Noname manuscript No.
(will be inserted by the editor)

On the costs of nonclassical logic

Volker Halbach · **Carlo Nicolai**

Received: date / Accepted: date

Abstract Solutions to semantic paradoxes often involve restrictions of classical logic for semantic vocabulary. In the paper we investigate the costs of these restrictions in a model case. In particular, we fix two systems of truth capturing the same conception of truth: (a variant) of the system KF of [Feferman 1991] formulated in classical logic, and (a variant of) the system PKF of [Halbach & Horsten 2006], formulated in basic De Morgan logic. The classical system is known to be much stronger than the nonclassical one. We assess the reasons for this asymmetry by showing that the truth theoretic principles of PKF cannot be blamed: PKF with induction restricted to non-semantic vocabulary coincides in fact with what the restricted version of KF proves true.

Keywords Formal Theories of Truth · Semantic Paradoxes · Logical Pluralism

1 Comparing the incomparable

There have been many attempts to solve or block the semantic paradoxes by restricting or weakening classical logic when it is applied to semantic vocabulary and, in particular, the truth predicate. Our aim is to assess the costs of making such incisions to classical logic. Any solution has its disadvantages and thus has to be revisionist to some extent. We do not expect that there is one correct solution, but rather that we face a choice. We can ban truth as a primitive notion and then only admit notions

We would like to thank Leon Horsten and an anonymous referee. The work of the second author was supported by the European Commission, Grant 658285 FOREMOTIONS.

Volker Halbach
New College, University of Oxford

Carlo Nicolai
Munich Center for Mathematical Philosophy, LMU Munich

of truth that can be defined from the chosen primitives. However if we opt for truth as a primitive notion, there are many options: truth can be treated as a predicate in a classical framework and many different ways of doing so are known. If classical logic is abandoned or restricted, possibilities abound; there is a plethora of logics, and for each logic there can be many ways to deal with the paradoxes. Hence it will be very hard to compare classical with nonclassical conceptions of truth. However, we can still try to compare some popular approaches and investigate their properties.

In the present paper we provide a case study: We consider a popular – perhaps the most popular – solution of the paradoxes: Kripke's [Kripke 1975] conception of truth can be axiomatized in a classical or a nonclassical framework. For the same conception of truth we specify two formal systems: The system KF in [Feferman 1991] is couched in classical logic, while PKF in [Halbach & Horsten 2006] is formulated in a nonclassical logic.[1]

With the main results of this paper we intend to show that, in a sense to be specified, both systems capture the same solution of the paradoxes. Of course, both systems differ in their theorems: One will feature all classical tautologies as theorems, while the other does not even prove all sentences of the form $\varphi \rightarrow \varphi$. Although it may be reasonably held that KF and PKF capture the same conception of truth, they differ in their deductive strength not only with respect to their truth-theoretic content, but also with respect to their truth-free consequences, that is, their arithmetical consequences, as [Halbach & Horsten 2006] showed.

The deductive weakness of PKF arises from the mutilation of classical logic, which invalidates certain patterns of mathematical reasoning that cannot be regained in PKF in any way. To justify this diagnosis of the source of deductive weakness, we show that it doesn't arise from other sources. The deductive weakness of PKF cannot be blamed on its logical rules and axioms. The underlying logical system of PKF is in fact complete with respect to the intended logic DM – basic De Morgan logic: a proof of this claim is contained in Appendix 3. But we also show that Peano arithmetic and the truth-theoretic axioms and rules formulated in the respective logics determine exactly the same models. Moreover, they both prove the same sentences to be true; that is, they prove the same theorems of the form $T\ulcorner\varphi\urcorner$. Only when we expand the mathematical schema of induction to the language with the truth predicate, the deductive weakness of PKF emerges.

These results enable us to assess the real costs of restricting classical logic, at least when axiomatizing this particular conception of truth: Surprisingly, the use of the nonclassical logic of PKF does not affect the truth-theoretic content, as will be shown in Theorem 1. However, the incisions to classical logic, when applied to sentences with the truth predicate, severely impede schematic reasoning with the truth predicate. More specifically, classical patterns of mathematical reasoning are no longer licensed in PKF. In KF the classical patterns can be used to prove new mathematical results that do not contain the truth predicate; in PKF the possibilities are restricted

[1] [Feferman 1991] doesn't use the name KF for the system. See [Halbach 2014, chapter 15] for details.

and thus PKF has far fewer theorems that are formulated in the purely arithmetical language without the truth predicate.

One might object that expanding the induction rule to the language with the truth predicate is in fact a strengthening by truth-theoretic axioms and that induction should not be considered to be a purely mathematical principle. However, we believe that expanding the induction rule to the language with the truth predicate doesn't involve any truth-theoretic justification. Mathematicians apply induction whatever the language might be. When we go from number theory to analysis, logic, or algebra, we don't ask for a new justification of induction. Thus, in sum, the restriction of classical logic imposes restrictions on our mathematical reasoning. It puts semantics apart from all other disciplines: Nobody would hesitate to apply mathematical tools including induction in any field of science. If we adopted a PKF-like theory of semantics, we couldn't apply these tools to semantics.

Our arguments will not convince those who think that , generally, classical logic is unsound and PKF gives the correct account of truth. Generally, arguments like the one expounded here will not impress the staunch proponent of some nonclassical solution of the semantic paradoxes, who believes that his or her account is the only correct account. On such a view, the general principle of induction is just incorrect and thus has to be rejected. The results on KF and PKF, however, should give a pause to those who largely accept classical logic, but think that the restriction of classical logic only for the semantic vocabulary is merely a tiny, negligible incision that comes at a low price well worth paying. As we shall show, the incision doesn't hit our truth-theoretic reasoning; it hits mathematical reasoning at its heart.

2 Conditions for truth

We start by describing the conception of truth that is to be captured by KF and PKF for the language $\mathscr{L}$ of arithmetic that has only sentences of the form $s = t$ as atomic sentences for closed terms s and t. We assume that $\mathscr{L}$ contains, besides the function symbols $\overline{0}, S, +, \times$, finitely many additional function symbols. For instance, $\dot{\neg}$ is a function symbol expressing the syntactic operation of negation. It is also convenient to have in our language a function symbol for the operation of replacing a term z for a free variable v in a formula φ. Occasionally, we will also employ functional notation without actually having corresponding function symbols in the language. This is the case of the arithmetical evaluation function $(\cdot)^\circ$ – definable in Peano arithmetic – that, when applied to a closed term, yields its value. We will be mainly interested in the language $\mathscr{L}_T := \mathscr{L} \cup \{T\}$ obtained by expanding the object language $\mathscr{L}$ with a new unary predicate T for truth. We assume a canonical Gödel numbering for $\mathscr{L}_T$-expressions: For each $\mathscr{L}_T$-expression e there is in the language a closed term $\ulcorner e \urcorner$ – the numeral of its code $\#e$ – such that $\ulcorner e \urcorner \neq \ulcorner f \urcorner$ is provable in Peano arithmetic for different expressions e, f. We will often not distinguish between $\mathscr{L}_T$ expressions and their codes. Finally, we stipulate that falsity is the truth of the negation; that is, saying that a sentence is false is saying that its negation is true.

For the different kinds of sentences in the language $\mathscr{L}$ the following truth conditions are obviously correct: An atomic sentence $s = t$ is true iff the terms s and t coincide in their values; $s = t$ is false (i.e. $\neg s = t$ is true) iff they differ. A conjunction is true iff both conjuncts are true; a conjunction is false iff at least one conjunct is false. A universally quantified sentence $\forall x \varphi(x)$ is true iff all its instances $\varphi(t)$ are true for all closed terms t; it's false iff $\varphi(t)$ is false for at least one closed term t. A negated sentence $\neg\varphi$ is true iff φ is false; and $\neg\varphi$ is false iff φ is true.

Once we move to $\mathscr{L}_T$, the following conditions seem in line with the ones for the connectives and quantifiers. A sentence $T\ulcorner\varphi\urcorner$ is true iff φ is true; a sentence $T\ulcorner\varphi\urcorner$ is false iff φ is false.[2] These conditions are adjoined to the ones above, but with *sentence* now understood as sentence of the expanded language with T.

The clauses can be taken as definition of a predicate *is true* in the metalanguage. It is not hard to bring the clauses into the form of a positive inductive definition in the sense of [Moschovakis 1974]. There are various sets satisfying the definitional clauses. If we also require that no sentence is true and false, then the conditions define the fixed points of Kripke's theory of truth with the Strong Kleene evaluation schema with some minor deviations. Here we don't rule out the possibility that for some sentence both the sentence and its negation are true.

If λ is a liar sentence and *is true* satisfies the conditions above, then λ is either both true and false (that is, λ and $\neg\lambda$ are true), or λ is neither true nor false. Hence the extension of *is true* cannot be closed under classical logic. The conception of truth just sketched is therefore nonclassical.

After this informal sketch, we consider formal systems that capture this conception of truth. The systems will be formulated in the sequent calculus. For the language $\mathscr{L}$ without the truth predicate we have the sequent calculus for full classical logic (cf. Appendix 1). All systems will contain Peano arithmetic, which is given by the usual 'recursive' equations for all function symbols and the induction rule:

$$\text{(IND)} \qquad \frac{\Gamma, \varphi(x) \Rightarrow \varphi(Sx), \Delta}{\Gamma, \varphi(\overline{0}) \Rightarrow \varphi(t), \Delta}$$

for all formulae $\varphi(x)$ of $\mathscr{L}$ and finite sets of formulae Γ and Δ.[3] The systems KF and PKF both extend Peano arithmetic. Axioms and rules differ between KF and PKF only in the extended language $\mathscr{L}_T$.

In KF and PKF the truth predicate is taken to be primitive. The clauses for truth sketched above are no longer taken as a definition of *is true* on the background of set theory that allows us to prove the existence of a set of sentences satisfying the conditions. In KF and PKF this primitive truth predicate is identified with the truth predicate T of $\mathscr{L}_T$.

[2] These conditions are somewhat simplified. Besides numerals $\ulcorner\varphi\urcorner$ for (codes of) sentences there will also be other closed terms with the code of φ as their value. So the clause should be: A sentence Tt is true iff the value of t is (the code of) a true sentence. Or one closes also under substitution of identicals. For the sake of the introduction here, we decided not to clutter the truth and falsity conditions with these subtleties.

[3] In (IND), x is not free in $\Gamma, \Delta, \varphi(\overline{0})$ and t is arbitrary.

If we turn the clauses in the above definition into axioms and extend all logical schemata of classical logic and the induction rule to the new language $\mathscr{L}_T$ with the truth predicate, then a variant of Feferman's system KF from [Feferman 1991] is obtained (see Appendix 2 for its axioms). KF is a classical system that describes a nonclassical conception of truth.

The system KF is a formalization of the definition above with the clauses of the definition turned into axioms for a primitive notion of truth. The use of classical logic in the axiomatization of truth in KF reflects the classicality of our metatheory. This, however, creates a mismatch between the theory we are using and the concept of type-free truth that is axiomatized by the theory.

If we opt for type-free concepts of truth and falsity, which are no longer classical, we should perhaps also be prepared to go all the way and adjust the logic we are using to reason about these concepts. Various authors have thus opted to axiomatize truth and falsity in a nonclassical logic. Among others, [Horsten 2009], [Horsten 2011], and [Kremer 1988] have advocated formal systems based on the Strong Kleene logic; [Field 2008] defends a similar system, but with an additional conditional.

These systems and the systems akin to the Kripke-Feferman system KF axiomatize the same concept of truth in the light of various criteria (see §4.1), and there is a deep disagreement for which direction one should decide. In section 4 we sketch some arguments that have been put forward for and against both options. Before this, however, we introduce PKF and its logic.

3 Partial Kripke-Feferman

In the present section we formally define the nonclassical approach to the conception of truth and falsity sketched above. We define the system PKF, standing for partial Kripke-Feferman.[4] In the recent literature several, slightly different versions of PKF have been employed under the same label. In presenting our version of PKF we will also highlight some of the differences occurring between these different versions and establish some basic but crucial facts concerning their models. In particular, these observations aim at fixing sufficient formal conditions for the satisfiability of sequents across models of PKF and its variants (see Propostions 1 and 3). They will be essential for the proof of our main results in §4 and §5. The reader who is acquainted with the basics of PKF and variants thereof may skip this section and move directly to §4.

In discussions of KF, PKF and related systems it is often claimed that a falsity predicate is not required and the truth predicate is sufficient because the falsity of a sentence can be defined as the truth of its negation. In KF-like systems with a falsity predicate we can drop the falsity predicate and define Fx as $T\dot{\neg}x$. However, usually this is not what it is meant, because the truth predicate and, consequently, the defined falsity predicate will still be truth and falsity predicates for the language *with* the primitive falsity predicate. What is usually intended when one defines falsity is the

[4] To keep the notation in line with [Halbach & Horsten 2006] and [Halbach 2014], we employ the label PKF although the logic DM that we employ is not partial but four-valued.

substitution of falsity with the truth of the negation also within codes sentences. A suitable translation function can be defined using the recursion theorem as described in [Halbach 2014, p. 37, Lemma 5.2]. In what remains, we stick with $\mathscr{L}_T$ that features a truth predicate only.

A four-valued model is a triple $(\mathscr{M}, S_1, S_2)$ where $\mathscr{M}$ is a model of the language $\mathscr{L}$ of arithmetic and S_1 and S_2 are subsets of the domain M of $\mathscr{M}$. S_1 is called *the extension of* T and S_2 *the antiextension*. We assume $\mathscr{M}$ to be a model of $\mathscr{L}$. We enrich $\mathscr{L}_T$ with constants $\overline{a}$ for all elements a of the domain of $\mathscr{M}$. This gives rise to the language $\mathscr{L}_T^+$. These constants do not belong to the language $\mathscr{L}$ or $\mathscr{L}_T$ and only expressions of $\mathscr{L}_T$ are coded, not expressions of $\mathscr{L}_T^+$ that are not in $\mathscr{L}_T$. We inductively define the relation $\vDash_{\mathsf{DM}}$ holding between models $(\mathscr{M}, S_1, S_2)$ and sentences of $\mathscr{L}_T^+$. For subsets S_1, S_2 of M, variables x, closed terms s and t of $\mathscr{L}_T^+$, $\mathscr{L}_T^+$-sentences φ, ψ, and $\mathscr{L}_T^+$-formulas ξ with exactly one free variable, we set:

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} s = t$ iff $s^{\mathscr{M}} = t^{\mathscr{M}}$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} s \neq t$ iff $s^{\mathscr{M}} \neq t^{\mathscr{M}}$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} Tt$ iff $t^{\mathscr{M}} \in S_1$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg Tt$ iff $t^{\mathscr{M}} \in S_2$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\neg\varphi$ iff $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi \wedge \psi$ iff $\big((\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi$ and $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \psi\big)$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg(\varphi \wedge \psi)$ iff $\big((\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\varphi$ or $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\psi\big)$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi \vee \psi$ iff $\big((\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi$ or $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \psi\big)$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg(\varphi \vee \psi)$ iff $\big((\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\varphi$ and $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\psi\big)$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \forall x\xi$ iff for all $a \in M$, $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \xi(\overline{a})$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg(\forall x\xi)$ iff for at least one $a \in M$, $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\xi(\overline{a})$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \exists x\xi$ iff for at least one $a \in M$, $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \xi(\overline{a})$

$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\exists x\xi$ iff for all $a \in M$, $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\xi(\overline{a})$

For any closed term s, $s^{\mathscr{M}}$ is the value of the term s in the model $\mathscr{M}$. If $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi$, we say that φ is true in $(\mathscr{M}, S_1, S_2)$; if $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\varphi$, we say that φ is false in $(\mathscr{M}, S_1, S_2)$.

We extend the definition of $\vDash_{\mathsf{DM}}$ to sequents $\Gamma \Rightarrow \Delta$, where Γ and Δ are finite sets of sentences of $\mathscr{L}_T^+$. $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \Gamma \Rightarrow \Delta$ if and only if the following two conditions are satisfied:

(a) If $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \gamma$ for all $\gamma \in \Gamma$, then there is at least one $\delta \in \Delta$ such that $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \delta$.

(b) If $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\delta$ for all $\delta \in \Delta$, then there is at least one $\gamma \in \Gamma$ such that $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\gamma$.

We write $\Gamma \models_{\mathsf{DM}} \Delta$ to express that (a) and (b) are satisfied by all four-valued models for $\mathscr{L}_T^+$. If DM is extended with additional initial sequents of $\mathscr{L}_T^+$-sentences and rules $\mathscr{A}$, we write $\Gamma \models_{\mathscr{A}} \Delta$ to express that (a) and (b) are satisfied by all four-valued models of $\mathscr{A}$.

A model $(\mathscr{M}, S_1, S_2)$ is *consistent* iff $S_1 \cap S_2 = \emptyset$; it is *complete* iff $M = S_1 \cup S_2$, where M is the domain of $\mathscr{M}$. In a consistent model there are no truth-value gluts; in a complete model there are no gaps in the following sense:

Lemma 1

(i) Assume $(\mathscr{M}, S_1, S_2)$ is consistent. Then there is no sentence such that $(\mathscr{M}, S_1, S_2) \models_{\mathsf{DM}} \varphi$ and $(\mathscr{M}, S_1, S_2) \models_{\mathsf{DM}} \neg\varphi$.

(ii) Assume $(\mathscr{M}, S_1, S_2)$ is complete. Then either $(\mathscr{M}, S_1, S_2) \models_{\mathsf{DM}} \varphi$ or $(\mathscr{M}, S_1, S_2) \models_{\mathsf{DM}} \neg\varphi$ holds for any given sentence φ.

Our reasoning with truth is now governed by the logic DM. This logic admits gluts and gaps, so it is a generalization of the Strong Kleene logic as conceived, for instance, in [Kripke 1975]. [Field 2008] calls our logic *basic De Morgan logic*. Those authors privilege a logic that results from adding a version of *ex-falso quodlibet* to basic De Morgan logic DM, ruling out complete models of $\mathscr{L}_T$. DM, by contrast, not only lacks sufficient resources to force only consistent or complete models, but it cannot even rule out or force models with simultaneous occurrences of truth-value gaps and gluts. We call these models *mixed models*. One way to generate mixed models is, for instance, by considering the fixed point of the operator Φ defined on page 9 applied to the pair $(\{\tau, \neg\tau\}, \varnothing)$ – where τ is a truth-teller sentence stating its own truth. The next proposition fixes some of these properties of DM.

Proposition 1

(i) For any mixed model $(\mathscr{M}, S_1, S_2)$ there are consistent and complete models satisfying all $\mathscr{L}_T^+$-sequents of the form $\varphi \Rightarrow \psi$ satisfied by $(\mathscr{M}, S_1, S_2)$.

(ii) For any consistent (complete) model $(\mathscr{M}, S_1, S_2)$ there is a complete (consistent) model $(\mathscr{M}, S_1', S_2')$ in which the same $\mathscr{L}_T^+$-sequents as in $(\mathscr{M}, S_1, S_2)$ hold.

As the reader can easily check, Proposition 1(i) cannot be generalized to arbitrary sequents $\Gamma \Rightarrow \Delta$, at least if one follows the strategy adopted below. The sequents $0 = 0 \Rightarrow \lambda, \tau$ and $\tau, \lambda \Rightarrow 0 \neq 0$, for instance, where λ and τ are a gap and a glut respectively, will hold in $(\mathscr{M}, S_1, S_2)$ but not in the corresponding consistent (complete) model we are about to construct.

Proof We first prove (i). Given a mixed model $(\mathscr{M}, S_1, S_2)$, we construct a consistent model $(\mathscr{M}, S_1', S_2')$ satisfying all sequents of the form $\varphi \Rightarrow \psi$ that hold in $(\mathscr{M}, S_1, S_2)$. Complete models can be achieved in a similar way. Essentially, we turn all gluts in the sense of $(\mathscr{M}, S_1, S_2)$ into gaps.
Let

$$S_1' := S_1 \setminus \{a \in M \mid a \in S_1 \cap S_2\}$$
$$S_2' := S_2 \setminus \{a \in M \mid a \in S_1 \cap S_2\}$$

$(\mathscr{M},S_1',S_2')$ is clearly consistent. We prove

(1) $\quad$ if $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}\gamma\Rightarrow\delta$, then $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\gamma\Rightarrow\delta$

Let us assume, seeking a contradiction, that $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\gamma\Rightarrow\delta$. Therefore one of the following must obtain:

(2) $\quad$ $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\gamma$ and $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\delta$

(3) $\quad$ $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\neg\delta$ and $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\neg\gamma$

If (2), then $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}\gamma$, by the monotonicity of our logic. By assumption, $\gamma\Rightarrow\delta$ holds in $(\mathscr{M},S_1,S_2)$, therefore the latter will also satisfy δ. By (2), δ must be a glut. Therefore by definition of $\models_{\mathsf{DM}}$ and assumption, $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}\neg\gamma$. However, by Lemma 1 one can prove by induction on the complexity of the standard $\mathscr{L}_T^+$-sentence γ, that

(4) $\quad$ if $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\gamma$ then $(\mathscr{M},S_1,S_2)\not\models_{\mathsf{DM}}\neg\gamma$

(4) gives us the desired contradiction.
The reasoning in the case in which (3) holds is analogous.

Next we move to the proof of (ii). Let a consistent model $(\mathscr{M},S_1,S_2)$ with domain M be given. Now define

$$S_1' := S_1\cup(M\setminus(S_1\cup S_2)) \text{ and } S_2' := S_2\cup(M\setminus(S_1\cup S_2)).$$

That is, all gaps are changed into gluts. Clearly, $(\mathscr{M},S_1',S_2')$ is complete. Again using Lemma 1 we have the equivalence of the following claims (5) and (6) for all $\mathscr{L}_T^+$-sentences φ:

(5) $\quad$ $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\varphi$ and $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\neg\varphi$

(6) $\quad$ $(\mathscr{M},S_1,S_2)\not\models_{\mathsf{DM}}\varphi$ and $(\mathscr{M},S_1,S_2)\not\models_{\mathsf{DM}}\neg\varphi$

This fact implies the following:

(7) $\quad$ If $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\varphi$, then $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}\neg\varphi$

(8) $\quad$ If $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}\neg\varphi$ then $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\varphi$

We prove:

(9) $\quad$ $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\Gamma\Rightarrow\Delta$ if and only if $(\mathscr{M},S_1,S_2)\not\models_{\mathsf{DM}}\Gamma\Rightarrow\Delta$

Let us assume $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\Gamma\Rightarrow\Delta$ but $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}\Gamma\Rightarrow\Delta$. Then one of the two following possibilities must obtain:

(10) $\quad$ For all $\gamma\in\Gamma$, $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\gamma$ and for all $\delta\in\Delta$, $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\delta$.

(11) $\quad$ For all $\delta\in\Delta$, $(\mathscr{M},S_1',S_2')\models_{\mathsf{DM}}\neg\delta$ and for all $\gamma\in\Gamma$, $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\neg\gamma$

In case (10), from the assumption that, for all $\delta\in\Delta$, $(\mathscr{M},S_1',S_2')\not\models_{\mathsf{DM}}\delta$ obtains, we conclude $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}\neg\delta$ for all $\delta\in\Delta$ by (7). Since by assumption $(\mathscr{M},S_1,S_2)\models_{\mathsf{DM}}$

$\Gamma \Rightarrow \Delta$, we conclude that there is a $\gamma \in \Gamma$ such that $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\gamma$ and therefore also $(\mathscr{M}, S_1', S_2') \nvDash_{\mathsf{DM}} \gamma$ by (8).
If (11), for all $\gamma \in \Gamma$, we have $(\mathscr{M}, S_1', S_2') \nvDash_{\mathsf{DM}} \neg\gamma$. Therefore, by (7), $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \gamma$ for each such γ. By the assumption $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \Gamma \Rightarrow \Delta$, there must be a $\delta_0 \in \Delta$ such that $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \delta_0$. By monotonicity, $(\mathscr{M}, S_1', S_2') \vDash_{\mathsf{DM}} \delta_0$. Thus $(\mathscr{M}, S_1', S_2') \vDash_{\mathsf{DM}} \delta_0$ and $(\mathscr{M}, S_1', S_2') \vDash_{\mathsf{DM}} \neg\delta_0$; finally, by the equivalence of (5) and (6), also $(\mathscr{M}, S_1, S_2) \nvDash_{\mathsf{DM}} \delta_0$.

The proof of the converse direction of (9) is left to the reader.

Up to this point, there are no restrictions on the extension S_1 and antiextension S_2 of the predicate T. A well-known method to obtain suitable pairs (S_1, S_2) is via (a version of) the construction in [Kripke 1975]. For simplicity, we describe this method with reference to the standard model of arithmetic; this enables us to focus on $\mathscr{L}_T$ and dispense with the additional constants of the enriched language $\mathscr{L}_T^+$. For an extension of this method to arbitrary models, the reader may consult [Cantini 1989, p. 108]. In the following, $\mathrm{Sent}_{\mathscr{L}_T}$, $\mathrm{NSent}_{\mathscr{L}_T}$, $\mathrm{Cterm}_{\mathscr{L}_T}$ will stand for the set of (codes of) sentences of $\mathscr{L}_T$, its complement, and the set of (codes of) closed terms of $\mathscr{L}_T$. The $\mathscr{L}$-formulas $\mathrm{Sent}_{\mathscr{L}_T}(x), \mathrm{NSent}_{\mathscr{L}_T}(x), \mathrm{Cterm}_{\mathscr{L}_T}(x)$ will strongly represent these sets in PA.

Let $S_1, S_2 \subseteq \omega$. The Kripke-jump Φ on (S_1, S_2) is defined as

$$\begin{aligned}\Phi(S_1, S_2) := \langle &\{\varphi \in \mathrm{Sent}_{\mathscr{L}_T} \mid (\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi\}, \\ &\mathrm{NSent}_{\mathscr{L}_T} \cup \{\varphi \in \mathrm{Sent}_{\mathscr{L}_T} \mid (\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\varphi\}\rangle\end{aligned}$$

Fixed-points of Φ are pairs (S_1, S_2) such that $\Phi(S_1, S_2) = (S_1, S_2)$. The model $(\mathbb{N}, S_1, S_2)$ is a fixed-point model iff (S_1, S_2) is a fixed point of Φ. Fixed points of Φ have been considered as qualified candidates for the interpretation of the truth predicate of $\mathscr{L}_T$. In particular, if (S_1, S_2) is a fixed point of Φ, we have, for all $\mathscr{L}_T$-sentences φ:

$$(\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} T^{\ulcorner}\varphi^{\urcorner} \text{ iff } (\mathbb{N}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi. \tag{12}$$

This property renders the truth predicate *transparent*: $T^{\ulcorner}\varphi^{\urcorner}$ is intersubstitutable with φ in any context, modulo satisfiability in a four-valued model.

In fixed points of Φ, S_2 can be defined from S_1 in the following way:

$$S_2 := \mathrm{NSent}_{\mathscr{L}_T} \cup \{\varphi \in \mathrm{Sent}_{\mathscr{L}_T} : \neg\varphi \in S_1\}$$

Hence, when dealing with models $(\mathbb{N}, S_1, S_2)$ where (S_1, S_2) is a fixed point of Φ, we can simply drop S_2 and recover the conception of truth from the beginning of the paper.

$$n \in \Phi(S_1) \Leftrightarrow \begin{cases} n \text{ is } s = t \text{ and } s^{\mathbb{N}} = t^{\mathbb{N}} \\ n \text{ is } \neg s = t \text{ and } s^{\mathbb{N}} \neq t^{\mathbb{N}} \\ n \text{ is } Tt \text{ and } t^{\mathbb{N}} \in S_1 \\ n \text{ is } \neg Tt \text{ and (the negation of } t^{\mathbb{N}} \text{ is in } S_1 \cap \mathrm{Sent}_{\mathscr{L}_T} \text{ or } t^{\mathbb{N}} \notin \mathrm{Sent}_{\mathscr{L}_T}) \\ n \text{ is } \neg\neg\varphi \text{ and } \varphi \in S_1 \\ n \text{ is } \varphi \wedge \psi \text{ and } (\varphi \in S_1 \text{ and } \psi \in S_1) \\ n \text{ is } \neg(\varphi \wedge \psi) \text{ and } (\neg\varphi \in S_1 \text{ or } \neg\psi \in S_1) \\ n \text{ is } \varphi \vee \psi \text{ and } (\varphi \in S_1 \text{ or } \psi \in S_1) \\ n \text{ is } \neg(\varphi \vee \psi) \text{ and } (\neg\varphi \in S_1 \text{ and } \neg\psi \in S_1) \\ n \text{ is } \forall x\chi \text{ and for all closed terms } t, \chi(t) \in S_1 \\ n \text{ is } \neg\forall x\chi \text{ and there is a closed term } t \text{ such that } \neg\chi(t) \in S_1 \\ n \text{ is } \exists x\chi \text{ and there is a closed term } t \text{ such that } \chi(t \in S_1 \\ n \text{ is } \neg\exists x\chi \text{ and for all closed terms } t, \neg\chi(t) \in S_1 \end{cases}$$

What we called the transparency property of the truth predicate, encompassed in (12), is doomed to failure if one replaces the DM satisfaction relation with the classical satisfaction relation. Classical models are obtained by *closing-off* a model $(\mathscr{M}, S_1, S_2)$, that is, by putting all sentences that are not in $S_1 \cup S_2$ into the anti-extension of the truth predicate. If we work in a model of KF, for instance, the Liar sentence $\lambda :\Leftrightarrow \neg T^{\ulcorner}\lambda^{\urcorner}$ is either in both the extension and the anti-extension of a fixed-point model or it is in neither of them; the closing-off of $(\mathscr{M}, S_1, S_2)$ will therefore decide λ in one way or the other. Either a closed-off model accepts λ but declares it untrue, or it refutes λ but deems false its negation.

The presence of a transparent truth predicate is one of the main motivations behind the formulation of nonclassical deductive systems validated by Kripke's construction. As anticipated, here we mainly focus on the system PKF. It has three main components: logic, arithmetic, and truth. The underlying logical system of PKF, that we will still call DM, is a variant of the systems described in [Scott 1975] and [Blamey 2002]. It is formulated in a sequent calculus where sequents are expressions of the form $\Gamma \Rightarrow \Delta$ with Γ, Δ finite sets of sentences. The initial sequents and rules of DM are displayed in Appendix 1. DM is sound with respect to the intended notion of logical consequence: If all sentences are true in a four-valued model, then at least one sentence in Δ is true in the model; if all sentences in Δ are false in a four-valued model, then at least one sentence in Γ is false in that model. DM is also complete with respect to the intended semantics. A Henkin-style argument for its completeness can be found in Appendix 3.

As we have mentioned in Section 2, the theory that generates truths for the T-free fragment of $\mathscr{L}_T$ is Peano arithmetic. PKF will then contain initial sequents $\Rightarrow \varphi$ where φ is a basic axiom of Peano arithmetic and the rule of induction (IND). In classical logic the schema of induction is easily derived from the rule. The induction rule of PKF is obtained from this rule by admitting any formula φ of $\mathscr{L}_T$ as an

instance. In the nonclassical logic DM the induction schema is no longer derivable from the rule (IND). In fact, the schema can be shown to be unsound.

The genuinely truth-theoretic initial sequents of PKF are the sequent versions of the conditions for truth described in Section 2:

PKF1 1. $s^\circ = t^\circ \Rightarrow T(s \doteq t)$
2. $T(s \doteq t) \Rightarrow s^\circ = t^\circ$

PKF2 1. $\mathrm{Sent}_{\mathscr{L}_T}(x \mathbin{\dot\wedge} y), T(x \mathbin{\dot\wedge} y) \Rightarrow Tx \wedge Ty$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x \mathbin{\dot\wedge} y), Tx \wedge Ty \Rightarrow T(x \mathbin{\dot\wedge} y)$

PKF3 1. $\mathrm{Sent}_{\mathscr{L}_T}(x \mathbin{\dot\vee} y), T(x \mathbin{\dot\vee} y) \Rightarrow Tx \vee Ty$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x \mathbin{\dot\vee} y), Tx \vee Ty \Rightarrow T(x \mathbin{\dot\vee} y)$

PKF4 1. $\mathrm{Sent}_{\mathscr{L}_T}(\dot\forall vx), \forall t\,(Tx(t/v)) \Rightarrow T(\dot\forall vx)$
2. $\mathrm{Sent}_{\mathscr{L}_T}(\dot\forall vx), T(\dot\forall vx) \Rightarrow \forall t\,(Tx(t/v))$

PKF5 1. $\mathrm{Sent}_{\mathscr{L}_T}(\dot\exists vx), \exists t\,(Tx(t/v)) \Rightarrow T(\dot\exists vx)$
2. $\mathrm{Sent}_{\mathscr{L}_T}(\dot\exists vx), T(\dot\exists vx) \Rightarrow \exists t\,(Tx(t/v))$

PKF6 1. $Tt^\circ \Rightarrow TTt$
2. $TTt \Rightarrow Tt^\circ$

PKF7 1. $\mathrm{Sent}_{\mathscr{L}_T}(x), \neg Tx \Rightarrow T\dot\neg x$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x), T\dot\neg x \Rightarrow \neg Tx$

PKF8 $Tx \Rightarrow \mathrm{Sent}_{\mathscr{L}_T}(x)$

The notation follows [Halbach 2014]. We say that PKF proves φ iff PKF proves the sequent $\Rightarrow \varphi$. The system PKF$\restriction$ is obtained from PKF by allowing only $\mathscr{L}$-formulas in instances of the induction rule (IND) and adding to it clauses for the subsitution of identicals into the scope of the truth predicate that are proved in PKF via (IND):

(Reg1) $\mathrm{Sent}_{\mathscr{L}_T}(x), s^\circ = t^\circ, Tx(t/v) \Rightarrow Tx(s/v)$

(Reg2) $\mathrm{Sent}_{\mathscr{L}_T}(x), s^\circ = t^\circ, Tx(s/v) \Rightarrow Tx(t/v)$

As [Halbach 2014, Lemma 16.9] observes, PKF is sound with respect to four-valued, fixed-point models based on the standard model $\mathbb{N}$ of arithmetic in the following sense.

Lemma 2 (soundness) *Assume Γ and Δ are set of $\mathscr{L}_T$-sentences. If $\Gamma \Rightarrow \Delta$ is derivable in* PKF*, then the two following conditions obtain for all fixed-point models $(\mathbb{N}, S_1, S_2)$:*

(i) If all sentences in Γ are true in $(\mathbb{N}, S_1, S_2)$, then at least one sentence in Δ is true in $(\mathbb{N}, S_1, S_2)$.

(ii) If all sentences in Δ are false (that is, their negations are true) in $(\mathbb{N}, S_1, S_2)$, then at least one sentence in Γ is false in $(\mathbb{N}, S_1, S_2)$.

That is, theorems of PKF are valid in those models; this can be established by a routine inductive argument. Furthermore, if (S_1, S_2) is a fixed point of Φ, then $(\mathbb{N}, S_1, S_2)$

is a model of PKF. These facts will be relevant later on when we will discuss in what sense PKF can be considered an axiomatization of Kripke's fixed-point construction.

Truth and falsity in a four-valued model are determined by the logic DM. The initial sequents and rules are also complete with respect to DM. From Proposition 4 in Appendix 3, we have:

Proposition 2 *Let Γ, Δ be finite sets of sentences of $\mathscr{L}_T$. Then we have:*

$$\textit{If } \Gamma \vDash_{\mathsf{PKF}} \Delta, \textit{ then } \mathsf{PKF} \vdash \Gamma \Rightarrow \Delta.$$

Proposition 2 obviously extends to the case of $\mathsf{PKF}\restriction$. Truth theoretic initial sequents and rules of PKF are also complete in another sense. We will make this precise in §4.2.

The semantic notion of transparency for truth predicates described by (12) above has an obvious syntactic counterpart. A system in $\mathscr{L}_T$ is transparent if it proves φ precisely if it proves $T^{\ulcorner}\varphi^{\urcorner}$ for all $\varphi \in \mathrm{Sent}_{\mathscr{L}_T}$. PKF and $\mathsf{PKF}\restriction$ are transparent systems. By induction on the complexity of the $\mathscr{L}_T$-sentence φ, one can prove the following:

Lemma 3 ([Halbach & Horsten 2006, Thm. 22]) *For all $\varphi \in \mathrm{Sent}_{\mathscr{L}_T}$ the following obtains:*

PKF (*resp.* $\mathsf{PKF}\restriction$) *proves* φ *if and only if* PKF (*resp.* $\mathsf{PKF}\restriction$) *proves* $T^{\ulcorner}\varphi^{\urcorner}$

Although PKF is a nonclassical theory, it behaves classically for sentences belonging to the ground language $\mathscr{L}$. In particular, it derives the sequent

$$\Rightarrow \varphi, \neg\varphi \tag{LEM}$$

for any $\mathscr{L}$-sentence φ. Once $\Rightarrow \varphi, \neg\varphi$ is derived, φ can be shown to behave completely classically. But this observation can be sharpened even further: The proof of (LEM) requires in fact a metatheoretic induction the complexity of the $\mathscr{L}$-sentence φ; The induction rule (IND) of PKF is not needed. Thus (LEM) is also provable in $\mathsf{PKF}\restriction$ for φ in $\mathscr{L}$.

Like KF, PKF proves the T-schema involving truth-free sentences, that is, the schema

$$\forall t \left(T^{\ulcorner}\varphi^{\urcorner}(t/^{\ulcorner}v^{\urcorner}) \leftrightarrow \varphi(t^{\circ})\right) \tag{utb}$$

where φ is a formula of the arithmetical language $\mathscr{L}$ with only v.[5] As before, the claim follows from a metatheoretic induction on the complexity of $\varphi(v)$; (utb) is therefore provable already in $\mathsf{PKF}\restriction$.

As we shall see in the next section, some authors advocate the view that, whereas classical logic is fine-tuned for mathematical or syntactic reasoning, it should be abandoned when from mathematics we move to different subject-matters such as semantics. Such view is defended for instance in [Field 2008]. Such authors have therefore welcome the fact that PKF behaves classically for the truth-free part of the language; if the power of classical logic can always be restored when dealing with

[5] We employ the lower-case label (utb) to distinguish it from the *theory* UTB from [Halbach 2014].

mathematical facts, there is no loss with respect to mathematical reasoning in adopting a nonclassical approach in semantics. We will argue against such view in the concluding section.

We conclude this section with extensions of Proposition 1 applied to models of $\mathsf{PKF}\restriction$. First we state an obvious corollary.

Corollary 1 *For any consistent model* $(\mathscr{M},S_1,S_2)$ *of* $\mathsf{PKF}\restriction$ (PKF) *there is a complete model of* $\mathsf{PKF}\restriction$ (PKF) *in which the same sequents as in* $(\mathscr{M},S_1,S_2)$ *hold. Similarly, for any complete model of* $\mathsf{PKF}\restriction$ (PKF) *there is a consistent model of* $\mathsf{PKF}\restriction$ (PKF) *that satisfies the same sequents as* $(\mathscr{M},S_1,S_2)$.

In proving Proposition 1 we also noticed that, in general, we cannot transform a mixed model of $\mathscr{L}_T$ into a consistent or a complete model that satisfies all sequents holding in the original model by changing gluts into gaps or gaps into gluts respectively. Satisfiability of PKF-provable sequents of the form $\varphi \Rightarrow \psi$, by contrast, is preserved by these transformations.

Proposition 3 *Let* $(\mathscr{M},S_1,S_2)$ *be a mixed model of* PKF (*resp.* $\mathsf{PKF}\restriction$). *Then there is a consistent model* $(\mathscr{M},S_1',S_2')$ *of* PKF (*resp.* $\mathsf{PKF}\restriction$) *satisfying all* $\mathscr{L}_T$*-sequents of the form* $\varphi \Rightarrow \psi$ *that hold in* $(\mathscr{M},S_1,S_2)$. *Similarly, there is a complete model* $(\mathscr{M},S_1^\star,S_2^\star)$ *of* PKF (*resp.* $\mathsf{PKF}\restriction$) *satisfying all* $\mathscr{L}_T$*-sequents of the form* $\varphi \Rightarrow \psi$ *that hold in* $(\mathscr{M},S_1,S_2)$.

Proof We only deal with the first part concerning consistent models. The second part of the claim is obtained in an analogous way.

The bulk of the proof is to show that we can transform a mixed model $(\mathscr{M},S_1,S_2)$ of PKF into a consistent model $(\mathscr{M},S_1',S_2') \vDash_{\mathsf{DM}} \mathsf{PKF}$. This will yield the claim for $\mathsf{PKF}\restriction$ and, by applying Proposition 1(i), also the satisfiability of all sequents of the form $\varphi \Rightarrow \psi$ that hold in the original model.

Again we turn gluts into gaps, that is we let

$$S_1' := S_1 \setminus \{a \in M \mid a \in S_1 \cap S_2\}$$
$$S_2' := S_2 \setminus \{a \in M \mid a \in S_1 \cap S_2\}$$

The resulting model $(\mathscr{M},S_1',S_2')$ is consistent. We show that it is a model of PKF by induction on the length of the derivation in PKF.

Logical initial sequents are obviously satisfied. For truth-theoretic initial sequents, we consider some instructive cases. Again it is useful to consider the language $\mathscr{L}_T^+$ featuring constants $\bar{a},\bar{b},\bar{c},\ldots$ for all elements of the domain M of $\mathscr{M}$, although in principle everything could be reformulated using variable assignments as expressions of the new language are not coded. Also, if f is a function symbol of $\mathscr{L}_T$ expressing a syntactic operation, we write $f^{\mathscr{M}}$ for the corresponding function in $\mathscr{M}$.

$\mathsf{PKF}3(1)$. Seeking a contradiction, we assume that one of the following holds:

(13) $(\mathscr{M},S_1',S_2') \vDash_{\mathsf{DM}} \mathrm{Sent}_{\mathscr{L}_T}(\bar{a}\dot{\vee}\bar{b}) \wedge T\bar{a} \vee T\bar{b}$ and $(\mathscr{M},S_1',S_2') \nvDash_{\mathsf{DM}} T(\bar{a}\dot{\vee}\bar{b})$,

(14) $(\mathscr{M},S_1',S_2') \vDash_{\mathsf{DM}} \neg T(\bar{a}\dot{\vee}\bar{b})$ and $(\mathscr{M},S_1',S_2') \nvDash_{\mathsf{DM}} \neg(T\bar{a} \vee T\bar{b}) \vee \neg\mathrm{Sent}_{\mathscr{L}_T}(\bar{a}\dot{\vee}\bar{b})$

If (13), either $a \in S_1'$ or $b \in S_1'$. Let us assume the former. Then also $a \in S_1$. By construction of $(\mathscr{M}, S_1', S_2')$, if $a \dot{\vee}^{\mathscr{M}} b \in S_1 \setminus S_1'$, also $a \dot{\vee}^{\mathscr{M}} b \in S_1 \cap S_2$. By assumption and the fact that $(\mathscr{M}, S_1, S_2)$ models PKF, $a \in S_1 \cap S_2$ and therefore $a \notin S_1' \cup S_2'$ by construction of $(\mathscr{M}, S_1', S_2')$. This, however, contradicts our assumption on a. The reasoning in the case in which $b \in S_1'$ is symmetric.
If (14), $a \dot{\vee}^{\mathscr{M}} b \in S_2$ and both a, b are in S_2 by our assumption on $(\mathscr{M}, S_1, S_2)$ – notice that we can safely assume that $a, b \in \mathrm{Sent}_{\mathscr{L}_T}^{\mathscr{M}}$. Also, $a \notin S_2'$ and $b \notin S_2'$. Thus $a, b \in S_1 \cap S_2$ and also $a \dot{\vee}^{\mathscr{M}} b \in S_1 \cap S_2$. But then $a \dot{\vee}^{\mathscr{M}} b \notin S_2'$, contradicting (14).

$\mathsf{PKF6}(1)$. We recall that the complete formulation of $\mathsf{PKF6}(1)$ is

$$\mathrm{Cterm}_{\mathscr{L}_T}(x^\circ), Tx^\circ \Rightarrow T\dot{T}x$$

If $\mathsf{PKF6}(1)$ fails to hold in $(\mathscr{M}, S_1', S_2')$, we have

(15) either $(\mathscr{M}, S_1', S_2') \vDash_{\mathsf{DM}} \mathrm{Cterm}_{\mathscr{L}_T}(\bar{a}^\circ) \wedge T\bar{a}^\circ$ and $(\mathscr{M}, S_1', S_2') \nvDash_{\mathsf{DM}} T\dot{T}\bar{a}$,

(16) or $(\mathscr{M}, S_1', S_2') \vDash_{\mathsf{DM}} \neg T\dot{T}\bar{a}$ and $(\mathscr{M}, S_1', S_2') \nvDash_{\mathsf{DM}} \neg T\bar{a}^\circ \vee \neg \mathrm{Cterm}_{\mathscr{L}_T}(\bar{a}^\circ)$

Let us assume (15). If $\dot{T}^{\mathscr{M}} a \in S_1 \setminus S_1'$, also $\dot{T}^{\mathscr{M}} a \in S_1 \cap S_2$. Therefore, by our assumptions on $(\mathscr{M}, S_1, S_2)$, $T\dot{\neg}\dot{T}\bar{a}$ will hold in it. We can safely assume that $\mathscr{M} \vDash_{\mathsf{DM}} \mathrm{Sent}_{\mathscr{L}_T}(\bar{a}^\circ)$: therefore also $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} T\dot{\neg}\bar{a}^\circ$. But then $a^{\circ^{\mathscr{M}}} \in S_1 \cap S_2$, thus $a^{\circ^{\mathscr{M}}} \notin S_1'$ against (15).
If (16), also $\mathscr{M} \vDash_{\mathsf{DM}} \mathrm{Cterm}_{\mathscr{L}_T}(\bar{a}^\circ)$. Since $a^{\circ^{\mathscr{M}}} \in S_1 \setminus S_1'$, also $a^{\circ^{\mathscr{M}}} \in S_1 \cap S_2$. Therefore $\dot{T}^{\mathscr{M}} a \in S_1 \cap S_2$ by (16) and the fact that $(\mathscr{M}, S_1, S_2)$ models PKF. Thus $\dot{T}^{\mathscr{M}} a \notin S_2'$, contradicting (16).

The cases of the other initial sequents are similar. For the induction step, logical rules are clearly preserved. The induction rule is also unproblematic since $(\mathscr{M}, S_1, S_2)$ satisfies it.

4 Truth-first versus logic-first

Many logicians have opted for solutions to the semantic paradoxes that rely on a restriction of classical logic to the nonsemantic vocabulary, or, to put it more favourably, a relaxation of classical logic for semantic vocabulary. Once classical logic is no longer fully applicable, as outlined in the previous section, it is possible to consider rules for truth that are fully transparent.

Classical logic and full transparency seem both desirable of a truth theory. The liar paradox shows that one cannot have both without triviality. Thus, it seems, that there is a decision to be made. Either one keeps classical logic across the board even for semantic notions and gives up transparency, *or* one restricts classical logic when semantic vocabulary is involved and keeps transparency. We refer to the former as the *logic-first* view and to the latter as the *truth-first* view.

What could be the reasons for a sacrificing classical logic in favour of transparent truth? Isn't classical logic at the centre of Quine's web of belief? Doesn't ripping out classical logic from the centre of this web affect almost all parts of the web? It

isn't so bad, or so proponents of the transparency option have argued. If we trade in classical logic for transparency, we don't have to abandon any part of the sciences or mathematics. Full classical logic can be retained for these theories. Truth is not a notion used in the sciences or mathematics; an unrestricted, general, fully transparent notion of truth is required only in some areas of philosophy where some cleaning up will be required anyway. So what needs to be done to preserve transparency is some clipping at the fringes of the web – at least according to some advocates of the truth-first view.

The proponents of the logic-first view, among them [Halbach 2014] and [Williamson 2016], emphasize the universality of logic. They see the restriction of classical logic in some area as an outright rejection. If some pattern of reasoning doesn't hold across the board, then it's no longer a logical principle. For instance, intuitionists do not accept the principle of excluded middle. The fact that they don't deny it and actually accept many instances of it, doesn't mean that intuitionists can be described as accepting classical logic with some restrictions. Rejecting some instances of the law of excluded middle means that classical logic is rejected. That PKF relies on classical logic in some of its parts doesn't mean that it's a classical theory. Its logic is DM, and in the parts that are not semantic classical logic applies in the same way classical logic applies to decidable sentences in intuitionistic logic.

The debate between the defenders of classical logic and the proponents of some nonclassical alternatives often tends to be ideological. The reasons put forward in favour and against nonclassical logics are usually highly theoretical and are often derived from deep and contentious methodological principles. When more concrete reasons are given, they commonly revolve around the paradoxes and their solutions; on them people are unlikely to agree.

It is also possible to compare classical and nonclassical theories using more concrete logical results that do not focus on the paradoxes. In the following we consider results of this kind.

4.1 Comparing the Classical and the Nonclassical

We seek to compare classical and nonclassical theories that share a common conception of truth. However, it is not straightforward to spell out what it means for a deductive system of truth to capture a semantic construction.

To compare theories formulated in different logics, one can consider their standard models: On this view an axiomatic theory captures a semantic construction exactly if the standard models of the theory belong to the class of models determined by the semantic theory. In more precise terms, we consider a categoricity criterion for the ω-models of the theory: a truth system W in the logic l captures a semantic construction $\mathfrak{S}$ if and only if, for all $S \subseteq \omega$,

(17) $(\mathbb{N}, S) \vDash_l W$ if and only if $(\mathbb{N}, S)$ belongs to $\mathfrak{S}$,

Both classical, KF-like theories and nonclassical, PKF-like theories are adequate – in the sense just described – with respect to Kripke's fixed-point construction based on the operator Φ as defined on page 9.[6]

Lemma 4 ([Feferman 1991, Halbach & Horsten 2006])

(i) $(\mathbb{N}, S) \vDash \mathsf{KF}$ *iff* *S is a fixed point of* Φ;[7]
(ii) $(\mathbb{N}, S) \vDash_{\mathsf{DM}} \mathsf{PKF}$ *iff* *S is a fixed point of* Φ.

From Lemma 4 one immediately obtains, for a fixed point $S \subseteq \omega$ of Φ,

$$(\mathbb{N}, S) \vDash \mathsf{KF} \text{ iff } (\mathbb{N}, S) \vDash_{\mathsf{DM}} \mathsf{PKF} \tag{18}$$

This suggests that the conception of truth described in the introduction is captured equally well by the nonclassical theory PKF and the classical theory KF. Our main result below will indicate that this observation can be in a sense generalized to nonstandard models for the theories with restricted induction. This further supports the view that KF and PKF actually do capture the same conception of truth.

Once it can be reasonably maintained that a nonclassical and a classical theory embody the same conception of truth, one has to find suitable means to compare the two theories qua formal systems. Some methods for comparing theories are not very promising for our purposes. For instance, when comparing classical theories, one can simply compare their theorems and investigate whether one theory is contained in the other. This simple method isn't useful for comparing theories in different logics, because a nonclassical truth theory (based on the truth-first view) will not contain some classical tautologies and thus never contain any classical theory. Similar remarks apply to more sophisticated methods such as relative interpretability and relative truth definability (cf. [Fujimoto 2010]); they are not suitable, at least in their standard definitions, for comparing theories formulated in different logics.

Although we cannot sensibly compare classical and nonclassical theories by determining whether one is contained in the other, we can still compare their internal theories; in particular, we can investigate whether the internal theory of one is contained in the other. In other words, we can ask whether all sentences that can be proved to be true in one theory can be proved to be true in the other theory.

We make this more formal. The internal theory IS of a system of truth S is defined in the following way:

$$\mathsf{IS} := \{\varphi : \mathsf{S} \vdash T^{\ulcorner}\varphi^{\urcorner}\}$$

Assume that we have a classical theory CL and a nonclassical theory NC. To compare the two, we can ask whether the internal logic of CL is contained in the inner logic

[6] [Fischer et alii 2015] study the criterion of ω-categoricity in depth and consider also other criteria such as similarity with a semantic construction and proof theoretic strength that are, in many respect, less refined than categoricity for ω-models. Having said that, it is also clear that ω-categoricity can at best be a necessary condition for an axiomatic theory to capture a semantic construction: by replacing the truth theoretic axioms of PKF with sequents of the form $T^{\ulcorner}\varphi^{\urcorner} \Rightarrow \varphi$ and $\varphi \Rightarrow T^{\ulcorner}\varphi^{\urcorner}$, in fact, one obtains again an ω-categorical axiomatization of Kripke's theory.

[7] This result extends to the cases in which S is a consistent or a complete fixed point and (CONS) or (COMP) are added to KF, as we shall see in §5.

of NC or whether the inverse inclusion holds. As mentioned above, we are especially interested in transparent nonclassical theories of truth, that is systems in which $T^{\ulcorner}\varphi^{\urcorner}$ is fully intersubstitutable with φ for all sentences φ. For such transparent systems, the inner logic and the outer logic, that is, the set of provable sentences, will coincide with the inner logic and, more formally expressed, the following will obtain for all sentences:

$$\varphi \in \mathsf{IS} \text{ if and only if } \mathsf{S} \vdash \varphi$$

Thus in transparent systems a sentence is provable if and only if it is provably true. Thus the comparison of the inner logics of CL and NC can be carried out by investigating whether all closed theorems of NC are in the inner logic ICL of the classical system CL (or the other way round).

Comparing what the theories prove true will also yield a comparison of their non-semantic consequences as in [Halbach & Horsten 2006]. This is a consequence of the provability, both in KF-like theories and in PKF-like theories, of the Tarskian biconditionals for arithmetical sentences. This analysis will yield information about the usefulness of a theory outside semantics. In fact looking at the consequences of a truth theory outside the more contentious field of truth theory may be a more promising way to compare truth theories than focusing on the way the theories handle the paradoxes.

Obviously, focusing on this sort of comparison leaves untouched many deep theoretical themes raised by the opposition between a classical and a nonclassical approach, such as the status of paradoxical sentences. However, this move enables us to discuss issues that are often set aside in the debate between the classical and the nonclassical logician, such as the status of the truth-free consequences of the systems.

4.2 A Completeness Theorem

We are now going to compare the internal theories of KF, PKF and some of their variants. [Halbach & Horsten 2006] showed that the internal theory IKF of KF is much stronger than PKF and that PKF is properly contained in IKF. IKF contains also more purely arithmetical sentences than PKF.

The system PKF is sound in the sense that, if a sentence is provable in PKF, then it is in the extension of the truth predicate in a four-valued, fixed-point model. The inductive proof of soundness can be formalized in a proper subsystem of KF (cf. also Lemma 5 below). This subsystem BT is obtained by replacing the induction rule of KF with the weaker rule of internal induction

$$\frac{\Gamma, T^{\ulcorner}\varphi(\dot{x})^{\urcorner} \Rightarrow T^{\ulcorner}\varphi(S\dot{x})^{\urcorner}, \Delta}{\Gamma, T^{\ulcorner}\varphi(0)^{\urcorner} \Rightarrow T^{\ulcorner}\varphi^{\urcorner}(t/v), \Delta}$$

for all $\mathscr{L}_T$-sentences $\varphi(v)$ with one free variable. Here the dot stands for the result of formally substituting the variable v with the numeral for x (or Sx).[8]

[8] [Cantini 1989] introduced the system BT, although his definition differs slightly from the one adopted here.

BT can prove iterations of the truth predicate up to the ordinal ω^ω, therefore we cannot go beyond provable truth iterations up to the ordinal ω^ω in PKF. This gives also an upper bound to the provable arithmetical sentence of PKF: If an $\mathscr{L}$-sentence φ is provable in PKF, then BT will prove $T^\ulcorner\varphi^\urcorner$; but BT proves the T-schema for arithmetical formulas, thus BT will also prove φ.

Furthermore, truth iterations up to ω^ω are precisely what we can prove in PKF. Not only in fact PKF behaves classically for formulas of $\mathscr{L}$; for sentences φ containing up to α self-applications of the truth predicate with $\alpha < \omega^\omega$, PKF proves

$$\Rightarrow T^\ulcorner\varphi^\urcorner \vee \neg T^\ulcorner\varphi^\urcorner;$$

in other words, such formulas are determinately true or false, provably in PKF. For such formulas, therefore, PKF can define a hierarchy of Tarskian truth predicates: PKF in fact proves the same arithmetical sentences as ramified truth up to ω^ω.[9] This difference in strength can be cashed out in terms of provable instances of transfinite induction: PKF proves exactly transfinite induction for $\mathscr{L}_T$ up to any ordinal smaller than ω^ω. This is in sharp contrast with KF, which proves transfinite induction for $\mathscr{L}_T$ up to any ordinal smaller than ε_0.

These results imply that IKF and PKF differ also in their purely arithmetical content. This difference can be illustrated in various ways by specifying purely arithmetical sentences that are provable in IKF but not in PKF. For instance, IKF contains the consistency statement and iterated reflection principles for PKF. Moreover, this asymmetry is also reflected by the amount of *arithmetical* transfinite induction available in PKF and IKF. The system PKF proves exactly all arithmetical instances of the transfinite induction schema for the language $\mathscr{L}$ up to any ordinal smaller than $\varphi_\omega(0)$, whereas KF and IKF prove arithmetical instances of transfinite induction up to any ordinal smaller than $\varphi_{\varepsilon_0}(0)$.[10]

Here we are interested in the source of the asymmetry between PKF and IKF. The two theories may for instance diverge in some truth principles that are present in KF and that are 'lost' in the transition from the classical to the nonclassical theory. In the following we exclude this possibility. To achieve this, we focus on the subsystems of PKF and KF obtained by extending the arithmetical base theory with purely truth theoretic principles, that is PKF$\restriction$ and KF$\restriction$. In the results of [Halbach & Horsten 2006], in fact, the induction rule extended with semantic vocabulary plays an important role.

Once the induction rules of KF and PKF are restricted to truth-free formulas, the two theories coincide in their theorems in the arithmetical language $\mathscr{L}$.[11] Moreover, by the classical behaviour of PKF on the truth-free fragment of PKF (cf. (LEM) on p. 12), the purely arithmetical reasoning available in the two theories coincide. By

[9] For a definition of the systems of ramified truth up to the Feferman-Schütte ordinal Γ_0, see [Halbach 2014, §9].

[10] For a definition of the Veblen functions, see again [Halbach 2014, §9].

[11] KF$\restriction$ is in fact conservative over PA. Any model of $\mathscr{M}$ of PA can be expanded to a model $(\mathscr{M}, S_1, S_2)$ of KF$\restriction$ where (S_1, S_2) is a fixed point of Φ. This also yields the conservativity of PKF over PA as if PKF$\restriction$ proved some formula φ of $\mathscr{L}$ not provable in PA, φ would also be provable in KF$\restriction$ by the monotonicity of the logic DM.

restricting induction we therefore rule out the possibility that our comparison of the truth theoretic content of KF and PKF may be vitiated by the interplay of truth and non-semantic (or only partially semantic) principles.

Also the source of the difference in strength between PKF and KF may reside in the absence, in PKF, of some essential logical or truth-theoretic principle. However, Lemma 2 tells us that PKF is complete with respect to logic, that is, the logical rules and axioms of PKF are complete with respect to basic De Morgan logic. We now show that the system is also complete, relative to IKF, with respect to the truth-theoretic axioms. In particular, we show that the truth theoretic initial sequents and rules of PKF are sufficient to capture the sentences that are provably true in KF. More precisely, we prove that $\mathsf{KF}\restriction \vdash T^{\ulcorner}\varphi^{\urcorner}$ implies $\mathsf{PKF}\restriction \vdash \varphi$. This will yield the equivalence of $\mathsf{PKF}\restriction$ and $\mathsf{IKF}\restriction$.[12]

We first recall the argument employed by Halbach & Horsten to establish that $\mathsf{PKF}\restriction$ is contained in the internal theory of $\mathsf{KF}\restriction$. An extension of this result to natural extensions of PKF and KF can be found in §5.

Lemma 5 *If* $\mathsf{PKF}\restriction$ *proves* φ*, then* $\varphi \in \mathsf{IKF}\restriction$.

Proof (Proof Sketch) The proof is by induction on the length of the derivation in $\mathsf{PKF}\restriction$. It suffices to prove, for $\mathsf{PKF}\restriction \vdash \Gamma \Rightarrow \Delta$, the claims

$$\mathsf{KF}\restriction \vdash \forall \mathbf{t}\left(T^{\ulcorner}\bigwedge\Gamma^{\urcorner}(\mathbf{t}/\mathbf{v}) \rightarrow T^{\ulcorner}\bigvee\Delta^{\urcorner}(\mathbf{t}/\mathbf{v})\right) \tag{19}$$

$$\mathsf{KF}\restriction \vdash \forall \mathbf{t}\left(T^{\ulcorner}\neg\bigvee\Delta^{\urcorner}(\mathbf{t}/\mathbf{v}) \rightarrow T^{\ulcorner}\neg\bigwedge\Gamma^{\urcorner}(\mathbf{t}/\mathbf{v})\right) \tag{20}$$

where $\bigwedge\varnothing :\leftrightarrow 0=0$ and $\bigvee\varnothing :\leftrightarrow 0\neq 0$, and $\bigwedge\Gamma(\bigvee\Gamma)$ is the conjunction (disjunction) of all sentences in Γ.

We now show that, given a model $(\mathscr{M},S_1,S_2)$ of $\mathsf{PKF}\restriction$, the closed-off model $(\mathscr{M},S_1)$ classically satifies $\mathsf{KF}\restriction$. We recall that for a model $\mathscr{M}$ to classically satisfy a sequent $\Gamma \Rightarrow \Delta$ it suffices to have

$$\text{if } \mathscr{M} \models \bigwedge\Gamma, \text{ then } \mathscr{M} \models \bigvee\Delta.$$

Lemma 6 *If* $(\mathscr{M},S_1,S_2) \vDash_{\mathsf{DM}} \mathsf{PKF}\restriction$*, then* $(\mathscr{M},S_1) \models \mathsf{KF}\restriction$.

Proof The proof is by induction on the length of the derivation in $\mathsf{KF}\restriction$. We check that the initial sequents of $\mathsf{KF}\restriction$ are satisfied by $(\mathscr{M},S_1)$.

The purely logical initial sequents of $\mathsf{KF}\restriction$ are satisfied by $(\mathscr{M},S_1)$ by the definition of the relation $\vDash$. For the truth-theoretic initial sequents, we consider three cases. Again for simplicity we work with the language $\mathscr{L}_T^+$.

Case 1: KF1. Assume $\mathscr{M} \models \mathrm{Cterm}_{\mathscr{L}_T}(\bar{a}) \wedge \mathrm{Cterm}_{\mathscr{L}_T}(\bar{b})$. Clearly we also have, for $\mathscr{M} \models \mathrm{Cterm}_{\mathscr{L}_T}(\bar{a})$,

$$(\mathscr{M},S_1) \vDash T\bar{a} \quad \text{iff} \quad (\mathscr{M},S_1,S_2) \vDash_{\mathsf{DM}} T\bar{a} \tag{21}$$

[12] The picture that arises from our result is complemented by the results in [Nicolai forth.], in which it is shown that PKF = IBT and that IKF = PKF plus transfinite induction up to ε_0.

Therefore we have the following:

$$(\mathscr{M},S_1) \vDash T(\bar{a} \doteq \bar{b}) \text{ iff } (\mathscr{M},S_1,S_2) \vDash_{\mathsf{DM}} T(\bar{a} \doteq \bar{b}) \tag{22}$$

Since $(\mathscr{M},S_1,S_2)$ is a model of $\mathsf{PKF}\restriction$, (22) obtains precisely if $a^{\circ\mathscr{M}} = b^{\circ\mathscr{M}}$. This implies the following:

$$(\mathscr{M},S_1) \models T(\bar{a} \doteq \bar{b}) \Rightarrow \bar{a}^\circ = \bar{b}^\circ \quad \text{and} \quad (\mathscr{M},S_1) \models \bar{a}^\circ = \bar{b}^\circ \Rightarrow T(\bar{a} \doteq \bar{b}).$$

Case 2: KF12. Assume $\mathscr{M} \models \mathrm{Cterm}_{\mathscr{L}_T}(\bar{a})$. We have that $(\mathscr{M},S_1) \vDash T\dot{T}\bar{a}$ if and only if $\dot{T}^{\mathscr{M}} a \in S_1$. This is the case exactly if $(\mathscr{M},S_1,S_2) \vDash_{\mathsf{DM}} T\dot{T}\bar{a}$ by definition of $\vDash_{\mathsf{DM}}$. Again, since $(\mathscr{M},S_1,S_2)$ is a model of $\mathsf{PKF}\restriction$, the former obtains if and only if $a^{\circ\mathscr{M}} \in S_1$. Therefore, with $a^{\circ\mathscr{M}} \in \mathrm{Cterm}^{\mathscr{M}}_{\mathscr{L}_T}$,

$$(\mathscr{M},S_1) \vDash T\dot{T}\bar{a} \Rightarrow T\bar{a}^\circ \tag{23}$$

$$(\mathscr{M},S_1) \vDash T\bar{a}^\circ \Rightarrow T\dot{T}\bar{a} \tag{24}$$

Case 3: KF13. We assume again $\mathscr{M} \models \mathrm{Cterm}_{\mathscr{L}_T}(\bar{a}^\circ)$. We recall that KF13 amounts to the sequents:

$$T\dot{\neg}\dot{T}t \Rightarrow T\dot{\neg}t^\circ, \neg\mathrm{Sent}_{\mathscr{L}_T}(t^\circ) \tag{25}$$

$$T\dot{\neg}t^\circ \vee \neg\mathrm{Sent}_{\mathscr{L}_T}(t^\circ) \Rightarrow T\dot{\neg}\dot{T}t \tag{26}$$

For (25), if $(\mathscr{M},S_1) \models T\dot{\neg}\dot{T}\bar{a}$, also $(\mathscr{M},S_1,S_2) \vDash_{\mathsf{DM}} T\dot{\neg}\dot{T}\bar{a}$ from (21). In $\mathsf{PKF}\restriction$, we have

$$\dfrac{\dfrac{\dfrac{\dfrac{\neg Tt^\circ \Rightarrow \neg Tt^\circ}{\neg T\dot{T}t \Rightarrow \neg Tt^\circ}\,\text{by PKF6(1)}}{T\dot{\neg}\dot{T}t \Rightarrow \neg Tt^\circ}\,\text{by PKF7(2)} \qquad \mathrm{Sent}_{\mathscr{L}_T}(t^\circ), \neg Tt^\circ \Rightarrow T\dot{\neg}t^\circ}{\mathrm{Sent}_{\mathscr{L}_T}(t^\circ), T\dot{\neg}\dot{T}t \Rightarrow T\dot{\neg}t^\circ}\,\text{Cut}}{T\dot{\neg}\dot{T}t \Rightarrow T\dot{\neg}t^\circ, \neg\mathrm{Sent}_{\mathscr{L}_T}(t^\circ)}\,\text{since } \mathrm{Sent}_{\mathscr{L}_T}(t^\circ) \in \mathscr{L}$$

Therefore either $\dot{\neg}^{\mathscr{M}} a^{\circ\mathscr{M}} \in S_1$ or $a^{\circ\mathscr{M}} \notin \mathrm{Sent}^{\mathscr{M}}_{\mathscr{L}_T}$. This suffices to conclude that $(\mathscr{M},S_1)$ classically satisfies (25).

For (26), if $(\mathscr{M},S_1) \models T\dot{\neg}\bar{a}^\circ \vee \neg\mathrm{Sent}_{\mathscr{L}_T}(\bar{a}^\circ)$, either $\dot{\neg}^{\mathscr{M}} a^{\circ\mathscr{M}} \in S_1$ or $a^{\circ\mathscr{M}}$ is not in $\mathrm{Sent}^{\mathscr{M}}_{\mathscr{L}_T}$. If the latter, we reason in PKF as follows:

$$\dfrac{\dfrac{\neg\mathrm{Sent}_{\mathscr{L}_T}(t^\circ) \Rightarrow \neg Tt^\circ}{\neg\mathrm{Sent}_{\mathscr{L}_T}(t^\circ) \Rightarrow \neg T\dot{T}t}\,\text{by PKF6(2)} \qquad \begin{array}{c}\vdots\\ \neg T\dot{T}t \Rightarrow T\dot{\neg}\dot{T}t^\circ\end{array}}{\neg\mathrm{Sent}_{\mathscr{L}_T}(t^\circ) \Rightarrow T\dot{\neg}\dot{T}t^\circ}\,\text{Cut}$$

In the right branch of the tree it is used the fact that $\mathsf{PKF}\restriction$ proves $\mathrm{Sent}_{\mathscr{L}_T}(\dot{T}t^\circ)$. The last line, together with our assumptions on $(\mathscr{M},S_1,S_2)$ and (21), forces that

$$(\mathscr{M},S_1) \models T\dot{\neg}\dot{T}\bar{a}$$

as required.

If by contrast $\mathscr{M} \models \mathrm{Sent}_{\mathscr{L}_T}(\bar{a}^\circ)$ and $\dot{\neg}^{\mathscr{M}} a^{\circ\mathscr{M}} \in S_1$, also

(27) $$(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \mathrm{Sent}_{\mathscr{L}_T}(\bar{a}^\circ) \wedge T\dot{\neg}\bar{a}^\circ$$

We can then reason in $\mathsf{PKF}\restriction$ as follows:

$$\dfrac{\dfrac{\mathrm{Sent}_{\mathscr{L}_T}(t^\circ), T\dot{\neg}t^\circ \Rightarrow \neg Tt^\circ}{\mathrm{Sent}_{\mathscr{L}_T}(t^\circ), T\dot{\neg}t^\circ \Rightarrow \neg T\dot{T}t}\ \text{by PKF7(2)}}{\mathrm{Sent}_{\mathscr{L}_T}(t^\circ), T\dot{\neg}t^\circ \Rightarrow T\dot{\neg}\dot{T}t}\ \text{by PKF7(1)}$$

The last line, by our assumption on $(\mathscr{M}, S_1, S_2)$, yields $(\dot{\neg}\dot{T}\bar{a})^{\mathscr{M}} \in S_1$ and, by (21), the desired conclusion.

For the induction step, the logical rules of $\mathsf{KF}\restriction$ are clearly truth-preserving in $(\mathscr{M}, S_1)$ by definition of $\vDash$. The induction rule of IND of classical PA is a rule of $\mathsf{PKF}\restriction$ by the classical behaviour of $\mathsf{PKF}\restriction$ described on page 12, therefore it is unproblematic.

We have now sufficient information to proceed with the proof of our main result.

Theorem 1 *If* $\mathsf{KF}\restriction \vdash T\ulcorner\varphi\urcorner$, *then* $\mathsf{PKF}\restriction \vdash \varphi$.

Proof We prove the counterpositive. If $\mathsf{PKF}\restriction \nvdash \varphi$ that is, if the sequent $\Rightarrow \varphi$ is not derivable in $\mathsf{PKF}\restriction$, then by the completeness of DM (see Proposition 4, Appendix 3),

(28) either there is a model $(\mathscr{M}, S_1, S_2)$ of $\mathsf{PKF}\restriction$ such that $(\mathscr{M}, S_1, S_2) \nvDash_{\mathsf{DM}} \varphi$

(29) or there is a model $(\mathscr{M}, S_1, S_2)$ of $\mathsf{PKF}\restriction$ such that $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \neg\varphi$

Assume (28) holds. By Lemma 6 we find a classical model $(\mathscr{M}, S_1)$ satisfying $\mathsf{KF}\restriction$. If, in addition, $\mathsf{KF}\restriction \vdash T\ulcorner\varphi\urcorner$, then $\ulcorner\varphi\urcorner^{\mathscr{M}} \in S_1$ and $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \varphi$ by the assumption on $(\mathscr{M}, S_1, S_2)$ and Lemma 3. Therefore if (28), $T\ulcorner\varphi\urcorner$ is not provable in $\mathsf{KF}\restriction$.

If (29) holds, $\ulcorner\neg\varphi\urcorner^{\mathscr{M}} \in S_1$ – and $\ulcorner\varphi\urcorner^{\mathscr{M}} \in S_2$ – by Lemma 3. Now the model $(\mathscr{M}, S_1, S_2)$ may be either consistent, or complete, or mixed. If the latter, by Proposition 1(i) and Proposition 3 the consistent model $(\mathscr{M}, S_1', S_2')$ obtained by turning truth-value gluts in $(\mathscr{M}, S_1, S_2)$ into truth-value gaps is indeed a model of $\mathsf{PKF}\restriction$ and satisfies all sequents of the form $\varphi \Rightarrow \psi$, for $\varphi, \psi \in \mathscr{L}_T$ that are satisfied in $(\mathscr{M}, S_1, S_2)$. Therefore $(\mathscr{M}, S_1', S_2') \vDash_{\mathsf{DM}} \neg\varphi$ and $\ulcorner\neg\varphi\urcorner^{\mathscr{M}} \in S_1'$. Now if again $\mathsf{KF}\restriction$ proves $T\ulcorner\varphi\urcorner$, by Lemma 6 also $\ulcorner\varphi\urcorner^{\mathscr{M}} \in S_1'$, contradicting the consistency of $(\mathscr{M}, S_1', S_2')$.

The reasoning for the case in which $(\mathscr{M}, S_1, S_2)$ is consistent is contained in the previous case. If $(\mathscr{M}, S_1, S_2)$ is complete, by Corollary 1 we transform $(\mathscr{M}, S_1, S_2)$ into a consistent model $(\mathscr{M}, S_1'', S_2'')$ of $\mathsf{PKF}\restriction$. Therefore, by (29), $\varphi \in S_2''$. Again we consider the classical model $(\mathscr{M}, S_1'')$ satisfying $\mathsf{KF}\restriction$ by Lemma 6. If $\mathsf{KF}\restriction$ proved $T\ulcorner\varphi\urcorner$, we would have $\ulcorner\varphi\urcorner^{\mathscr{M}} \in S_1''$, again contradicting the consistency of $(\mathscr{M}, S_1'', S_2'')$.

Lemma 5 already told us that the internal theory of $\mathsf{KF}\restriction$ contains $\mathsf{PKF}\restriction$: the former is strong enough to formalize the fact that a proof in $\mathsf{PKF}\restriction$ of a sentence φ of $\mathscr{L}_T$ guarantees the truth of φ. Theorem 1 gives us a converse: in determining the truth or falsity of a sentence φ in the classical universe of $\mathsf{KF}\restriction$ we are not pushed away from what can be determined true and false in $\mathsf{PKF}\restriction$.

Corollary 2 *The sentences provable* $\mathsf{PKF}\restriction$ *coincide with the internal theory of* $\mathsf{KF}\restriction$.

5 Extensions

The results introduced in the previous section are stable under extensions of PKF and KF obtained by restricting the class of their models in some natural way. In this section we focus on the extensions of PKF and KF that force either consistent or complete models.

In the definition of a four-valued model $(\mathscr{M}, S_1, S_2)$, we have imposed no condition on the pair (S_1, S_2). In this section we require them to be either exhaustive or disjoint (but not both). We will call these models *coc-models* from 'consistent or complete'.

Definition 1

(i) *The logic* DM^* *is obtained by adding the sequent*

(GG) $\qquad \varphi, \neg\varphi \Rightarrow \psi, \neg\psi$

to the initial sequents of DM.

(ii) PKF^* *is obtained by adding truth-theoretic initial sequents and rules of* PKF, *including the extended induction rule, to* DM^*. $\mathsf{PKF}^*\restriction$ *is obtained via the usual restriction and the addition of the regularity axioms.*

(iii) KF^* *is obtained by adding to the initial sequents and rules of* KF *the combination of the following principles in disjunctive form forcing either complete or consistent models but not both (cf. Appendix 2).*

(COMP) $\qquad \forall x(\mathrm{Sent}_{\mathscr{L}_T}(x) \wedge \neg Tx \rightarrow T\dot{\neg}x)$

(CONS) $\qquad \forall x(\mathrm{Sent}_{\mathscr{L}_T}(x) \wedge T\dot{\neg}x \rightarrow \neg Tx)$

Again $\mathsf{KF}^*\restriction$ *is defined in the obvious way.*

It's clear that we cannot have both COMP and CONS *sine contradictione*.

The system DM^* is sound with respect to coc-models. It is clear in fact that (GG) forces either complete or consistent models. By adapting the argument given in Proposition 4 of Appendix 3, one verifies that DM^* is also complete with respect to the intended semantics, coc-models in this case. We state this fact for PKF^*:

Lemma 7 *Let* Γ, Δ *be finite sets of* $\mathscr{L}_T$*-sentences. Then*

$$\Gamma \vDash_{\mathsf{PKF}^*} \Delta \text{ if and only if } \mathsf{PKF}^* \vdash \Gamma \Rightarrow \Delta.$$

Here $\vDash_{\mathsf{PKF}^*}$ *is defined as on page 6 with a suitable relativization to coc-models.*

Obviously, PKF^* will also be sound with respect to suitable fixed-point models in the sense of Lemma 2 above. In particular, if $\Gamma \Rightarrow \Delta$ is derivable in PKF^*, then for all fixed-point models $(\mathbb{N}, S_1, S_2)$ with $S_1 \cap S_2 = \varnothing$ or $S_1 \cup S_2 = \omega$ the following two properties obtain:

(i) If all sentences in Γ are true in $(\mathbb{N}, S_1, S_2)$, there is a sentence in Δ that is true in $(\mathbb{N}, S_1, S_2)$.
(ii) If all sentences in Δ are false in $(\mathbb{N}, S_1, S_2)$, there is at least one sentence in Γ that is false in $(\mathbb{N}, S_1, S_2)$.

Obviously these observations apply also to $\mathsf{PKF}^*\restriction$.

Like in the case of KF and PKF, there is a straightforward sense in which KF^* (resp. $\mathsf{KF}^*\restriction$) and PKF^* (resp. $\mathsf{PKF}^*\restriction$) may be taken to embody the conception of truth described at the beginning of the paper. Lemma 4 yields in fact that, for a consistent or complete fixed point $S \subseteq \omega$,

(30) $\qquad (\mathbb{N}, S) \vDash \mathsf{KF}^*$ if and only if $(\mathbb{N}, S) \vDash_{\mathsf{DM}} \mathsf{PKF}^*$

The same holds for the restricted versions of the theories.

Like DM, also DM^* is therefore slightly different from the logic originally employed by Kripke. It also differs from the logic of PKF as described in [Horsten 2009], [Horsten 2011], and [Halbach 2014]; these presentations do not feature the initial sequent

(GG) $\qquad \varphi, \neg\varphi \Rightarrow \psi, \neg\psi$

They admit the possibility of 'mixed' models, that is models that contain both gaps and gluts.

[Halbach & Horsten 2006] force consistent or complete models by considering a truth-theoretic initial sequent of the form $Tt, \neg Tt \Rightarrow \lambda$, provided that the value of the closed term t is a sentence of $\mathscr{L}_T$. By our definition of logical consequence for DM, if λ is a glut in some model $(\mathscr{M}, S_1, S_2)$ of PKF, then it will be a fortiori in S_2, thus by the λ-sequent and transparency either φ or $\neg\varphi$ will be in S_2 for any sentence φ. Similarly, if λ is a gap in some $(\mathscr{M}, S_1, S_2) \vDash_{\mathsf{DM}} \mathsf{PKF}$, then there cannot be gluts in $(\mathscr{M}, S_1, S_2)$ by a similar reasoning.

We now verify that our results are stable under the variations considered, in particular we check that $\mathsf{IKF}^*\restriction$ is identical with the set of theorems of $\mathsf{PKF}^*\restriction$. Lemma 5 can be extended to the new setting. This extension involves proving cases not covered by the proof in [Halbach & Horsten 2006], that we now consider.

Lemma 8 *If* $\mathsf{PKF}^*\restriction$ *proves* φ*, then* $\varphi \in \mathsf{IKF}^*\restriction$.

Proof We recall that it suffices to prove

(31) $\qquad \mathsf{KF}^*\restriction \vdash \forall \mathbf{t}\left(T^\ulcorner \bigwedge \Gamma^\urcorner(\mathbf{t}/\mathbf{v}) \rightarrow T^\ulcorner \bigvee \Delta^\urcorner(\mathbf{t}/\mathbf{v})\right)$

(32) $\qquad \mathsf{KF}^*\restriction \vdash \forall \mathbf{t}\left(T^\ulcorner \neg \bigvee \Delta^\urcorner(\mathbf{t}/\mathbf{v}) \rightarrow T^\ulcorner \neg \bigwedge \Gamma^\urcorner(\mathbf{t}/\mathbf{v})\right)$

We treat the case of the new initial sequent (GG) not included in the proofs in [Halbach & Horsten 2006] and [Halbach 2014]. We prove (31) and (32) for (GG) in $\mathsf{KF}+\mathsf{CONS}$. The cases for $\mathsf{KF}+\mathsf{COMP}$ are easily adapted. Reasoning in $\mathsf{KF}+\mathsf{CONS}$, we have

$$\dfrac{\dfrac{\dfrac{T^{\ulcorner}\neg\varphi^{\urcorner}\Rightarrow\neg T^{\ulcorner}\varphi^{\urcorner}}{T^{\ulcorner}\varphi^{\urcorner},T^{\ulcorner}\neg\varphi^{\urcorner}\Rightarrow}\,(\neg\mathrm{k1})}{T^{\ulcorner}\varphi\wedge\neg\varphi^{\urcorner}\Rightarrow}\,\text{by logic, KF4 and multiple cuts}}{T^{\ulcorner}\varphi\wedge\neg\varphi^{\urcorner}\Rightarrow T^{\ulcorner}\psi\vee\neg\psi^{\urcorner}}\,(\mathrm{W1})$$

For (20) it is worth recalling that for all $\varphi\in\mathscr{L}_T$,

$$\mathsf{KF}+\mathsf{CONS}\vdash T^{\ulcorner}\varphi^{\urcorner}\Rightarrow\varphi \tag{33}$$

Again in KF+CONS, we have

$$\dfrac{\dfrac{T^{\ulcorner}\neg(\psi\vee\neg\psi)^{\urcorner}\Rightarrow T^{\ulcorner}\neg(\psi\vee\neg\psi)^{\urcorner}}{T^{\ulcorner}\neg(\psi\vee\neg\psi)^{\urcorner}\Rightarrow\neg(\psi\vee\neg\psi)}\,\text{by (33) and Cut}}{T^{\ulcorner}\neg(\psi\vee\neg\psi)^{\urcorner}\Rightarrow\neg T^{\ulcorner}\varphi\wedge\neg\varphi^{\urcorner}}\,\text{by logic and (W2)}$$

We now move to the analogue of Theorem 1 for $\mathsf{KF}^*\restriction$ and $\mathsf{PKF}^*\restriction$.

Corollary 3 *If* $\mathsf{KF}^*\restriction\vdash T^{\ulcorner}\varphi^{\urcorner}$, *also* $\mathsf{PKF}^*\restriction\vdash\varphi$.

Proof The argument provided in the proof of Theorem 1 essentially suffices to obtain the result. In particular, one again proves the counterpositive. By Lemma 7, one of the two following conditions must obtain:

(34) Either there is a model $(\mathscr{M},S_1,S_2)$ of $\mathsf{PKF}^*\restriction$ such that $(\mathscr{M},S_1,S_2)\nvDash_{\mathsf{DM}}\varphi$,

(35) or there is a model $(\mathscr{M},S_1,S_2)$ of $\mathsf{PKF}^*\restriction$ such that $(\mathscr{M},S_1,S_2)\vDash_{\mathsf{DM}}\neg\varphi$

If (34), one simply proceeds as in Theorem 1 under the assumption (28). If (35), since (GG) is an initial sequent of $\mathsf{PKF}^*\restriction$, one does not have to deal with mixed models of $\mathsf{PKF}^*\restriction$: this means that in the case of complete models an application of Corollary 1 yields the result.

6 Conclusion

Our results explain why IKF is much richer than PKF and contains more true arithmetical sentences than PKF. It puts the blame on the logic and its effects on mathematical reasoning. The weakness of PKF could have its source in different aspects of PKF: PKF could be weaker than IKF because it may not feature the strongest possible axioms and rules for the nonclassical logic. The completeness of PKF with respect to the logic rules out this possibility: The logical rules and axioms of PKF are complete and cannot be strengthened in a sound way.

This leaves open the possibility that PKF does not include some rule for truth that should have been included. In the setting of the nonclassical sequent calculus

of PKF it's not so easy to see whether some rules can be strengthened and in the course of the reformulation of the truth axiom into the rules of PKF a version that is too weak might have been chosen inadvertently. Unlike in the case of the logic, we cannot prove a semantic completeness theorem because there always will be true truth-theoretic statements such as $T^{\ulcorner}\mathsf{Con}_{\mathsf{ZF}}^{\urcorner}$ that are true (for all we know) but cannot be expected to be provable in any truth theory based on Peano arithmetic.

However, as we are interested in a comparison with IKF, we can investigate whether the truth-theoretic rules of PKF are incomplete compared to what can be proved to be in IKF by using all the truth-theoretic axioms of KF. Theorem 1 rules out this possibility as well. In fact, it shows more. If we can combine all logical axioms of PKF, all arithmetical axioms, all its truth rules and just omit the induction rule, then the resulting theorems are exactly the same as those in IKF. Only when we add the induction rule to KF and PKF, the systems come apart. PKF is weaker than IKF not because using the nonclassical logic cripples our reasoning about truth, but rather because we cannot exploit or capture the full strength of arithmetical induction when our logic is restricted.

Induction forms part of our mathematical reasoning patterns. Hence, when we restrict classical logic for the language with the truth predicate, we do not only affect our reasoning about truth but also our mathematical reasoning. In a sense our completeness result shows that our reasoning about truth is not really affected by employing the logic of PKF, because using classical logic to think about the same concept of truth doesn't yield any additional insights, that is, $\mathsf{PKF}\restriction$ proves all sentences in the inner logic of $\mathsf{KF}\restriction$.

We have mentioned the strategy of restricting classical logic for semantic reasoning, while retaining classical logic for all sentences not involving the truth predicate. This strategy is often defended by claiming that the restriction of classical logic doesn't affect mathematics, because only reasoning with the truth predicate is nonclassical and in mathematics we don't employ the truth predicate. We think that this defence fails for two reasons.

First, mathematics is not tied to a fixed language. Of course we use mathematics in many other areas, most notably in the sciences. Induction can be applied to any condition whether it contains non-mathematical vocabulary or not. Reasoning patterns such as induction is mathematical even if the vocabulary involved should belong to physics or chemistry. Reasoning about truth would be the only exception if the defence of nonclassical logic succeeded.

The defence fails for a further reason. KF (and IKF) contains more purely arithmetical sentences than PKF. Thus the change in logic does affect mathematics. Now one could object that these additional arithmetical theorems in KF can only be established by proofs involving the truth predicate. This is correct, but nobody would say that a claim doesn't belong to physics because it can only be established by going through mathematical or chemical reasoning. Furthermore one could defend the view that reasoning in KF is mathematical. And thus one never leaves mathematics even when reasoning about truth.

Our results might also be employed to revive some aspects of the program sketched in [Reinhardt 1986]. There it was envisaged an instrumentalist interpretation of KF as a tool to unravel theorems of PKF: if theorems of PKF coincided with what KF proves true, the clumsiness of four-valued logic could be sidestepped. We could simply use KF while maintaining that its trustworthy core is captured in a direct and natural way by PKF. This possibility, as we have mentioned, is excluded by the results of [Halbach & Horsten 2006]. IKF is much stronger than PKF.

Corollary 2 seems to suggest the success of a more limited version of Reinhardt's program that focuses on $\mathsf{KF}\restriction$ and $\mathsf{PKF}\restriction$. Our results support in fact an instrumentalist understanding of $\mathsf{KF}\restriction$ as a device for generating classical proofs of theorems of PKF. This approach could be defended, for instance, by holding that it is $\mathsf{PKF}\restriction$, and not PKF, that captures the core of the conception of truth outlined at the beginning of the paper. To articulate such approach, in other words, one would have to argue for the restriction of basic principles belonging to the toolbox of science such as mathematical induction.[13] We have already argued against such restrictions. It is nonetheless reasonable to believe that the instrumentalist reading of $\mathsf{KF}\restriction$ just outlined may be attractive to authors that do not share our stance on the nature and role of mathematical schemata.

Of course, our results here are only a case study. The base theory can varied, the assumptions on truth, and the way classical logic is restricted can be varied. In other base theories different schemata may be employed. We don't intend to embark on the enterprise of browsing through all possible combinations. We think the burden of proof is with those who advocate a restriction of classical logic.

The results by [Halbach & Horsten 2006] and the results of this paper can serve as a warning: We shouldn't expect that the effects of restricting classical logic for use with the truth predicate can be contained.[14]

References

Blamey 2002. Blamey, S. (2002). 'Partial Logic'. In Gabbay D. and F. Guenther. *Handbook of Philosophical Logic*. Kluwer, Dordrecht, Vol. 5, 261-253, second edition.

Cantini 1989. Cantini, A. (1989). Notes on Formal Theories of Truth. *ZeitSchrift für matematische Logik und Grundlagen der Mathematik* 35, 97–130.

Feferman 1991. Feferman, S. (1991). Reflecting on Incompleteness. *The Journal of Symbolic Logic* 56: 1–49.

Field 2008. Field, H. (2008). *Saving truth from paradox*. Oxford University Press.

Fischer et alii 2015. Fischer, M, V. Halbach, J. Kriener, J. Stern (2015). Axiomatizing Semantic Theories of Truth. *The Review of Symbolic Logic* 8(02), pp 257-278.

Halbach 2014. Halbach, V. (2014). *Axiomatic Theories of Truth*. CUP.

Feferman 1984. Feferman, S. (1984). Towards useful type-free theories. *The Journal of Symbolic Logic* 49, 75–111.

Fujimoto 2010. Fujimoto, K. (2010). Relative truth definability of axiomatic theories of truth. *The Bulletin of Symbolic Logic* 16, 305–344.

[13] Equivalently, restriction to the least number principles or to forms of collection.

[14] See also [Halbach 2014, chapter 20].

Halbach & Horsten 2006. Halbach, V. and L. Horsten (2006). Axiomatizing Kripke's Theory of Truth. *The Journal of Symbolic Logic* 71: 677-712.

Horsten 2009. Horsten, L. (2009). Levity. *Mind*, 118, 555–581.

Horsten 2011. Horsten, L. (2012). *The Tarskian Turn: deflationism and axiomatic truth.* Princeton University Press 2012.

Kremer 1988. Kremer, M. (1988). Kripke and the Logic of Truth. *Journal of Philosophical Logic* 17, 225-278.

Kripke 1975. Kripke, S. (1975). Outline of a Theory of Truth. *Journal of Philosophy* 72: 690–712.

McGee 1991. McGee, V. (1991). *Truth, Vagueness and Paradox: An Essay in the Logic of Truth*. Hackett, Cambridge.

Moschovakis 1974. Moschovakis, Y. (1974). Elementary Induction on Abstract Structures. North-Holland, Amsterdam.

Nicolai forth.. Nicolai, C. Provably true sentences across axiomatizations of Kripke's theory of truth. Unpublished Manuscript.

Reinhardt 1986. Reinhardt, W. (1986). Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth. *Journal of Philosophical Logic* 15: 219-251.

Scott 1975. Scott, D. (1975). 'Combinators and Classes'. In Böhm, C. *.λ-Calculus in Computer Science*. Springer, Berlin 1975: 1–26.

Williamson 2016. Williamson, T. (2016). Semantic Paradoxes and Abductive Methodology. To appear in: Armour-Garb, B. (forthcoming). *The Relevance of the Liar*. Oxford University Press.

Appendix 1: the logic DM

The initial sequents of the system DM are:

$$\text{(IN)}\quad \Gamma \Rightarrow \Delta, \text{ if } \Gamma \cap \Delta \neq \varnothing \qquad \text{(Cut)}\ \frac{\Gamma \Rightarrow \varphi, \Delta \qquad \Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

$$\text{(W1)}\ \frac{\Gamma \Rightarrow \Delta}{\Gamma, \varphi \Rightarrow \Delta} \qquad \text{(W2)}\ \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \varphi, \Delta}$$

$$\text{(Sub)}\ \frac{\Gamma \Rightarrow \Delta}{\Gamma(t/x) \Rightarrow \Delta(t/x)} \qquad (\neg)\ \frac{\Gamma \Rightarrow \Delta}{\neg\Delta \Rightarrow \neg\Gamma}$$

$$(\mathsf{d}\neg 1)\quad \varphi \Rightarrow \neg\neg\varphi \qquad (\mathsf{d}\neg 2)\quad \neg\neg\varphi \Rightarrow \varphi$$

$$(=1)\ \Rightarrow t = t \qquad (=2)\ \ s = t, \varphi(s/x) \Rightarrow \varphi(t/x)$$

$$(\wedge 1)\ \frac{\Gamma, \varphi, \psi \Rightarrow \Delta}{\Gamma, \varphi \wedge \psi \Rightarrow \Delta} \qquad (\wedge 2)\ \frac{\Gamma, \varphi \wedge \psi \Rightarrow \Delta}{\Gamma, \varphi, \psi \Rightarrow \Delta}$$

$$(\vee 1)\ \frac{\Gamma \Rightarrow \varphi, \psi, \Delta}{\Gamma \Rightarrow \varphi \vee \psi, \Delta} \qquad (\vee 2)\ \frac{\Gamma \Rightarrow \varphi \vee \psi, \Delta}{\Gamma \Rightarrow \varphi, \psi, \Delta}$$

$$(\forall 1)\ \frac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma \Rightarrow \forall x\varphi, \Delta} \qquad (\forall 2)\ \frac{\Gamma \Rightarrow \forall x\varphi, \Delta}{\Gamma \Rightarrow \varphi(t/x), \Delta}$$

$$(\exists 1)\ \frac{\Gamma, \exists x\varphi \Rightarrow \Delta}{\Gamma, \varphi(t/x) \Rightarrow \Delta} \qquad (\exists 2)\ \frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma, \exists x\varphi \Rightarrow \Delta}$$

In $(\neg)$, $\neg\Gamma$ denotes the set of all negations of sentences in Γ. In $(\forall 1)$ and $(\exists 2)$, x is not free in the lower sequent.

Classical logic K is obtained by replacing, in DM, $(\neg)$ with the stronger

$$(\neg\mathsf{k}1)\ \frac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma, \neg\varphi \Rightarrow \Delta} \qquad (\neg\mathsf{k}2)\ \frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \neg\varphi, \Delta}$$

Appendix 2: Axioms of KF

PAT is the theory formulated in K (cf. Appendix 1) extended with initial sequents $\Rightarrow \varphi$, for φ a basic axiom of PA, and the induction rule

$$\text{(IND)}\qquad \frac{\Gamma, \varphi(x) \Rightarrow \varphi(Sx), \Delta}{\Gamma, \varphi(\overline{0}) \Rightarrow \varphi(t), \Delta}$$

with x not free in $\Gamma, \Delta, \varphi(\overline{0})$ and t arbitrary, for all formulas φ of $\mathscr{L}_T$. KF is obtained from PAT by adding the following initial sequents, where s, t range over (codes of) closed terms of $\mathscr{L}_T$:

KF1 1. $s^\circ = t^\circ \Rightarrow T(s \doteq t)$ (where $(\cdot)^\circ$ is the arithmetical evaluation function)
2. $T(s \doteq t) \Rightarrow s^\circ = t^\circ$

KF2 1. $s^\circ \neq t^\circ \Rightarrow T(\dot{\neg} s = t)$
2. $T(\dot{\neg} s = t) \Rightarrow s^\circ \neq t^\circ$

KF3 1. $\mathrm{Sent}_{\mathscr{L}_T}(x), T\dot{\neg}\dot{\neg}x \Rightarrow Tx$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x), Tx \Rightarrow T\dot{\neg}\dot{\neg}x$

KF4 1. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\wedge} y), T(x \dot{\wedge} y) \Rightarrow Tx \wedge Ty$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\wedge} y), Tx \wedge Ty \Rightarrow T(x \dot{\wedge} y)$

KF5 1. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\wedge} y), T\dot{\neg}(x \dot{\wedge} y) \Rightarrow T\dot{\neg}x \vee T\dot{\neg}y$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\wedge} y), T\dot{\neg}x \vee T\dot{\neg}y \Rightarrow T\dot{\neg}(x \dot{\wedge} y)$

KF6 1. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\vee} y), T(x \dot{\vee} y) \Rightarrow Tx \vee Ty$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\vee} y), Tx \vee Ty \Rightarrow T(x \dot{\vee} y)$

KF7 1. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\vee} y), T\dot{\neg}(x \dot{\vee} y) \Rightarrow T\dot{\neg}x \wedge T\dot{\neg}y$
2. $\mathrm{Sent}_{\mathscr{L}_T}(x \dot{\vee} y), T\dot{\neg}x \wedge T\dot{\neg}y \Rightarrow T\dot{\neg}(x \dot{\vee} y)$

KF8 1. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\forall} vx), \forall t\, T(x(t/v)) \Rightarrow T(\dot{\forall} vx)$
2. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\forall} vx), T(\dot{\forall} vx) \Rightarrow \forall t\, T(x(t/v))$

KF9 1. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\forall} vx), \exists t\, T(\dot{\neg} x(t/v)) \Rightarrow T(\dot{\neg}\dot{\forall} vx)$
2. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\forall} vx), T(\dot{\neg}\dot{\forall} vx) \Rightarrow \exists t\, T(\dot{\neg} x(t/v))$

KF10 1. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\exists} vx), \exists t\, T(x(t/v)) \Rightarrow T(\dot{\exists} vx)$
2. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\exists} vx), T(\dot{\exists} vx) \Rightarrow \exists t\, T(x(t/v))$

KF11 1. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\exists} vx), \forall t\, (T\dot{\neg} x(t/v)) \Rightarrow T(\dot{\neg}\dot{\exists} vx)$
2. $\mathrm{Sent}_{\mathscr{L}_T}(\dot{\exists} vx), T(\dot{\neg}\dot{\exists} vx) \Rightarrow \forall t\, (T\dot{\neg} x(t/v))$

KF12 $Tt^\circ \Leftrightarrow T\dot{T}t$

KF13 $T\dot{\neg}\dot{T}t \Leftrightarrow T\dot{\neg}t^\circ \vee \neg\mathrm{Sent}_{\mathscr{L}_T}(t^\circ)$

KF14 $Tx \Rightarrow \mathrm{Sent}_{\mathscr{L}_T}(x)$

KF$\restriction$ is obtained by restricting (IND) to formulas of $\mathscr{L}$ and adding the sequents (Reg1) and (Reg2). We consider the extension of KF – that we call KF* – via the disjunction of the following sentences, forcing complete or consistent extensions of the truth predicate respectively:

(COMP) $\quad \forall x(\mathrm{Sent}_{\mathscr{L}_T}(x) \wedge \neg Tx \rightarrow T\dot{\neg}x)$

(CONS) $\quad \forall x(\mathrm{Sent}_{\mathscr{L}_T}(x) \wedge T\dot{\neg}x \rightarrow \neg Tx)$

Appendix 3: Completeness of DM

The bulk of this appendix is to prove Proposition 4 below. We recall that we write $\Gamma \models_{\mathscr{S}} \Delta$ to mean that for all four-valued models $\mathscr{M}$ of a set of sequents and rules $\mathscr{S} \supseteq \mathsf{DM}$, the following are jointly satisfied:[15]

- if $\mathscr{M} \models_{\mathsf{DM}} \varphi$ for all $\varphi \in \Gamma$, then $\mathscr{M} \models_{\mathsf{DM}} \psi$ for at least one $\psi \in \Delta$;
- if $\mathscr{M} \models_{\mathsf{DM}} \neg\psi$ for all $\psi \in \Delta$, then $\mathscr{M} \models_{\mathsf{DM}} \neg\varphi$ for at least one $\varphi \in \Gamma$.

For notational simplicity, we do not deal with variable assignments.

The informal sketch of the strategy is as follows: we start with a pair (Γ, Δ) such that $\Gamma \Rightarrow \Delta$ is not derivable in $\mathscr{S}$, and build 'maximal' sets Γ^+ and Δ^+ such that $\Gamma^+ \Rightarrow \Delta^+$ is still not $\mathscr{S}$-derivable. We can in fact extend our notion of derivability (and underivability) in DM to sequents $\Lambda \Rightarrow \Theta$ where Λ, Θ are infinite: $\mathscr{S} \vdash \Lambda \Rightarrow \Theta$ iff for some finite $\Lambda' \subseteq \Lambda$, $\Theta' \subseteq \Theta$, we have $\mathscr{S} \vdash \Lambda' \Rightarrow \Theta'$. From Γ^+, Δ^+ a four-valued model $\mathscr{M}$ of $\mathscr{S}$ can be then read off, one that satisfies at least one of the following:

- if $\mathscr{M} \models_{\mathsf{DM}} \varphi$ for all $\varphi \in \Gamma^+$, then $\mathscr{M} \not\models_{\mathsf{DM}} \psi$ for $\psi \in \Delta^+$;
- if $\mathscr{M} \models_{\mathsf{DM}} \neg\psi$ for all $\psi \in \Delta^+$, then $\mathscr{M} \not\models_{\mathsf{DM}} \neg\varphi$ for $\varphi \in \Gamma^+$.

Henkin-style arguments of the kind given below can be found in [Kremer 1988] for a different sequent calculus and in [McGee 1991] for a system of natural deduction. [Blamey 2002] hints at a more general strategy.

Proposition 4 *Let $\mathscr{S}$ be a set of sequents in a language $\mathscr{L}$ and Γ, Δ finite sets of $\mathscr{L}$-formulas. Then:*

$$\textit{if } \Gamma \models_{\mathscr{S}} \Delta, \textit{ then } \vdash_{\mathscr{S}} \Gamma \Rightarrow \Delta.$$

We prove the counterpositive. If $\not\vdash_{\mathscr{S}} \Gamma \Rightarrow \Delta$, by (Cut), there cannot be a φ such that

$$\vdash_{\mathscr{S}} \Gamma \Rightarrow \Delta, \varphi \text{ and } \vdash_{\mathscr{S}} \Gamma, \varphi \Rightarrow \Delta.$$

Let $\mathscr{L}^+ := \mathscr{L} \cup C$ the language $\mathscr{L}$ expanded with a countable set of new constants. Let $\{\varphi_n \mid n \in \omega\}$ and $\{c_i \mid i \in \omega\}$ enumerations of the formulae of $\mathscr{L}^+$ and constants in C respectively.

Next we inductively define Γ^+ and Δ^+ as follows:

- $\Gamma_0 := \Gamma$ and $\Delta_0 := \Delta$
- if $\not\vdash_{\mathscr{S}} \Gamma_n, \varphi_n \Rightarrow \Delta_n$, then $\Delta_{n+1} = \Delta_n$ and

$$\Gamma_{n+1} := \begin{cases} \Gamma_n \cup \{\varphi_n, \chi(c_i/x)\}, & \text{if } \varphi_n \text{ is } \exists x\chi \text{ and } c_i \text{ is the least constant in } C \\ & \quad \text{not occurring in } \Gamma_n, \Delta_n, \varphi_n. \\ \Gamma_n \cup \{\varphi_n, \neg\chi(c_i/x)\}, & \text{if } \varphi_n \text{ is } \neg\forall x\chi \text{ and } c_i \text{ the least constant in } C \text{ not} \\ & \quad \text{occurring in } \Gamma_n, \Delta_n, \varphi_n. \\ \Gamma_n \cup \{\varphi_n\}, & \text{otherwise} \end{cases}$$

[15] By writing '$\mathscr{M}$ is a model of $\mathscr{S}$' we mean that $\mathscr{M}$ is a model of its derivable sequents.

– if $\vdash_{\mathscr{S}} \Gamma_n, \varphi_n \Rightarrow \Delta_n$, we have $\Gamma_{n+1} = \Gamma_n$ and

$$\Delta_{n+1} = \begin{cases} \Delta_n \cup \{\varphi_n, \chi(c_i/x)\}, & \text{if } \varphi_n \text{ is } \forall x\chi \text{ and } c_i \text{ is the least constant in } C \text{ not occurring in } \Gamma_n, \Delta_n, \varphi_n. \\ \Delta_n \cup \{\varphi_n, \neg\chi(c_i/x)\}, & \text{if } \varphi_n \text{ is } \neg\exists x\chi \text{ and } c_i \text{ the least constant in } C \text{ not occurring in } \Gamma_n, \Delta_n, \varphi_n. \\ \Delta_n \cup \{\varphi_n\}, & \text{otherwise} \end{cases}$$

Let $\Gamma^+ = \bigcup_{n\in\omega} \Gamma_n$ and $\Delta^+ = \bigcup_{n\in\omega} \Delta_n$. Obviously we have $\Gamma \subset \Gamma^+$ and $\Delta \subset \Delta^+$.

Next we state some properties of the pair (Γ^+, Δ^+).

Lemma 9

1. *If $\Gamma' \subset \Gamma^+$, $\Delta' \subset \Delta^+$, with Γ', Δ' finite, and $\mathscr{S} \vdash \Gamma' \Rightarrow \Delta'$, then there is an $n \in \omega$ such that*
$$\mathscr{S} \vdash \Gamma_n \Rightarrow \Delta_n$$
2. *$\nvdash_{\mathscr{S}} \Gamma_n \Rightarrow \Delta_n$ for all n;*
3. *$\Gamma^+ \cap \Delta^+ = \varnothing$*
4. *$t = t \in \Gamma^+$ and $t \neq t \in \Delta^+$;*
5. *If $\varphi \notin \Gamma^+$, then $\vdash_{\mathscr{S}} \Gamma^+ \cup \{\varphi\} \Rightarrow \Delta^+$ (symmetrically for $\varphi \notin \Delta^+$);*
6. *$s = t \in \Gamma^+$ and $\varphi(s/x) \in \Gamma^+$, then $\varphi(t/x) \in \Gamma^+$;*
7. *$s \neq t$ and $\varphi(s/x) \in \Delta^+$, then $\varphi(t/x) \in \Delta^+$;*
8. *$s = t \in \Gamma^+$ if and only if $s \neq t \in \Delta^+$;*
9. *there is no $\varphi \in \mathscr{L}^+$ such that*
$$\Gamma^+ \Rightarrow \Delta^+, \varphi \text{ and } \varphi, \Gamma^+ \Rightarrow \Delta^+.$$

A four-valued model $\mathscr{M}$ can now be defined as follows:

$$M := \{[\![t]\!] \mid t \in Term_{\mathscr{L}^+}\}$$

where

$$[\![t]\!] := \{s \in Term_{\mathscr{L}^+} \mid s = t \in \Gamma^+\}$$

that is the equivalence class of terms of $\mathscr{L}^+$ determined by $s = t \in \Gamma^+$. Moreover,

$$P^+_{\mathscr{M}} := \{([\![t_1]\!], ..., [\![t_n]\!]) \mid P(t_1, ..., t_n) \in \Gamma^+\}$$
$$P^-_{\mathscr{M}} := \{([\![t_1]\!], ..., [\![t_n]\!]) \mid \neg P(t_1, ..., t_n) \in \Gamma^+\}$$

The term model $\mathscr{M}$ so-defined enables us to prove the following lemma, which will in turn yields the desired result:[16]

Lemma 10 *For all $\varphi \in \mathscr{L}^+$,*

1. *$\varphi \in \Gamma^+$ if and only if $\mathscr{M} \vDash_{\mathsf{DM}} \varphi$;*
2. *if $\varphi \in \Delta^+$, then $\mathscr{M} \nvDash_{\mathsf{DM}} \varphi$.*

[16] Of course in the other case in which we focus on Δ^+, the Lemma would need to be suitably modified.

Proof By induction on the positive complexity of φ.[17] We mostly focus on the base and on the existential quantifier cases.

Case 1 : $\varphi := (s = t)$. $(s = t) \in \Gamma^+$ iff $[\![s]\!] = [\![t]\!]$ iff $\mathscr{M} \vDash_{\mathsf{DM}} s = t$. If $(s = t) \in \Delta^+$ and $\mathscr{M} \vDash_{\mathsf{DM}} s = t$, then $[\![s]\!] = [\![t]\!]$ and $s = t \in \Gamma^+$, which would imply the $\mathscr{S}$-derivability of $\Gamma^+ \Rightarrow \Delta^+$, against Lemma 9.

Case 2 : $\varphi := (s \neq t)$. If $(s \neq t) \in \Gamma^+$, by Lemma 9(8), $s = t \in \Delta^+$. By Lemma 9(2), $s = t \notin \Gamma^+$, therefore $[\![s]\!] \neq [\![t]\!]$ and $\mathscr{M} \vDash_{\mathsf{DM}} s \neq t$. If $s \neq t \in \Delta^+$, again by Lemma 9(8), $s = t \in \Gamma^+$ and so $\mathscr{M} \nvDash_{\mathsf{DM}} s \neq t$.

Case 3 : $\varphi := P(t_1, ..., t_n)$. By construction of $\mathscr{M}$, $\varphi \in \Gamma^+$ iff $([\![t_1]\!], ..., [\![t_n]\!]) \in P^+_{\mathscr{M}}$, that is $\mathscr{M} \vDash_{\mathsf{DM}} P(t_1, \ldots, t_n)$. If $P(t_1, ..., t_n) \in \Delta^+$, then by Lemma 9(3), $P(t_1, ..., t_n) \notin \Gamma^+$; therefore $([\![t_1]\!], ..., [\![t_n]\!]) \notin P^+_{\mathscr{M}}$.

Case 4 : $\varphi := \neg P(t_1, \ldots, t_n)$. Similar to the previous one.

Case 5 : $\varphi := \neg\neg\chi$. It's easy to check that $\varphi \in \Gamma^+$ iff $\chi \in \Gamma^+$ (otherwise $\Gamma^+, \chi \Rightarrow \Delta^+$ or $\Gamma^+, \neg\neg\chi \Rightarrow \Delta^+$ by Lemma 9, thus $\Gamma^+ \Rightarrow \Delta^+$ by the DM-rules). We can then apply the induction hypothesis.
If $\varphi \in \Delta^+$, then $\chi \in \Delta^+$ (otherwise $\Gamma^+ \Rightarrow \chi, \Delta^+$, and thus $\Gamma^+ \Rightarrow \Delta, \varphi$). Again we apply the induction hypothesis.

Case 6 : $\varphi := \exists v_i \chi$.
By construction of Γ^+, $\varphi \in \Gamma^+$ if and only if $\chi(c) \in \Gamma^+$ for suitable c. This is then equivalent to $\mathscr{M} \vDash_{\mathsf{DM}} \chi(c)$ and $\mathscr{M} \vDash_{\mathsf{DM}} \exists v_i \chi$ by induction hypothesis and definition of $\vDash_{\mathsf{DM}}$.
If $\varphi \in \Delta^+$, then $\chi(t/v_i) \in \Delta^+$ for any t (otherwise $\vdash_{\mathscr{S}} \Gamma^+ \Rightarrow \Delta^+, \chi(t/v_i)$, by Lemma 9 and $\vdash_{\mathscr{S}} \Gamma^+ \Rightarrow \Delta^+$ since $\mathscr{S} \vdash \chi(t) \Rightarrow \exists v \chi$). By induction hypothesis, $\mathscr{M} \nvDash_{\mathsf{DM}} \chi(t/v_i)$.

Case 7 : $\varphi := \neg\exists v_i \chi$. If $\varphi \in \Gamma^+$, then $\neg\chi(t/v_i) \in \Gamma^+$ for any t (as $\neg\chi(t) \notin \Gamma^+$ for some t entails $\mathscr{S} \vdash \Gamma^+, \varphi \Rightarrow \Delta^+$, thus $\Gamma^+ \Rightarrow \Delta^+$, quod non). Also, if $\neg\chi(t/v_i) \in \Gamma^+$ for any t, then $\neg\exists v_i \chi \in \Gamma^+$ (as if $\varphi \notin \Gamma^+$, then $\varphi \in \Delta^+$, thus $\neg\chi(c) \in \Delta^+$ for suitable $c \in C$). By induction hypothesis, this is equivalent to $\mathscr{M} \vDash_{\mathsf{DM}} \neg\chi(t)$ for any t and $\mathscr{M} \vDash_{\mathsf{DM}} \varphi$.
By definition of Δ^+, if $\varphi \in \Delta^+$, then $\neg\chi(c/v_i) \in \Delta^+$ for suitable $c \in C$. By induction hypothesis, $\mathscr{M} \nvDash_{\mathsf{DM}} \neg\chi(c/v_i)$. Thus $\mathscr{M} \nvDash_{\mathsf{DM}} \neg\exists v_i \chi$.

With Lemma 10 at hand one can check that all initial sequents in $\mathscr{S}$ are satisfied by $\mathscr{M}$.

This conclude the proof of Proposition 4. From the assumption $\nvdash_{\mathscr{S}} \Gamma \Rightarrow \Delta$ we constructed (Γ^+, Δ^+) and the term model $\mathscr{M}$. The reduct $\mathscr{M} \restriction \mathscr{L}$ of $\mathscr{M}$ to $\mathscr{L}$ is such that $\mathscr{M} \restriction \mathscr{L}$ is a model of $\mathscr{S}$ and

$$\text{for all } \varphi \in \Gamma,\ \mathscr{M} \restriction \mathscr{L} \vDash_{\mathsf{DM}} \varphi \text{ and } \mathscr{M} \restriction \mathscr{L} \nvDash_{\mathsf{DM}} \psi \text{ for all } \psi \in \Delta.$$

Thus $\Gamma \nvDash_{\mathscr{S}} \Delta$. This completes the proof of Proposition 4.

[17] The positive complexity of a formula φ of $\mathscr{L}$ is 0 for atomic and negated atomic sentences; and $n+1$ for sentences of the form $\neg\neg\varphi, \varphi \wedge \psi, \forall x \varphi$ with n the maximum of the complexities of φ and ψ.