner$.

[19] In doing so, I follow roughly Halbach's idea in proving that closing the theory of positive truth **PUTB** under NEC and CONEC results in a consistent theory. However, my proof strategy is slightly different than the proof in (Halbach 2014, Theorem 19.21, p. 271).

We know from Proposition 2 that closing $(\mathbb{N}, A, S_2)$ under Γ reaches fixed points. From Propositions 3 and 4 we have that $\mathfrak{M}^* \models \mathbf{KF_{pos}}$.

Now we show that NEC is valid in $\mathfrak{M}^*$. We want to show the following:

(i) If $\mathbf{KF^{nec}_{pos}} \vdash \varphi$, then $\mathfrak{M}^* \models \varphi$ and $\mathfrak{M}^* \models \mathsf{T}\ulcorner\varphi\urcorner$.

We only focus on applications of NEC to φ that are not positive, because for any φ in $sent^+_{\mathsf{TF}}$ NEC is derivable from the biconditionals for the truth predicate.[20]

We reason by standard induction on the number of application of NEC. We reason about some derivation in $\mathbf{KF^{nec}_{pos}}$, and we let some application of NEC to a non-positive sentence be given. We focus on a sub-proof Q of this derivation, such that Q ends with an application of NEC

$$\frac{\varphi}{\mathsf{T}\ulcorner\varphi\urcorner}.$$

If the above application of NEC is the first application of NEC, then we can conclude that everything up to and including φ is provable in $\mathbf{KF_{pos}}$.

By the fact that $\mathfrak{M}^*$ is a model of $\mathbf{KF_{pos}}$ we have that $\mathfrak{M}^* \models \varphi$. Moreover, since we assumed that φ is not positive and provable in $\mathbf{KF^{nec}_{pos}}$,[21] we conclude by definition of A that the code of φ is in A. By the fact that the pair (A, S_2) is a fixed point of Γ we can conclude that $\mathfrak{M}^* \models \mathsf{T}\ulcorner\varphi\urcorner$.

Now, assume that in $\mathbf{KF^{nec}_{pos}}$, n applications of NEC are satisfied in $\mathfrak{M}^*$. We reason about some sub-derivation Q' ending with the $n + 1$ application of NEC:

$$\frac{\varphi}{\mathsf{T}\ulcorner\varphi\urcorner}$$

By our induction hypothesis we have that $\mathfrak{M}^* \models \varphi$. Moreover, by the fact that φ is non-positive by assumption and provable in $\mathbf{KF^{nec}_{pos}}$, we can reason analogously to the case of the first application of NEC and argue that $\mathfrak{M}^* \models \mathsf{T}\ulcorner\varphi\urcorner$.

Therefore, we can conclude that the chosen model $\mathfrak{M}^*$ satisfies NEC.

Proof of Proposition 6 We want to show that any application of CONEC is admissible, i.e., that for any φ proved in $\mathbf{KF^*_{pos}}$ with any number of application of CONEC, φ is also provable in $\mathbf{KF^{nec}_{pos}}$. We do so by induction of the number of applications of CONEC. We reason about some arbitrary derivation R in $\mathbf{KF^*_{pos}}$, and let some applications of CONEC be given. Similarly to previous case, we also focus on applications of CONEC to non-positive sentences, since CONEC is derivable from the biconditionals for the truth predicate for positive sentences. We focus on some sub-derivation P of R in $\mathbf{KF^*_{pos}}$ ending with an application of CONEC

$$\frac{\mathsf{T}\ulcorner\varphi\urcorner}{\varphi}$$

[20] For this one employs the fact, mentioned ealier in the sketch of Proposition 4 that **UTFB** is a sub-theory of $\mathbf{KF_{pos}}$.

[21] Trivially, since φ is provable in $\mathbf{KF_{pos}}$ and $\mathbf{KF^{nec}_{pos}}$ extends $\mathbf{KF_{pos}}$.

If the above application of CONEC is the first application of CONEC in R, then we can conclude that everything up to and including $\mathsf{T}\ulcorner\varphi\urcorner$ is provable in $\mathbf{KF}^{\mathbf{nec}}_{\mathbf{pos}}$. Therefore, we have by Proposition 5 that $\mathfrak{M}^* \models \mathsf{T}\ulcorner\varphi\urcorner$. From this we conclude that the code of φ is in $\Gamma(A, S_2)$, and by the fact that φ is not positive by assumption we conclude that the code of φ has to be in A. By the definition of A we conclude that $\mathbf{KF}^{\mathbf{nec}}_{\mathbf{pos}} \vdash \varphi$. Therefore, there is a derivation in $\mathbf{KF}^{\mathbf{nec}}_{\mathbf{pos}}$ such that φ is provable without the application of CONEC in the sub-derivation P of R.

Now, we assume that in $\mathbf{KF}^{*}_{\mathbf{pos}}$, n applications of CONEC are admissible. We reason about some sub-derivation of P' ending with the $n+1$ application of CONEC:

$$\frac{\mathsf{T}\ulcorner\varphi\urcorner}{\varphi}$$

By our induction hypothesis we have that $\mathsf{T}\ulcorner\varphi\urcorner$ is provable in $\mathbf{KF}^{\mathbf{nec}}_{\mathbf{pos}}$, and by Proposition 5 we have that $\mathfrak{M}^* \models \mathsf{T}\ulcorner\varphi\urcorner$. Therefore, the code of φ is in $\Gamma(A, S_2)$ and by the assumption that φ is not positive, we conclude that the code of φ has to be in A. By the definition of A we conclude that φ is provable in $\mathbf{KF}^{\mathbf{nec}}_{\mathbf{pos}}$, so the $n+1$ application of CONEC is also admissible. □

From Propositions 5 and 6 , we can observe that we have standard models of $\mathbf{KF}^{\mathbf{nec}}_{\mathbf{pos}}$, which are also models of $\mathbf{KF}^{*}_{\mathbf{pos}}$. Morever, we can see that $\mathbf{KF}^{*}_{\mathbf{pos}}$ is consistent and internally consistent; it is consistent because it has models. The internal consistency follows from the fact that $\mathbf{KF}^{*}_{\mathbf{pos}}$ is closed under CONEC; if $\mathbf{KF}^{*}_{\mathbf{pos}}$ were to be internally inconsistent, then we would have a sentence φ, such that $\mathbf{KF}^{*}_{\mathbf{pos}}$ proves $\mathsf{T}\ulcorner\varphi \wedge \neg\varphi\urcorner$. The closure under CONEC would imply that $\mathbf{KF}^{*}_{\mathbf{pos}}$ proves $\varphi \wedge \neg\varphi$. However, this contradicts the fact that $\mathbf{KF}^{*}_{\mathbf{pos}}$ has models. From Propositions 5 and 6 it is almost straightforward to prove that $\mathbf{KF}^{*}_{\mathbf{pos}}$ is consistent and internally consistent, if closed under global reflection. We formulate global reflection unrestrictedly:

$$\forall x(sent_{\mathsf{TF}}(x) \wedge prov_{KF^*_{pos}}(x) \rightarrow \mathsf{T}(x)) \qquad (GRP_{KF^*_{pos}})$$

Theorem 1 $\mathfrak{M}^* \models \mathbf{GRP}[\mathbf{KF}^{*}_{\mathbf{pos}}]$.

Proof We reason about $\mathfrak{M}^*$, take some sentence φ, such that $\mathfrak{M}^* \models prov_{KF^*_{pos}}\ulcorner\varphi\urcorner$ and reason in the following manner: since $prov_{KF^*_{pos}}\ulcorner\varphi\urcorner$ is true in $\mathfrak{M}^*$ and is an arithmetical sentence, we can conclude that $prov_{KF^*_{pos}}\ulcorner\varphi\urcorner$ is true in $\mathbb{N}$. This gives us—by the meaning of the provability predicate—that $\mathbf{KF}^{*}_{\mathbf{pos}} \vdash \varphi$. By the fact that $\mathbf{KF}^{*}_{\mathbf{pos}}$ is closed under NEC we conclude that $\mathbf{KF}^{*}_{\mathbf{pos}} \vdash \mathsf{T}\ulcorner\varphi\urcorner$. By the fact that NEC is satisfied in $\mathfrak{M}^*$ we have that $\mathfrak{M}^* \models \mathsf{T}\ulcorner\varphi\urcorner$. That is, $\mathfrak{M}^*$ is a model of $\mathbf{GRP}[\mathbf{KF}^{*}_{\mathbf{pos}}]$. □

Moreover, we have the following Corollary:

Corollary 1 $\mathbf{GRP}[\mathbf{KF}^{*}_{\mathbf{pos}}]$ *is consistent and internally consistent.*

4 Philosophical Discussion

The aim of this section is to investigate the philosophical significance of the results presented in Sect. 3. In Sect. 4.1, I introduce the context of *cognitive projects*[22] with their *authenticity-conditions* and *norms*, that I will employ for my epistemological analysis.[23] I present the distinction between theories of *scientific truth* and of *full disquotational truth*, made in Fischer et al. (2019). After that, I introduce the notion of *trustworthiness* and argue for my claim that the theories of positive truth and falsity are indeed trustworthy theories to be employed in our cognitive projects.

In the remainder of the section, I am going to present some possible worries and problems for the proponents of theories of positive truth. I will address them, and suggest possible replies to those worries available—as I will argue—to the proponent of positive truth.[24]

4.1 Cognitive Projects, Authenticity-Conditions and Norms

Wright introduces the notion of cognitive project in Wright (2004). In a nutshell, a cognitive project is defined by a pair: a question, and a procedure one might competently execute in order to answer it. Basically, any cognitive enquiry can be seen as some cognitive project. We can think of very small-scale cognitive enquiries, such as the following:

$$\langle \text{ What is the weather like outside?, Sense perception } \rangle$$

In this enquiry, the subject wants to figure out what the weather is like and as procedure to answer the question in her enquiry the subject has sense perception. So for instance, she might just look outside and answer the question of her enquiry. There is a plethora of cognitive projects: there are enquiries in the empirical sciences, in philosophy, and also in mathematics.[25]

An *authenticity-condition*[26] is—relative to a cognitive project—any proposition doubt about which would rationally require doubt about the efficacy or the significance of the cognitive project. Looking at our previous example of cognitive enquiry, an authenticity-condition of that enquiry is for instance the proposition that sense perception is reliable. We can see that it is so by looking at the subject's procedures to answer the question of her enquiry; if the subject would rationally doubt that sense perception is reliable, she would doubt the significance of her project. This is so because—by

[22] This has been discussed by Wright and in mainstream epistemology in many places. See for instance Wright (2004, 2012).

[23] Therefore, my claims about the epistemic value of theories of positive truth depend on the framework that I use, which of course can be questioned. However, here I am taking this framework for granted.

[24] However, I will only be able to sketch the responses to those worries.

[25] The fact that cognitive enquiries are ubiquitous is fairly uncontroversial. For the recognition of cognitive projects in mathematics see for instance Galinon (2014), Fischer et al. (2019), Horsten (2021), Wright (2016), Pedersen (2016, 2021).

[26] This is Wright's terminology in Wright (2012). Authenticity-conditions are called cornerstone propositions in Wright (2004).

the definition of this cognitive project—sense perception is the chosen procedure to be competenty executed in that particular enquiry. In general, authenticity-conditions include propositions expressing the normal and proper functioning of relevant cognitive faculties, the reliability of instruments utilised, the correctness of relevant theory, the soundness of relevant principles of inference, and so on. The authenticity-condition of any cognitive enquiry into the world involving sense perception is the proposition that, quite trivially, that there is an external world. An authenticity-condition of cognitive enquiries about an arithmetical subject matter, employing some theory **S**, is that **S** is non-trivial. The idea behind non-triviality being an authenticity condition is that, if a theory is trivial, then the theory is not reliable as a source of warrants in our enquiry.[27] For any rational agent, who employs some theory **S**, the soundness of **S**, and the good standing of the concepts of **S**, are authenticity-conditions.[28] With Wright—and with Fischer et al.—I also hold that a necessary condition for concepts to be in good standing is that they are consistent.

Like any epistemic practice, cognitive projects have aims and goals, which are pursued by the agents engaging in them. For a given epistemic practice, we let an *epistemic norm* be any proposition that regulates the practice. Roughly, epistemic norms are rules of the epistemic practice: an epistemic norm could be for instance 'One should not believe a proposition p against compelling evidence that p is false'. Similarly, in an epistemic practice investigating some mathematical subject matter, an epistemic norm could be for instance 'One should not believe truly inconsistent theories'.[29] As we can see, epistemic norms have normative force; they regulate how individuals and communities *should* pursuit their aims in their epistemic practice.

In an epistemic practice X with epistemic norms $R_1, \ldots, R_n$ (relative to X), agents engaging with X are bound to follow the epistemic norms; if a rational agent engages with X, then she is *epistemically obligated* to act in accordance with the norms. Of course, when I say that an agent has an epistemic obligation, I do not suggest that she cannot act otherwise. What I mean with 'an agent is epistemically obligated' is that an agent acting not in accordance with the epistemic norms of the epistemic practice can be held *epistemically blameworthy* for her commitments by other agents.[30] For example, let an epistemic practice X be given, and assume that X has the norm $R =$ 'For any proposition p, one should not believe p against compelling evidence that p is false'. Moreover, assume that some agent is engaging with X and believes some proposition q, although there is compelling evidence that q is false. Then she can be held epistemically blameworthy for her belief that q.

[27] See for instance Pedersen (2021).

[28] Clearly, one could question Wright's list of authenticity-conditions. However, here I follow—with Fischer et al.—Wright and take for granted that these propositions are in fact authenticity-conditions.

[29] For my purposes, it is not important here to investigate the nature of such norms, or to spell out what norms a specific cognitive project has. Here I only want to make the (rather trivial) fact explicit that epistemic practices are regulated by epistemic norms.

[30] For a recent account of epistemic blame and its relation to epistemic norms see for instance Brown (2018).

4.2 Trustworthiness and Epistemic Value

Fischer et al. (2019) make a distinction between two concepts of truth in the context of cognitive projects: a concept of *scientific* truth and a concept of *full disquotational* truth.[31] Fischer et al. describe the concept of scientific truth as a theoretical concept, employed in scientific theories with the aim of explaining non-semantic facts. They argue that the scientific concept of truth is not different from other scientific, theoretical concepts that might be employed in science. A fundamental characteristics of the scientific concept of truth is that the logic of truth should inherit the logic of the non-semantic language. In other words, in our context, where the non-semantic theory is first-order **PA** formulated in classical logic, the theory of scientific truth should also be formulated in classical logic.[32]

On the other hand, disquotational truth only intends to be a device of quotation and disquotation. This informal concept follows the intuition that it should be unproblematic to assert that if some state of affairs is so and so, then it is true that some state of affairs is so and so, and *vice versa*.[33] Fischer et al. argue that, since the full disquotational concept of truth intends to be a device of naive, i.e. unrestricted, quotation and disquotation, this concept should be governed by some non-classical logic (to avoid triviality). Fischer et al. identify the theories $\mathbf{TS_0}$ and $\mathbf{UTS_0}$ presented in Sect. 2.2 as theories of full disquotational truth.

Consider some rational agent, who is engaging in some cognitive project. Suppose that the agent employs some theory of truth **S** in her enquiry. From our previous section, we have that within this cognitive project both the soundness of **S** and the consistency of **S**'s concept of truth are authenticity-conditions.

Trustworthiness is an adequacy condition on theories of truth that arises from a reflection on the importance of the authenticity-conditions mentioned above. They argue in the following manner: if the soundness of **S** is made explicit, by adding GRP_S to **S**, then **S**'s concept of truth should remain consistent. Conversely, if **S** is either inconsistent of internally inconsistent with GRP_S, then **S** is *untrustworthy*: **S** cannot be trusted as a tool to be employed in our cognitive enquiry to pursue our epistemic aims. Otherwise, the theory is trustworthy. The thought behind the *trustworthiness* requirement is that any acceptable theory (of truth) **S** should remain (internally) consistent, if the agent employing **S** makes the authenticity-conditions of her enquiry

[31] Fischer et al. remain quite open about whether this distinction amounts to Field's distinction in Field (1994) between *inflationary* and *deflationary* truth, or to McGee's distinction in McGee (2005) between *disquotational* and *causally explanatory*—other times called *correspondence*—truth. My focus here is simply to present their distinction and I won't further investigate this question.

[32] It is interesting to point out that Fischer et al. do not propose the stronger requirement that the logic inside the scope of the truth predicate should be classical.

[33] This is the informal intuition that Tarski has at the beginning of Tarski (1936), which goes back to Aristotle.

explicit.[34] In addition, they argue for the importance of formulating the soundness of **S** with global reflection.[35]

From the results of Sect. 2.2, we have that the theories $\mathbf{TS_0}$ and $\mathbf{UTS_0}$ investigated in Fischer et al. (2017) are *trustworthy*; from Proposition 1 we can correctly infer that these theories of full disquotational truth in non-classical logic are (internally) consistent, if the authenticity-condition of soundness is made explicit with global reflection. On the other hand, many theories of scientific truth in classical logic are not trustworthy. Fischer et al. claim the following:

> Theories of [scientific] truth do not sit well with statements of their own soundness. [...] Scientific notions of truth, are inadequate if such a requirement is adopted. [...] In theories of classical truth we cannot consistently hold that what they prove is true, and not false. This entails that scientific theories of truth suffer the same fate, by our assumption that only theories of classical truth can be considered theories of scientific truth. (Fischer et al. 2019, pp. 7 - 8)

I agree with Fischer et al. that *many* theories of scientific truth in classical logic are not trustworthy. As we saw in Sect. 2.2, theories such as **KF** are internally inconsistent with global reflection and therefore not trustworthy by the standards set in Fischer et al. (2019). There are other theories that are not trustworthy by the same standards: **FS** is an an example of such theories.[36] It is well-known that **FS** is incompatible with global reflection:

Proposition 7 (Horsten et al. 2012, Proposition 4.6) **GRP[FS]** *is inconsistent.*[37]

From these results the following observation follows:

Observation 1 $\mathbf{TS_0}$ and $\mathbf{UTS_0}$ are trustworthy, whereas **KF** and **FS** are not.

The situation looks promising for the proponent of theories of positive truth; the results of Sect. 3 allow us to make a similar observation:

Observation 2 The theories **TFB**, **UTFB** and $\mathbf{KF}^*_{\mathbf{pos}}$ are trustworthy.

Just by the informal understanding of trustworthiness, we can conclude that theories of full disquotational truth and theories of positive truth can be trusted as tools to be employed in our cognitive enquiries. This suggests that these theories have a better epistemic status than for instance **KF** and **FS**; after all, the latter theories cannot be trusted.

Moreover, we can see that, for a given cognitive enquiry X, *if* X has the epistemic norm ‘One should only accept trustworthy theories’, then proponents of theories such

[34] Although Fischer et al. do not state this explicitly, it is plausible to take the trustworthiness requirement as a necessary condition for a theory’s adequacy. I agree with Fischer et al. that trustworthiness is a natural and desirable property. The desirability of some form of coherence of theories of truth as been already pointed out—although only informally—by Halbach and Horsten (2015) and by Leitgeb (2007).

[35] In doing so, they follow the idea already expressed by Kreisel and Lévy (1968). I agree with Fischer et al. that global reflection is natural and the intended way of formulating the soundness of **S**.

[36] See Halbach (2014) and Friedman and Sheard (1987) for two presentations of **FS**. Another example is the theory **VF**. For a presentation of **VF** see for instance Cantini (1990).

[37] This is so because **FS** is ω-inconsistent. A (recursively axiomatisable) theory **S** is ω-inconsistent just in case there is a φ, such that $\mathbf{S} \vdash \varphi(n)$ for all n and $\mathbf{S} \vdash \neg\forall x \varphi(x)$.

as **KF** or **FS** find themselves in an epistemically worse position than the proponents of full disquotational non-classical theories, or proponents of theories of positive truth: let a rational agent, proponent of (for instance) **KF** be given. Suppose that the agent is committed to **KF** in her mathematical cognitive enquiry. Then, simply by the assumption that the acceptance of trustworthy theories is an epistemic norm we have that she is epistemically obligated to revise her commitments to **KF**, insofar as other agents can hold her blameworthy for her commitments to **KF**. The situation is quite different in the case of positive truth.[38] Suppose that a rational agent is a proponent of $\mathbf{KF}^*_{\mathbf{pos}}$ and that she is committed to $\mathbf{KF}^*_{\mathbf{pos}}$ in her cognitive enquiry. If the agent was warranted to accept $\mathbf{KF}^*_{\mathbf{pos}}$ to start with, of course she is not to blame for her commitments to $\mathbf{KF}^*_{\mathbf{pos}}$ from the perspective of the epistemic norm ‘One should only accept trustworthy theories’.[39]

Although we can conclude that theories of positive truth are as good as theories of full disquotational truth if we consider the trustworthiness requirement, there are still some clarifications to be made and also some possible worries and problems for the proponent of positive truth.

4.3 Positive Truth and Cognitive Projects: The Worries

As we saw in Sect. 2.2, **GRP[KF]** proves that the liar is both true and false. This happens because global reflection works as a bridge principle between the *external* and the *internal* logic, where the former is the logic outside the scope of T and the latter is the logic inside the scope of T. Importantly, **KF**’s external logic is classical, whereas its internal logic is non-classical. Global reflection is problematic because it pushes **KF**’s classical external negation inside the scope of **KF**’s non-classical truth predicate.

On the other hand, the proof of internal inconsistency is blocked in the case of $\mathbf{GRP}[\mathbf{KF}^*_{\mathbf{pos}}]$; in the proof of the internal inconsistency in **GRP[KF]** *unrestricted* principles of compositionality for the truth predicate are employed, that is, truth principles also for statements that are not strictly positive. Compositionality together with $\mathsf{KF11}$ and $\mathsf{KF12}$ are enough to derive the internal inconsistency in **GRP[KF]**. However, in $\mathbf{GRP}[\mathbf{KF}^*_{\mathbf{pos}}]$ compositionality is accepted for strictly positive statements only. Basically, in $\mathbf{KF}^*_{\mathbf{pos}}$ the external negation does not interact with the internal negation—the falsity predicate–, and this is essential to block the proof of the internal inconsistency. However, as Nicolai points out in (Nicolai 2021, p. 736) one can define a translation $(*)$ from the language $\mathcal{L}_{\mathsf{T}}$ into the strictly positive language $\mathcal{L}^+$ that essentially replaces negative occurrences of the truth predicate with the falsity predicate F of $\mathcal{L}^+$.[40] Crucially, if we were to accept the translation function $(*) : \mathcal{L}_{\mathsf{T}} \to \mathcal{L}^+$, then via global reflection the internal inconsistency would arise again.[41]

[38] Also the proponent of full disquotational non-classical truth finds herself in this epistemic situation.

[39] Of course, the situation might change if other epistemic norms are accepted.

[40] The details of the translation are not important for our purposes. The crucial idea involved in the translation that I am interested in, is the fact that via the translation F is understood as $\neg\mathsf{T}$. For the interested reader, the details of the translation can be found in Nicolai (Nicolai 2021, p. 751).

[41] This is Nicolai’s proof in (Nicolai 2021, Proposition 1).

This result *prima facie* threatens the philosophical importance of the internal consistency of $\mathbf{GRP}[\mathbf{KF}^*_{\mathbf{pos}}]$, insofar as one can argue that—via the translation $(*)$—the proponent of theories of positive truth finds herself in the same position of the proponent of theories such as **KF**. This worry is pressing, since if the proponent of positive truth must accept the translation $(*)$, then she is going to end up with an untrustworthy theory, resulting in an epistemically blameworthy position.[42] Let me state this worry fully explicitly, for a given agent and some theory of positive truth **S**:

> (translation) Can the proponent of positive truth have a warrant for her acceptance of **S** and nevertheless being warranted in rejecting the translation $(*)$?

As we can see, (translation) targets a proponent of a theory of positive truth **S**, who is warranted in her acceptance of **S**. However, the following worry, which concerns the question of warrant to accept theories of positive truth to start with, is even more pressing:[43]

> (warrant) Is there any warrant to accept theories of positive truth to start with, and if there is such, what is the force of such warrant?

This addresses the well-known problem of providing a positive argument for theories of positive truth; theories of positive truth are usually conceived as a response to the paradoxes, insofar as the restriction to positive biconditionals simply and straightforwardly retains consistency, without loss of generality and proof-theoretic strength for the arithmetical language.[44] However, such restriction is taken to be artificial. Halbach (2009), Horsten and Leigh (2017) and Cieśliński (2017, 2015) all independently argue that the theories of positive truth and falsity are well-motivated, via a careful analysis and diagnosis of the paradoxes of truth.[45] However, the force of the argument for positive truth still needs to be spelled out: for instance, one might (and should) ask how good the warrant for the acceptance of positive truth, by means of the analysis of the paradoxes, is. To my knowledge, the only place where the question about the warrant's goodness is explicitly discussed is (Cieśliński 2015, Section 5.5.). There, he claims the following:

> Restoration of the consistency of disquotational theory is a natural aim. Naive, unrestricted T-schema generates a contradiction—that's a fact to which all truth theorists must react and the disquotationalist is no exception. Restoring the con-

[42] I am deeply grateful to an anonymous referee of this journal for expressing this worry and for suggesting an explicit discussion of it.

[43] I am grateful to an anonymous referee of this journal for expressing this worry.

[44] This is well-known. See for instance (Halbach 2014, Corollary 19.18) or (Cieśliński 2015, Theorem 11).

[45] The idea is that, by analysing the paradoxes of truth, one formulates the hypothesis that paradoxes necessarily involve occurrences of the truth predicate that are not strictly positive. Curry's paradox fits into this hypothesis, if implication is not taken as a primitive: if '→' is taken to be defined as usually, then Curry's paradox also involves a negative occurrence of the truth predicate. See for instance (Cieśliński 2017, pp. 53–54) for a discussion of this issue.

> sistency of a theory of truth should be treated as a permissible motivation for the disquotationalist to proceed. The question is only how far it can take us.[46]

I agree with Cieśliński that in principle, restoring consistency should be treated as a permissible motivation to proceed. However, how far can this take us? A way of addressing this question might involve explaining the relations and dependencies between the respective answers to (warrant) and (translation).

A third worry involves the informal claim that theories of positive truth are theories of scientific truth. Roughly, this addresses the worry that theories of positive truth are too restrictive and because of that inadequate for scientific cognitive projects. To address this worry, we can state the following question explicitly:

> (project) What cognitive project can a theory of positive truth be associated with, so that positive truth embodies the concept of scientific truth?[47]

The aim of the remainder of this section is to address these worries, spelled out as (project), (warrant), and (translation). I will suggest possible responses to these worries and argue that these are available to the proponent of positive truth.[48]

4.4 Responses

Let's start with (project). It is well-known that theories of positive truth are restrictive with respect to their truth-theoretic principles. On the other hand, the truth-theoretic principles of the theories of full disquotational truth are fully unrestricted. And even the classical theory **KF** has fully general compositional truth-theoretic principles. So, in order to provide an answer to (project) the proponent of theories of positive truth has to provide a cognitive project, where positive truth plays the theoretical role embodied by scientific truth.

I think that the proponent of positive truth has a good response to (project): positive truth can be employed in cognitive projects, in which truth is employed as a tool to investigate mathematical, non-semantic facts. These are enquiries into some mathematical, non-semantic subject matter. In the case of an enquiry into an arithmetical subject matter, it is known that $\mathbf{KF}^*_{\mathbf{pos}}$ is proof-theoretically equivalent to its negative formulation, **KF**, for the full arithmetical language. Within these cognitive projects—so the proponent of positive truth can argue—positive truth is as general as the scientific truth predicate of **KF**. I don't see why the proponent of positive truth shouldn't be allowed or be able to claim that positive truth is a scientific notion of truth, *in purely mathematical cognitive projects*: it is a theoretical concept employed in the investigation of mathematical facts.

However, we have to recognise that positive truth can hardly be a scientific concept of truth in cognitive projects, which aim at investigating *semantic facts involving truth*.

[46] Cieśliński briefly addresses a version of (warrant)—his question (3) in Cieśliński (2015). However, he only focuses on this question in the context of disquotationalism about truth and he also doesn't consider the relation to the additional worries that I discuss.

[47] I am deeply grateful to an anonymous referee of this journal to point out that these issues have been left open and were not properly addressed by a previous version of this article.

[48] However, a thorough defence of each response would exceed the scope of this paper.

When investigating some fully general notion of truth, the choice of positive truth needs an independent motivation. In other words, the warrant provided by the usefulness of positive truth for an arithmetical sub-language does not justify the choice of positive truth as a tool to investigate fully unrestricted, self-referential truth. For such projects, I agree with Fischer et al. that the proponent of full disquotational truth is in a much better position.[49] I think that this is enough, as a tentative answer to (project); after all, an answer to (project) only amounts to providing some cognitive project, where positive truth plays the theoretical role of scientific truth.

Looking at (warrant), we can see that, *if* we focus on the proposed cognitive projects, where positive truth is a theoretical tool to investigate purely mathematical subject matters, then the proponent of positive truth should be able to employ Cieśliński's argument from the analysis of the paradoxes as a motivation to choose theories of positive truth in their cognitive enquiry: the restriction of the unrestricted biconditionals to some subsets thereof is motivated by the argument from the paradoxes, and this seems to be enough to warrant the *instrumental* acceptance of positive truth. After all, positive truth is just a useful tool. Cieśliński's reasoning suggests that the careful analysis of the paradoxes provides the agent with a warrant to accept *some* restriction of the T-biconditionals. However, the proponent of positive truth still has to provide a motivation for positive truth explicitly.[50] I believe that in the context of the proposed cognitive projects it is very hard to see what the philosophical motivation for positive truth would be.

Fortunately, the proponent of positive truth doesn't need any such philosophical motivation; in the context of these cognitive projects the proponent of positive truth only needs a warrant for instrumental acceptance of positive truth, and for such warrant simple pragmatic considerations about the virtues of positive truth for the success of the project should be available to the proponent of positive truth.

With respect to (translation): given the instrumental acceptance of positive truth, the proponent of positive truth in such purely mathematical cognitive projects has good reasons to reject the translation $(*)$ given by pragmatic considerations; the translation would bring the inconsistency back, threatening the success of the cognitive enquiry.

Let me conclude with a clarification: although sketched, I think that these responses are good options available to the proponent of positive truth. However, it should be added that by focussing on instrumental acceptance the proponent of positive truth may have too easy answers to (translation) and (warrant): at the moment, she is only able to respond to these challenges via pragmatic considerations. On ther other hand, the proponent of full disquotational truth might even have a *philosophical* argument for her warrant to acceptance the theories of full disquotational truth: the truth predicate

[49] As an anonymous referee of this journal pointed out, considerations about the models of positive truth suggest that the proponent of positive truth cannot claim the generality of its concept of truth as a tool to investigate fully unrestricted, self-refential truth; theories such as $\mathbf{UTS_0}$ do not exclude any natural model of truth. On the other hand, theories such as $\mathbf{KF}^*_{\mathbf{pos}}$ do exclude natural models, such as the minimal model. However, I don't think that this worry remains, if positive truth is accepted as a theoretical tool to investigate purely mathematical subject matters. After all, for this purpose it is crucial that the theory of truth allows for standard interpretations.

[50] This is so because Cieśliński's reasoning would also motivate the choise of some typed notion of truth, with no need to opt for a type-free positive truth predicate.

of full disquotational truth theories embodies, or captures, some informal concept of full disquotational truth.[51]

The challenge to provide a philosophical argument for the choice of positive truth and falsity, which is not simply pragmatic, is still open. I think that the proponent of positive truth and falsity might at least have the following option: she could try to understand the concept of truth as expressing some epistemic notions similar to *warranted assertibility*.[52] Alternatively, one could understand truth as the stronger notion of *super-assertibility*.[53] Under these interpretations, the proponent of positive truth (and falsity) should have good *philosophical reasons* to reject the translation (*) and would therefore have a philosophical answer to (translation).[54]

However, one would have to explain how the notion of warranted assertibility (resp. super-assertibility) motivates the choice of the positive biconditionals. Moreover, the proponent of positive truth would also have to address the usual objections against the thesis that truth can be understood as an epistemic concept.[55] After that, the proponent of positive truth would have to *at least* assess whether this philosophical argument for positive truth provides good answers to (warrant), (translation) *and* (project).

Acknowledgements Earlier versions of this paper have been presented at conferences in Bristol, Munich and Moscow. Thanks to all the participants for their valuable comments and questions. Moreover, I thank Volker Halbach, Martin Fischer, Graham Leigh, Hartry Field, Marianna Antonutti Marfori, Benedict Eastaugh, Johannes Stern, Catrin Campbell-Moore and two anonymous referees of this journal for their helpful comments. I am deeply grateful to Leon Horsten and Kentaro Fujimoto for their continuous support. Finally, I am grateful to the South, West and Wales Doctoral Training Partnership and the Arts and Humanities Research Council for supporting my research project.

Funding My research is funded by the South, West and Wales Doctoral Training Partnership and by the Arts and Humanities Research Council (grant reference AH/L503939/1).

Availability of data and material Not applicable

Declarations

Conflict of interest There is no conflict of interest.

Code availability Not applicable.

[51] However, let me point out that the very idea that *philosophical arguments* are epistemically better than *pragmatic arguments* is questionable. I think that, in order to make some clarity and progress with respect to this distinction, an investigation of cognitive projects with their epistemological questions should be connected to the recent literature and problems of *naturalism* in mathematics..

[52] This has been investigated for instance by Kvanvig (1999), and by Tennant (1995).

[53] This has been proposed for instance by Wright (1996).

[54] The argument against $(*)$ follows from the so-called problem of neutral states of information, discussed for instance by Kvanvig (1999).

[55] To provide such fully fleshed out philosophical interpretation of positive truth and falsity in terms of assertibility is left open for another occasion.

References

Brown, J. (2018). What is epistemic blame? *Noûs, 54*(2), 389–407.

Cantini, A. (1989). Notes on formal theories of truth. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 35*, 97–130.

Cantini, A. (1990). A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic, 55*, 244–259.

Cieśliński, C. (2015). Typed and untyped disquotational truth. *Unifying the Philosophy of Truth* (pp. 307–320). Netherlands: Springer.

Cieśliński, C. (2017). *The Epistemic Lightness of Truth*. Deflationism and its Logic: Cambridge University Press.

Feferman, S. (1962). Transfinite recursive progressions of axiomatic theories. *The Journal of Symbolic Logic, 27*(3), 259–316.

Feferman, S. (1964). Systems of predicative analysis. *The Journal of Symbolic Logic, 29*(1), 1–30.

Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic, 56*, 1–47.

Field, H. (1994). Deflationist views of meaning and content. *Mind, 103*(411), 249–285.

Fischer, M., Nicolai, C., & Horsten, L. (2017). Iterated reflection over full disquotational truth. *Journal of Logic and Computation, 27*(8), 2631–2651.

Fischer, M., Horsten, L., & Nicolai, C. (2019). Hypatia's silence truth justification and entitlement. *Noûs, 55,* 62–85. https://doi.org/10.1111/nous.12292

Franzen, T. (2004). Inexhaustibility. Lecture Notes in Logic. CRC Press.

Friedman, H., & Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic, 33*, 1–21.

Friedman, H., & Sheard, M. (1988). The disjunction and existence properties for axiomatic systems of truth. *Annals of Pure and Applied Logic, 40*(1), 1–10.

Galinon, H. (2014). Acceptation, cohérence et responsabilité. In Liber Amicorum Pascal Engel 320–333.

Halbach, V. (2009). Reducing compositional to disquotational truth. *The Review of Symbolic Logic, 2*(4), 786–798.

Halbach, V. (2014). Axiomatic Theories of Truth. Cambridge University Press, Cambridge, UK, revised edition.

Horsten, L. (2021). On reflection. *The Philosophical Quarterly, 71*(4), 738–757.

Halbach, V. & Horsten, L. (2015). Norms for theories of reflexive truth. In K. Fujimoto, J. M. Fernandez, H. Galinon, & T. Achourioti (Eds.), *Unifying the Philosphy of Truth, volume Unifying the Philosophy of Truth.* Springer Verlag. 263–280

Horsten, L., & Leigh, G. E. (2017). Truth is simple. *Mind, 126*(501), 195–232.

Horsten, L., Leigh, G. E., Leitgeb, H., & Welch, P. (2012). Revision revisited. *The Review of Symbolic Logic, 5*(4), 642–664.

Horsten, L. and Zicchetti, M. (2021). Truth, reflection and commitment. In Stern, J. and Nicolai, C., editors, Modes of Truth. The Unified Approach to Truth, Modalities, and Paradox. Routledge. 69–87

Kreisel, G., & Lévy, A. (1968). Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 14*, 97–142.

Kvanvig, J. (1999). Truth and Superassertibility. *Philosophical Studies, 93*(1), 1–19.

Leigh, G. E. (2016). Reflecting on truth. *IfCoLog Journal of Logics and their Applications, 3*(4), 557–594.

Leitgeb, H. (2007). What theories of truth should be like (but cannot be). *Philosophy Compass, 2*(2), 276–290.

McGee, V. (2005). Afterword: Trying (with limited success) to demarcate the disquotation-correspondence intuition. In J. Beall & B. Armour-Garb (Eds.), *Deflationary Truth.* Open Court. 143–152

Nicolai, C. (2021). Fix, express, quantify: Disquotation after its logic. *Mind, 130*(519), 727–757.

Pedersen, N. (2021). Cornerstone epistemology: scepticism, mathematics, non-evidentialism, consequentualism, pluralism. *Non-Evidentialist Epistemology, Brill Studies in Skepticism* (pp. 230–264). Netherlands: Brill.

Pedersen, N. J. L. L. (2016). Hume's principle and entitlement: On the epistemology of the neo-fregean programme. In P. Ebert & M. Rossberg (Eds.), 186–202 *Abstractionism.* Oxford University Press.

Tarski, A. (1936). Über den Begriff der logischen Folgerung. *Actes du Congrès International de Philosophie Scientifique, 7*, 1–11.

Tennant, N. (1995). On negation, truth and warranted assertibility. *Analysis, 55*(2), 98–104.

Turing, A. (1939). Systems of logic based on ordinals. *Proceedings of the London Mathematical Society, 2*, 161–228.

Wright, C. (1996). Precis of truth and objectivity. *Philosophy and Phenomenological Research, 56*(4), 863.

Wright, C. (2004). Warrant for nothing (and foundations for free)? *Proceedings of the Aristotelian Society, Supplementary Volumes, 78*, 167–245.

Wright, C. (2012). Replies part iv: Warrant, transmission and entitlement. In Coliva, A., editor, Mind, Meaning, and Knowledge. Themes from the Philosophy of Crispin Wright, pages 451–486. Oxford University Press.

Wright, C. (2016). Abstraction and epistemic entitlement: On the epistemological status of hume's principle. In P. A. Ebert & M. Rossberg (Eds.), *Abstractionism: Essays in Philosophy of Mathematics* (pp. 161–185). Oxford University Press.