PROVABLY TRUE SENTENCES ACROSS AXIOMATIZATIONS OF KRIPKE'S THEORY OF TRUTH

CARLO NICOLAI
LMU MUNICH

Abstract. We study the relationships between two clusters of axiomatizations of Kripke's fixed-point models for languages containing a self-applicable truth predicate. The first cluster is represented by what we will call 'PKF-like' theories, originating in recent work Halbach and Horsten, whose axioms and rules (in Basic De Morgan Logic) are all valid in fixed-point models; the second by 'KF-like' theories first introduced by Solomon Feferman, that lose this property but reflect the classicality of the metatheory in which Kripke's construction is carried out. We show that to any natural system in one cluster – corresponding to natural variations on induction schemata – there is a corresponding system in the other proving the same sentences true, addressing a problem left open by Halbach and Horsten and accomplishing a suitably modified version of the project sketched by Reinhardt aiming at an instrumental reading of classical theories of self-applicable truth.

1. Introduction

Kripke's *Outline of a Theory of Truth* [13] can be reasonably considered one of the most influential works on truth and the liar paradox since Tarski's seminal paper [22]. By employing the metamathematics of inductive definitions, Kripke provides a class of models (fixed-point models) for a language containing a truth predicate in which any sentence ϕ has the same semantic value as the sentence 'ϕ is true'. This approximates the fundamental intuition that the thought expressed by ϕ coincides with the thought expressed by 'ϕ is true'. In Kripke's original construction, however, a liar sentence λ will be neither (definitely) true nor (definitely) false. The 'logic' of the models is forced to be nonclassical.[1]

The recent literature offers two main alternatives to axiomatize Kripke's semantic construction: on the one hand, one can remain faithful to the reasoning available in fixed-point models and resort only to rules and axioms that are sound with respect to these models. We will be interested in a particular collection of theories of this sort that we will call 'the PKF-like cluster' from the name of its most studied representative PKF – from 'Partial Kripke-Feferman' – first introduced in [9]. Alternatively, one may remain faithful to the classicality of the metatheory in which the class of models is given: to this route corresponds a collection of theories that we call 'the KF-cluster' from the name of the well-known system KF – from 'Kripke-Feferman' – introduced in [3, 6].

KF and PKF are, in a sense, on a par: they share compositional truth-theoretic principles and feature open-ended induction principles. What they don't share, of course, is the underlying logic, for the reasons sketched above. It is therefore at least surprising that PKF and KF spectacularly diverge in their proof-theoretic strength (see again [6, 8]): PKF proves the same arithmetical

[1]We will employ a more general variant of the construction which can be found, for instance, in [8], in which λ will be neither true nor false or both true and false.

sentences as PA plus arithmetical transfinite induction up to any ordinal smaller than $\varphi_\omega 0$,[2] or as the theory of ramified truth up to ω^ω. KF, by contrast, matches the arithmetical consequences of PA plus arithmetical induction up to any ordinal smaller than $\varphi_{\varepsilon_0}0$, or of ramified truth up to ε_0. This also means that PKF, although preserving the 'internal' reasoning available in a fixed-point model, cannot deem true many sentences that are nonetheless forced by KF into the extension of the truth predicate. This already leads to the natural question of finding a subtheory of KF whose theorems of the form $\mathsf{T}\ulcorner\phi\urcorner$ coincide with claims of the same form derivable in PKF. This problem was already considered by Halbach and Horsten in [9], where it was also asked whether the theory in question could be KFI, a version of KF featuring only internal truth induction ([9, p. 701]). Corollary 3 answers this question positively.

Similarly, it seems natural to investigate whether there is a system, in the PKF-cluster, that declares true the same sentences as KF. Corollary 4, in combination with Corollary 1, identifies this system in PKF extended with a transfinite induction rule for the truth language up to any ordinal smaller than ε_0. The technique employed to obtain Corollary 3 and Corollary 4 amounts to a formalization, in suitable PKF-like theories, of a variant of the asymmetric interpretation for normalized derivations in KF-like theories given by Cantini in [3].

In the next section we introduce PKF- and KF-like theories in their logical, arithmetical, and truth-theoretic components. §3 surveys and extends previous work on the inclusion of the provably true sentences in PKF-like theories in suitable KF-like theories. §4 is devoted to the proofs of Corollaries 3 and 4. §5 summarizes the results obtained in previous sections and suggests that the resulting picture partially accomplishes a variant of a program sketched by Reinhardt in [19].

2. Logic, arithmetic, truth

In the present section we introduce the main components of the system PKF of [9] (see also [8]) and of some of its natural variants. This amounts to introducing their underlying logic BDM, their arithmetical (or syntactic) part, and their truth-theoretic principles. We also introduce the classical theory KF from [6] and recall why PKF, KF and their extensions can be considered axiomatizations of Kripke's fixed-point models.

We start with the language $\mathcal{L}$ of arithmetic equipped with finitely many function symbols for suitable primitive recursive functions and expand it with a primitive predicate T: we call the resulting language $\mathcal{L}_\mathsf{T}$.

Definition 1.

(i) A *four-valued model* of $\mathcal{L}_\mathsf{T}$ is a tuple $(\mathcal{M}, S_1, S_2)$ where $\mathcal{M}$ is a model of the language $\mathcal{L}$ and $S_1, S_2 \subseteq M$ (the domain of $\mathcal{M}$) are the extension and antiextension of T. We allow $S_1 \cap S_2 \neq \varnothing$ and $S_1 \cup S_2 \neq M$.

(ii) If $S_1 \cup S_2 = M$, we call $(\mathcal{M}, S_1, S_2)$ *complete.* If $S_1 \cap S_2 \neq \varnothing$, we call it *consistent.*

(iii) A four-valued model $(\mathcal{M}, S_1, S_2)$ is called *symmetric* when $S_1 \cap S_2 = \varnothing$ or $S_1 \cup S_2 = M$.

We define the relation $\vDash_{\mathsf{DM}}$ obtaining between four-valued models and formulas of $\mathcal{L}_\mathsf{T}$. For simplicity, we extend $\mathcal{L}_\mathsf{T}$ to a language $\mathcal{L}_\mathsf{T}^\mathcal{M}$ featuring new constants $\bar{c}$ for all $c \in M$. When $\mathbb{N}$ is the standard model, we can safely take $\mathcal{L}_\mathsf{T}^\mathcal{M} = \mathcal{L}_\mathsf{T}$. The mapping $\cdot^\mathcal{M} \colon \mathsf{Term}_{\mathcal{L}_\mathsf{T}^\mathcal{M}} \to M$ is defined in the usual way.

[2]See §2 for the definition of the Veblen functions.

Definition 2 (Basic De Morgan evaluation schema)**.** Let $(\mathcal{M}, S_1, S_2)$ be a four-valued model, s, t be closed terms of $\mathcal{L}_{\mathsf{T}}^{\mathcal{M}}$, ϕ, ψ sentences of $\mathcal{L}_{\mathsf{T}}^{\mathcal{M}}$, and $\chi(v)$ formulas of $\mathcal{L}_{\mathsf{T}}^{\mathcal{M}}$ with only the variable displayed free.

$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} s = t \text{ iff } s^{\mathcal{M}} = t^{\mathcal{M}}$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} s \neq t \text{ iff } s^{\mathcal{M}} \neq t^{\mathcal{M}}$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \mathsf{T}t \text{ iff } t^{\mathcal{M}} \in S_1$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\mathsf{T}t \text{ iff } t^{\mathcal{M}} \in S_2$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\neg\phi \text{ iff } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \phi$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \phi \wedge \psi \text{ iff } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \phi \text{ and } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \psi$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg(\phi \wedge \psi) \text{ iff } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\phi \text{ or } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\psi$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \phi \vee \psi \text{ iff } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \phi \text{ or } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \psi$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg(\phi \vee \psi) \text{ iff } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\phi \text{ and } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\psi$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \forall v\chi \text{ iff for all } c \in M, \ (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \chi(\bar{c})$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\forall v\chi \text{ iff there exists a } c \in M \text{ such that } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\chi(\bar{c})$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \exists v\chi \text{ iff there is a } c \in M \text{ such that } (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \chi(\bar{c})$$
$$(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\exists v\chi \text{ iff for all } c \in M, \ (\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\chi(\bar{c})$$

The label 'four-valued models' derives from the properties of propositional connectives that can be easily extracted from Definition 2. A sequent is an expression of the form $\Gamma \Rightarrow \Delta$, where Γ, Δ are finite sets of formulas. Definition 2 can be extended to sequents. For simplicity, I give the definition of BDM-satisfaction for sequents involving sets of sentences but it should be clear how to extend it to sets of formulas by means of variable assignments.

Definition 3. Let $(\mathcal{M}, S_1, S_2)$ be again a four-valued model. Then $(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \Gamma \Rightarrow \Delta$ if and only if both the following conditions hold:

(i) if for all $\phi \in \Gamma$, $(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \phi$, then there is a $\psi \in \Delta$ such that $(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \psi$;
(ii) if for all $\psi \in \Delta$, $(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\psi$, then there is a $\phi \in \Gamma$ such that $(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\phi$.

To the satisfaction relation $\vDash_{\mathsf{DM}}$ we associate the following notion of logical consequence with multiple conclusion. Again it is convenient to consider finite sets of sentences.

Definition 4. Let Γ, Δ be finite sets of sentences and $\mathcal{C}$ a class of four-valued models. Then $\Gamma \vDash_{\mathcal{C}} \Delta$ if and only if, for all $\mathcal{M} \in \mathcal{C}$, also $\mathcal{M} \vDash_{\mathsf{DM}} \Gamma \Rightarrow \Delta$.

We will often be concerned with the case in which $\mathcal{C}$ is a class of four-valued models of a set $\mathcal{S}$ of appropriate non-logical inferences of the form

$$\frac{\{\Gamma_i, \phi_i \Rightarrow \psi_i, \Delta_i\}_{i=1}^k}{\Gamma, \phi \Rightarrow \psi, \Delta} \tag{1}$$

where $\phi, \phi_i, \psi, \psi_i$ are formulas and Γ_i, Δ_i finite sets of formulas. In this case we will directly write $\Gamma \vDash_{\mathcal{S}} \Delta$ instead to referring to the class of models $\mathcal{C}$. In (1) k can be 0: in this case the inference is simply a nonlogical initial sequent.

The logic associated with this notion of logical consequence can be seen as a multiple-conclusion variant of the four-valued logic introduced in [1] that, besides the preservations of the values 'true' and 'both true and false', also prescribes the anti-preservation of 'false' and 'both true and false'. Following [7], we call it Basic De Morgan Logic as it guarantees the standard De Morgan transformations of the positive connectives $\wedge, \vee$. If we focus on *symmetric* models, we obtain the logic considered in [12, 9] that cannot distinguish between gluts (sentences in $S_1 \cap S_2$) and gaps (sentences not in $S_1 \cup S_2$).

We now introduce a two-sided sequent calculus that is sound and complete with respect to the notion of logical consequence introduced in Definition 4: we call this system BDM.

Definition 5 (BDM). *Initial sequents* are of the form $\phi \Rightarrow \phi$, where ϕ is an atomic or negated atomic formula of $\mathcal{L}_{\mathsf{T}}$ (literal). We have the following logical rules of inference:

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\ (\text{CUT})$$

$$\frac{\Gamma \Rightarrow \Delta}{\Gamma, \phi \Rightarrow \Delta}\ (\text{WL}) \qquad \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \phi, \Delta}\ (\text{WR})$$

$$\frac{\Gamma \Rightarrow \neg\Delta}{\Delta \Rightarrow \neg\Gamma}\ (\neg\text{R}) \qquad \frac{\neg\Gamma \Rightarrow \Delta}{\neg\Delta \Rightarrow \Gamma}\ (\neg\text{L})$$

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \phi \wedge \psi}\ (\wedge\text{R}) \qquad \frac{\phi, \psi, \Gamma \Rightarrow \Delta}{\phi \wedge \psi, \Gamma \Rightarrow \Delta}\ (\wedge\text{L})$$

$$\frac{\Gamma \Rightarrow \Delta, \phi, \psi}{\Gamma \Rightarrow \Delta, \phi \vee \psi}\ (\vee\text{R}) \qquad \frac{\phi, \Gamma \Rightarrow \Delta \qquad \psi, \Gamma \Rightarrow \Delta}{\phi \vee \psi, \Gamma \Rightarrow \Delta}\ (\vee\text{L})$$

$$\frac{\chi(t), \Gamma \Rightarrow \Delta}{\forall x\chi, \Gamma \Rightarrow \Delta}\ (\forall\text{L}) \qquad \frac{\Gamma \Rightarrow \Delta, \phi(x)}{\Gamma \Rightarrow \Delta, \forall x\phi}\ (\forall\text{R})$$

$$\frac{\Gamma \Rightarrow \Delta, \phi(t)}{\Gamma \Rightarrow \Delta, \exists x\phi}\ (\exists\text{R}) \qquad \frac{\chi(x), \Gamma \Rightarrow \Delta}{\exists x\chi, \Gamma \Rightarrow \Delta}\ (\exists\text{L})$$

Remark **1.**

(i) rules on the first line are *structural inferences*;

(ii) in $\neg$R and $\neg$L, $\neg\Gamma$ and $\neg\Delta$ denote the finite sets of negations of formulas in Γ and Δ respectively;

(iii) in $\forall$R and $\exists$L, the variable x is not free in the lower sequent;

(iv) complex formulas introduced in the lower sequent are called *principal*, the corresponding formulas in the upper sequents are called *auxiliary*, and Γ, Δ are the *side* formulas. We write $\vdash_{\mathcal{S}}$ for derivability in BDM plus an additional set of nonlogical inference $\mathcal{S}$;

(v) a *symmetric* variant of BDM, that we call SDM, can be obtained by adding to BDM the initial sequent

(GG) $$\phi, \neg\phi \Rightarrow \psi, \neg\psi$$

A version of Lemma 1 below can be obtained for SDM relative to symmetric models;

(vi) classical logic K is obtained by extending BDM with the negation rules

$$\frac{\phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\phi}\ (\neg 1) \qquad \frac{\Gamma \Rightarrow \Delta, \phi}{\neg\phi, \Gamma \Rightarrow \Delta}\ (\neg 2)$$

BDM is closed under $(\neg 1)$, $(\neg 2)$ for formulas ϕ provided that it can also prove $\Rightarrow \phi, \neg\phi$;

(vii) the usual substitution, inversion, cut-elimination lemmata hold for BDM;

(viii) for $\mathcal{S}$ a set of non-logical rules of inference of the form (1), we write $S \vdash \phi$ for 'the sequent $\Rightarrow \phi$ is provable in from S'.

Lemma 1 (Adequacy)**.** $\vdash_{\mathcal{S}} \Gamma \Rightarrow \Delta$ *if and only if* $\Gamma \vDash_{\mathcal{S}} \Delta$.

The soundness direction of Lemma 1 is straighforward. For a proof of the completeness direction we refer to [2, 17].

Definition 6. The system $\mathsf{BDM}_=$ (similarly for $\mathsf{SDM}_=$) is obtained by adding to BDM the nonlogical initial sequents

$$(=1) \qquad \Rightarrow r_0 = r_0$$

$$(=2) \qquad r_0 = r_1, \phi(r_0) \Rightarrow \phi(r_1)$$

for r_0, r_1 arbitrary terms of $\mathcal{L}_\mathsf{T}$.

Once identity is in the language, the arithmetical component of PKF is readily introduced:

Definition 7 (basic, basik)**.** The system basic consists of $\mathsf{BDM}_=$ in $\mathcal{L}_\mathsf{T}$ plus all initial sequents $\Rightarrow \phi$ with ϕ an axiom of Peano Arithmetic PA different from the induction schema or a defining equation for the (finitely many) additional symbols for primitive recursive functions that we require. We call basik the result of adding the arithmetical initial sequents to classical logic K plus the identity initial sequents.

Variations on induction schemata will play an important role in what follows: for simplicity to deal with the relationships between internal and external induction, we formulate induction schemata and rules for formulas with exactly one free variable, but nothing essential rests on this assumption. More specifically, we will mainly focus on induction *rules* over basic, as the schema

$$(\mathsf{IND0}) \qquad \Rightarrow \phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(x+1)) \rightarrow \forall x\, \phi(x)$$

for $\phi(v)$ a formula of $\mathcal{L}_\mathsf{T}$ may fail to be sound when we have BDM or subclassical variants in the background.[3]

Definition 8.

(i) Full $\mathcal{L}_\mathsf{T}$ induction has the form

$$\frac{\phi(x), \Gamma \Rightarrow \Delta, \phi(x+1)}{\phi(0), \Gamma \Rightarrow \Delta, \phi(r)}\ (\mathsf{IND1})$$

with $\phi(0)$ a sentence of $\mathcal{L}_\mathsf{T}$, x not occurring free in $\Gamma, \Delta, \phi(0)$, and r an arbitrary term.

(ii) We will also consider the following rule of *internal* induction whose formulation requires the representation in basic of some syntactic notions and operations (cf. the comprehensive list in §2.2). In particular, $\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(\cdot,\cdot)$ represents in basic (and of course also in basik) the relation obtaining between $\mathcal{L}_\mathsf{T}$-formulas with one free variable and their free variable:

$$\frac{\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(x,u), \mathsf{T}\, x(\dot{y}/u), \Gamma \Rightarrow \Delta, \mathsf{T}x(y \dot{+} 1/u)}{\mathsf{T}x(\ulcorner 0 \urcorner/u), \Gamma \Rightarrow \Delta, \mathsf{T}x(\dot{z}/u)}\ (\mathsf{IND2})$$

In ($\mathsf{IND2}$) z is an eigenvariable and $x(\dot{y}/u)$ expresses formal substitution in x of u with the numeral for y.

We notice however that, over basik, the schemata ($\mathsf{IND0}$) and ($\mathsf{IND1}$) are equivalent due to the classical rules for the negation and the conditional.

[3] To show that there are unsound instances of the induction schema, one can for instance employ the formula $\lambda \wedge y = y$ as relevant instance of $\mathsf{IND0}$.

2.1. **Ordinals.** For our purposes it is sufficient to assume a notation system of ordinals up to the least strongly critical ordinal Γ_0 as it is carried out in, e.g., [18, Ch. 3]. We recall the main properties of a suitable such notation.

A non-zero ordinal α is *principal* if it cannot be expressed as $\zeta + \eta$ for $\zeta, \eta < \alpha$. Starting with principal ordinals, one can define the class $\mathsf{C}(\alpha)$ of α-critical ordinals by transfinite induction by letting

$$\begin{aligned}\mathsf{C}(0) &:= \text{`the class of principal ordinals'}\\ \mathsf{C}(\alpha+1) &:= \text{`the class of fixed points of the function enumerating } \mathsf{C}(\alpha)\text{'}\\ \mathsf{C}(\lambda) &:= \bigcap_{\zeta<\lambda} \mathsf{C}(\zeta) \ \text{ for } \lambda \text{ a limit ordinal}\end{aligned}$$

The *Veblen functions* φ_α are the enumerating functions of $\mathsf{C}(\alpha)$. The class of *strongly critical* ordinals SC contains precisely the ordinals α that are themselves α-critical. Γ_ζ indicates the ζ-th strongly critical ordinal. Principal ordinals α that are not themselves strongly critical are such that $\alpha = \varphi_\zeta \eta$ for $\eta, \zeta < \alpha$ (see [18, Lem. 3.4.17]). Therefore, by this fact and Cantor's normal form theorem, ordinals $< \Gamma_0$ can be uniquely determined as words of the alphabet $(0, +, \varphi \cdot \cdot)$.

Lemma 2. *Assuming a standard primitive recursive encoding of finite sequences in natural numbers:*

(i) There are primitive recursive notations $\mathcal{O} \subseteq \mathbb{N}$ (for ordinals $< \Gamma_0$), $\mathcal{P} \subseteq \mathcal{O}$ (for principal ordinals $< \Gamma_0$), a primitive recursive relation $\prec \subseteq \mathbb{N} \times \mathbb{N}$, and an evaluation function $|\cdot|$ defined as:
- $0 \in \mathcal{O}$ *and* $|0| = 0$;
- $n_1, \ldots, n_m \in \mathcal{P}$ *and* $|n_1| \geq \ldots \geq |n_m|$, *then* $(1, n_1, \ldots, n_m) \in \mathcal{O}$ *and* $|(1, n_1, \ldots, n_m)| = |n_1| + \ldots + |n_m|$;
- *if* $n_1, n_2 \in \mathcal{O}$, *then* $(2, n_1, n_2) \in \mathcal{P}$ *and* $|(2, n_1, n_2)| = \varphi_{|n_1|}(|n_2|)$
- $n \prec m :\leftrightarrow n \in \mathcal{O}$ *and* $m \in \mathcal{O}$ *and* $|n| < |m|$

(ii) For every $o \in \mathcal{O}$, $|o| < \Gamma_0$; for every $\alpha < \Gamma_0$, there is an $o \in \mathcal{O}$ such that $|o| = \alpha$.

For ordinals $\alpha < \Gamma_0$, we denote with a the corresponding numeral in the representation of $\mathcal{O}$ and we do not distinguish between ordinal functions such as the Veblen functions and their representations. Given the well-known identities $\varphi_0 \alpha = \omega^\alpha$, and $\varphi_1 0 = \varepsilon_0$ we often employ, also in formal contexts, the more familiar formulation of ordinals $\leq \varepsilon_0$.

The system $(\mathcal{O}, \mathcal{P}, \prec)$ enables us to formulate the following principles of transfinite induction:

$$(\mathsf{TI}^{\varepsilon_0}_{\mathcal{L}_\mathsf{T}}) \qquad \frac{\forall a \prec b\, \phi(a), \Gamma \Rightarrow \Delta, \phi(b)}{\Gamma \Rightarrow \Delta, \forall a \prec \varepsilon_0\, \phi(a)}$$

$$(\mathsf{TI}^{<\omega^\omega}_{\mathcal{L}_\mathsf{T}}) \qquad \frac{\forall a \prec b\, \phi(a), \Gamma \Rightarrow \Delta, \phi(b)}{\Gamma \Rightarrow \Delta, \forall a \prec c\, \phi(a)} \quad \text{for all } \gamma (= |c|) < \omega^\omega$$

$$(\mathsf{TI}^{<\varepsilon_0}_{\mathcal{L}_\mathsf{T}}) \qquad \frac{\forall a \prec b\, \phi(a), \Gamma \Rightarrow \Delta, \phi(b)}{\Gamma \Rightarrow \Delta, \forall a \prec c\, \phi(a)} \quad \text{for all } \gamma < \varepsilon_0$$

2.2. **Truth principles.** The truth-theoretic components of our systems amount to the initial sequents displayed in Table 1. In the table, the evaluation function $\mathsf{val}(\cdot)$ can be taken to be primitive recursive; we employ the usual primitive recursive syntactic predicates $\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x)$ ('x is a closed term of $\mathcal{L}_\mathsf{T}$'), $\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x)$ ('x is a sentence of $\mathcal{L}_\mathsf{T}$'), $\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(x)$ (x is a formula of $\mathcal{L}_\mathsf{T}$ with one free variable') and the following primitive recursive syntactic operations that we list on the first column of the chart below accompanied by the corresponding intensional abbreviations:

FUNCTION	MEANING	ABBREVIATION
$\mathsf{sub}(x, y, \mathsf{num}(z))$	'substituting y with the numeral for z in x'	$x(\dot{z}/y)$
$\mathsf{id}(x, y)$	'equating x and y'	$x = y$
$\mathsf{nid}(x, y)$	'negating the equality between x and y'	$x \neq y$
$\mathsf{ng}(x)$	'negating x'	$\neg x$
$\mathsf{dn}(x)$	'negating twice x'	$\neg\neg x$
$\mathsf{and}(x, y)$	'conjoining x and y'	$x \wedge y$
$\mathsf{or}(x, y)$	'disjoining x and y'	$x \vee y$
$\mathsf{all}(x, y)$	'universally quantifying x w.r.t y'	$\forall xy$
$\mathsf{ex}(x, y)$	'existentially quantifying x w.r.t y'	$\exists xy$
$\mathsf{tr}(x)$	'prefixing a truth predicate to x'	$\mathsf{T}x$
$\mathsf{sub}(\ulcorner\phi(v)\urcorner, \ulcorner v\urcorner, \mathsf{num}(x))$	'replacing in $\phi(v)$ the variable v with x'	$\ulcorner\varphi(\dot{x})\urcorner$

Definition 9.

(i) PKF is obtained by extending basic with (At1-2), (T1-2), (T6), ($\neg$1-2), ($\wedge$1-2), ($\vee$1-2), ($\forall$1-2), ($\exists$1-2), (IND1);

(ii) PKF$\restriction$ is obtained from PKF by allowing only formulas of $\mathcal{L}$ in instances of (IND1) and adding (Reg1-2) to the resulting system;

(iii) PKF^+ is obtained by extending PKF with $\mathsf{TI}^{<\varepsilon_0}_{\mathcal{L}_\mathsf{T}}$;

(iv) PKF_S is obtained by adding (GG) to PKF. $\mathsf{PKF}_\mathsf{S}\restriction$ and $\mathsf{PKF}^+_\mathsf{S}$ are defined accordingly.

My goal in this paper is to relate the theories just introduced with suitable variants of Feferman's axiomatization of Kripke's theory of truth. These variants are presented in the next definition.

Definition 10.

(i) KF results from adding to basik: (At1-4), (T1-6), ($\neg$3-4), ($\wedge$1-4), ($\vee$1-4), ($\forall$1-4), ($\exists$1-4);

(ii) KF$\restriction$ is obtained from KF by restricting (IND1) to $\mathcal{L}$-formulas and adding (Reg1-2) to the resulting system;

(iii) KFI results from KF$\restriction$ once the restricted version of (IND1) is replaced with (IND2);

(iv) KF_S the extension of KF with the combination, in disjunctive form, of (cm) and (cs), that is:

$$\Rightarrow \forall x(\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \to \mathsf{T}x \vee \mathsf{T}\neg x) \vee \forall x(\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \wedge \mathsf{T}\neg x \to \neg \mathsf{T}x)$$

Similarly for $\mathsf{KF}_\mathsf{S}\restriction$ and KFI_S.

As anticipated, the way in which we will relate these two clusters of theories is by comparing what PKF-like and KF-like theories prove true in the precise sense of comparing their theorems of the form $\mathsf{T}\ulcorner\phi\urcorner$.

EQUATIONS	(At1)	$\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{T}(x{=}y) \Rightarrow \mathsf{val}(x) = \mathsf{val}(y)$
	(At2)	$\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{val}(x) = \mathsf{val}(y) \Rightarrow \mathsf{T}(x{=}y)$
	(At3)	$\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{T}(x{\neq}y) \Rightarrow \mathsf{val}(x) \neq \mathsf{val}(y)$
	(At4)	$\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{val}(x) \neq \mathsf{val}(y) \Rightarrow \mathsf{T}(x{\neq}y)$
TRUTH ASCRIPTIONS	(T1)	$\mathsf{T}x \Rightarrow \mathsf{T}\ulcorner \mathsf{T}\dot{x} \urcorner$
	(T2)	$\mathsf{T}\ulcorner \mathsf{T}\dot{x} \urcorner \Rightarrow \mathsf{T}x$
	(T3)	$\mathsf{T}\neg x \Rightarrow \mathsf{T}\ulcorner \neg\mathsf{T}\dot{x} \urcorner$
	(T4)	$\mathsf{T}\ulcorner \neg\mathsf{T}\dot{x} \urcorner \Rightarrow \mathsf{T}\neg x, \neg\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x)$
	(T5)	$\neg\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \Rightarrow \mathsf{T}\ulcorner \neg\mathsf{T}\dot{x} \urcorner$
	(T6)	$\mathsf{T}x \Rightarrow \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x)$
	(Reg1)	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{T}x(\dot{\mathsf{val}}(y)/v) \Rightarrow \mathsf{T}x(y/v)$
	(Reg2)	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{T}x(y/v) \Rightarrow \mathsf{T}x(\dot{\mathsf{val}}(y)/v)$
	(cm)	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \Rightarrow \mathsf{T}x, \mathsf{T}\neg x$
	(cs)	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \Rightarrow \neg\mathsf{T}x, \neg\mathsf{T}\neg x$
	(GG)	$\phi, \neg\phi \Rightarrow \psi, \neg\psi$
PROPOSITIONAL CONNECTIVES	$(\neg 1)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x), \neg\mathsf{T}x \Rightarrow \mathsf{T}\neg x$
	$(\neg 2)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{T}\neg x \Rightarrow \neg\mathsf{T}x$
	$(\neg 3)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{T}\neg\neg x \Rightarrow \mathsf{T}x$
	$(\neg 4)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x), \mathsf{T}x \Rightarrow \mathsf{T}\neg\neg x$
	$(\wedge 1)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\wedge}y), \mathsf{T}(x{\wedge}y) \Rightarrow \mathsf{T}x \wedge \mathsf{T}y$
	$(\wedge 2)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\wedge}y), \mathsf{T}x, \mathsf{T}y \Rightarrow \mathsf{T}(x{\wedge}y)$
	$(\wedge 3)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\wedge}y), \mathsf{T}\neg(x{\wedge}y) \Rightarrow \mathsf{T}\neg x, \mathsf{T}\neg y$
	$(\wedge 4)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\wedge}y), \mathsf{T}\neg x \vee \mathsf{T}\neg y \Rightarrow \mathsf{T}\neg(x{\wedge}y)$
	$(\vee 1)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\vee}y), \mathsf{T}(x{\vee}y) \Rightarrow \mathsf{T}x, \mathsf{T}y$
	$(\vee 2)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\vee}y), \mathsf{T}x \vee \mathsf{T}y \Rightarrow \mathsf{T}(x{\vee}y)$
	$(\vee 3)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\vee}y), \mathsf{T}\neg(x{\vee}y) \Rightarrow \mathsf{T}\neg x \wedge \mathsf{T}\neg y$
	$(\vee 4)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x{\vee}y), \mathsf{T}\neg x, \mathsf{T}\neg y \Rightarrow \mathsf{T}\neg(x{\vee}y)$
QUANTIFIERS	$(\forall 1)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall vx), \mathsf{T}(\forall vx) \Rightarrow \forall u\, \mathsf{T}x(\dot{u}/v)$
	$(\forall 2)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall vx), \forall u\, \mathsf{T}x(\dot{u}/v) \Rightarrow \mathsf{T}(\forall vx)$
	$(\forall 3)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall vx), \mathsf{T}\neg(\forall vx) \Rightarrow \exists u\, \mathsf{T}\neg x(\dot{u}/v)$
	$(\forall 4)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall vx), \exists u\, \mathsf{T}\neg x(\dot{u}/v) \Rightarrow \mathsf{T}\neg(\forall vx)$
	$(\exists 1)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\exists vx), \mathsf{T}(\exists vx) \Rightarrow \exists u\, \mathsf{T}x(\dot{u}/v)$
	$(\exists 2)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\exists vx), \exists u\, \mathsf{T}x(\dot{u}/v) \Rightarrow \mathsf{T}(\exists vx)$
	$(\exists 3)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\exists vx), \mathsf{T}\neg(\exists vx) \Rightarrow \forall u \mathsf{T}\neg x(\dot{u}/v)$
	$(\exists 4)$	$\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\exists vx), \forall u\, \mathsf{T}\neg x(\dot{u}/v) \Rightarrow \mathsf{T}\neg\exists vx)$

TABLE 1. INITIAL SEQUENTS FOR TRUTH

Definition 11 (Internal theory)**.** For any system S in $\mathcal{L}_{\mathsf{T}}$ considered in Definitions 9 and 10, the 'internal theory' of S is

$$\mathsf{I}S := \{\phi \in \mathcal{L}_{\mathsf{T}} \mid S \vdash \mathsf{T}\ulcorner\phi\urcorner\}$$

For the sake of readability, we do not distinguish between derivability '$\vdash$' in PKF-like and KF-like theories although, strictly speaking, the meaning of the turnstile shifts between the two clusters.

Let U be a PKF-like theory, that is one of the theories introduced in Def. 9. We have, by induction on the complexity of $\phi \in \mathcal{L}_{\mathsf{T}}$ and $\psi \in \mathcal{L}$:

Lemma 3.

(i) (Intersubstitutivity) $U \vdash \Gamma \Rightarrow \Delta, \mathsf{T}\ulcorner\phi\urcorner$ *if and only if* $U \vdash \Gamma \Rightarrow \Delta, \phi$*, for any* $\mathcal{L}_{\mathsf{T}}$*-sentence* ϕ*;*

(ii) $U \vdash \psi \vee \neg\psi$*, for* $\psi \in \mathcal{L}$*.*

Remark **2.** Lemma 3 tells us that $\mathsf{I}U = U$ for U as above. By contrast, in KF-like theories (ii) is trivially obtained for all $\phi \in \mathcal{L}_{\mathsf{T}}$; therefore the sentence $\lambda \vee \neg\lambda$, for λ provably equivalent in basik to $\neg\mathsf{T}\ulcorner\lambda\urcorner$, is already sufficient to establish that the set of theorems of a KF-like theory does not coincide with its internal theory.

Although full intersubstitutivity is out of reach for KF-like theories, the full T-schema is provable in them for formulas of $\mathcal{L}$ with free variables. This again by external induction of the complexity of the formula involved:

Lemma 4. *Let* V *be a* KF*-like theory as specified in Definition 10. Then for all formulas* $\phi(\vec{v})$ *of* $\mathcal{L}$*,* V *proves*

$$\forall\vec{x}\,(\mathsf{T}\ulcorner\phi(\dot{\vec{x}})\urcorner \leftrightarrow \phi(\vec{x}))$$

We conclude this section by providing a formal rendering of the claim that PKF-like and KF-like theories amount to two ways of axiomatizing the conception of truth associated to the framework in [13]. By restricting our attention to the standard model $\mathbb{N}$ of $\mathcal{L}$, it is well-known that the set of (codes of) $\mathcal{L}_{\mathsf{T}}$-sentences BDM-true in $(\mathbb{N}, S_1, S_2)$ enjoys a positive inductive definition in the parameters S_1 and S_2.[4] To choose pairs (S_1, S_2) compatible with an intended interpretation of the extension and antiextension of T, one can employ this definition to introduce the monotone operator

$$\begin{aligned}\Phi(S_1, S_2) := \big\langle &\{\phi \in \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}} \mid (\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \phi\},\\ &(\omega \setminus \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}) \cup \{\phi \in \mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}} \mid (\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\phi\}\big\rangle\end{aligned}$$

Crucially, fixed-points of Φ enjoy the following property, for all $\mathcal{L}_{\mathsf{T}}$-sentences ϕ:

(2) $(\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \mathsf{T}\ulcorner\phi\urcorner$ if and only if $(\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \phi$ and

(3) $(\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\mathsf{T}\ulcorner\phi\urcorner$ if and only if $(\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\phi$

The following lemma gives us at least a necessary condition for considering PKF and KF-like theories as axiomatizations of fixed-point models. In this work it is harmless to consider it also as a sufficient condition, although a thorough discussion of this criterion would likely lead to a revision of this claim (see also [15]). We state the lemma for PKF and KF, but Lemma 5 clearly holds for any other combination with the obvious adjustments to the satisfaction relation.

[4]Clearly so does the set of sentences SDM-true in $(\mathbb{N}, S_1, S_2)$ with S_1 and S_2 disjoint or exhaustive.

Lemma 5. *For (S_1, S_2) a fixed point of Φ,*

$$(\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \mathsf{PKF} \quad \textit{iff} \quad (\mathbb{N}, S_1) \vDash \mathsf{KF}$$

This concludes the section devoted to the required preliminaries.

3. Soundness proofs in classical systems

Halbach and Horsten in [9] showed that the soundness proof for PKF with respect to fixed-point models that lies at the root of Lemma 5 can be formalized in KFI. In doing so they proved the inclusion of PKF in IKFI. In this section we show that this reasoning can be extended to other PKF-like and KF-like theories from Definitions 9, 10. In what follows we write $\bigwedge \Gamma(\vec{x})$ ($\bigvee \Gamma(\vec{x})$) for the conjunction (disjunction) of the members of Γ, and $\vec{x}$ for the free variables appearing in them.

Proposition 1. *The claim 'if $U \vdash \Gamma \Rightarrow \Delta$, then*

$$V \vdash \forall \vec{x} \Big(\big(\mathsf{T}\ulcorner \bigwedge \Gamma(\dot{\vec{x}}) \urcorner \rightarrow \mathsf{T}\ulcorner \bigvee \Delta(\dot{\vec{x}}) \urcorner \big) \wedge \big(\mathsf{T}\ulcorner \neg \bigvee \Delta(\dot{\vec{x}}) \urcorner \rightarrow \mathsf{T}\ulcorner \neg \bigwedge \Gamma(\dot{\vec{x}}) \urcorner \big) \Big) \text{'}$$

holds for the following pairs of theories U, V: $(\mathsf{PKF}\restriction, \mathsf{KF}\restriction)$, $(\mathsf{PKF}, \mathsf{KFI})$, $(\mathsf{PKF}^+, \mathsf{KF})$, $(\mathsf{PKF_S}\restriction, \mathsf{KF_S}\restriction)$, $(\mathsf{PKF_S}, \mathsf{KFI_S})$, $(\mathsf{PKF}^+_{\mathsf{S}}, \mathsf{KF_S})$.

Proof. The proof is by induction on the length of a derivation in the PKF-like theories involved. As in the proof of [9, Thm. 27], we only mention few paradigmatic cases. We also focus on some cases that characterize $\mathsf{PKF}\restriction$ and PKF^+ and their extensions with (GG) and that are therefore not covered by Halbach and Horsten's proof.

Let us consider for instance the case of ($\neg$1). We reason in $\mathsf{KF}\restriction$ so that it is clear that the same reasoning goes through in the cases of its extensions. For the first conjunct, we have:

$$\Rightarrow \mathsf{T}\ulcorner \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\dot{x}) \urcorner \wedge \mathsf{T}\neg \ulcorner \mathsf{T}\dot{x} \urcorner \rightarrow \mathsf{T}\neg x \qquad \text{by Lemma 4, (T4)}$$

$$\Rightarrow \mathsf{T}\ulcorner \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\dot{x}) \urcorner \wedge \neg \mathsf{T}\dot{x} \urcorner \rightarrow \mathsf{T}\ulcorner \mathsf{T}\neg \dot{x} \urcorner \qquad \text{by (T1-2)}$$

For the second conjunct, we reason as follows:

$$\Rightarrow \mathsf{T}\neg \ulcorner \mathsf{T}\neg \dot{x} \urcorner \rightarrow \mathsf{T}\neg\neg x \vee \mathsf{T}\ulcorner \neg \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\dot{x}) \urcorner \qquad \text{by (T4), Lemma 4}$$

$$\Rightarrow \mathsf{T}\neg \ulcorner \mathsf{T}\neg \dot{x} \urcorner \rightarrow \mathsf{T}\neg\neg \ulcorner \mathsf{T}\dot{x} \urcorner \vee \mathsf{T}\ulcorner \neg \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\dot{x}) \urcorner \qquad \text{by (T1), } (\neg 4)$$

$$\Rightarrow \mathsf{T}\neg \ulcorner \mathsf{T}\neg \dot{x} \urcorner \rightarrow \mathsf{T}\ulcorner \neg (\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\dot{x}) \wedge \neg \mathsf{T}\dot{x}) \urcorner \qquad \text{by } (\wedge 4)$$

In theories forcing symmetric models we need to deal with the initial sequent (GG) (cf. Def. 9(iv)), that is, we need to show, in a suitable KF-like theory augmented with either (cs) or (cm),

$$\Rightarrow (\mathsf{T}\ulcorner \phi \wedge \neg \phi \urcorner \rightarrow \mathsf{T}\ulcorner \psi \vee \neg \psi \urcorner) \wedge (\mathsf{T}\ulcorner \neg (\psi \vee \neg \psi) \urcorner \rightarrow \mathsf{T}\ulcorner \neg (\phi \wedge \neg \phi) \urcorner) \tag{4}$$

We consider the second conjunct only as the reasoning for the first is simpler. If (cs), then

$$\mathsf{T}\ulcorner \neg (\psi \vee \neg \psi) \urcorner \Rightarrow \exists x (\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \wedge \mathsf{T}\neg x \wedge \mathsf{T}x) \qquad \text{by } (\vee 3), (\neg 3)$$

$$\mathsf{T}\ulcorner \neg (\psi \vee \neg \psi) \urcorner \Rightarrow \mathsf{T}\ulcorner \neg (\phi \wedge \neg \phi) \urcorner \qquad \text{by (cs) and K}$$

The case in which (cm) holds is particularly easy as the desired conclusion follows from the instance $\Rightarrow \mathsf{T}\ulcorner \phi \urcorner \vee \mathsf{T}\ulcorner \neg \phi \urcorner$ of (cm) by ($\neg$4), ($\wedge$4) and logic.

As an example of a rule of inference, we consider ($\neg$R), which also highlights the role of both conjuncts in the argument. For simplicity, we omit free variables. By induction hypothesis, a KF-like theory proves

$$\Rightarrow \mathsf{T}\ulcorner \bigwedge \Gamma \urcorner \rightarrow \mathsf{T}\ulcorner \bigvee \neg \Delta \urcorner \tag{5}$$

$$\Rightarrow \mathsf{T}\ulcorner \neg \bigvee \neg \Delta \urcorner \rightarrow \mathsf{T}\ulcorner \neg \bigwedge \Gamma \urcorner \tag{6}$$

We require

$$\Rightarrow \mathsf{T}\ulcorner \bigwedge \Delta \urcorner \to \mathsf{T}\ulcorner \bigvee \neg\Gamma \urcorner \tag{7}$$

$$\Rightarrow \mathsf{T}\ulcorner \neg \bigvee \neg\Gamma \urcorner \to \mathsf{T}\ulcorner \neg \bigwedge \Delta \urcorner \tag{8}$$

For (7), we have in a KF-like theory:

$$\begin{aligned}
\Rightarrow \mathsf{T}\ulcorner \delta_0 \wedge \ldots \wedge \delta_{n-1}\urcorner &\to \mathsf{T}\ulcorner \neg\neg\delta_0 \urcorner \wedge \ldots \wedge \mathsf{T}\ulcorner \neg\neg\delta_{n-1}\urcorner && \text{by } (\wedge 1), (\neg 4)\\
&\to \mathsf{T}\dot\neg(\ulcorner \neg\delta_0 \vee \ldots \vee \neg\delta_{n-1}\urcorner) && \text{by } (\vee 4)\\
&\to \mathsf{T}\ulcorner \neg(\gamma_0 \wedge \ldots \wedge \gamma_{m-1})\urcorner && \text{by } (6)\\
&\to \mathsf{T}\ulcorner \neg\gamma_0 \vee \cdots \vee \neg\gamma_{m-1}\urcorner && \text{by } (\wedge 3), (\vee 2)
\end{aligned}$$

The reasoning for (8) is analogous, with the crucial contribution of (5).

We now deal with the induction rule of $\mathsf{PKF}\!\restriction$. The same argument clearly goes through for $\mathsf{PKF_S}\!\restriction$ and $\mathsf{KF_S}\!\restriction$. By induction hypothesis, $\mathsf{KF}\!\restriction$ proves

$$\Rightarrow \mathsf{T}\ulcorner \bigwedge \Gamma \wedge \phi(\dot x)\urcorner \to \mathsf{T}\ulcorner \phi(x \dot{+} 1) \vee \bigvee \Delta \urcorner \tag{9}$$

$$\Rightarrow \mathsf{T}\ulcorner \neg(\bigvee \Delta \vee \phi(x \dot{+} 1))\urcorner \to \mathsf{T}\ulcorner \neg(\bigwedge \Gamma \wedge \phi(\dot x))\urcorner \tag{10}$$

for a standard $\mathcal{L}$-formula $\phi(v)$. To deal with the first conjunct of the main claim, by $(\wedge 2)$, $(\vee 1)$, and Lemma 4 applied to (9), we obtain

$$\Rightarrow \mathsf{T}\ulcorner \bigwedge \Gamma \urcorner \wedge \phi(x) \to \phi(x+1) \vee \mathsf{T}\ulcorner \bigvee \Delta \urcorner$$

Therefore also

$$\Rightarrow \mathsf{T}\ulcorner \bigwedge \Gamma \wedge \phi(0)\urcorner \to \mathsf{T}\ulcorner \phi(\dot y) \vee \bigvee \Delta \urcorner$$

by the induction rule of $\mathsf{KF}\!\restriction$, Lemma 4 and the truth rules for propositional connectives. For the second conjunct, we want to obtain

$$\Rightarrow \mathsf{T}\ulcorner \neg(\bigvee \Delta \vee \phi(\dot y))\urcorner \to \mathsf{T}\ulcorner \neg(\bigwedge \Gamma \wedge \phi(0))\urcorner \tag{11}$$

By (10), $(\vee 4)$, $(\vee 3)$ and Lemma 4, $\mathsf{KF}\!\restriction$ proves

$$\Rightarrow \mathsf{T}\ulcorner \neg \bigvee \Delta \urcorner \wedge \neg\phi(x+1) \to \mathsf{T}\ulcorner \neg \bigwedge \Gamma \urcorner \vee \neg\phi(x) \tag{12}$$

By propositional logic and the induction schema of $\mathsf{KF}\!\restriction$,

$$\Rightarrow \mathsf{T}\ulcorner \neg \bigvee \Delta \urcorner \wedge \neg \mathsf{T}\ulcorner \neg \bigwedge \Gamma \urcorner \to (\phi(0) \to \phi(y)) \tag{13}$$

from which we obtain (11) by propositional logic, Lemma 4 and $(\wedge 4)$, $(\vee 3)$.

We conclude with the rule $\mathsf{TI}^{<\varepsilon_0}_{\mathcal{L}_\mathsf{T}}$ of PKF^+ and $\mathsf{PKF}^+_\mathsf{S}$ by crucially employing the fact that KF (and therefore $\mathsf{KF_S}$) proves $\mathsf{TI}^{<\varepsilon_0}_{\mathcal{L}_\mathsf{T}}$ by a simple adaptation of Gentzen's lower bound proof for PA. Assuming that the following sequents are derivable in KF,

$$\Rightarrow \mathsf{T}\ulcorner \bigwedge \Gamma \wedge\ \forall z \prec \dot a\, \phi(z)\urcorner \to \mathsf{T}\ulcorner \phi(\dot a)\ \vee\ \bigvee \Delta \urcorner \tag{14}$$

$$\Rightarrow \mathsf{T}\ulcorner \neg(\bigvee \Delta \vee \phi(\dot a))\urcorner \to \mathsf{T}\ulcorner \neg(\forall z \prec \dot a\, \phi(z) \wedge\ \bigwedge \Gamma)\urcorner \tag{15}$$

we show

$$\Rightarrow \mathsf{T}\ulcorner \bigwedge \Gamma \urcorner \to \mathsf{T}\ulcorner \forall z \prec \dot c\, \phi(z)\urcorner \vee\ \bigvee \Delta \urcorner \tag{16}$$

$$\Rightarrow \mathsf{T}\ulcorner \neg(\bigvee \Delta\ \vee \forall z \prec \dot c\, \phi(z))\urcorner \to \mathsf{T}\ulcorner \neg \bigwedge \Gamma \urcorner \tag{17}$$

for all c such that $|c| < \varepsilon_0$.

For (16), we obtain by the induction hypothesis, Lemma 4, $(\wedge 2)$, $(\vee 1)$, $(\forall 2)$, and $\mathsf{TI}^{<\varepsilon_0}_{\mathcal{L}_\mathsf{T}}$,

$$\mathsf{T}\ulcorner \bigwedge \Gamma \urcorner \Rightarrow \forall z \prec c\, \mathsf{T}\ulcorner \phi(\dot z)\urcorner \vee \mathsf{T}\ulcorner \bigvee \Delta \urcorner \tag{18}$$

for all suitable c. The desired conclusion is then obtained by Lemma 4, ($\vee$2), ($\forall$2) and propositional logic. For (17), we start with (15) and obtain by logic and the truth principles for $\vee, \forall$:

$$\mathsf{T}\ulcorner\neg\bigvee\Delta\urcorner, \mathsf{T}\ulcorner\neg\phi(\dot{a})\urcorner \Rightarrow \exists z \prec a\, \mathsf{T}\ulcorner\neg\phi(\dot{z})\urcorner, \mathsf{T}\ulcorner\neg\bigwedge\Gamma\urcorner \tag{19}$$

Therefore, by logic and $\mathsf{TI}^{<\varepsilon_0}_{\mathcal{L}_\mathsf{T}}$,

$$\mathsf{T}\ulcorner\neg\bigvee\Delta\urcorner \Rightarrow \forall z \prec c\, \neg\mathsf{T}\ulcorner\neg\phi(\dot{z})\urcorner, \mathsf{T}\ulcorner\neg\bigwedge\Gamma\urcorner \tag{20}$$

for all suitable c. Finally, by logic and the truth principles for $\vee, \forall$:

$$\Rightarrow \mathsf{T}\ulcorner\neg(\bigvee\Delta \vee \forall z \prec \dot{c}\, \phi(z))\urcorner \rightarrow \mathsf{T}\ulcorner\neg\bigwedge\Gamma\urcorner \tag{21}$$

□

Proposition 1 obviously yields an upper bound for the arithmetical sentences provable in a PKF-like theory, given the following correspondence between KF-like theories and systems of ramified analysis (cf. [4]) studied in [3]:

(i) $\mathsf{KF}\!\upharpoonright$ and $\mathsf{KF_S}\!\upharpoonright$ have the same $\mathcal{L}$-theorems as PA.
(ii) KFI and $\mathsf{KFI_S}$ have the same arithmetical theorems as ramified analysis (or ramified truth) up to ω^ω, or $\mathsf{PA} + \mathsf{TI}^{<\varphi_\omega 0}_{\mathcal{L}}$.
(iii) KF and $\mathsf{KF_S}$ have the same arithmetical theorems as ramified analysis (or ramified truth) up to ε_0, or $\mathsf{PA} + \mathsf{TI}^{<\varphi_{\varepsilon_0} 0}_{\mathcal{L}}$.

Moreover, we immediately obtain the inclusion of PKF-like theories in the internal theories of the corresponding KF-like theories.

Corollary 1.
(i) $\mathsf{PKF}\!\upharpoonright \subseteq \mathsf{IKF}\!\upharpoonright$ *and* $\mathsf{PKF_S}\!\upharpoonright \subseteq \mathsf{IKF_S}\!\upharpoonright$*;*
(ii) $\mathsf{PKF} \subseteq \mathsf{IKFI}$ *and* $\mathsf{PKF_S} \subseteq \mathsf{IKFI_S}$*;*
(iii) $\mathsf{PKF}^+ \subseteq \mathsf{IKF}$ *and* $\mathsf{PKF}^+_\mathsf{S} \subseteq \mathsf{IKF_S}$*.*

The next section addresses the question whether the above inclusions are proper or not.

4. Levels of truth in PKF-like theories

The present section settles the question of the relationships between PKF-like theories and the internal theories of the corresponding KF-like theories. The resulting picture provides an answer to a series of questions left open in [9]. We begin by sketching a result proved in [17] about theories with restricted induction.

Lemma 6. *For any $\phi \in \mathcal{L}_\mathsf{T}$, if $\mathsf{KF}\!\upharpoonright$ ($\mathsf{KF_S}\!\upharpoonright$) proves $\mathsf{T}\ulcorner\phi\urcorner$, then $\mathsf{PKF}\!\upharpoonright$ ($\mathsf{PKF_S}\!\upharpoonright$) proves ϕ.*

Proof Sketch. For details, we refer to [17]. We reason for $\mathsf{KF}\!\upharpoonright$ and $\mathsf{PKF}\!\upharpoonright$ but it should be clear how to proceed in the case of $\mathsf{KF_S}\!\upharpoonright$.

Assuming $\mathsf{PKF}\!\upharpoonright \nvdash \phi$, we show $\mathsf{KF}\!\upharpoonright \nvdash \mathsf{T}\ulcorner\phi\urcorner$. By the completeness theorem for BDM (and extensions thereof) one of the following must hold:

(1) there is a four-valued model $(\mathcal{M}, S_1, S_2)$ of $\mathsf{PKF}\!\upharpoonright$ such that $(\mathcal{M}, S_1, S_2) \nvDash_{\mathsf{DM}} \phi$;
(2) there is a four-valued model $(\mathcal{M}, S_1, S_2)$ of $\mathsf{PKF}\!\upharpoonright$ such that $(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\phi$.

The case in which (1) holds is the easiest. We first notice, by induction on the length of the derivation in $\mathsf{KF}\!\upharpoonright$, that $(\mathcal{M}, S_1) \vDash \mathsf{KF}\!\upharpoonright$. If $\mathsf{KF}\!\upharpoonright \vdash \mathsf{T}\ulcorner\phi\urcorner$, therefore, $\ulcorner\phi\urcorner \in S_1$ and $(\mathcal{M}, S_1, S_2) \vDash_{\mathsf{DM}} \phi$ by Lemma 3, quod non by (1).
If (2) holds, there are three sub-cases.

Sub-case 1: $(\mathcal{M}, S_1, S_2)$ is consistent (cf. Def. 1(ii)). By assuming $\mathsf{KF}\!\upharpoonright \vdash \mathsf{T}\ulcorner\phi\urcorner$, we obtain $\ulcorner\phi\urcorner \in S_1$ because $(\mathcal{M}, S_1) \vDash \mathsf{KF}\!\upharpoonright$, contradicting the consistency of $(\mathcal{M}, S_1, S_2)$.

Sub-case 2: $(\mathcal{M}, S_1, S_2)$ is complete. In this case we can consider $(\mathcal{M}, R_1, R_2)$ obtained from $(\mathcal{M}, S_1, S_2)$ by turning truth-value gluts into truth-value gaps. This consistent model satisfies $\mathsf{PKF}\!\upharpoonright$ and all sequents satisfied by $(\mathcal{M}, S_1, S_2)$. We can then reason as in *sub-case 1*.

Sub-case 3: $(\mathcal{M}, S_1, S_2)$ contains both truth-value gaps and truth-value gluts. Again we move to the consistent model $(\mathcal{M}, R_1, R_2)$ obtained by turning gluts into gaps. This model, besides being a model of $\mathsf{PKF}\!\upharpoonright$, satisfies all sequents of the form $\Rightarrow \phi$ satisfied by $(\mathcal{M}, S_1, S_2)$. Again we can now reason in the first sub-case.

□

4.1. **Classical truth predicates in PKF-like theories.** We now recall the well-known fact that there is a strict correspondence between the amount of transfinite induction for $\mathcal{L}_{\mathsf{T}}$-formulas provable in PKF-like theories and the amount of classical, Tarskian truth predicates definable in them.

Definition 12. We let, for $\alpha < \Gamma_0$,

$$\mathcal{L}_0 := \varnothing \qquad \mathcal{L}_1 := \mathcal{L}$$

$$\mathcal{L}_{\alpha+1} := \mathcal{L}_\alpha \cup \{\mathsf{T}_a\} \text{ with } \alpha \geq 1 \qquad \mathcal{L}_\lambda := \bigcup_{\beta<\lambda} \mathcal{L}_\beta$$

with $\mathsf{T}_a x :\leftrightarrow \mathsf{Sent}_{\mathcal{L}_a}(x) \wedge \mathsf{T}x$, again $\alpha \geq 1$.

Lemma 7 (essentially [9, Lem. 32-33])**.**

(i) $\mathsf{PKF} \vdash \forall x(\mathsf{Sent}_{\mathcal{L}}(x) \to (\mathsf{T}x \vee \neg\mathsf{T}x))$

(ii) For $\alpha < \Gamma_0$,

$$\mathsf{PKF} \vdash \forall z < a(\mathsf{Sent}_{\mathcal{L}_z}(x) \to (\mathsf{T}x \vee \neg\mathsf{T}x)) \Rightarrow \mathsf{Sent}_{\mathcal{L}_a}(x) \to (\mathsf{T}x \vee \neg\mathsf{T}x)$$

Proof Sketch. (i) is proved by formal induction on the complexity of the sentence involved. The claim is trivial if x is atomic, i.e. of the form $u{=}v$, for u, v formal closed terms. For the induction step, one notices that PKF proves sequents of the form (cf. [8, Lem. 16.15])[5]

(22) $$\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x), \mathsf{T}x \vee \neg\mathsf{T}x \Rightarrow \mathsf{T}\neg x \vee \neg\mathsf{T}\neg x$$

(23) $$\mathsf{Sent}_{\mathcal{L}_{\mathsf{T}}}(x \wedge y), \mathsf{T}x \vee \neg\mathsf{T}x, \mathsf{T}y \vee \neg\mathsf{T}y \Rightarrow \mathsf{T}(x \wedge y) \vee \neg\mathsf{T}(x \wedge y)$$

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

For (ii), it suffices to assume that $1 < \alpha < \Gamma_0$. If α is a successor ordinal, we perform again an induction on the length of y to show

(25) $$\mathsf{Sent}_{\mathcal{L}_a}(x) \to \mathsf{T}x \vee \neg\mathsf{T}x \Rightarrow \mathsf{Sent}_{\mathcal{L}_{a+1}}(y) \to \mathsf{T}y \vee \neg\mathsf{T}y$$

In particular, to prove (25) we need (22)-(24) and

(26) $$\mathsf{T}x \vee \neg\mathsf{T}x \Rightarrow \mathsf{T}\ulcorner\mathsf{T}\dot{x}\urcorner \vee \neg\mathsf{T}\ulcorner\mathsf{T}\dot{x}\urcorner$$

which is in turn obtained by (T1), (T2). The case in which α is a limit ordinal is immediate by definition of $\mathcal{L}_\lambda$. □

[5] Similar facts are provable for the omitted cases of $\vee$ and $\exists$.

Lemma 7 immediately entails that , if $S \supseteq \mathsf{PKF}$ proves $\mathsf{TI}^{<\alpha}_{\mathcal{L}_\mathsf{T}}$, then S also proves $\mathsf{T}_b x \vee \neg \mathsf{T}_b x$ for $\beta < \alpha$. As a consequence, in such a theory S we can (i) reason classically with all truth predicates $\mathsf{T}_b x$ (see Remark 1); (ii) define classical ramified truth predicates for all $\beta < \alpha$. We have, in particular:

Corollary 2.

(i) $\mathsf{PKF} \vdash \mathsf{T}_a x \vee \neg \mathsf{T}_a x$ *for all* $\alpha < \omega^\omega$. *The same holds for* $\mathsf{PKF_S}$.
(ii) $\mathsf{PKF}^+ \vdash \mathsf{T}_a x \vee \neg \mathsf{T}_a x$ *for all* $\alpha < \varepsilon_0$. *The same holds for* $\mathsf{PKF}^+_\mathsf{S}$.

4.2. **The asymmetric interpretation formalized.** In this section we formalize, in suitable PKF-like theories, a method introduced by Cantini in [3, §5] to assign indices to truth predicates occurring in normalized KFI- and KF-derivations. Again we focus on the theories without the initial sequents (cs) or (cm), but the proofs adapt to these extensions without essential modifications.

We first reformulate KF-like theories in a one-sided Tait-style calculus. As usual, negation $\neg$ is now defined via the De Morgan laws starting from the *atomic* formulas $r = s, r \neq s, \mathsf{T}r, \neg\mathsf{T}r$ – we refer to these latter two kinds of formulas as *truth ascriptions*. We still call the resulting language $\mathcal{L}_\mathsf{T}$. $\Gamma, \Delta \ldots$ range over finite sets of formulas and $r, s, t \ldots$ over arbitrary terms of $\mathcal{L}_\mathsf{T}$.

Definition 13.

(i) TKFI is formulated in $\mathcal{L}_\mathsf{T}$ and its axioms are:

Logic and identity:

$\Gamma, \phi, \neg\phi$ with ϕ atomic

$\Gamma, s = s$ $\quad\quad$ $\Gamma, s \neq t, \neg\phi(s), \phi(t)$ with ϕ atomic

$\Gamma, s \neq t, t = s$ $\quad\quad$ $\Gamma, s \neq t, t \neq r, s = r$

Arithmetic:

$\Gamma, \mathsf{S}r \neq 0$ $\quad\quad$ $\Gamma, \mathsf{S}r \neq \mathsf{S}t, r = t$

Equations for suitable p.r. function symbols

Truth axioms:

$$\Gamma, \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \neg\mathsf{T}(x \dot{=} y), \mathsf{val}(x) = \mathsf{val}(y)$$
$$\Gamma, \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{val}(x) \neq \mathsf{val}(y), \mathsf{T}(x \dot{=} y)$$
$$\Gamma, \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \neg\mathsf{T}(x \dot{\neq} y), \mathsf{val}(x) \neq \mathsf{val}(y)$$
$$\Gamma, \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \mathsf{val}(x) = \mathsf{val}(y), \mathsf{T}(x \dot{\neq} y)$$
$$\Gamma, \neg\mathsf{T}x, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x)$$
$$\Gamma, \neg\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(x, v), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \neg\mathsf{T}x(\dot{\mathsf{val}}(y)/v), \mathsf{T}x(y/v)$$
$$\Gamma, \neg\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(x, v), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \neg\mathsf{T}x(y/v), \mathsf{T}x(\dot{\mathsf{val}}(y)/v)$$

Its rules of inference are:

Standard logical principles including (cf. [21]*)* (Cut)

$$\frac{\Gamma, \phi \qquad \Gamma, \neg\phi}{\Gamma}\ (\mathrm{Cut}) \qquad\qquad \frac{\Gamma, (\neg)\mathsf{T}z}{\Gamma, (\neg)\mathsf{T}\ulcorner \mathsf{T}\dot{z}\urcorner}\ (\mathsf{Trp})$$

$$\frac{\Gamma, (\neg)(\mathsf{T}\neg z \vee \neg\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(z))}{\Gamma, (\neg)\mathsf{T}\neg\ulcorner \mathsf{T}\dot{z}\urcorner}\ (\mathsf{Tnrp}) \qquad\qquad \frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \qquad \Gamma, (\neg)\mathsf{T}x}{\Gamma, (\neg)\mathsf{T}\neg\neg x}\ (\mathsf{Tdn})$$

$$\frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x \wedge y) \qquad \Gamma, (\neg)(\mathsf{T}x \wedge \mathsf{T}y)}{\Gamma, (\neg)\mathsf{T}(x \wedge y)}\ (\mathsf{T\&}) \qquad\qquad \frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x \wedge y) \qquad \Gamma, (\neg)(\mathsf{T}\neg x \vee \mathsf{T}\neg y)}{\Gamma, (\neg)\mathsf{T}\neg(x \wedge y)}\ (\mathsf{T\neg\&})$$

$$\frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x \vee y) \qquad \Gamma, (\neg)(\mathsf{T}x \vee \mathsf{T}y)}{\Gamma, (\neg)\mathsf{T}(x \vee y)}\ (\mathsf{Tor}) \qquad\qquad \frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x \vee y) \qquad \Gamma, (\neg)(\mathsf{T}\neg x \wedge \mathsf{T}\neg y)}{\Gamma, (\neg)\mathsf{T}\neg(x \vee y)}\ (\mathsf{Tnor})$$

$$\frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall vx) \qquad \Gamma, (\neg)\forall y\, \mathsf{T}x(\dot{y}/v)}{\Gamma, (\neg)\mathsf{T}\forall vx}\ (\mathsf{Tall}) \qquad\qquad \frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\forall vx) \qquad \Gamma, (\neg)\exists y\, \mathsf{T}\neg x(\dot{y}/v)}{\Gamma, (\neg)\mathsf{T}\neg\forall vx}\ (\mathsf{Tnall})$$

$$\frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\exists vx) \qquad \Gamma, (\neg)\exists y\, \mathsf{T}x(\dot{y}/v)}{\Gamma, (\neg)\mathsf{T}\exists vx}\ (\mathsf{Tex}) \qquad\qquad \frac{\Gamma, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(\exists vx) \qquad \Gamma, (\neg)\forall y\, \mathsf{T}\neg x(\dot{y}/v)}{\Gamma, (\neg)\mathsf{T}\neg\exists vx}\ (\mathsf{Tnex})$$

$$\frac{\Gamma, \mathsf{T}x(\ulcorner 0\urcorner/v) \qquad \Gamma, \forall y\, (\mathsf{T}x(\dot{y}/v) \to \mathsf{T}x(y \dot{+} 1/v))}{\Gamma, \mathsf{T}x(\dot{u}/v)}\ (\mathsf{Tind})$$ with u not free in $\Gamma, \mathsf{T}x(\ulcorner 0\urcorner/v)$.

(ii) TKF is obtained by replacing (Tind) with the schema, for $\phi(v)$ an arbitrary formula of $\mathcal{L}_\mathsf{T}$ with only v free:[6]

$$\frac{\Gamma, \phi(0) \qquad \Gamma, \forall y\, (\phi(y) \to \phi(y+1))}{\Gamma, \forall x\phi}\ (\mathsf{Ind})$$

It is not difficult to see that TKFI and TKF contain KFI and KF respectively.

Notation. The *length* of a derivation-tree $\mathcal{D}$ is defined inductively as the supremum of the lengths of its sub-derivations plus 1. The *complexity* of a formula of $\mathcal{L}_\mathsf{T}$ is 0 for atomic formulas and is extended to complex formulas in the usual way. The *cut rank* of $\mathcal{D}$ is defined as the maximum complexity of its cut-formulas. As usual, $S \vdash^k_l \Gamma$ expresses that Γ is derivable in S with a proof of finite length $\leq k$ and cut-rank $\leq l$: therefore $S \vdash^k \Gamma$ conveys the information that Γ has a proof with cuts only applied to atomic formulas (*normal cuts*). This relation can be canonically represented in arithmetic via a recursively enumerable predicate $\mathsf{Bew}_S(k, l, \ulcorner\Gamma\urcorner)$.

Lemma 8 (Partial cut-elimination)**.** *The following is provable in* basic *plus (*IND1*), with* $2^x_0 := x$ *and* $2^x_{n+1} := 2^{2^x_n}$*:*

$$\mathsf{Bew}_{\mathsf{TKFI}}(y, u, \ulcorner\Gamma\urcorner) \to \mathsf{Bew}_{\mathsf{TKFI}}(2^y_u, 0, \ulcorner\Gamma\urcorner)$$

We now assign indices to truth predicates involved in derivations with only normal cuts by formalizing [3, Thm. 9.9] in PKF. In particular, we introduce the predicates $\mathsf{Tr}^b_a(x)$ that formalize the capture the idea of an *asymmetric* interpretation of the construction of the minimal fixed-point of Kripke's theory of truth. The role of these predicates is essential in evaluating truth ascriptions and their intended meaning can be described as follows: if $\mathsf{Tr}^b_a(\mathsf{T}\ulcorner\phi\urcorner)$, then the sentence ϕ is *definitely true* at the level β of the construction of the minimal fixed point; if $\mathsf{Tr}^b_a(\neg\mathsf{T}\ulcorner\phi\urcorner)$, then ϕ is *not yet* true at level α of the construction.

Definition 14. We inductively define the predicate $\mathsf{Tr}^b_a(z)$ for $\alpha, \beta < \Gamma_0$ applying to (codes of) sentences of $\mathcal{L}_\mathsf{T}$. We make use of the Tarskian, hierarchical truth predicates introduced in

[6]In this case the last two truth axioms become redundant.

Definition 12.

$$\begin{aligned}
&\mathsf{Tr}^b_a(u{=}v) :\leftrightarrow \mathsf{val}(u) = \mathsf{val}(v) && \mathsf{Tr}^b_a(u{\neq}v) :\leftrightarrow \mathsf{val}(u) \neq \mathsf{val}(v)\\
&\mathsf{Tr}^b_a(\mathsf{T}u) :\leftrightarrow \mathsf{T}_b\mathsf{val}(u) && \mathsf{Tr}^b_a(\neg\mathsf{T}u) :\leftrightarrow \neg\mathsf{T}_a\mathsf{val}(u)\\
&\mathsf{Tr}^b_a(u{\wedge}v) :\leftrightarrow \mathsf{Tr}^b_a(u) \wedge \mathsf{Tr}^b_a(v) && \mathsf{Tr}^b_a(\neg(u{\wedge}v)) :\leftrightarrow \mathsf{Tr}^b_a(\neg u) \vee \mathsf{Tr}^b_a(\neg v)\\
&\mathsf{Tr}^b_a(u{\vee}v) :\leftrightarrow \mathsf{Tr}^b_a(u) \vee \mathsf{Tr}^b_a(v) && \mathsf{Tr}^b_a(\neg(u{\vee}v)) :\leftrightarrow \mathsf{Tr}^b_a(\neg u) \wedge \mathsf{Tr}^b_a(\neg v)\\
&\mathsf{Tr}^b_a(\forall vu) :\leftrightarrow \forall y\, \mathsf{Tr}^b_a(u(\dot{y}/v)) && \mathsf{Tr}^b_a(\neg\forall vu) :\leftrightarrow \exists y\, \mathsf{Tr}^b_a(\neg u(\dot{y}/v))\\
&\mathsf{Tr}^b_a(\exists vu) :\leftrightarrow \exists y\, \mathsf{Tr}^b_a(u(\dot{y}/v)) && \mathsf{Tr}^b_a(\neg\exists vu) :\leftrightarrow \forall y\, \mathsf{Tr}^b_a(\neg u(\dot{y}/v))
\end{aligned}$$

The definition extends to (codes of) sets Γ of substitutional instances of $\mathcal{L}_\mathsf{T}$-formulas by considering the disjunction of the elements of Γ.

Lemma 9. *For $\delta < \Gamma_0$ and $1 \leq \alpha < \beta < \gamma < \delta$, if $U \supseteq \mathsf{PKF}$ and $U \vdash \mathsf{TI}^\delta_{\mathcal{L}_\mathsf{T}}$, then*

$$U \vdash \forall x\, (\mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x) \wedge \mathsf{Tr}^c_b(x) \to \mathsf{Tr}^d_a(x))$$

Proof. The proof is by formal induction on the construction of x. We consider the crucial cases of truth ascriptions. The others are easily obtained. We reason in U.

If $\mathsf{Tr}^c_b(\mathsf{T}y)$, then $\mathsf{T}_c\mathsf{val}(y)$ by Def. 14. Also, for $\alpha < \beta < \delta$, we can prove by induction on x:

$$\Rightarrow \neg\mathsf{T}_a x, \mathsf{T}_b x \tag{27}$$

Therefore $\mathsf{T}_d\mathsf{val}(y)$. Similarly, if $\mathsf{Tr}^c_b(\neg\mathsf{T}y)$, we obtain $\neg\mathsf{T}_b\mathsf{val}(y)$ and, again by (27), $\neg\mathsf{T}_a\mathsf{val}(y)$. □

Proposition 2. *For $n \in \omega$, $1 \leq \alpha < \omega^\omega$, $\mathsf{PKF} \vdash \mathsf{Bew}_{\mathsf{TKFI}}(n, 0, \ulcorner\Gamma(\dot{\vec{y}})\urcorner) \to \mathsf{Tr}^{a+\omega^n}_a(\ulcorner\Gamma(\dot{\vec{y}})\urcorner)$.*

Proof. The proof is by meta-induction on n. We argue informally in PKF: in particular, when referring to claims derivable in PKF of the form $\Rightarrow \Gamma$, we often simply write Γ. For readability, we also omit the additional parameters in side formulas.

$n = 0$. If Γ is of the form $\Gamma', \phi, \neg\phi$, with ϕ atomic, then either Γ is of the form $\Gamma', r = s, r \neq s$ or of the form $\Gamma', \mathsf{T}x, \neg\mathsf{T}x$. In the former case we are done by Lemma 3(ii); in the latter, we employ Corollary 2 and (27) to conclude $\neg\mathsf{T}_a x, \mathsf{T}_{a+1}x$ and therefore $\mathsf{Tr}^{a+1}_a(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_a x, \mathsf{T}_{a+1}x$. The desired conclusion then follows from Definition 14. The case of the identity axioms is trivial.

If Γ is an instance of the first *truth axiom*, one simply notices that

$$\Rightarrow \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(x), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \neg\mathsf{T}_a(x{=}y), \mathsf{val}(x) = \mathsf{val}(y)$$

holds for all $\alpha < \omega^\omega$. We are then done by applying Definition 14. If Γ is for the form $\Gamma', \neg\mathsf{T}x, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x)$, it suffices to notice that $\neg\mathsf{T}_a x, \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x)$ holds. If Γ is of the form

$$\Gamma', \neg\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(x, v), \neg\mathsf{Cterm}_{\mathcal{L}_\mathsf{T}}(y), \neg\mathsf{T}x(\dot{\mathsf{val}(y)}/v), \mathsf{T}x(y/v),$$

then we have for all $\alpha < \omega^\omega$,

$$\begin{aligned}
&\Rightarrow \neg\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(x, v), \neg\mathsf{T}_a x(\dot{\mathsf{val}(y)}/v), \mathsf{T}_a x(y/v) && \text{by } (\mathsf{Reg1\text{-}2})\\
&\Rightarrow \neg\mathsf{Fml}^1_{\mathcal{L}_\mathsf{T}}(x, v), \neg\mathsf{T}_a x(\dot{\mathsf{val}(y)}/v), \mathsf{T}_{a+1} x(y/v) && \text{by Lemma 9}
\end{aligned}$$

The desired conclusion follows then by Definition 14.

$n > 0$. We first consider a proof π in TKFI of length n ending with an application of (Trp):

$$\pi : \quad \dfrac{\pi_0 : \quad \begin{matrix}\vdots\\ \Gamma', \mathsf{T}z\end{matrix}}{\Gamma', \mathsf{T}\ulcorner\mathsf{T}\dot{z}\urcorner}$$

By induction hypothesis, for $\alpha < \omega^\omega$, since π_0 has length $\leq n_0 < n$, we have $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner, \ulcorner\mathsf{T}\dot{z}\urcorner)$, that is $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner)$, $\mathsf{T}_{a+\omega^{n_0}}z$. But then already PA suffices to prove $\mathsf{Sent}_{\mathcal{L}_{a+\omega^n}}(\ulcorner\mathsf{T}\dot{z}\urcorner)$ and $\mathsf{T}\ulcorner\mathsf{T}\dot{z}\urcorner$ holds by the axioms for truth ascriptions. Therefore also $\mathsf{Tr}_a^{a+\omega^n}(\ulcorner\Gamma'\urcorner)$, $\mathsf{T}_{a+\omega^n}\ulcorner\mathsf{T}\dot{z}\urcorner$ by Lemma 9. By applying Definition 14 we obtained the desired conclusion for each external $\alpha < \Gamma_0$. To complete the treatment of (Trp), we now consider a proof π in TKFI of length n ending with:

$$\pi : \cfrac{\pi_0 : \begin{array}{c}\vdots\\ \Gamma', \neg\mathsf{T}z\end{array}}{\Gamma', \neg\mathsf{T}\ulcorner\mathsf{T}\dot{z}\urcorner}$$

Again, since the length of π_0 is $\leq n_0 < n$ we have, for all $\alpha < \omega^\omega$, $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma', \neg\mathsf{T}\dot{z}\urcorner)$, that is $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner)$, $\neg\mathsf{T}_a z$. However, it is an arithmetical fact that $\mathsf{Sent}_{\mathcal{L}_a}(\ulcorner\mathsf{T}\dot{z}\urcorner)$ entails $\mathsf{Sent}_{\mathcal{L}_a}(z)$, therefore, by Lemma 9 and (T1-2),

$$\Rightarrow \mathsf{Tr}_a^{a+\omega^n}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}\ulcorner\mathsf{T}\dot{z}\urcorner, \neg\mathsf{Sent}_{\mathcal{L}_a}(\ulcorner\mathsf{T}\dot{z}\urcorner)$$

which yields the claim by Definition 14. The cases for (Tnrp) are treated similarly.

For the Boolean connectives, we consider (T&), first focusing on a proof π of length n ending with:

$$\pi : \cfrac{\pi_0 : \cfrac{\vdots}{\Gamma', \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x\dot\wedge y)} \qquad \pi_1 : \cfrac{\vdots}{\Gamma', \mathsf{T}x \wedge \mathsf{T}y}}{\Gamma', \mathsf{T}(x\dot\wedge y)}$$

Let π_0, π_1 be of length $\leq n_0 < n$. By induction hypothesis and Definition 14, $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner), \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x\dot\wedge y)$, and $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^{n_0}}x\wedge\mathsf{T}_{a+\omega^{n_0}}y$ are derivable in PKF; therefore $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^{n_0}}(x\dot\wedge y)$ by the behaviour of $\mathsf{T}_{a+\omega^{n_0}}$, and $\mathsf{Tr}_a^{a+\omega^n}(\ulcorner\Gamma', \mathsf{T}\dot{x}\dot\wedge\dot{y}\urcorner)$ by Definition 14 and Lemma 9. If π ends with:

$$\pi : \cfrac{\pi_0 : \cfrac{\vdots}{\Gamma', \mathsf{Sent}_{\mathcal{L}_\mathsf{T}}(x\dot\wedge y)} \qquad \pi_1 : \cfrac{\vdots}{\Gamma', \neg(\mathsf{T}x \wedge \mathsf{T}y)}}{\Gamma', \neg\mathsf{T}(x\dot\wedge y)}$$

As before, since π_0, π_1 have length $\leq n_0 < n$, by induction hypothesis $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_a x, \neg\mathsf{T}_a y$ is a derivable sequent in PKF. Therefore $\mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner)$, $\neg\mathsf{T}_a(x\dot\wedge y)$ will also be derivable since $\neg\mathsf{Sent}_{\mathcal{L}_a}x, \neg\mathsf{T}x$ entails $\neg\mathsf{Sent}_{\mathcal{L}_a}(x\dot\wedge y), \neg\mathsf{T}(x\dot\wedge y)$. Therefore $\mathsf{Tr}_a^{a+\omega^n}(\ulcorner\Gamma', \neg\mathsf{T}(\dot{x}\dot\wedge\dot{y})\urcorner)$ by Definition 14 and Lemma 9.

Let a proof in TKFI end with an application of (Tind) – where, for readability, we have suppressed the substituted variable:

$$\pi : \cfrac{\pi_0 : \cfrac{\vdots}{\Gamma', \mathsf{T}x(\ulcorner 0\urcorner)} \qquad \pi_1 : \cfrac{\vdots}{\Gamma', \neg\mathsf{T}x(\dot{y}), \mathsf{T}x(y \dot{+} 1)}}{\Gamma', \mathsf{T}x(\dot{u})}$$

with u arbitrary. Since π_0, π_1 have length $\leq n_0 < n$, the induction hypothesis gives us, for all $\alpha < \omega^\omega$:

$$\Rightarrow \mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^{n_0}}x(\ulcorner 0\urcorner) \tag{28}$$

$$\Rightarrow \mathsf{Tr}_a^{a+\omega^{n_0}}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_a x(\dot{y}), \mathsf{T}_{a+\omega^{n_0}}x(y \dot{+} 1) \tag{29}$$

By Definition 14, it suffices to obtain $\mathsf{Tr}_a^{a+\omega^n}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^n} x(\dot{u})$ for an arbitrary u; this, however, would follow by applying Lemma 9 to

$$\Rightarrow \mathsf{Tr}_a^{a+\omega^{n_0+1}}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^{n_0}(\dot{u}+1)} x(\dot{u}) \tag{30}$$

again for u arbitrary. To derive (30), we employ (IND1). (28) gives us directly the claim for $u = 0$. Let $\beta := \alpha + \omega^{n_0}(y+1)$. We have, in PKF,

$$\begin{aligned}
&\Rightarrow \mathsf{Tr}_b^{b+\omega^{n_0}}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_b x(\dot{y}), \mathsf{T}_{b+\omega^{n_0}} x(y \dot{+} 1) && \text{by (29)}\\
&\Rightarrow \mathsf{Tr}_a^{a+\omega^{n_0+1}}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_b x(\dot{y}), \mathsf{T}_{b+\omega^{n_0}} x(y \dot{+} 1) && \text{by Lemma 9}\\
\mathsf{T}_{a+\omega^{n_0}(\dot{y}+1)} x(\dot{y}) &\Rightarrow \mathsf{Tr}_a^{a+\omega^{n_0+1}}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^{n_0}(\dot{y}+2)} x(y \dot{+} 1) && \text{Cor. 2, def. } \beta\\
&\Rightarrow \mathsf{Tr}_a^{a+\omega^{n_0+1}}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^{n_0}(\dot{u}+1)} x(\dot{u}) && \text{(IND1)}\\
&\Rightarrow \mathsf{Tr}_a^{a+\omega^{n_0+1}}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{a+\omega^{n_0+1}} x(\dot{u}) && \text{Lemma 9}
\end{aligned}$$

The last line, as anticipated, yields the claim by Definition 14 and, possibly, Lemma 9 as $\omega^{n_0+1} \leq \omega^n$.

The case of a proof ending with an application of the cut rule is the crucial case when the presence of the *negative* clause of our formalization of the asymmetric interpretation matters – that is the ordinal subscript a in $\mathsf{Tr}_a^b(x)$. So if our proof π ends with an application of the cut rule

$$\pi : \cfrac{\pi_0 : \cfrac{\vdots}{\Gamma', \mathsf{T}s} \qquad \pi_1 : \cfrac{\vdots}{\Gamma', \neg\mathsf{T}s}}{\Gamma'}$$

we have that $\pi_0, \pi_1 \leq m < n$. By induction hypothesis, for $\beta < \omega^\omega$,

$$\mathsf{Tr}_b^{b+\omega^m}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{b+\omega^m}\mathsf{val}(s) \tag{31}$$

$$\mathsf{Tr}_b^{b+\omega^m}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_b\mathsf{val}(s) \tag{32}$$

By applying Lemma 9 to (31) we obtain

$$\mathsf{Tr}_b^{b+\omega^m\cdot 2}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{b+\omega^m}\mathsf{val}(s) \tag{33}$$

Moreover, by letting β to be $\beta + \omega^m$ in (32) and applying Lemma 9 to the result, we obtain

$$\mathsf{Tr}_b^{b+\omega^m\cdot 2}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_{b+\omega^m}\mathsf{val}(s) \tag{34}$$

Since $\omega^m \cdot 2 < \omega^n$, we obtain the desired result by Lemma 9 and cut. □

Corollary 3. *For ϕ an $\mathcal{L}_\mathsf{T}$-sentence, if* $\mathsf{KFI} \vdash \mathsf{T}\ulcorner\phi\urcorner$*, then* $\mathsf{PKF} \vdash \phi$.

Proof. If $\mathsf{KFI} \vdash \mathsf{T}\ulcorner\phi\urcorner$, by Lemma 8 PKF also proves $\mathsf{Bew}_{\mathsf{TKFI}}(n, 0, \ulcorner\mathsf{T}\ulcorner\phi\urcorner\urcorner)$ for some $n \in \omega$. Therefore, still in PKF, Proposition 2 gives us $\mathsf{T}_{\omega^n}\ulcorner\phi\urcorner$, and hence ϕ by Lemma 3. □

Next we apply a modification of this strategy to KF. Since an analogue of Lemma 8 (partial cut elimination) is not available for TKF, we move to the infinitary system KF^∞.

Definition 15. The theory KF^∞ is obtained from TKF by

(i) omitting free variables
(ii) replacing arithmetical axioms with the initial sequents

$$\Gamma, r = s \qquad\qquad \Gamma, r \neq s$$

whenever r, s are closed terms and, respectively, $r^\mathbb{N} = s^\mathbb{N}$ and $r^\mathbb{N} \neq s^\mathbb{N}$.

(iii) replacing (Ind) and the logical rule of introduction of the universal quantifier with the infinitary rule:

$$\frac{\Gamma, \phi(s) \qquad \text{for any closed term } s}{\Gamma, \forall x\, \phi(x)}\ (\omega)$$

Derivations in KF^∞ are now possibly infinite as (ω) has infinitely many premises. However, it is well-known (cf. [21, 20]) that we can restrict our attention to recursive applications of the ω-rule by considering (the index for) a primitive recursive enumeration of the premises: derivations become well-founded trees whose nodes are either (codes of) the root, or codes of instances of axioms, or the result of applying a unary, binary, or ω-rule to codes of the premises. In all these cases codes also contain information about the length, cut-rank of the coded derivations. This enables us to find a predicate $\mathsf{Bew}_\infty(a, n, \ulcorner\Gamma\urcorner)$ expressing that there is a tree whose nodes are 'locally correct' in the sense just specified and that $\mathsf{KF}^\infty \vdash^\alpha_n \Gamma$, that is Γ is derivable in KF^∞ with a derivation of ordinal length $\leq \alpha$ and cut rank $\leq n$.

The following lemma states the well-known fact that the transfinite induction of PKF^+ enables us to formalize the embedding and partial cut elimination for externally given proofs, where $\varphi_0^m\alpha$ stands for m iterations of $\varphi_0\cdot$ on α.

Lemma 10 (Cf. [21, 3, 18])**.**

(i) (Embedding). $\mathsf{PKF}^+ \vdash \mathsf{Bew}_{\mathsf{KF}}(n, m, \ulcorner\Gamma\urcorner) \to \mathsf{Bew}_\infty(\omega^2, m, \ulcorner\Gamma\urcorner)$*, for all* $n, m \in \omega$*.*

(ii) (Partial Cut-Elimination). For $\alpha < \varepsilon_0$*,* $\mathsf{PKF}^+ \vdash \mathsf{Bew}_\infty(a, m, \ulcorner\Gamma\urcorner) \to \mathsf{Bew}_\infty(\varphi_0^m a, 0, \ulcorner\Gamma\urcorner)$

Lemma 10 tells us that we can restrict our attention to derivations of cut-rank 0. We can finally proceed with an analogue of Proposition 2. We notice that there are no free variables now in the claim.

Proposition 3. *For* $1 \leq \alpha, \beta < \varepsilon_0$*,* $\mathsf{PKF}^+ \vdash \mathsf{Bew}_\infty(a, 0, \ulcorner\Gamma\urcorner) \to \mathsf{Tr}_b^{b+2^a}(\ulcorner\Gamma\urcorner)$

Proof. The proof proceeds by external, transfinite induction on $\alpha < \varepsilon_0$. As most of the cases follow the blueprint of Proposition 3, I only focus on the cut rule: the argument is structurally similar to the case of the cut rule in Proposition 2, but the change on the ordinal bounds and its relevance to the main claim makes it worth repeating. The case of the rule (ω) for limit ordinals is also immediate given the definition of the predicate $\mathsf{Tr}^{\cdot}_{\cdot}(\cdot)$.

Let's assume that our proof π ends with an application of the cut rule:

$$\pi : \quad \frac{\pi_0 : \dfrac{\vdots}{\Gamma', \mathsf{T}s} \qquad \pi_1 : \dfrac{\vdots}{\Gamma', \neg\mathsf{T}s}}{\Gamma'}$$

with π_0, π_1 of length $\leq \gamma < \alpha < \varepsilon_0$. By induction hypothesis, for $\beta < \varepsilon_0$,

$$\mathsf{Tr}_b^{b+2^c}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{b+2^c}\mathsf{val}(s) \tag{35}$$

$$\mathsf{Tr}_b^{b+2^c}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_b\mathsf{val}(s) \tag{36}$$

Therefore we obtain

$$\mathsf{Tr}_b^{b+2^c\cdot 2}(\ulcorner\Gamma'\urcorner), \mathsf{T}_{b+2^c}\mathsf{val}(s) \tag{37}$$

$$\mathsf{Tr}_b^{b+2^c\cdot 2}(\ulcorner\Gamma'\urcorner), \neg\mathsf{T}_{b+2^c}\mathsf{val}(s) \tag{38}$$

by applying, respectively, Lemma 9 to (35), and by letting β to be $\beta + 2^\gamma$ in (36) before applying Lemma 9 to the result. An application of the cut rule – and possibly Lemma 9 since $2^\gamma \cdot 2 \leq 2^\alpha$ – yields $\mathsf{Tr}_b^{b+2^a}(\ulcorner\Gamma'\urcorner)$ as desired. □

Corollary 4. *For all* $\mathcal{L}_{\mathsf{T}}$*-sentences* ϕ*, if* $\mathsf{KF} \vdash \mathsf{T}\ulcorner\phi\urcorner$*, then* $\mathsf{PKF}^{+} \vdash \phi$*.*

Proof. If $\mathsf{KF} \vdash \mathsf{T}\ulcorner\phi\urcorner$ then, for some $n \in \omega$, PKF^{+} proves $\mathsf{Bew}_{\infty}(\omega^2, n, \ulcorner\mathsf{T}\ulcorner\varphi\urcorner\urcorner)$ and therefore, by Lemma 10(ii), also $\mathsf{Bew}_{\infty}(a, 0, \ulcorner\mathsf{T}\ulcorner\varphi\urcorner\urcorner)$ for some $\alpha < \varepsilon_0$. By Proposition 3, $\mathsf{T}_c\ulcorner\phi\urcorner$ for some $\gamma < \varepsilon_0$ is provable in PKF^{+} and so is ϕ by Lemma 3. □

It is worth stressing again here that the strategy pursued in Proposition 2 and Proposition 3 clearly goes through without changes in the ordinal bounds even if, instead of KFI and KF, we focus on the case of $\mathsf{KFI}_{\mathsf{S}}$ and KF_{S}.

5. Summary of the results and philosophical assessment

Corollary 1, together with Lemma 6, Corollaries 3 and 4, yields the following theorem.

Theorem 1.

(i) $\mathsf{PKF}\!\restriction\, = \mathsf{IKF}\!\restriction$ *and* $\mathsf{PKF}_{\mathsf{S}}\!\restriction\, = \mathsf{IKF}_{\mathsf{S}}\!\restriction$*;*
(ii) $\mathsf{PKF} = \mathsf{IKFI}$ *and* $\mathsf{PKF}_{\mathsf{S}} = \mathsf{IKFI}_{\mathsf{S}}$*;*
(iii) $\mathsf{PKF}^{+} = \mathsf{IKF}$ *and* $\mathsf{PKF}^{+}_{\mathsf{S}} = \mathsf{IKF}_{\mathsf{S}}$*.*

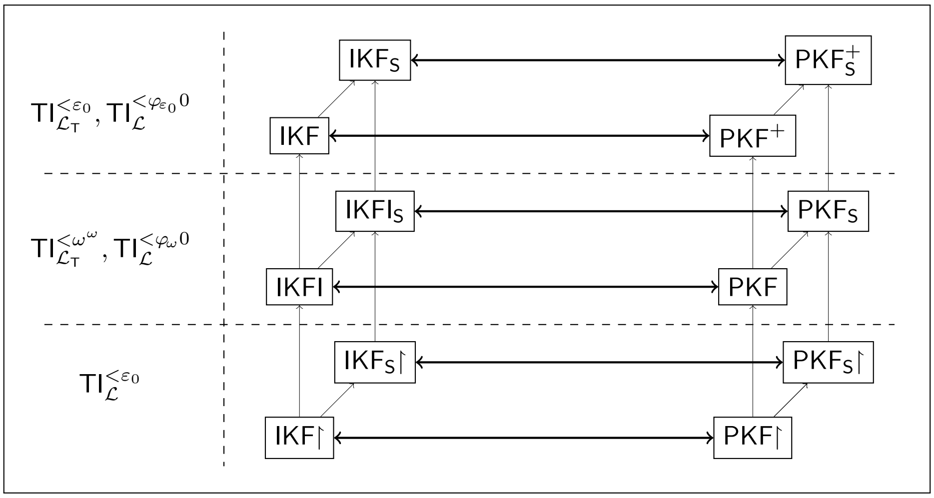

Figure 1. The content of Thm. 1

The content of Theorem 1 is summarized in Figure 1. The arrows stand for theory inclusion: on the leftmost column one can find the levels of transfinite induction with and without the truth predicate corresponding to the relevant PKF-like and internal KF-like theories. Here I briefly sketch some direction for a philosophical evaluation of the results presented above: a proper philosophical assessment is deferred to forthcoming work.

One might think, as Reinhardt [19] did, of KF-like theories as devices to grasp the core of Kripke's construction. This project is clearly doomed if it is read in analogy with Hilbert's program as reducing the *reasoning* available in fixed-point models to proofs in KF-like theories. This is because, due to the close connections between the sequent arrow and the material conditional in the axioms of KF-like theories, none of these axioms will be sound with respect to fixed-point models.[7] Reinhardt's project, in this sense, is completely superseded by the formulation of Halbach's and Horsten's PKF and its variants.

But there may be another sense in which Reinhardt's original program can be rescued. Starting with Feferman's famous remark 'nothing like sustained ordinary reasoning can be carried on in. . . [Basic De Morgan Logics]' ([5, p. 95]), several authors have advocated the view that the costs of abandoning classical logic – in our case, preferring PKF-like theories over KF-like theories – surpass the advantages given by the intersubstitutivity of ϕ and $\mathsf{T}\ulcorner\phi\urcorner$ in all contexts [16, 8, 17, 23].[8] According to this line of reasoning, abandoning classical logic cripples the links between semantics and other chunks of science and philosophy. The asymmetry of provable transfinite induction between KF and PKF amounts therefore to a clear example of non-semantic patterns of reasoning that one has to give up when endorsing logics such as BDM. Other authors, such as Reinhardt [19], defend classical theories, and in particular KF, as devices to retrieve IKF that, besides being a set of sentences true in any fixed-point model, displays some virtues in dealing with strengthened liar challenges.

For both kinds of supporters of KF-like solutions Theorem 1 offers the possibility of refining their positions in a way that also seems to rescue the analogy with Hilbert's programme suggested by Reinhardt. Just like Hilbert [10], who intended to give a solid basis to the mathematicians' work in 'the paradise which Cantor has created for [them]' by admitting, for any proof involving ideal elements, a real counterpart, the advocate of KF (in both forms) may look at any of its theorems of the form $\mathsf{T}\ulcorner\phi\urcorner$ and take for granted the existence of a proof of ϕ in PKF^+, therefore involving only principles of reasoning sound with respect to fixed point models. Under this reading, one appears to be free to employ KF and enjoy its alleged advantages, therefore, without being pushed away from the paradise Kripke has pointed us at.

By contrast, according to the proponent of PKF-like theories, the intersubstitutivity of ϕ and $\mathsf{T}\ulcorner\phi\urcorner$ and the 'silence' of PKF with respect to paradoxical sentences is preferable to keeping classical logic, which can nonetheless be recaptured for relevant portions of the language (cf. Lemma 3). For such authors the way in which intersubstitutivity fails in KF-theories, i.e. via the provability of claims of the form $(\neg\mathsf{T}\ulcorner\lambda\urcorner\wedge\lambda)\vee(\mathsf{T}\ulcorner\lambda\urcorner\wedge\neg\lambda)$, is a sufficient reason to consider classical theories as inadequate accounts of truth ([7, 11]). Also for such positions, Theorem 1 proves to be relevant: if challenged with the alleged clumsiness and weakness of the inferential structure of, say, PKF, its advocate may reply that theorems of PKF can be reached in a classical manner, that is via KFI. Furthermore, this classical detour is by no means ad hoc, as it is adequate in the sense of Lemma 5.

We conclude by pointing at some possible extensions of our study. It is natural to ask, for instance, whether a KF-like theory that is proof-theoretically stronger than KF proves the same sentences true as a natural extension of PKF with more transfinite induction. A pair of theories to test in this respect may be on the one hand John Burgess' strengthening of KF (called BKF in [8, §17]) that is as strong as ramified truth up to Γ_0 or ID_1, on the other PKF plus $\mathsf{TI}^{<\Gamma_0}_{\mathcal{L}_\mathsf{T}}$. Furthermore, it seems reasonable to look for a uniform and justified way to reach IKF without departing from the hierarchy of PKF-like theories: to this end, a hierarchy of reflection principles on a version of BDM extended with weak arithmetical axioms and initial sequents of the form

[7] This is essentially the content of of [9, Thm. 8].

[8] Elsewhere I will attempt to consider the cases of PKF-like theories and KF-like theories and the content of Theorem 1 as case studies for isolating the abductive virtues and drawbacks of both sides and apply them to the broader issue of logical pluralism. Here I limit myself to sketch how the picture resulting from Theorem 1 can be relevant for the positions considered above.

$\mathsf{T}\ulcorner\phi\urcorner \Rightarrow \phi$ and $\phi \Rightarrow \mathsf{T}\ulcorner\phi\urcorner$ may be tested. Some initial steps in this direction have been carried out in [14], although it is still an open question whether finitely many applications of reflection may lead from this basic starting point to IKF.

References

[1] N. Belnap. A useful four-valued logic. In J.M. Dunn and G. Epstein, editors, *Modern Uses of Multiple-Valued Logic*. Dordrecht, 1977.

[2] S. Blamey. Partial logic. In Dov. M Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, volume 5, pages 261–353. Kluwer, Dordrecht, 2 edition, 2002.

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

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

[5] S. Feferman. Towards useful type-free theories i. *Journal of Symbolic Logic*, 49(1):75–111, 1984.

[6] S. Feferman. Reflecting on incompleteness. *Journal of Symbolic Logic*, 56: 1–49, 1991.

[7] H. Field. *Saving truth from paradox*. Oxford University Press, Oxford, 2008.

[8] V. Halbach. *Axiomatic theories of truth. Revised edition*. Cambridge University Press, 2014.

[9] V. Halbach and L. Horsten. Axiomatizing Kripke's theory of truth in partial logic. *Journal of Symbolic Logic*, 71: 677–712, 2006.

[10] D. Hilbert. Über das Unendliche. In J. Van Heijenoort, editor, *From Frege to Gödel. A source book in mathematical logic*. Harvard University Press, 1967.

[11] L. Horsten. *The Tarskian Turn*. MIT University Press, Oxford, 2012.

[12] M. Kremer. Kripke and the logic of truth. *Journal of Philosophical Logic*, 17: 225–278, 1988.

[13] S. Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–712, 1975.

[14] C. Nicolai M. Fischer and L. Horsten. Iterated reflection over full disquotational truth. *Submitted*.

[15] J. Stern M. Fischer, V. Halbach and J. Kriener. Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, 8(2):257–278, 2015.

[16] V. McGee. *Truth, vagueness, and paradox*. MIT University Press, 1991.

[17] C. Nicolai and V. Halbach. On the costs of nonclassical logic. *Journal of Philosophical Logic*. Online first: DOI: 10.1007/s10992-017-9424-3.

[18] W. Pohlers. *Proof Theory. A firt step into impredicativity*. Springer, 2009.

[19] W. Reinhardt. Some remarks on extending and interpreting theories with a partial predicate for truth. *Journal of Philosophical Logic*, 15:219–251, 1986.

[20] U. R. Schmerl. Iterated reflection principle and the ω-rule. *Journal of Symbolic Logic*, 4:721–733, 1982.

[21] H. Schwichtenberg. Proof theory: some applications of cut-elimination. In Jon Barwise, editor, *Handbook of Mathematical Logic*. North-Holland, 1977.

[22] A. Tarski. Der Wahrhetisbegriff in den formalisierten Sprachen. In *Logic, semantics, metamathematics*, pages 152–278. Clarendon Press, Oxford, 1956.

[23] T. Williamson. Semantic paradoxes and abductive methodology. In B. Armour-Garb, editor, *The relevance of the liar*. Oxford University Press, Oxford.