Iterated reflection over full disquotational truth

MARTIN FISCHER, *MCMP, LMU Muenchen.*
E-mail: M.Fischer@lrz.uni-muenchen.de

CARLO NICOLAI, *Department of Philosophy and Religious Studies, Utrecht University, Janskerkhof 13, 3512 BL Utrecht, The Netherlands.*
E-mail: carlo.nicolai6@gmail.com

LEON HORSTEN, *Cotham House, Bristol BS6 6JL, United Kingdom.*
E-mail: Leon.Horsten@bristol.ac.uk

Abstract

Iterated reflection principles have been employed extensively to unfold epistemic commitments that are incurred by accepting a mathematical theory. Recently this has been applied to theories of truth. The idea is to start with a collection of Tarski-biconditionals and arrive by iterated reflection at strong compositional truth theories. In the context of classical logic, it is incoherent to adopt an initial truth theory in which A and 'A is true' are inter-derivable. In this article, we show how in the context of a weaker logic, which we call Basic De Morgan Logic, we can coherently start with such a fully disquotational truth theory and arrive at a strong compositional truth theory by applying a natural uniform reflection principle a finite number of times.

Keywords: Truth, semantic Paradoxes, Basic de Morgan logic, reflection principles.

1 Introduction

In the article we pursue the strategy of iterating reflection principles on a basic truth theory TS_0 that encapsulates, or so we argue, the fundamental building blocks of truth-theoretic reasoning. Its components are: a theory of the objects of truth (syntax theory in our case), basic truth-theoretic principles that enable us to infer $\mathsf{T}\ulcorner\phi\urcorner$ from a sentence ϕ and vice-versa. To remain faithful to these assertability conditions for truth ascriptions, classical logic cannot be used on account of the liar paradox. In particular, one should be prepared not to have the rule of conditionalization for all sentences. Also, it is desirable to work at the right level of generality by, for instance, reasoning without committing oneself to paracomplete or paraconsistent options. We employ a logic that does this and call it, following [9], *Basic De Morgan logic* (cf. §2 for the definition).

The theory TS_0, however, may not be all there is to truth. Many authors have discussed further desiderata for truth — such as full compositionality — that are out of reach for our basic theory TS_0. Recently, Leon Horsten and Graham Leigh have studied iterations of reflection over a basic theory in classical logic [18]. Their classical starting point, however, leads to a loss of the intimate connection between truths and their assertability that is present in TS_0. In the following we therefore extend Horsten and Leigh's strategy in the framework of Basic De Morgan logic.

Reflection is rooted in the fundamental intuition that we are committed to the truth of the sentences that are provable in a theory that we accept. This operation can be expressed in different ways: as *global reflection*, where we make explicit use of the truth predicate, or as *uniform reflection*, where

Published online August 8, 2017 doi:10.1093/logcom/exx023

we express this intuition schematically without direct reference to truth. In the classical case, global and uniform reflection are provably different operations: a variety of well-known theories of truth can be closed under uniform reflection but not under global reflection.[1] However, this is not so in the case of Basic De Morgan logics and variations thereof, where global and uniform reflection coincide for a wide class of theories containing the truth principles of TS_0.

The theories that we study in this work, resulting from iterating reflection over TS_0, can be characterized as *internal* axiomatizations of Kripke's fixed point construction — specifically, its four-valued version based on Basic De Morgan Logic as it can be found in [14] — because their principles and rules are all sound with respect to this class of models. The paradigmatic case of an internal theory is the theory PKF introduced in [15]. These theories are faithful to the fixed-point models and interact well with the process of reflection. There is an alternative array of theories capturing Kripke's construction *externally*. They are couched in classical logic and they are meant to be faithful to the set-theoretic definition of this class of models. Among the external theories we find the well-known classical axiomatization KF of Kripke's construction [8], and also the iteration of reflection studied by Horsten and Leigh [18]. Although external axiomatizations invoke principles that are not valid in the intended semantics and they usually cannot be closed under global reflection, they are proof-theoretically stronger than the corresponding internal axiomatizations. As a consequence, they also deem true more sentences that belong to the intended extension of the truth predicate than the corresponding internal axiomatizations; for some authors (cf. [14, 16]), this is considered a clear advantage over internal theories.

For instance, KF is proof-theoretically much stronger than its natural internal counterpart PKF. The former proves transfinite induction for all sentences of the language with the truth predicate up to any ordinal smaller than ε_0, the latter only up to any ordinal smaller than ω^ω. The main result of the present article is that two steps of reflection over TS_0 enbles us to recapture all principles of PKF and prove significantly more transfinite induction than what is available in PKF. Moreover, iterated reflection on TS_0 enables us to reach the strength of KF.

2 The core laws of truth

In this section we introduce the main components of the theory TS_0. We first introduce a two-sided sequent version of Basic De Morgan logic and state some simple properties of this calculus. We then introduce the principles governing the objects to which truth is ascribed, which will amount to the axioms of a very weak arithmetical theory. Finally we state the truth theoretic principles of TS_0.

2.1 Basic De Morgan logic

We employ a two-sided sequent calculus BDM reminiscent of the one employed in [14]; sequents are expressions of the form $\Gamma \Rightarrow \Delta$ where Γ, Δ are finite sets of formulas. We write $\neg\Gamma$ for $\{\neg A \mid A \in \Gamma\}$. BDM is a subsystem of a suitable two-sided classical calculus; its axioms and rules are listed in Table 1. Intuitively, BDM is obtained from classical logic by replacing the usual clauses for negation with (CP1) and (CP2) below. However, the general negation rules (CP1-2) enable us to derive the sequents $A \Rightarrow \neg\neg A$ and $\neg\neg A \Rightarrow A$ for all formulas A and make the following contraposition rule admissible in BDM:

$$\frac{\Gamma \Rightarrow \Delta}{\neg\Delta \Rightarrow \neg\Gamma}\ (\mathsf{Cont}).$$

[1] Examples are the system KF from [8] plus the axiom stating that the extension of the truth predicate is consistent and the system FS from [13] originating in [11].

TABLE 1. The system BDM

$$A \Rightarrow A \qquad \text{for } A \in \mathcal{L}_{\mathsf{T}}$$

$$\frac{\Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta}\ (\mathsf{LW}) \qquad \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A}\ (\mathsf{RW})$$

$$\frac{\Gamma \Rightarrow \Delta, A \qquad A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\ (\mathsf{Cut})$$

$$\frac{\neg\Gamma \Rightarrow \Delta}{\neg\Delta \Rightarrow \Gamma}\ (\mathsf{CP}_1) \qquad \frac{\Gamma \Rightarrow \neg\Delta}{\Delta \Rightarrow \neg\Gamma}\ (\mathsf{CP}_2)$$

$$\frac{A, B, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta}\ (\mathsf{L}\wedge) \qquad \frac{\Gamma \Rightarrow \Delta, A \qquad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}\ (\mathsf{R}\wedge)$$

$$\frac{A, \Gamma \Rightarrow \Delta \qquad B, \Gamma \Rightarrow \Delta}{A \vee B, \Gamma \Rightarrow \Delta}\ (\mathsf{L}\vee) \qquad \frac{\Gamma \Rightarrow \Delta, A, B}{\Gamma \Rightarrow \Delta, A \vee B}\ (\mathsf{R}\vee)$$

$$\frac{\Gamma, A(t) \Rightarrow \Delta}{\Gamma, \forall x A \Rightarrow \Delta}\ (\mathsf{L}\forall) \qquad \frac{\Gamma \Rightarrow \Delta, A(t)}{\Gamma \Rightarrow \Delta, \exists x A}\ (\mathsf{R}\exists)$$

$$\frac{\Gamma \Rightarrow \Delta, A(x)}{\Gamma \Rightarrow \Delta, \forall\, x\, A}\ (\mathsf{R}\forall) \qquad \frac{\Gamma, A(x) \Rightarrow \Delta}{\Gamma, \exists\, x\, A \Rightarrow \Delta}\ (\mathsf{L}\exists)$$

x not free in Γ, Δ $\qquad$ x not free in Γ, Δ

The following closely related lemma will be extensively used in what follows:

LEMMA 1
For a language $\mathcal{L}_0$ with signature $\mathcal{S} = \{P_1, \ldots, P_n\}$, if for all atomic formulas A of $\mathcal{L}_0$ we can prove $\Rightarrow A, \neg A$, then BDM formulated in $\mathcal{L}_0$ is closed under the following classical rules for negation:

$$\frac{\Gamma \Rightarrow \Delta, A}{\neg A, \Gamma \Rightarrow \Delta}\ (\neg\mathsf{L}) \qquad \frac{A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A}\ (\neg\mathsf{R}).$$

BDM enjoys standard properties of Gentzen-type sequent calculi such as substitution, inversion and cut elimination.

A natural semantics for BDM is given in terms of four-valued models, that is we also allow predicates with a partial or a paraconsistent behaviour (gaps and gluts).[2] The intended satisfaction relation has a double clause: a sequent is satisfied in a model $\mathcal{M}$ just in case if all formulas in the antecedent are true in $\mathcal{M}$ there is a formula in the consequent true in $\mathcal{M}$, *and* if all formulas in

[2]Our logic is close to what is sometimes called FDE. We follow, however, Field's terminology in [9, Ch. 3].

the succedent are false in $\mathcal{M}$, there is a false in $\mathcal{M}$ formula in the antecedent. BDM is sound and complete with respect to the semantics just hinted at (see [3]).

2.2 *The theory* TS_0

The basic theory TS_0 and all its extensions will be formulated in the language of arithmetic $\mathcal{L} = \{0, \mathsf{S}, +, \times, \mathsf{exp}\}$ with exponentiation. With x not occurring in the $\mathcal{L}$-term t, expressions of the form $(\forall x \leq t)\ \varphi(x)$ and $(\exists x \leq t)\ \varphi(x)$ are said to be obtained from $\varphi(v)$ by *bounded quantification*: Δ_0-formulas (or *elementary*) of $\mathcal{L}$ are formulas that contain only bounded quantifiers. The classes of Σ_n and Π_n formulas of $\mathcal{L}$ are then defined in the usual way. In practice, we work in the expansion of $\mathcal{L}$ with finitely many function symbols corresponding to suitable elementary operations and the truth predicate T. We call the resulting language $\mathcal{L}_\mathsf{T}$. We call a formula of $\mathcal{L}_\mathsf{T}$ *arithmetical* if it does not contain occurrences of T.

To formulate TS_0, we first consider identity, which is governed by usual principles:

$$\Rightarrow t = t \tag{Id1}$$

$$s = t, A(s) \Rightarrow A(t). \tag{Id2}$$

TS_0 will also contain initial sequents $\Rightarrow A$ for all basic axioms A of a suitable system of arithmetic, in our case Kalmar's elementary arithmetic EA formulated in $\mathcal{L}_\mathsf{T}$ (cf. [2, 12]).[3] In addition, our basic theory features an induction rule

$$\frac{\Gamma, A(x) \Rightarrow A(x+1), \Delta}{\Gamma, A(0) \Rightarrow A(t), \Delta} \tag{Δ_0-IND}$$

where x is not free in $A(0), \Delta, \Gamma$, t is arbitrary, and A is a Δ_0-formula of the language $\mathcal{L}$ of arithmetic without the truth predicate. We call the resulting system Basic.

The core principles of truth capture the fundamental idea that one is justified in asserting a sentence A precisely when she is justified in asserting that A is true.

DEFINITION 1 (The system TS_0)
TS_0 is obtained by extending Basic with the initial sequents

$$\mathsf{T}(\ulcorner A \urcorner) \Rightarrow A \tag{T1}$$

$$A \Rightarrow \mathsf{T}(\ulcorner A \urcorner) \tag{T2}$$

for all $\mathcal{L}_\mathsf{T}$-sentences A.

TS_0 stands for 'truth sequents'. The subscript 0 indicates a restriction of induction to Δ_0-formulas; its absence indicates full induction. The semantic conservativeness — and therefore the

[3] The class of *elementary functions* $\mathcal{E}$ is obtained by closing the initial functions $\mathsf{zero}(\cdot)$, $\mathsf{suc}(\cdot)$, $+$, $\times$, 2^x, $\mathsf{P}^n_i(x_1, \ldots, x_n) = x_i$ with $(1 \leq i \leq n)$, truncated subtraction $x \dot{-} y$ under the operations of composition and bounded minimalization:

$$H(\vec{x}) = F(G_1(\vec{x}), \ldots, G_n(\vec{x})); \qquad (\mu t \leq y) P(\vec{x}, t) = \begin{cases} \text{the least } t \leq y \text{ s.t. } P(\vec{x}, t) \\ 0, \quad \text{if there is no such } t \end{cases}$$

where F, G are elementary functions and P an elementary predicate. EA has sufficient resources to naturally introduce new relations corresponding to the elementary functions by proving their defining equations. In particular, the functions in $\mathcal{E}$ are exactly the functions that can be Σ_1-defined in EA (see [25, §3.1]).

consistency — of TS_0 over Basic can be obtained by expanding any model of the latter with an interpretation of the truth predicate resulting from a positive inductive definition along the lines of Kripke's fixed-point construction (cf. [4, §5]). The following observation can be found in [15, Lem. 16].

LEMMA 2
$\Rightarrow A, \neg A$ is derivable in TS_0 for A arithmetical.

The principles of TS_0 are just right to capture the desired assertability conditions for truth ascriptions: its basic truth-theoretic principles (T1)–(T2) are in fact not as strong as the classical Tarski-biconditionals: otherwise they would lead to inconsistency. But they are also stronger than mere inference rules, as the latter do not allow for conditionalization for arithmetical sentences.

We can think of TS_0 as a minimal internal axiomatization of a fixed-point construction along the lines of Kripke's [20]. The fixed-points we are interested in are in fact fixed-points of a monotone operator $\mathcal{J}$ associated with the Basic De Morgan evaluation scheme.[4] The crucial property of Kripke style fixed points S, i.e. sets of sentences S such that $\mathcal{J}(S)=S$, is that every sentence A is in S iff $\mathsf{T}(\ulcorner A \urcorner)$ is in S. By combining this fact with the notion of satisfaction introduced on p. 2633 we can easily see that for a fixed point S, the model $(\mathbb{N}, S)$ satisfies TS_0 when S is taken to be the extension of the truth predicate. Moreover, for $(\mathbb{N}, S)$ to satisfy TS_0, S has to contain the same sentences as $\mathcal{J}(S)$. This means that S is a fixed-point of $\mathcal{J}$ iff $(\mathbb{N}, S)$ satisfies TS_0, and therefore it is an internal axiomatization of the fixed-point construction in the sense of §1. Moreover any $\mathcal{L}_\mathsf{T}$-theory in BDM satisfying the adequacy condition just considered will contain the principles of TS_0.

2.3 *Intermezzo on arithmetization*

In what follows, we assume a canonical Gödel numbering for $\mathcal{L}_\mathsf{T}$-expressions. For a fixed expression e of $\mathcal{L}_\mathsf{T}$, we will use the usual Gödel corners for the closed term of $\mathcal{L}_\mathsf{T}$ representing Gödel number $\#e$ of e. Therefore, for formulas A of $\mathcal{L}_\mathsf{T}$, we will have $\ulcorner A \urcorner = \overline{\#A}$. Similarly, for sequents $\Gamma \Rightarrow \Delta$, $\ulcorner \Gamma \Rightarrow \Delta \urcorner = \overline{\#\Gamma \Rightarrow \Delta}$, where the Gödel code of $\Gamma \Rightarrow \Delta$ is taken to be an ordered pair whose components are the codes of the finite sets Γ and Δ.[5] Closed terms standing for specific Gödel codes of $\mathcal{L}_\mathsf{T}$-expressions contrast with open terms standing for templates to generate such closed terms: a well-known example of such a template is the open $\mathcal{L}_\mathsf{T}$-term $\mathsf{sub}(\ulcorner A(v) \urcorner, \ulcorner v \urcorner, \mathsf{num}(x))$, standing for the result of formally substituting, in the formula $A(v)$, the free variable v with the numeral for x. To distinguish these open terms from specific codes, we use square brackets instead of Gödel codes, so that, for instance, $[A(x)]$ stands for $\mathsf{sub}(\ulcorner A(v) \urcorner, \ulcorner v \urcorner, \mathsf{num}(x))$.[6] This distinction clearly generalizes to sequents and formulas with more than one free variable: $[\Gamma \vec{x} \Rightarrow \Delta \vec{x}]$ refers to the simultaneous substitution in $\ulcorner \Gamma \Rightarrow \Delta \urcorner$ of the variables in the strings $\vec{x}$ with their corresponding numerals, where of course $[\Gamma x \Rightarrow \Delta x]$ is short for $\mathsf{sub}((\ulcorner \Gamma \urcorner, \ulcorner \Delta \urcorner), \ulcorner x \urcorner, \mathsf{num}(x))$. When it is clear from the context which free variable we are formally substituting, we will omit it and treat sub as a binary function.

In EA we can easily carry out an elementary arithmetization of the standard syntactic notions and operations such as the notion of being a closed term of $\mathcal{L}_\mathsf{T}$ (formally, $\mathsf{ct}(x)$), the notion of being a sentence of the language $\mathcal{L}_T$ (formally, $\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x)$), the operation of prepending a truth predicate to x (formally, $\mathsf{tr}(x)$), and so on (see e.g. [2]). Theories will be elementary presented sets of axioms

[4] For a definition of the operator and the evaluation scheme we refer the reader to Halbach [14, section 15.1].

[5] We assume that the code of the finite set Γ is the code of the sequence of codes of formulas in Γ in ascending order.

[6] The square brackets notation is often replaced by the so-called Feferman dot notation, in which, for instance, $\mathsf{sub}(\ulcorner A(v) \urcorner, \ulcorner v \urcorner, \mathsf{num}(x))$ is abbreviated with $\ulcorner A(\dot{x}) \urcorner$.

and rules, and we write $\mathsf{Ax}_T(x)$ for 'x is an axiom of T'. Variations of canonical provability will be particularly relevant in what follows, especially in our discussion of reflection. We list them here accompanied by their intuitive meaning:

$\mathsf{Prf}_T(x,y)$	'x is a proof in T of the sequent y'
$\mathsf{Prv}_T(x,y)$	'the sequent y is provable in T with a proof of length no greater than x'
$\mathsf{Pr}_T(x)$	'x is a provable sequent in T'
$\mathsf{Thm}_T(x)$	'x is a theorem of T, that is the 'sequent' $(\ulcorner\varnothing\urcorner, x)$ is provable in T'
$\mathsf{Pr}^2_T(x,y)$	'the inference from the sequent x to the sequent y is provably admissible in T'

We conclude by commenting on the predicate $\mathsf{Pr}^2_T(x,y)$. In addition to being admissible, a *provably admissible rule* in a theory T requires the existence of a T-provable proof transformation of the proof of the premise into a proof of the conclusion of the rule.[7] Pr^2_T enjoys generalized versions of some of the properties usually ascribed to provability predicates:

If $\dfrac{\Gamma(x)\Rightarrow\Delta(x)}{\Theta(x)\Rightarrow\Lambda(x)}$ is admissible in T, provably in Basic, then (Pr1)

$$\mathsf{Basic}\vdash\Rightarrow\mathsf{Pr}^2_T([\Gamma(x)\Rightarrow\Delta(x)],[\Theta(x)\Rightarrow\Lambda(x)])$$

If the sequents (Pr2)

$$\Rightarrow\mathsf{Pr}^2_T([\Gamma(x)\Rightarrow\Delta(x)],[\Theta(x)\Rightarrow\Lambda(x)]);$$

and

$$\Rightarrow\mathsf{Pr}_T([\Gamma(x)\Rightarrow\Delta(x)])$$

are derivable in Basic, then also

$$\Rightarrow\mathsf{Pr}_T([\Theta(x)\Rightarrow\Lambda(x)])$$

is derivable in Basic.

2.4 *The weakness of* TS_0 *and the advantages of reflection in* BDM

To evaluate the theories of truth considered in this work we list a number of desiderata. Many of them have been already proposed and discussed by truth-theorists — see for instance [21] and [17]. As already emphasized, we consider the intersubstitutivity of φ and $\mathsf{T}^\ulcorner\varphi^\urcorner$ as guiding principle and therefore we do not argue for it but take it as primitive.[8] In addition, we require our truth predicate to be *compositional*, therefore enabling us to explain how we can understand complex sentences only on the basis of an understanding of its compounds and its logical structure. In particular, we

[7] For instance, although the rule of cut applied to 'geometric' formulations of Robinson's arithmetic Q is admissible in it, cut is not provably admissible in Q as this procedure is of hyperexponential growth rate. (For a geometric presentation of Robinson's arithmetic and for the cut elimination for it, see [22].)

[8] This immediately yields the identity of inner and outer logic as defined in [14, 21].

prefer theories that, given a set of logical constants, allow the truth predicate to commute with these constants for *all* sentences of the language, that is we require compositionality in quantified form and not only compositionality in schematic form. Finally, we aim at *proof-theoretically strong* theories of truth. The desideratum of strength can be motivated in several ways: following the broadly abductive picture sketched in [27], one might simply argue that a theory that proves more, all other things being equal, has a higher scientific status than a weaker one. Moreover, truth has served a foundational role in several philosophical programmes, such as Feferman's predicativism or Aczel's attempt to recover Frege's foundations [1, 8], and therefore a stronger theory might provide alternative, and perhaps more appealing, formulations of theories of sets or of other mathematical objects.[9]

In the light of these criteria TS_0, although natural, simple, and fully disquotational, is clearly not completely adequate. For one thing, it is not compositional in the required sense. For a universally quantified sentence such as $\forall x\, A(x)$, for instance, TS_0 cannot how its truth value depends on the truth values of its compounds $A(t)$ because the sequent $\forall x\, \mathsf{T}([Ax]) \Rightarrow \mathsf{T}(\ulcorner \forall x\, A(x) \urcorner)$ is not derivable for all A.

Closely related to the criterion of strength is the capability of a theory of proving the soundness of its base theory. This claim, in our setting, takes the form of the global reflection principle for Basic, i.e.

$$\mathsf{Thm}_{\mathsf{Basic}}(x) \Rightarrow \mathsf{T}(x), \tag{$\mathsf{GRF}_{\mathsf{Basic}}$}$$

which is not derivable in TS_0. The underivability of the global reflection principle directly follows from the fact that TS_0 is a conservative extension of Basic. Therefore, in terms of strength measured with respect to Basic, TS_0 is as bad as it can get. Conservativeness, however, is only a boundary that one can use to partition truth theories into conservative and non-conservative ones. To distinguish between non-conservative extensions of Basic, finer-grained measures are required: one option is to consider how many arithmetical sentences a theory of truth proves (or equivalently, how much arithmetical transfinite induction the theory proves). This, however, does not directly take into account general claims involving the truth predicate that we would like to consider. Therefore, to measure the strength of our theories of truth we will consider the amount of transfinite induction for the language $\mathcal{L}_{\mathsf{T}}$ that is provable in them. This criterion has the advantage of giving us direct information about how many truth iterations are provable in a theory (see §4).

Following a strategy already proposed and defended in [18] for theories formulated in classical logic, one may think of TS_0 as *implicitly* containing stronger principles, including compositional ones and principles of transfinite induction. This relation of implicit containment can be unfolded via postulating a hierarchy of reflection principles over TS_0. Traditionally, reflection principles for a theory T are explicit soundness assertions ('*whatever is provable in T, is true*'). The soundness of T is naturally expressed via GRF_T.[10] However, by Tarski's undefinability theorem, GRF_T can only be formulated if the expressive resources of T are increased with a fresh truth predicate. Therefore, if one wants to express soundness in an arithmetical language, one must resort to schemata. A well-known candidate is what is widely known as the *uniform reflection principle* for T:

$$\forall x(\mathsf{Thm}_T([A(x)]) \rightarrow A(x)). \tag{RFN_T}$$

RFN_T states that, for every number x, if A is satisfied by the numeral for x, provably in T, then A is satisfied by x. However, we are mainly concerned with languages that *do* contain a truth predicate.

[9] Our criteria more or less correspond to the ones listed in [21], except of course the one requiring classical logic.

[10] This reading of reflection is ubiquitous in the literature. See for instance the classical handbook entry [26] and [14]. Kreisel and Lévy in [19] clearly states that global reflection is the intended soundness claim for a theory T.

In this context, therefore, the most natural way to express the soundness of a theory is by means of global reflection. In fact, if the truth predicate satisfies minimal conditions, the global reflection principles implies all instances of uniform reflection for a theory T.

There is therefore 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 Leigh and Horsten in [18]. For T an axiomatization of Kripke's fixed point construction in classical logic, in fact, the result of adding GRF_T 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 $T+\mathsf{GRF}_T$ with S a fixed point, then S is inconsistent.[11] In contrast, $T+\mathsf{RFN}_T$ can have models of the form $(\mathbb{N}, S)$ for S a consistent fixed point (in fact all consistent fixed points).

There is a natural explanation for the internal inconsistency of $T+\mathsf{GRF}_T$: classical theories T of the sort just mentioned are in fact *unsound* with respect to the notion of truth captured by T, and GRF_T makes this explicit. In fact, many theorems involving the truth predicate in a classical axiomatization T of the fixed-point construction based on Basic De Morgan Logic are outside the extension of the truth predicate given by consistent fixed points. The classical tautology $\lambda \vee \neg\lambda$ involving a liar sentence is one such example. Uniform reflection alone, in theories such as T, does not suffice to uncover their unsoundness:[12] this is the sense in which the intimate connection between global and uniform reflection is lost in the classical setting considered here.[13]

The close connections between the two forms of reflection just considered, however, can be restored by moving to internal axiomatizations of (Basic De Morgan) Kripke fixed points such as extensions of TS_0. To see this, we first reformulate both principles in rule form and adapt them to the sequent-style formulation of TS_0 that we have chosen.

$$\frac{\Rightarrow \mathsf{Thm}_T([Ax])}{\Rightarrow A(x)}\ (\mathsf{RFN}_T^R) \qquad\qquad \frac{\Rightarrow \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \wedge \mathsf{Thm}_T(x)}{\Rightarrow \mathsf{T}x}\ (\mathsf{GRF}_T^R).$$

Finally, we introduce an extension of TS_0 obtained by replacing the axioms ($\mathsf{T}1$) and ($\mathsf{T}2$) with

(i) $A(x) \Rightarrow \mathsf{T}[Ax]$;
(ii) $\mathsf{T}[Ax] \Rightarrow A(x)$.

We call the resulting system UTS_0 ('*uniform* TS_0). We can now establish that not only uniform and global reflection are connected in Basic De Morgan logic, but that they actually *coincide*.

PROPOSITION 1
Let T contain UTS_0. Then $T+\mathsf{RFN}_T^R$ and $T+\mathsf{GRF}_T^R$ are identical theories.

PROOF. We start by showing that global reflection entails uniform reflection. Reasoning in $T+\mathsf{GRF}_T^R$, we assume that the sequent $\Rightarrow \mathsf{Thm}_T([Ax])$ is derivable in it. Then, by GRF_T^R, we have $\mathsf{T}[Ax]$ and therefore Ax by (ii) above.

[11] By the diagonal lemma, the arithmetical part of T already proves $(\lambda \wedge \neg\mathsf{T}\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge \mathsf{T}\ulcorner\lambda\urcorner)$ for λ a liar sentence. Therefore $T+\mathsf{GRF}_T$ proves $\mathsf{T}\big(\ulcorner(\lambda \wedge \neg\mathsf{T}\ulcorner\lambda\urcorner) \vee (\neg\lambda \wedge \mathsf{T}\ulcorner\lambda\urcorner)\urcorner\big)$. Since T is an axiomatization of the class of Kripke fixed points, we can use compositional and truth-iteration principles to obtain, still in $T+\mathsf{GRF}_T$, $\mathsf{T}(\ulcorner\lambda \wedge \neg\lambda\urcorner)$. A well-known example of such a theory T is a the theory KF from [8] — see also [14].

[12] This is also the reason why Horsten and Leigh could consider iterations of uniform reflection without restrictions on the fixed-points models.

[13] This phenomenon is not confined to axiomatizations of Kripke fixed points based on De Morgan Logic. Also in supervaluational fixed points uniform and global reflection come apart.

For the other direction, we reason in $T+\mathsf{RFN}_T^R$ and assume that the sequent $\Rightarrow \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \wedge \mathsf{Thm}_T(x)$ is derivable in it. Also, we know that (i) and (ii) are derivable sequents of UTS_0 — and then also of T —, and therefore $\Rightarrow \mathsf{Pr}_T(x, [\mathsf{T}x])$ is a derivable sequent of T. By combining this latter fact with our assumption, we obtain $\Rightarrow \mathsf{Thm}_T([\mathsf{T}x])$. By RFN_T^R, therefore, we can conclude $\Rightarrow \mathsf{T}x$, as desired. ■

Proposition 1 suggests that Leigh and Horsten's project can be more coherently carried out in the context of *non-classical* theories of truth. In the next section, we will in fact employ strengthenings of the reflection principles considered in Proposition 1 to unfold the truth-theoretic and mathematical content implicit in the acceptance of TS_0.

3 Reflecting on TS_0

This section introduces the main results of the article: in Section 3.1 we discuss several alternative reflection rules and motivate the choice of a particular form of reflection on admissible rules that turns out to be stronger than simple reflection on derivable sequents. In Section 3.2 we show that the closure under two applications of our rule of reflection suffices to recover the strong internal axiomatization of Kripke's fixed point PKF. Finally, in Section 3.3 we show that the result of reflecting twice on TS_0 proves more transfinite induction for the language with the truth predicate than PKF itself. We conclude the section by investigating further iterations of reflection.

3.1 Reflection on sequents and rules

As we have seen, in the classical setting the uniform reflection schema and rule take the form:

$$\mathsf{Thm}_T([A(x)]) \Rightarrow A(x) \qquad (\mathsf{RFN}_T^R)$$

$$\frac{\Rightarrow \mathsf{Thm}_T([A(x)])}{\Rightarrow A(x)}. \qquad (\mathsf{URFN}_T^R)$$

Over EA, URFN_T^R and RFN_T^R are equivalent, as shown by Feferman in [5]. In the non-trivial direction, i.e. going from the rule to the initial sequent, one shows that Basic suffices to formalize the fact that the sentence $\mathsf{Prf}_T(\bar{n}, \ulcorner \Rightarrow A(\bar{m}) \urcorner) \rightarrow A(\bar{m})$ is provable in T for any $m, n \in \omega$. Therefore one application of (URFN_T^R) yields (RFN_T^R).

In the non-classical setting the situation is different. Whereas with classical logic on the background we can formulate (URFN_T^R) and (RFN_T^R) in a one-sided sequent calculus, there are good reasons to stick with a two-sided calculus for Basic De Morgan logic. In a one-sided classical system, in fact, sequents $A, \neg A$ play the role that initial sequents $A \Rightarrow A$ play in a two-sided setting. In our system this correspondence breaks down: first of all, $A, \neg A$ is not generally valid in our intended semantics — if A is a liar sentence, for instance —, whereas $A \Rightarrow A$ are initial sequents of our system. Moreover, there is no conditional naturally corresponding to the sequent arrow since $\Rightarrow A \rightarrow A$ is just a notational variant of $\Rightarrow A \vee \neg A$.

We therefore opt for a formulation of our first reflection principle as applying to provable (two-sided) sequents. As a consequence, basic syntactic considerations force us to formulate reflection in rule-form. The *uniform reflection principle for sequents* of T takes the following form:

$$\frac{\Rightarrow \mathsf{Pr}_T([\Gamma\vec{x} \Rightarrow \Delta\vec{y}])}{\Gamma \Rightarrow \Delta}. \qquad (\mathsf{r}_T)$$

But the simple rule of reflection (r_T) is not the only form of reflection that will be relevant for what follows. A suitable conditional — such as the classical or the intuitionistic conditional — enables one to compress in one sequent chains of reasoning featuring embedded implications. In our setting, the highly *meta-theoretic* nature of the sequent arrow forces us to capture these chains of reasoning explicitly via suitable extensions of the simple reflection rule (r_T). One way to achieve this is to focus not only on provable sequents, but also to take into account rules admissible in T via the provability predicate $\mathsf{Pr}^2_T(x,y)$ introduced above.

We define the *uniform reflection principle for provably admissible rules* in T:

$$\frac{\Rightarrow \mathsf{Pr}^2_T([\Gamma(x)\Rightarrow\Delta(x)],[\Theta(x)\Rightarrow\Lambda(x)]) \qquad \Gamma(x)\Rightarrow\Delta(x)}{\Theta(x)\Rightarrow\Lambda(x)}. \tag{R_T}$$

Obviously, in the context of any reasonable theory T, the theory $T+\mathsf{r}_T$ is a subtheory of $T+\mathsf{R}_T$. The rule R_T states that if there is a uniform T-provable proof transformation of a proof of $\Gamma(\overline{n})\Rightarrow\Delta(\overline{n})$ into a proof of $\Theta(\overline{n})\Rightarrow\Lambda(\overline{n})$ for each $n\in\omega$, and moreover $\Gamma(x)\Rightarrow\Delta(x)$ is provable, then also $\Theta(x)\Rightarrow\Lambda(x)$ is provable. The rule R_T is in a sense supplementing BDM with the possibility of dealing with chains of reasoning involving sequents. It is based on a formalization of admissible rules, which in the context of classical logic are easily compressed into sentence form and combined, informally, via conditionalization and transitivity. Although it increases the expressive power of the theories in BDM; however, it is clear that R_T does not amount to restoring classical logic, as these chains of reasoning can only be managed in the safe and controlled environment guaranteed by the formalization in the support theory: the rule R_T relies essentially on the syntactic capabilities of the background non-logical principles and is clearly not a logical principle. More importantly, the acceptability of our reflection principles, just like the acceptability of any other reflection principle, is based on preservation of soundness: as we shall see later on, the addition of R_T to our theories of truth preserves not only soundness but all semantic properties of the theories of truth in BDM, intersubstitutivity *in primis*, and therefore it can be safely assumed regardless of how it interacts with the underlying logic.

If T is an axiomatizable theory, then the *reflection on* T is the closure of Basic under the reflection rules $\mathsf{r}(T)$ and $\mathsf{R}(T)$:

$$\mathsf{r}(T) := \mathsf{Basic}+(\mathsf{r}_T)$$
$$\mathsf{R}(T) := \mathsf{Basic}+(\mathsf{R}_T).$$

Theories obtained by iterating our reflection rules are then defined in a standard manner: for instance, $\mathsf{R}(\mathsf{R}(T))$ is the result of closing $\mathsf{R}(T)$ under $\mathsf{R}_{\mathsf{R}(T)}$. We abbreviate $\mathsf{R}(\mathsf{R}(T))$ as $\mathsf{R}^2(T)$, and similarly for more iterations.

We have introduced (R_T) as a generalization of (r_T). A natural question is whether (R_T) is actually stronger than the simpler rule. We will not answer to this question in this article but we will prove some facts that may be relevant for a future answer. For instance, we now provide an upper bound for the strength of $\mathsf{r}(\mathsf{UTS}_0)$; later — cf. Proposition 3 — we will show that the resulting theory is a proper subtheory of $\mathsf{R}^2(\mathsf{TS}_0)$.

The upper bound for $\mathsf{r}(\mathsf{UTS}_0)$ that we now provide is given in terms of the theory PKF that was mentioned in the introduction. PKF is also formulated in the language $\mathcal{L}_\mathsf{T}$, and its axioms and rules are displayed in Table 2.

In $\mathsf{T}{=}_{1\text{-}2}$, the function symbol $\doteq$ represents the elementary syntactic operation of forming an identity statement out of two terms. A similar notation will be applied for other syntactic operations.

TABLE 2. The theory PKF

LOGIC	logical initial sequents and rules of BDM
IDENTITY	Id1, Id2
ARITHMETIC	$\Rightarrow A$ for A a basic axiom of EA
	plus the *full* induction rule for $\mathcal{L}_{\mathsf{T}}$:
	$\dfrac{\Gamma, A(x) \Rightarrow A(x+\bar{1}), \Delta}{\Gamma, A(0) \Rightarrow A(t), \Delta}$ (IND)
ATOMIC TRUTH	$(\mathsf{T}{=}_1)$ $\quad \mathsf{ct}(x), \mathsf{ct}(y), \mathsf{val}(x) = \mathsf{val}(y) \Rightarrow \mathsf{T}(x \doteq y)$
	$(\mathsf{T}{=}_2)$ $\quad \mathsf{ct}(x), \mathsf{ct}(y), \mathsf{T}(x \doteq y) \Rightarrow \mathsf{val}(x) = \mathsf{val}(y)$
	(TT_1) $\quad \mathsf{T}[\mathsf{T}x] \Rightarrow \mathsf{T}x$
	(TT_2) $\quad \mathsf{T}x \Rightarrow \mathsf{T}[\mathsf{T}x]$
TRUTH PRINCIPLES FOR CONNECTIVES	$(\mathsf{T}\wedge_1)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \dot{\wedge} y), \mathsf{T}(x) \wedge \mathsf{T}(y) \Rightarrow \mathsf{T}(x \dot{\wedge} y)$
	$(\mathsf{T}\wedge_2)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \dot{\wedge} y), \mathsf{T}(x \dot{\wedge} y) \Rightarrow \mathsf{T}(x) \wedge \mathsf{T}(y)$
	$(\mathsf{T}\vee_1)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \dot{\vee} y), \mathsf{T}(x) \vee \mathsf{T}(y) \Rightarrow \mathsf{T}(x \dot{\wedge} y)$
	$(\mathsf{T}\vee_2)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \dot{\vee} y), \mathsf{T}(x \dot{\vee} y) \Rightarrow \mathsf{T}(x) \vee \mathsf{T}(y)$
	$(\mathsf{T}\neg_1)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x), \mathsf{T}(\dot{\neg} x) \Rightarrow \neg \mathsf{T}(x)$
	$(\mathsf{T}\neg_2)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x), \neg \mathsf{T}(x) \Rightarrow \mathsf{T}(\dot{\neg} x)$
TRUTH PRINCIPLES FOR QUANTIFIERS	$(\mathsf{T}\forall_1)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(\dot{\forall} y x), \forall y \mathsf{T}x(y/v) \Rightarrow \mathsf{T}(\dot{\forall} y x)$
	$(\mathsf{T}\forall_2)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(\dot{\forall} y x), \mathsf{T}(\dot{\forall} y x) \Rightarrow \forall y \mathsf{T}x(y/v)$
	$(\mathsf{T}\exists_1)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(\dot{\exists} y x), \exists y \mathsf{T}x(y/v) \Rightarrow \mathsf{T}(\dot{\exists} y x)$
	$(\mathsf{T}\exists_2)$ $\quad \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(\dot{\exists} y x), \mathsf{T}(\dot{\exists} y x) \Rightarrow \exists y \mathsf{T}x(y/v)$

As mentioned earlier, PKF is an internal axiomatization of Kripke's theory of truth. Crucially, PKF is fully compositional as also negation commutes with the truth predicate. Halbach and Horsten in [15] have measured the proof-theoretic strength of PKF by showing that PKF proves arithmetical transfinite induction up to the ordinal $\varphi_\omega 0$. Therefore we can use PKF as means of comparison for our theories of iterated reflection.

PROPOSITION 2
$\mathsf{r}(\mathsf{UTS}_0)$ is a subtheory of PKF.

PROOF. To prove Proposition 2 we only need to check that PKF can handle reflection. We first establish that PKF is strong enough to prove the soundness of UTS_0. To this end, for ordinal codes α, we define a hierarchy of predicates $\mathsf{Tr}_\alpha(\cdot)$ as $\mathsf{T}(\cdot) \wedge \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}^{\dot{<}\alpha}}(\cdot)$, where $\mathcal{L}^{<\alpha}$ is defined as $\mathcal{L}$ plus all

predicates T_β for $\beta<\alpha$ (cf. [14, Ch. 9]).[14] Halbach and Horsten establish in [15] that PKF proves the predicates Tr_β to behave like Tarskian truth predicates for $\beta<\omega^\omega$: that is, for formulas of $\mathcal{L}_\mathsf{T}$ that are in $\mathcal{L}_\mathsf{T}^{<\omega}$, the classical commutation conditions for typed truth predicates hold, while for $\mathcal{L}_\mathsf{T}$-formulas A not in $\mathcal{L}_\mathsf{T}^{\omega}$, we can prove $\neg\mathsf{Tr}_\omega\ulcorner A\urcorner$.

We can extend this definition to sequents via the predicate $\mathsf{TR}_\alpha(\cdot)$ in the following way:

$$\mathsf{TR}_\alpha(\ulcorner\Gamma\Rightarrow\Delta\urcorner):\leftrightarrow\big((\mathsf{Tr}_\alpha(\ulcorner\wedge\Gamma\urcorner)\rightarrow\mathsf{Tr}_\alpha(\ulcorner\vee\Delta\urcorner))\wedge (\mathsf{Tr}_\alpha(\ulcorner\neg\vee\Delta\urcorner)\rightarrow\mathsf{Tr}_\alpha(\ulcorner\neg\wedge\Gamma\urcorner))\big).$$

From the proof-theoretic analysis of PKF in [15] we know that the truth predicates up to ω^ω behave classically in it, and that we can employ the material conditional to carry out the inductive proof of the following:

$$\mathsf{PKF}\vdash\Rightarrow\mathsf{Pr}_{\mathsf{UTS}_0}([\Gamma x\Rightarrow\Delta x])\rightarrow\mathsf{TR}_\omega([\Gamma x\Rightarrow\Delta x]). \tag{1}$$

The proof employs the induction rule of PKF. It suffices, therefore, to establish

$$\Rightarrow\mathsf{Prv}_{\mathsf{UTS}_0}(0,[\Gamma x\Rightarrow\Delta x])\rightarrow\mathsf{TR}_\omega([\Gamma x\Rightarrow\Delta x]) \tag{2}$$

$$\mathsf{Prv}_{\mathsf{UTS}_0}(u,[\Gamma x\Rightarrow\Delta x])\rightarrow\mathsf{TR}_\omega([\Gamma x\Rightarrow\Delta x])\Rightarrow \mathsf{Prv}_{\mathsf{UTS}_0}(u+1,[\Gamma x\Rightarrow\Delta x])\rightarrow\mathsf{TR}_\omega[\Gamma x\Rightarrow\Delta x]), \tag{3}$$

where $\mathsf{Prv}_T(x,y)$ expresses that y is provable in T with a proof of length less or equal to x (see page 2636).

We consider the crucial case of the characterizing principles of UTS_0, (i)–(ii) on page 2638. Reasoning classically in PKF, we assume

$$\Rightarrow\mathsf{Prv}_{\mathsf{UTS}_0}(0,[\mathsf{T}[Ax]\Rightarrow Ax]). \tag{4}$$

We need to show

$$\Rightarrow\mathsf{Tr}_\omega[\mathsf{T}[Ax]]\rightarrow\mathsf{Tr}_\omega[Ax] \tag{5}$$

$$\Rightarrow\mathsf{Tr}_\omega[\neg Ax]\rightarrow\mathsf{Tr}_\omega[\neg\mathsf{T}[Ax]]. \tag{6}$$

We start with (5) and reason informally: if $\mathsf{Tr}_\omega[\mathsf{T}[Ax]]$, then for some $n\in\omega$ and $m<n$, $\mathsf{Tr}_n[\mathsf{T}_m[Ax]]$. Therefore, since Tr_n and Tr_m are Tarskian truth predicates, also $\mathsf{Tr}_n[Ax]$.

Similarly for (6), if $\mathsf{Tr}_\omega[\neg Ax]$, then $\mathsf{Tr}_n[\neg Ax]$ for some $n\in\omega$, and, since $\mathsf{Tr}_n[Ax]$ is in $\mathcal{L}_\mathsf{T}^{n+1}$, also $\mathsf{Tr}_{n+1}\neg[\mathsf{T}_n[Ax]]$ and therefore $\mathsf{Tr}_\omega[\neg\mathsf{T}[Ax]]$. ■

3.2 Recovering compositionality by reflection

One of the goals of this section is to show that by reflecting on our core laws of truth we can recover desirable compositional principles. More specifically, reflecting on TS_0 is sufficient to recover the initial sequents and the full induction rule of PKF.

In a first step we show that adding the reflection principle for TS_0 to Basic allows us to derive the initial sequents of UTS_0.

[14] The truth predicates Tr_α can be defined for as many ordinals as we can code in our theory. In Section 3.3, in particular, we will employ a coding for ordinals smaller than Γ_0.

LEMMA 3
$\mathsf{UTS}_0 \subseteq \mathsf{r}(\mathsf{TS}_0)$.

PROOF. For all $n \in \omega$ we have $\mathsf{TS}_0 \vdash \mathsf{T}(\ulcorner A(n) \urcorner) \Rightarrow A(n)$. Therefore, Basic proves:

$$\Rightarrow \forall y(\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(y) \rightarrow \mathsf{Ax}_{\mathsf{TS}_0}(\mathsf{sub}(\ulcorner \mathsf{T}x \urcorner, \mathsf{num}(y)), y)) \tag{7}$$

and

$$\Rightarrow \forall x(\mathsf{Sent}_{\mathcal{L}}([Bx]), \tag{8}$$

where, we recall, $[Bx] := \mathsf{sub}(\ulcorner Bv \urcorner, \mathsf{num}(x))$ for all $\mathcal{L}_\mathsf{T}$-formulas B with one free variable. Therefore, by combining (7) and (8), we also have in Basic

$$\Rightarrow \mathsf{Ax}_{\mathsf{TS}_0}(\mathsf{sub}(\ulcorner \mathsf{T}x \urcorner, \mathsf{num}([Ax]), [Ax])). \tag{9}$$

Therefore, by definition of the canonical provability predicate $\mathsf{Pr}_{\mathsf{TS}_0}$,

$$\Rightarrow \mathsf{Pr}_{\mathsf{TS}_0}(\mathsf{sub}(\ulcorner \mathsf{T}x \urcorner, \mathsf{num}([Ax])), [Ax]). \tag{10}$$

Recall that $\mathsf{tr}(x)$ is the elementary function that formally prefixes $\ulcorner \mathsf{T} \urcorner$ to the numeral for x. Then Basic also proves the equation:

$$\Rightarrow \mathsf{sub}(\ulcorner \mathsf{T}x \urcorner, \mathsf{num}([Ax])) = \mathsf{sub}(\mathsf{tr}([Av]), \mathsf{num}(x)). \tag{11}$$

By performing the appropriate substitution in (10), we have

$$\Rightarrow \mathsf{Pr}_{\mathsf{TS}_0}(\mathsf{sub}(\mathsf{tr}([Av]), \mathsf{num}(x)), [Ax]). \tag{12}$$

The other direction is analogous.

In $\mathsf{r}(\mathsf{TS}_0)$ — and a fortiori in $\mathsf{R}(\mathsf{TS}_0)$ — therefore, we obtain

$$\mathsf{T}[Ax] \Rightarrow A(x)$$
$$A(x) \Rightarrow \mathsf{T}[Ax]$$

as desired. ■

As a consequence of the previous lemma, in $\mathsf{r}(\mathsf{TS}_0)$ we can already prove the full truth sequents for atomic arithmetical formulas and for truth ascriptions containing free variables ($\mathsf{T}{=}_1$), ($\mathsf{T}{=}_2$), (TT_1), and (TT_2). That (TT_1) and (TT_2) are direct instances of the initial truth sequents of UTS_0 is immediate. For the identity initial sequents, a slightly more general version of the Tarski sequents would be required, namely one in which at least two free variables appear. However, since we are working over Basic, we can always assume that the free variable in the truth sequents of UTS_0 stands for (the code of) a string of free variables of finite length.

Also the initial, compositional sequents of PKF for the propositional connectives $\wedge, \vee, \neg$ can be proved in $\mathsf{r}(\mathsf{TS}_0)$:

LEMMA 4
In $\mathsf{r}(\mathsf{TS}_0)$ we can derive ($\mathsf{T}\wedge_{1\text{-}2}$), ($\mathsf{T}\vee_{1\text{-}2}$), ($\mathsf{T}\neg_{1\text{-}2}$).

PROOF. In TS_0 we can directly prove the schematic form of the compositional clauses, for example $\mathsf{TS}_0 \vdash \mathsf{T}(\ulcorner A \urcorner) \wedge \mathsf{T}(\ulcorner B \urcorner) \Rightarrow (\mathsf{T}\ulcorner A \wedge B \urcorner)$ for all $\mathcal{L}_\mathsf{T}$-sentences A, B. By formalizing this fact in Basic, we obtain

$$\Rightarrow \mathsf{Pr}_{\mathsf{TS}_0}([\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x \dot{\wedge} y), \mathsf{T}x \wedge \mathsf{T}y \Rightarrow \mathsf{T}(x \dot{\wedge} y)]). \tag{13}$$

In $\mathsf{r}(\mathsf{TS}_0)$, therefore, we can then move from the formalization to the full quantifiable statement

$$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x \dot{\wedge} y), \mathsf{T}x \wedge \mathsf{T}y \Rightarrow \mathsf{T}(x \dot{\wedge} y) \tag{14}$$

as desired. The cases of the other connectives are analogous. ■

However, by looking at Table 2 one realizes that compositional initial sequents for the propositional connectives by themselves are not enough to capture all truth principles of PKF: we also need initial sequents for quantifiers and full induction for $\mathcal{L}_\mathsf{T}$ (IND). We will first show how to recover full induction.

It is a well-known result that (uniformly) reflecting on EA suffices to obtain the full induction schema for $\mathcal{L}$. Kreisel and Lévy, in [19], proved the equivalence of uniform reflection and full induction over EA — that is the equivalence of EA plus uniform reflection for EA and Peano Arithmetic (PA).[15] We will apply Kreisel and Levy's strategy to our setting. To do so, however, their original argument has to be modified in several respects. First of all, we allow formulas of $\mathcal{L}_\mathsf{T}$ and not just of $\mathcal{L}$ to appear in instances of the induction schema. In addition, we have to consider an induction *rule* because the induction axiom involving the material conditional fails to be sound in the setting of Basic De Morgan Logic. Finally, already in this step, we shall employ our generalized reflection rule R_T instead of the basic reflection rule r_T. In what follows we denote as PA_T the version of PA formulated in $\mathcal{L}_\mathsf{T}$ whose logic is BDM and in which the truth predicate can appear in instances of induction.

LEMMA 5
$\mathsf{PA}_\mathsf{T} \subseteq \mathsf{R}(\mathsf{Basic})$.

PROOF. Let $A(x)$ be a formula in $\mathcal{L}_\mathsf{T}$ with one free variable. We want to show that in R(Basic) the full induction rule

$$\frac{\Gamma, A(x) \Rightarrow A(x + \bar{1}), \Delta}{\Gamma, A(0) \Rightarrow A(t), \Delta} \tag{15}$$

for formulas of $\mathcal{L}_\mathsf{T}$ is admissible. The following inference is admissible in Basic – and in fact in predicate logic in $\mathcal{L}_\mathsf{T}$ only — for any $n \in \omega$:

$$\frac{\Gamma, A(x) \Rightarrow A(x + \bar{1}), \Delta}{\Gamma, A(0) \Rightarrow A(\bar{n}), \Delta}. \tag{16}$$

By (Pr1), since the proof transformation in (16) is elementary, Basic proves

$$\Rightarrow \mathsf{Pr}^2_{\mathsf{Basic}}(\ulcorner \Gamma, A(x) \Rightarrow A(x + \bar{1}), \Delta \urcorner, \ulcorner \Gamma, A(0) \Rightarrow A(\dot{y}), \Delta \urcorner). \tag{17}$$

Now by assumption, (17), and $\mathsf{R}_{\mathsf{Basic}}$ we conclude

$$\Gamma, A(0) \Rightarrow A(y), \Delta.$$

■

[15] See Beklemishev [2, Cor. 4.3] for a proof of this fact.

The full set of compositional sequents of PKF is obtained by complementing the clauses for the connectives by the ones for quantifiers. This can be achieved by closing the theory $\mathsf{R}(\mathsf{TS}_0)$ under $\mathsf{R}_{\mathsf{R}(\mathsf{TS}_0)}$, that is, by performing one iteration of the general reflection rule.

LEMMA 6
$\mathsf{R}^2(\mathsf{TS}_0)$ proves $(\mathsf{T}\forall_{1\text{-}2})$ and $(\mathsf{T}\exists_{1\text{-}2})$.

PROOF. We prove $\mathsf{T}\forall_1$; the other cases are treated similarly. For all $\mathcal{L}_\mathsf{T}$-formulas $A(v)$ with only v free, $\mathsf{R}(\mathsf{TS}_0)$ proves

$\mathsf{T}[Ay] \Rightarrow A(y)$	by Lemma 3
$\forall y \mathsf{T}[Ay] \Rightarrow \forall y A(y)$	by logic
$\forall y \mathsf{T}[Ay] \Rightarrow \mathsf{T}^\ulcorner \forall y A(y)^\urcorner$	by (T2).

The argument just carried out in $\mathsf{R}(\mathsf{TS}_0)$ can uniformly be formalized in Basic, i.e., Basic proves:

$$\Rightarrow \mathsf{Pr}_{\mathsf{R}(\mathsf{TS}_0)}(^\ulcorner \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall \dot{y} \dot{x}), \forall y \mathsf{T} \dot{x}(y/v) \Rightarrow \mathsf{T}(\forall y \dot{x}(y/v))^\urcorner).$$

Therefore $\mathsf{R}^2(\mathsf{TS}_0)$ suffices to conclude

$$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall \dot{y} x), \forall y \mathsf{T} x(y/v) \Rightarrow \mathsf{T}(\forall y x(y/v)),$$

as desired. ■

COROLLARY 1
$\mathsf{PKF} \subseteq \mathsf{R}^2(\mathsf{TS}_0)$.

Corollary 1 shows that two iterations of the generalized rule $\mathsf{R}_{\mathsf{TS}_0}$ over our basic theory TS_0 suffice to recover all compositional truth laws that were not immediately provable in the original theory as well as the full induction rule for the language $\mathcal{L}_\mathsf{T}$. If reflection is considered to be a procedure already implicit in the acceptance of TS_0, then the laws of PKF follow naturally from a few applications of this process. However, it is natural to ask whether the inclusion established in Cor. 1 is proper.

These questions translate, on the conceptual side, into the task of approximating the set of sentences that are valid in the intended models of our theories, which are the Kripke fixed-point models. In doing so, we gather information on how many truth iterations and general claims involving truth we are permitted to assert upon accepting TS_0 (after reflection) and how many mathematical patterns of reasoning we regain in the form of transfinite induction.

3.3 Recovering transfinite induction by reflection

In this section, we investigate the question of how much transfinite induction for $\mathcal{L}_\mathsf{T}$ can be recovered in iterations of the generalized reflection rule over TS_0. One of the upshots of our analysis will be that $\mathsf{R}^2(\mathsf{TS}_0)$ properly extends PKF.

To carry out our proofs, we need to assume a notation system $(\mathsf{OT}, \prec)$ for ordinals up to the Feferman–Schütte ordinal Γ_0 as it can be found, for instance, in [24]. OT is a primitive recursive set of ordinal codes and $\prec$ a primitive recursive relation on OT that is isomorphic to the usual ordering of ordinals up to Γ_0. We distinguish between fixed ordinal codes, which we denote with $\alpha, \beta, \gamma \ldots$, and $\zeta, \eta, \theta \ldots$ as abbreviations for variables ranging over elements of OT. From the results in [15] it

follows that PKF proves transfinite induction for $\mathcal{L}_\mathsf{T}$ only up to any ordinal smaller than ω^ω. If we focus only on $\mathcal{L}$-formulas, however, PKF proves transfinite induction for much higher ordinals. In particular, PKF proves the same arithmetical sentences as PA plus transfinite induction for $\mathcal{L}$ up to any ordinal smaller than $\varphi_\omega 0$.

Before analysing how much transfinite induction can be proved in $\mathsf{R}^2(\mathsf{TS}_0)$, we introduce some notation. The schema of transfinite induction up to α for the formula $A(v)$ of a language $\mathcal{L}_1$ containing $\mathcal{L}$ is the rule

$$\frac{\forall \xi \prec \eta A(\xi) \Rightarrow A(\eta)}{\Rightarrow \forall \xi \prec \alpha A(\xi)}\ \mathsf{TI}_{\mathcal{L}_1}(A,\alpha).$$

We then denote transfinite induction up to some ordinal α with $\mathsf{TI}_{\mathcal{L}_1}(\prec\alpha)$, standing for the closure under all rules $\mathsf{TI}_{\mathcal{L}_1}(A,\beta)$ for $A \in \mathcal{L}_1$ and $\beta \prec \alpha$. Analogously, we write $\mathsf{TI}_{\mathcal{L}_1}(\alpha)$ for the closure under all rules $\mathsf{TI}_{\mathcal{L}_1}(A,\alpha)$ for $A \in \mathcal{L}_1$. In what follows, we will only deal with the cases in which $\mathcal{L}_1$ is either $\mathcal{L}$ itself or $\mathcal{L}_\mathsf{T}$.

As a measure of strength of the theories obtained via iteration of reflection, we will mainly focus on how much transfinite induction for $\mathcal{L}_\mathsf{T}$ is derivable in such theories. However, there is often a direct connection between the amount of transfinite induction for $\mathcal{L}_\mathsf{T}$ and $\mathcal{L}$ derivable in a truth theory. Both in the case of KF and PKF, for instance, the amount of transfinite induction for $\mathcal{L}_\mathsf{T}$ available in the systems — that is $\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(\prec\varphi_1 0)$ and $\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(\prec\varphi_0\omega)$ respectively — can be used to define classical, Tarskian truth predicates indexed by these ordinals with the crucial contribution of the compositional truth principles of the two theories (see Fefeman's [8] for KF and Halbach and Horsten's [15] for PKF). This gives a lower bound for the systems in terms of ramified truth hierarchies up to $\varphi_1 0$ (or ε_0) and $\varphi_0\omega$ (or ω^ω) respectively, which — by a classical result by Feferman (cf. [7]) — yields that KF and PKF are proof-theoretically as strong as at least $\mathsf{PA}+\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(\prec\varphi_{\varepsilon_0}0)$ and $\mathsf{PA}+\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(\prec\varphi_\omega 0)$ respectively.

The following proposition shows that iterating the generalized reflection rule twice over TS_0 enables us to go beyond PKF. This also gives us more information about the question that was posed on page 2640 about the comparison between the rules (r_T) and (R_T). By Proposition 2, the theory $\mathsf{r}(\mathsf{UTS}_0)$ is a subtheory of PKF. The next will entail that $\mathsf{R}^2(\mathsf{TS}_0)$ is indeed stronger than PKF.

PROPOSITION 3
$\mathsf{R}^2(\mathsf{Basic}) \vdash \mathsf{TI}_{\mathcal{L}_\mathsf{T}}(\omega^\omega)$.

PROOF. We first prove in $\mathsf{R}(\mathsf{Basic})$ that, for all $n \in \omega$,

$$\frac{\Gamma, \forall \zeta \prec \eta A(\zeta) \Rightarrow A(\eta), \Delta}{\Gamma \Rightarrow \forall \zeta \prec \omega^n A(\zeta), \Delta}. \tag{18}$$

To prove (18), we first prove in $\mathsf{R}(\mathsf{Basic})$, for all $n \in \omega$:

$$\frac{\forall \zeta \prec \eta A(\zeta) \Rightarrow A(\eta)}{\forall \zeta \prec \eta A(\zeta) \Rightarrow \forall \zeta \prec \eta + \omega^n A(\zeta)}. \tag{19}$$

We reason as follows in $\mathsf{R}(\mathsf{Basic})$:

$$\forall\zeta\prec\eta A(\zeta)\Rightarrow A(\eta) \tag{20}$$

$$\forall\zeta\prec\eta A(\zeta)\Rightarrow\forall\zeta\prec\eta+\omega^0 A(\zeta) \quad \text{by (20)} \tag{21}$$

$$\forall\zeta\prec\eta A(\zeta)\Rightarrow\forall\zeta\prec\eta+\omega^n A(\zeta) \quad \text{external ind. hyp.} \tag{22}$$

$$\forall\zeta\prec\eta+(\omega^n\times x)A(\zeta)\Rightarrow\forall\zeta\prec\eta+(\omega^n\times x)+\omega^n A(\zeta) \quad \text{from (22)} \tag{23}$$

$$\forall\zeta\prec\eta+(\omega^n\times 0)A(\zeta)\Rightarrow\forall x\forall\zeta\prec\eta+(\omega^n\times x)A(\zeta) \quad \text{by (IND)} \tag{24}$$

$$\forall\zeta\prec\eta A(\zeta)\Rightarrow\forall\zeta\prec\eta+\omega^{n+1} A(\zeta) \tag{25}$$

The last two lines give us the induction step and therefore (19) by, possibly, a series of cuts.

Now in Basic,

$$\Rightarrow \mathsf{Pr}^2_{\mathsf{R}(\mathsf{Basic})}([\forall\zeta\prec\eta A(\zeta)\Rightarrow A(\eta)],[\forall\zeta\prec\eta A(\zeta)\Rightarrow\forall\zeta\prec\eta+\omega^x A(\zeta)]). \tag{26}$$

Therefore, in $\mathsf{R}^2(\mathsf{Basic})$,

$$\frac{\forall\zeta\prec\eta A(\zeta)\Rightarrow A(\eta)}{\forall\zeta\prec\eta A(\zeta)\Rightarrow\forall x\forall\zeta\prec\eta+\omega^x A(\zeta)}. \tag{27}$$

That is

$$\frac{\forall\zeta\prec\eta A(\zeta)\Rightarrow A(\eta)}{\forall\zeta\prec\eta A(\zeta)\Rightarrow\forall\zeta\prec\eta+\omega^\omega A(\zeta)}. \tag{28}$$

From (28), by letting η to be 0, we get

$$\frac{\Gamma,\forall\zeta\prec\eta A(\zeta)\Rightarrow A(\eta),\Delta}{\Gamma\Rightarrow\forall\zeta\prec\omega^\omega A(\zeta),\Delta}. \tag{29}$$

■

By the proof theoretic analysis of PKF we know that it can only prove transfinite induction for $\mathcal{L}_\mathsf{T}$ for ordinals smaller than ω^ω. But this fact is not dependent in any way on the truth theoretic principles of PKF: already PA_T, in fact, proves $\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(<\omega^\omega)$. This is also reflected by the fact that Proposition 3 does not rely on the truth principles of TS_0. However, by Corollary 1, we have:

COROLLARY 2
PKF is a proper subtheory of $\mathsf{R}^2(\mathsf{TS}_0)$.

Transfinite induction up to ω^ω, however, is clearly not the limit of what we can achieve in $\mathsf{R}^2(\mathsf{Basic})$. By using similar methods to the ones employed in Proposition 3, and starting from (28), we can verify that the following rule is admissible in $\mathsf{R}^2(\mathsf{Basic})$:

$$\frac{\forall\zeta\prec\eta A(\zeta)\Rightarrow A(\eta)}{\forall\zeta\prec\theta A(\zeta)\Rightarrow\forall\zeta\prec\theta+\omega^{\omega+k} A(\zeta)}.$$

Generalizing this strategy it is possible to show the following:

LEMMA 7
In $\mathsf{R}^{(n+1)}(\mathsf{Basic})$ the following rule is admissible:

$$\frac{\forall\zeta\prec\eta A(\zeta)\Rightarrow A(\eta)}{\forall\zeta\prec\theta A(\zeta)\Rightarrow\forall\zeta\prec\theta+\omega^{\omega\times n} A(\zeta)}.$$

PROOF. By external induction on n. We have established the claim for $n=1$. Assume that it holds for n. Then we can argue in $\mathsf{R}^{(n+1)}(\mathsf{Basic})$: Assume

$$\forall\zeta \prec \eta A(\zeta) \Rightarrow A(\eta).$$

Then by the induction hypothesis we have

$$\forall\zeta \prec \theta A(\zeta) \Rightarrow \forall\zeta \prec \theta+\omega^{\omega\times n} A(\zeta)$$

and

$$\forall\zeta \prec \theta+\omega^{\omega\times n}\times x\, A(\zeta) \Rightarrow \forall\zeta \prec \theta+\omega^{\omega\times n}\times x+\omega^{\omega\times n} A(\zeta).$$

By the induction principle (IND) we obtain

$$\forall\zeta \prec \theta+\omega^{\omega\times n}\times 0\, A(\zeta) \Rightarrow \forall x\forall\zeta \prec \theta+\omega^{\omega\times n}\times x\, A(\zeta)$$

giving us

$$\forall\zeta \prec \theta A(\zeta) \Rightarrow \forall\zeta \prec \theta+\omega^{\omega\times n}\times\omega\, A(\zeta)$$

which is

$$\forall\zeta \prec \theta A(\zeta) \Rightarrow \forall\zeta \prec \theta+\omega^{\omega\times n+1} A(\zeta).$$

Therefore, by iterating this argument m-times, we can obtain, for each m:

$$\forall\zeta \prec \theta A(\zeta) \Rightarrow \forall\zeta \prec \theta+\omega^{\omega\times n+m} A(\zeta).$$

In $\mathsf{R}^{(n+2)}(\mathsf{Basic})$, therefore, we can conclude

$$\forall\zeta \prec \theta A(\zeta) \Rightarrow \forall\zeta \prec \theta+\omega^{\omega\times(n+1)} A(\zeta).$$

■

Lemma 7 immediately entails that $\mathsf{R}^{n+1}(\mathsf{Basic})$ proves $\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(<\omega^{\omega\times n})$. Therefore, if we reflect on TS_0 instead of Basic, we are able to define in $\mathsf{R}^n(\mathsf{TS}_0)$ ramified truth predicates for any ordinal smaller than $\omega^{\omega\times n}$ by following the strategy employed by Halbach and Horsten in [15] and hinted at on page 2646.

The strategy employed in Lemma 7 can be iterated even further. Ideally, we would like to reach, by as little reflection iterations as possible, the amount of transfinite induction for $\mathcal{L}_\mathsf{T}$ — and therefore of ramified truth predicates — that are available in KF, the classical counterpart of PKF. However, we conclude this section by providing only a first, and presumably rather inefficient, approximation to this task.

By letting $\mathsf{R}^\omega(\mathsf{Basic}) := \bigcup_{n\in\omega}\mathsf{R}^n(\mathsf{Basic})$, a direct consequence of Lemma 7 is that

COROLLARY 3
In $\mathsf{R}^\omega(\mathsf{Basic})$ we have $\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(<\omega^{(\omega^2)})$

Therefore the theory $\mathsf{R}^\omega(\mathsf{TS}_0)$ can define ramified truth predicates indexed by all ordinals $\omega^{\omega\times n}$ for all natural numbers n.

Although ω may seem to be a natural stopping point, the procedure can be iterated even further into the transfinite. Following a well-known tradition initiated by Feferman in [6], the theories $\mathsf{R}^n(\mathsf{Basic})$ can all be shown to be recursively enumerable. Moreover, the notion of being a proof in $\mathsf{R}^n(\mathsf{Basic})$

is recursive. We can then find a primitive recursive function enumerating all those proof predicates. By employing the recursion theorem, therefore, we can find an index for this enumeration that can be used to formalize, via a recursive predicate, the notion of being a proof employing rules proper of one of the theories $\mathsf{R}^n(\mathsf{Basic})$. This, however, suffices to formulate the notion of being a proof in $\mathsf{R}^\omega(\mathsf{Basic})$: clearly, similar procedure can be extended at least to ordinals smaller than ε_0.

But once a recursive formalization of transfinite iterations of our reflection rules is available, it becomes clear that enough iterations of reflection over Basic will lead us to the amount of transfinite induction for $\mathcal{L}_\mathsf{T}$ available in KF. By letting $\omega_0 := 1$, and $\omega_{n+1} := \omega^{\omega_n}$, we have, rather unsurprisingly,

OBSERVATION 1
$\mathsf{R}^{\omega_n+1}(\mathsf{Basic}) \vdash \mathsf{TI}_{\mathcal{L}_\mathsf{T}}(\omega_n)$.

4 Conclusion

Starting with principles that are minimally constitutive of the notion of truth, such as the initial sequents of the theory TS_0, we have investigated the result of iterating reflection rules over them. A similar project, in the context of classical logic and therefore without the basic principles of TS_0, has been recently pursued by Horsten and Leigh [18]. We claim that our non-classical setting provides a more coherent framework for such a project for two main reasons. First, in a *classical* setting the intersubstitutivity of A and $\mathsf{T}\ulcorner A \urcorner$ (which is the defining characteristic of TS_0) cannot be consistently maintained. Second, following a theme by Kreisel, the global reflection GRF_T for a theory T is the *intended* soundness extension of T. Other proof theoretic reflection principles, including the uniform reflection principle RFN_T, are only justified by an appeal to global reflection. However, as shown in Section 2.3, in classical axiomatizations of Kripke's fixed point constructions, the use of the global reflection principle is at odds with the overall strategy of iterating reflection rules.

One way to understand the results of this article is by asking which statements TS_0 and the result of iterating reflection rules over it can prove to be true, i.e., by considering their provable sequents of the form $\Rightarrow \mathsf{T}\ulcorner A \urcorner$ for A in $\mathcal{L}_\mathsf{T}$ or, in short, their *truth theorems*. The logic BDM in itself — i.e. without identity — has no theorems at all. When initial sequents for identity are added to it as well as arithmetical initial sequents, even if the truth predicate is in the signature of the theory, one only obtains arithmetical theorems but no truth theorems. TS_0, in contrast, does prove truth theorems, but only truth theorems of the form

$$\underbrace{\mathsf{T}\ldots\mathsf{T}}_{\text{n-times}}\ulcorner A \urcorner,$$

where A is an arithmetical theorem of Basic. This shortcoming of TS_0 is accompanied by the lack of other desirable properties of the theory, such as full compositionality (see again Section 2.3). By adding a uniform or global *reflection rule* to TS_0, we restore our full capability of reasoning inductively with the truth predicate, and several compositional truth sequents. Full compositionality, together with the possibility of establishing theorems of the form

$$\underbrace{\mathsf{T}\ldots\mathsf{T}}_{\omega^{\omega+n}\text{-times}}\ulcorner A \urcorner$$

for A again an arithmetical theorem of Basic, is reached when we consider the theory $\mathsf{R}^2(\mathsf{TS}_0)$, i.e., via a further iteration of the generalized reflection rule R_T over TS_0. At this stage, we already recapture and surpass all truth theorems of the full compositional theory PKF. A natural goal for the

process of iteration may be to reach the truth theorems of the classical theory KF (or equivalently, $\mathsf{PKF}+\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(<\varepsilon_0)$, as shown in [23]). This can be achieved via suitable transfinite progressions of theories obtained by reflection over TS_0.

From a semantic perspective, there is a tight match between the truth theorems of our theories and the levels of the construction of the minimal fixed point of Kripke's construction from [20]. By extending TS_0 with an ω-rule, this connection can be made explicit: the theorems of TS_0 plus the ω-rule are exactly the $\mathcal{L}_\mathsf{T}$-sentences that are in the extension of the truth predicate in the minimal fixed point of Kripke's theory (see [10] for a recent proof). Uniform reflection principles are recursive approximations of the ω-rule. Therefore iterations of reflection, and the corresponding truth theorems of the resulting theories, can be seen as approximations to the full ω-rule added to TS_0 as they represent initial stages of the construction of the minimal fixed point. It is also clear that all the theories that we have considered are *internal* axiomatizations of Kripke fixed points. Therefore the hierarchy that we have studied can also be seen as an attempt to capture, via recursively axiomatized theories, the set of grounded sentences first isolated by Kripke.

Nonetheless our work leaves many open questions and possibilities for improvement. From a technical point of view, a sharper proof-theoretic analysis of the theories obtained by iterated reflection would be desirable to see clearly, for instance, how much one can obtain with finite iterations of reflection. Moreover, it would be interesting to see whether the reflection rules can be strengthened via 'higher-order' reflection rules in such a way that only finitely many iterations of them can suffice to reach the truth theorems of KF. Finally, there remains the question whether the gap between $\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(<\omega^\omega)$ and $\mathsf{TI}_{\mathcal{L}_\mathsf{T}}(<\varepsilon_0)$ — which is determined by whether PA in the signature of $\mathcal{L}_\mathsf{T}$ is formulated in BDM or classical logic respectively — can be closed by supplementing BDM with a suitable conditional in such a way that the conceptual advantages of the treatment of truth in TS_0 are preserved.

Acknowledgements

This research was supported by the DFG project "Syntactical Treatments of Interacting Modalities" and by the European Commission (grant 658285 - FOREMOTIONS). The authors would like to thank Volker Halbach, Graham Leigh, Johannes Stern, and the two anonymous referees for their comments and suggestions.

References

[1] P. Aczel. Frege structures and the notions of proposition, truth and set. In *Journal of Symbolic Logic*, J. Barwise, H. J. Keisler and K. Kunen, eds, pp. 244–246. North-Holland, 1980.

[2] L. Beklemishev. Reflection principles and provability algebras in formal arithmetic. *Russian Mathematical Surveys*, **60**, 197–268, 2005.

[3] S. Blamey. Partial logic. In *Handbook of Philosophical Logic*, Vol. 5, D. Gabbay and F. Guenthner, eds, pp. 261–353. 2 edn, 2002.

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

[5] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, **XLIX**, 35–92, 1960.

[6] S. Feferman. Transfinite recursive progression of axiomatic theories. *The Journal of Symbolic Logic*, **27**, 259–316, 1962.

[7] S. Feferman. Systems of predicative analysis. *Journal of Symbolic Logic*, **29**, 1–30, 1964.

[8] S. Feferman. Reflecting on incompleteness. *The Journal of Symbolic Logic*, **56**, 1–47, 1991.
[9] H. Field. *Saving Truth from Paradox*. Oxford University Press, 2008.
[10] M. Fischer and N. Gratzl. Infinitary proof systems and partial truth. Submitted.
[11] H. Friedman and M. Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, **33**, 1–21, 1987.
[12] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic*. Springer, 1993.
[13] V. Halbach. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, **35**, 311–27, 1994.
[14] V. Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, revised edition, 2014.
[15] V. Halbach and L. Horsten. Axiomatizing Kripke's theory of truth. *The Journal of Symbolic Logic*, **71**, 677–712, 2006.
[16] V. Halbach and C. Nicolai. On the costs of nonclassical logic. *Journal of Philosophical Logic*, 2017, doi:10.1007/s10992-017-9424-3.
[17] L. Horsten and V. Halbach. Norms for theories of reflexive truth. In *Unifying the Philosophy of Truth*, K. Fujimoto, J. M. Fernández, H. Galinon and T. Achourioti, eds. Springer, 2015.
[18] L. Horsten and G. E. Leigh. Truth is simple. *Mind*, **126**, 195–232, 2017.
[19] G. Kreisel and A. Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **14**, 97–142, 1968.
[20] S. Kripke. Outline of a theory of truth. *The Journal of Philosophy*, **72**, 690–716, 1975.
[21] H. Leitgeb. What theories of truth should be like (but cannot be). *Philosophy Compass*, **2**, 276–290, 2007.
[22] S. Negri and J. Von Plato. Cut elimination in the presence of axioms. *Bulletin of Symbolic Logic*, **4**, 418–435, 1998.
[23] C. Nicolai. Provably true sentences across axiomatizations of Kripke's theory of truth. *Studia Logica*, pp. 1–30, 2017. doi:10.1007/s11225-017-9727-y.
[24] W. Pohlers. *Proof Theory, The First Step into Impredicativity*. Springer, 2009.
[25] H. Schwichtenberg. *Proofs and Computations*. Cambridge University Press, 2012.
[26] C. Smorynski. The incompleteness theorems. In *Handbook of Mathematical Logic*, J. Barwise, ed., pp. 821–865. Dordrecht, 1977.
[27] T. Williamson. Semantic paradoxes and abductive methodology. In *Reflections on the Liar*, B. Armour-Garb, ed. Oxford University Press, 2017.

Received 6 March 2017