THE JOURNAL OF SYMBOLIC LOGIC
Volume 71, Number 2, June 2006

AXIOMATIZING KRIPKE'S THEORY OF TRUTH

VOLKER HALBACH AND LEON HORSTEN

Abstract. We investigate axiomatizations of Kripke's theory of truth based on the Strong Kleene evaluation scheme for treating sentences lacking a truth value. Feferman's axiomatization KF formulated in classical logic is an indirect approach, because it is not sound with respect to Kripke's semantics in the straightforward sense; only the sentences that can be proved to be true in KF are valid in Kripke's partial models. Reinhardt proposed to focus just on the sentences that can be proved to be true in KF and conjectured that the detour through classical logic in KF is dispensable. We refute Reinhardt's Conjecture, and provide a direct axiomatization PKF of Kripke's theory in partial logic. We argue that any natural axiomatization of Kripke's theory in Strong Kleene logic has the same proof-theoretic strength as PKF, namely the strength of the system $\mathrm{RA}_{<\omega^\omega}$ ramified analysis or a system of Tarskian ramified truth up to ω^ω. Thus any such axiomatization is much weaker than Feferman's axiomatization KF in classical logic, which is equivalent to the system $\mathrm{RA}_{<\varepsilon_0}$ of ramified analysis up to ε_0.

Wovon man nicht sprechen kann, darüber muß man schweigen.

Adapted from Wittgenstein,
Tractatus logico-philosophicus 7

§1. Introduction: Reinhardt's interpretation of Kripke's theory of truth. There have been various attempts to block the inconsistency arising from the liar paradox by allowing sentences not to have a single classical truth value. In particular, on these accounts the liar sentence will not receive a single classical truth value true or false. The presence of sentences that lack a truth value, that have more than one truth value or that have a non-classical truth value requires an alternative to classical logic to deal with these sentences. Presumably the best known approach in this vein is Kripke's [24] theory developed in *Outline of a theory of truth*. Kripke studies partial models for a language containing a truth predicate T. The models are partial in the sense that sentences containing the truth predicate may fail to receive a truth value.

More precisely, by an inductive construction Kripke defined models of partial logic extending classical standard models where φ and $T\ulcorner\varphi\urcorner$ have the same truth

Received July 11, 2005.

Key words and phrases. Truth, Kripke-Feferman theory, axiomatic theories of truth, reflective closure, Strong Kleene scheme, partial logic, liar paradox.

Thanks to Vann McGee, Andrea Cantini and two anonymous referees for valuable comments and suggestions for improvement of this paper.

0022-4812/06/7102-0018/$4.60

value for all sentences φ, even if φ contains the truth predicate. φ and $T\ulcorner\varphi\urcorner$ will be either both true or both false or they will both lack a truth value.[1] In these models the liar sentence λ as well as $\neg\lambda$ and $T\ulcorner\lambda\urcorner$ lack a truth value. Kripke considered various options for assigning truth values to complex sentences containing sentences that lack a truth value. Among these, the Strong Kleene valuation scheme appeared to be very natural; in this article we shall focus on Kripke's account based on Strong Kleene logic.

For various reasons it has been thought to be desirable to have an axiomatization of Kripke's theory. We do not go over those reasons here; we just mention that they range from attempts to explicate our intuitive notion of truth to assigning the truth predicate a rôle in the foundations of mathematics (Cantini [7], Feferman [12] and McGee [28]; Halbach [19] gives an overview).

However, Kripke's theory does not lend itself very easily to a straightforward axiomatization. As the models obtained by Kripke are partial, a formal system describing Kripke's models would have to be formulated not in classical logic but in a formal system of partial logic. Although this is feasible (see, e.g., Kremer [23]), such systems have not become very popular; they were often rejected as unpractical and unsuitable for the rôle that had been assigned to them in larger research programs.

Kripke had already hinted at a uniform method for obtaining a classical (non-partial) model from any partial model.[2] These classical models lend themselves to an axiomatization in classical logic. The method for obtaining classical models from partial models is known as 'closing off'. In the simplest case, in order to 'close off' a partial model, one adds all sentences lacking a truth value in the partial model to the set of sentences that are false in the model. Whereas the interpretation of the truth predicate in the partial model partitions the set of sentences into three sets (true, false and undefined), the new 'closed off' model is purely classical because it joins the sentences false or undefined in the partial model into one set and therefore partitions the set of all sentences merely in two sets.

Relying on this method of turning partial models of Kripke's account into classical models, Feferman [12] devised an elegant and very natural looking formalization of Kripke's fixed point models in classical logic.[3] The resulting system is commonly known as KF. Natural models for KF are the 'closed off' versions of Kripke's fixed point models. Thus the axioms of KF—at least without axiom CONS below—describe a positive inductive definition.

KF and its variants were closely scrutinized by various logicians (e.g., Feferman [12], Cantini [5, 7]) and defended as philosophically significant accounts of

[1] Kripke started from acceptable structures in the sense of Moschovakis [32]. See Kripke [24] and McGee [28] for more on Kripke's theory of truth. We assume that the reader is somewhat familiar with Kripke's theory.

[2] See Kripke [24, p. 80-81] (The page numbers refer to the reprinted version of Kripke's paper.) Kripke denied explicitly that he is departing from classical logic even with his partial models. See footnote 18 of Kripke [24]. Especially this last point has puzzled scores of commentators. See, e.g., Visser [43, p. 640–642].

[3] Feferman gives Kripke's minimal fixed point model as the main example of a model for his theory. His theory, however, does not feature an axiom excluding any other fixed point. Cantini pointed out that Feferman's system can be interpreted as formalizing validity in all fixed points of the Kleene valuation scheme. See Cantini [5, p. 111, Proposition 5.8(i)].

truth (e.g., McGee [28, 29], and Soames [40]). But adopting classical logic for axiomatizations of Kripke's theory yields certain awkward and unintended consequences. We shall look at the drawbacks of the use of classical logic in axiomatizing Kripke's theory below in some detail.

Reinhardt has asked whether there is an instrumentalist justification of KF. In particular, he has posed the problem whether KF can be viewed as a tool for producing theorems that would also be derivable in a direct formalization of Kripke's original theory in partial logic, if one focuses on the sentences that are provably true in KF. We shall answer Reinhardt's Problem negatively.

This result is still compatible with a much weaker claim concerning the innocence of the move to classical logic: it does not rule out that KF and a direct axiomatization of truth in partial logic share many properties. In particular, one might conjecture that both serve a similar rôle in foundational issues. Even if KF and an axiomatization in partial logic differ in their truth-theoretic content, both may still yield the same non-semantic consequences; more precisely, they might yield the same consequences or at least consequences of the same proof-theoretic strength, if all sentences containing the truth predicate are neglected. We shall demonstrate that also this hope has to be given up. To this end we shall present an axiomatization PKF of Kripke's theory of truth in partial logic. We claim that any natural axiomatizations of Kripke's theory in Strong Kleene logic will be equivalent to our system PKF.

We determine the proof-theoretic strength of PKF as that of the system $\mathsf{RA}_{<\omega^\omega}$ of ramified analysis up to any ordinal level smaller than ω^ω. In contrast, KF is much stronger; Feferman [12] has established that KF is equivalent to the system $\mathsf{RA}_{<\varepsilon_0}$ of ramified analysis up to ε_0. This result shows that axiomatizing Kripke's theory in the most natural way leads to a system that is much weaker than the classical system KF. In particular, the arithmetical content of both theories is far from identical.

From this we conclude that there is no justification of KF that relies on a reduction of KF to a natural axiomatization of Kripke's theory in partial logic. This does not preclude the existence of a justification of KF by other means, but seen as a natural theory of truth PKF seems to be preferable over KF.

Our proof-theoretic analysis also sheds some light on the classification of axiomatizations of Kripke's theory with respect to other theories of truth. For instance, autonomously iterated theories of classical truth are much stronger than axiomatizations of Kripke's theory.

Note on notation. We assume that $\neg$, $\wedge$, $\vee$, $\forall$ and $\exists$ are our only primitive logical symbols, while all other logical symbols are defined. $=$ is the only predicate symbol of the language $\mathscr{L}_{\mathrm{PA}}$ of arithmetic. Adding the unary symbol T for the truth predicate to $\mathscr{L}_{\mathrm{PA}}$ yields the language $\mathscr{L}_{\mathrm{T}}$. In contrast to several other authors, we do not employ an additional primitive predicate for falsity of a sentence, which can be defined in our framework as the truth of the negation of the respective sentence.

If φ is a formula, $\ulcorner\varphi\urcorner$ will be used for the numeral of the Gödel number of φ. If $\varphi(x)$ contains the free variable x, $\ulcorner\varphi(\dot{x})\urcorner$ stands for the result of formally replacing the variable x in $\varphi(x)$ by the numeral with value x.

The notation here differs only slightly from Feferman [12]. $\mathrm{Var}(x)$, $\mathrm{ClTerm}(x)$, $\mathrm{Sent}(x)$ and $\mathrm{For}(x, v)$ are natural representations of the sets of codes of all variables, closed terms, all sentences of $\mathscr{L}_T$, and formulas with only the variable v free. The function val(x) takes (codes of) terms to what they denote; the value of a number that is not the code of a closed term is never the code of an expression of $\mathscr{L}_T$. $\underset{\cdot}{\neg}$ represents the function that sends the code of a formula to the code of its negation. Thus PA will prove $\underset{\cdot}{\neg}\ulcorner\varphi\urcorner = \ulcorner\neg\varphi\urcorner$ for all formulas of $\mathscr{L}_T$. Moreover, we assume that $\mathsf{PA} \vdash \forall x(\text{Sent}(\underset{\cdot}{\neg}x) \rightarrow \text{Sent}(x))$. Analogous remarks apply to other underdotted symbols like $\underset{\cdot}{\wedge}$, $\underset{\cdot}{\vee}$ or $\underset{\cdot}{T}$. In particular, $\underset{\cdot}{\forall}$ represents the function that takes the code of a variable and of a formula and yields (the code of) the universal quantification of the formula by the variable. $x(\dot{y}/v)$ is a function expression for a function that takes the codes of a formula, a number and a variable and gives the code of the formula where the numeral of y is substituted for the variable in the formula. Thus, for instance, we can express using this function expression as in KF8 below that a universally quantified sentence $\forall x\varphi(x)$ of $\mathscr{L}_T$ is true if and only if all of its instances $\varphi(\overline{n})$ are true. Here $\overline{n}$ is the numeral of n.

To simplify the presentation, we may assume that the language of Peano arithmetic contains function symbols for all these primitive recursive operations and the operations used for defining them. Of course, all required defining equations must be added as axioms to Peano arithmetic.

In order to fix the notation, we state some features of Kripke's construction. $\mathbb{N}$ is the standard model arithmetic. $(\mathbb{N}, S_1, S_2)$ is the partial expansion of $\mathbb{N}$ to the language $\mathscr{L}_T$, where T is assigned the extension S_1 and the antiextension S_2. We always assume $S_1 \cup S_2 \subseteq \omega$. The model $(\mathbb{N}, S_1, S_2)$ is partial in the sense that there might be sentences not in $S_1 \cup S_2$.

As mentioned above, we use the Strong Kleene scheme for the evaluation of complex sentences in a partial model (see also, for instance, McGee [28]). This is inductively described as follows:

(i) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi$ if φ is a true atomic or negated atomic arithmetical sentence,
(ii) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} T\,t$ iff t is a closed term with value $n \in S_1$,
(iii) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\, T\,t$ iff t is a closed term with value $n \in S_2$,
(iv) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\neg\varphi$ iff $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi$,
(v) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi \wedge \psi$ iff $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi$ and $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \psi$,
(vi) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg(\varphi \wedge \psi)$ iff $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\varphi$ or $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\psi$,
(vii) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi \vee \psi$ iff $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi$ or $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \psi$,
(viii) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg(\varphi \vee \psi)$ iff $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\varphi$ and $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\psi$,
(ix) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \forall x\varphi(x)$ iff for all $n \in \omega$ $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi(\overline{n})$,
(x) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\forall x\varphi(x)$ iff for at least one $n \in \omega$ $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\varphi(\overline{n})$,
(xi) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \exists x\varphi(x)$ iff for at least one $n \in \omega$ $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi(\overline{n})$,
(xii) $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\exists x\varphi(x)$ iff for all $n \in \omega$ $(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\varphi(\overline{n})$.

In particular, we have the following equivalences for all numerals $\overline{n}$ of the respective numbers n:

$$(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} T\,\overline{n} \text{ iff } n \in S_1,$$
$$(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\, T\,\overline{n} \text{ iff } n \in S_2.$$

If $n \notin S_1 \cup S_2$, then neither $T\,\overline{n}$ nor $\neg\, T\,\overline{n}$ holds in the partial model $(\mathbb{N}, S_1, S_2)$. In this case n is said to lack a truth value.

The operator Φ (the Kripke jump) is defined on pairs of sets (S_1, S_2) of numbers as follows, if $\overline{Sent}$ is the set of all numbers that are not sentences of $\mathscr{L}_{\mathrm{T}}$: $\Phi(S_1, S_2) =$

$$(\{\varphi \in \mathscr{L}_{\mathrm{T}} : (\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi\}, \overline{Sent} \cup \{\varphi \in \mathscr{L}_{\mathrm{T}} : (\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \neg\varphi\}).$$

That is, applied to a pair (S_1, S_2), Φ yields a pair of sets of sentences with the set of all sentences valid in (S_1, S_2) under the Strong Kleene scheme as first component; the second component is the union of the set of all non-sentences and the set of all sentences false in (S_1, S_2).

A model $(\mathbb{N}, S_1, S_2)$ is a *fixed point model* iff $\Phi(S_1, S_2) = (S_1, S_2)$. Kripke showed that there are many different fixed point models.

Fixed points of the operator Φ are characterized by the following attractive property of the interpretation of the truth predicate:

PROPOSITION 1 (Kripke). *$(\mathbb{N}, S_1, S_2)$ is a fixed point model if and only if the following holds*:

$$(\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} T\ulcorner\varphi\urcorner \text{ iff } (\mathbb{N}, S_1, S_2) \models_{\mathsf{SK}} \varphi.$$

As has been mentioned above, by 'closing off' one can convert a partial model $(\mathbb{N}, S_1, S_2)$ into the classical model $(\mathbb{N}, S_1)$. The latter is simply the expansion of the standard model $\mathbb{N}$ to the language $\mathscr{L}_{\mathrm{T}}$ where T is assigned the interpretation S_1. In this model $(\mathbb{N}, S_1)$, $\neg T\overline{n}$ holds for all $n \notin S_1$ whether they are in S_2 or not. Thus 'closing off' is the generation of a classical model by merging the set of false sentences with the set of all 'gappy' sentences.

These 'closed off' models are the intended models for the Kripke-Feferman theory KF. KF is formulated in classical logic. KF comprises all axioms of Peano arithmetic PA including all induction axioms in the language *with* the truth predicate. The truth-theoretic axioms of KF are given in the following list:

(KF1) $\forall x, y\big(\mathrm{ClTerm}(x) \wedge \mathrm{ClTerm}(y) \rightarrow (T\, x \underset{\cdot}{=} y \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$,

(KF2) $\forall x, y\big(\mathrm{ClTerm}(x) \wedge \mathrm{ClTerm}(y) \rightarrow (T \underset{\cdot}{\neg} x \underset{\cdot}{=} y \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y))\big)$,

(KF3) $\forall x\big(\mathrm{Sent}(x) \rightarrow (T \underset{\cdot}{\neg}\underset{\cdot}{\neg} x \leftrightarrow T\, x)\big)$,

(KF4) $\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\wedge} y) \leftrightarrow T\, x \wedge T\, y)\big)$,

(KF5) $\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \underset{\cdot}{\neg}(x \underset{\cdot}{\wedge} y) \leftrightarrow T \underset{\cdot}{\neg} x \vee T \underset{\cdot}{\neg} y)\big)$,

(KF6) $\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T(x \underset{\cdot}{\vee} y) \leftrightarrow T\, x \vee T\, y)\big)$,

(KF7) $\forall x \forall y\big(\mathrm{Sent}(x) \wedge \mathrm{Sent}(y) \rightarrow (T \underset{\cdot}{\neg}(x \vee y) \leftrightarrow T \underset{\cdot}{\neg} x \wedge T \underset{\cdot}{\neg} y)\big)$,

(KF8) $\forall v \forall x\big(\mathrm{Var}(v) \wedge \mathrm{For}(x, v) \rightarrow (T \underset{\cdot}{\forall} v x \leftrightarrow \forall y\, T\, x(\dot{y}/v))\big)$,

(KF9) $\forall v \forall x\big(\mathrm{Var}(v) \wedge \mathrm{For}(x, v) \rightarrow (T \underset{\cdot}{\neg} \underset{\cdot}{\forall} v x \leftrightarrow \exists y\, T \underset{\cdot}{\neg} x(\dot{y}/v))\big)$,

(KF10) $\forall v \forall x\big(\mathrm{Var}(v) \wedge \mathrm{For}(x, v) \rightarrow (T \underset{\cdot}{\exists} v x \leftrightarrow \exists y\, T\, x(\dot{y}/v))\big)$,

(KF11) $\forall v \forall x\big(\mathrm{Var}(v) \wedge \mathrm{For}(x, v) \rightarrow (T \underset{\cdot}{\neg} \underset{\cdot}{\exists} v x \leftrightarrow \forall y\, T \underset{\cdot}{\neg} x(\dot{y}/v))\big)$,

(KF12) $\forall x\big(\mathrm{ClTerm}(x) \rightarrow (T\, \underset{\cdot}{T} x \leftrightarrow T\, \mathrm{val}(x))\big)$,

(KF13) $\forall x\big(\mathrm{ClTerm}(x) \rightarrow (T \underset{\cdot}{\neg} \underset{\cdot}{T} x \leftrightarrow (T \underset{\cdot}{\neg}\, \mathrm{val}(x) \vee \neg\, \mathrm{Sent}(\mathrm{val}(x))))\big)$,[4]

[4]Our version varies from other formulations of KF in axiom KF13, which says that $\neg\, T\, t$ (t a closed term) is true if and only if one of the two following conditions is satisfied:

(i) The negation of the value of t is true. (If this condition is satisfied, the value of t is a sentence of $\mathscr{L}_{\mathrm{T}}$ according to our assumption $\mathsf{PA} \vdash \forall x(\mathrm{Sent}(\underset{\cdot}{\neg} x) \rightarrow \mathrm{Sent}(x))$ and KF14).

(ii) The value of t is not a sentence of $\mathscr{L}_{\mathrm{T}}$.

In other versions the second condition is suppressed. Although our axiom KF13 makes the axioms slightly less elegant, it avoids problems in other places. However, not much depends on the exact formulation of this axiom and we could still prove our mains results if condition (ii) were dropped.

(KF14) $\forall x\big(T\,x \rightarrow \mathrm{Sent}(x)\big)$,
(Cons) $\forall x\big(\mathrm{Sent}(x) \rightarrow \neg(T\,x \wedge T\,\underset{\cdot}{\neg}x)\big)$.

There are several alternative versions of KF. Feferman's [12] own formulation appeared in print after other variants of KF had been published. His version lacks CONS; Feferman uses the label Ref(PA) for his system. As far as we know, Reinhardt [36] was the first to call the system KF. Cantini [5] uses the designation for a system similar to our KF but with restricted induction.

The quantifier axioms KF8–KF11 state that the truth of a quantified sentence depends on the truth or falsity of its numerical instances; as has been already explained, the antecedent $\mathrm{Var}(v) \wedge \mathrm{For}(x, v)$ expresses that v is a variable and x is a formula of $\mathscr{L}_{\mathrm{T}}$ with at most v free. Thus according to axiom KF8 a universally quantified sentence $\forall z\varphi$ is true if and only if for all *numerals* $\varphi(\overline{n})$ is true. One might want to strengthen this condition by requiring that for all closed terms t (not only numerals) $\varphi(t)$ is true. However, this would not strengthen the axiom, because one can show in KF that the truth or falsity of a sentence φ does not depend on the particular shape of the terms in φ but only on their value:

LEMMA 2.

$$\begin{aligned}\mathsf{KF} \vdash \forall x\forall y\forall z\forall v\,(\,&\mathrm{ClTerm}(x) \wedge \mathrm{ClTerm}(y) \wedge \mathrm{For}(z, v) \wedge \mathrm{val}(x) = \mathrm{val}(y)\\ &\rightarrow (T\,z(x/v) \leftrightarrow T\,z(y/v))).\end{aligned}$$

PROOF. The lemma is proved by an induction on the complexity of the formula coded by z in KF. The required instance of the induction scheme contains an instance of T. ⊣

Next we state two observations concerning the soundness of KF with respect to its intended semantics. In the first place we note that KF is sound with respect to 'closed off' models:

PROPOSITION 3 (Feferman, Reinhardt). *If* KF $\vdash \varphi$, *then* φ *holds in all* (*classical*) *closed off fixed point models.*

Therefore KF is sound with respect to Kripke's original theory in the following sense:

PROPOSITION 4. *If* KF $\vdash T\ulcorner\varphi\urcorner$, *then* φ *holds in all fixed point models.*

The axiomatization of a partial notion of truth in classical logic (or the 'closing-off' of Kripke's partial models) yields several oddities, which make KF in itself hardly acceptable as a natural and intuitively plausible theory of truth. In particular, KF proves that some of its own theorems are not true. KF, so to speak, disproves its own soundness. An example of such a theorem is the liar sentence λ:

REMARK 5. KF proves λ and $\neg\,T\ulcorner\lambda\urcorner$.

PROOF. The liar sentence is of the form $\neg\,T\,l$ where l is a closed term with PA $\vdash l = \ulcorner\neg\,T\,l\urcorner$.

$$\begin{aligned}\mathsf{KF} \vdash T\,l &\rightarrow T\ulcorner\neg\,T\,l\urcorner\\ &\rightarrow T\,\underset{\cdot}{\neg}l \qquad\qquad \text{KF13 and KF14}\end{aligned}$$

From CONS we also have $T\,l \rightarrow \neg\,T\,\underset{\cdot}{\neg}l$. Therefore KF proves $\neg\,T\,l$, that is, KF proves the liar sentence λ. Since PA $\vdash l = \ulcorner\neg\,T\,l\urcorner$ and λ is $\neg\,T\,l$, KF proves also $\neg\,T\ulcorner\lambda\urcorner$. ⊣

CONS is essential in the proof that KF proves the liar sentence λ. In §2 we shall argue that dropping CONS is not an attractive option either.

Obviously, the provability of the liar sentence renders KF implausible as a natural formalization of Kripke's theory. The attractiveness of Kripke's theory is due to the fact that it does not make the liar sentence true or false.[5]

Apart from these well known drawbacks of KF, the proof of the liar sentence in KF reveals that the notion of truth axiomatized by KF is not compositional. The truth of the liar sentence in KF does not supervene on the truth of a subsentence or somehow subordinated sentence; rather it is an artifact of the axiomatization of Kripke's theory of partial truth in classical logic.

Reinhardt [36, p. 242–243] pointed out that the status of the liar sentence in KF and several other odd consequences of the use of classical logic make the system KF thoroughly unsatisfactory as it stands, for it proves sentences (such as λ) which by its own lights are untrue. Reinhardt argued that one ought to unwaveringly adhere to the partial party line.[6] On the partial account, λ can be neither asserted nor denied, because it is neither true nor false. So an axiomatization of this account should prove neither λ (or, equivalently, $\neg T\ulcorner\lambda\urcorner$) nor $\neg\lambda$. In view of this, Reinhardt considered the theory

$$\mathsf{IKF} = \{\varphi : \mathsf{KF} \vdash T\ulcorner\varphi\urcorner\}.$$

This theory IKF is called the *inner logic* of KF. The unattractive aspect of KF that was discussed above can be generalized by saying that the *outer logic* of KF (i.e., KF itself) differs from its inner logic.[7] Reinhardt thought that IKF captures the thoroughly partial core of Kripke's theory of truth, which is underlined by the soundness result, that is, Proposition 4. At any rate, it is free of the problem that was seen to haunt KF, for we see from KF12 that for all sentences φ the following holds:

$$\varphi \in \mathsf{IKF} \text{ iff } T\ulcorner\varphi\urcorner \in \mathsf{IKF}.$$

Moreover, if a sentence φ is an element of IKF, then φ is true in all fixed point models. Reinhardt saw KF as no more than an instrument, as a machine for generating the honest-to-God theorems of IKF. He put this as follows:

> "The case of Hilbert's program is instructive. Hilbert viewed (or sought to view) infinitary mathematics (at some level) as a purely formal exercise; the only contentful mathematics was finitary. Hilbert's program attempted to justify the use of non-contentual mathematics. [...] Hilbert's formalism about infinitary mathematics is analogous to the proposal that is made in this paper about meaningful but nonsignificant sentences [i.e., sentences $\varphi \in \mathsf{KF}$ which do not belong to IKF].

[5]The fact that a theory leaves the liar sentence λ or its truth undecided does not imply that the theory suggests that the liar sentence is not true. Most sensible theories of truth over PA do not prove that, e.g., the consistency statement of ZF is true. This should not be taken as an evidence that the theory somehow suggests that the consistency statement of ZF is not true. Therefore we reject Glanzberg's criticism [15, p. 115] of the partial approach.

[6]In recent years, this interpretation was defended by Soames [40, chapter 6] who at the time of writing of his book must have been unaware of Reinhardt's work.

[7]Several authors have emphasized that the identity of the inner and outer logic of a system of truth is a desirable feature. One of them is Michael Sheard [39, p. 175-176].

There is also an analogue of Hilbert's program here: to justify the use of nonsignificant sentences entirely within the framework of significant sentences. I would like to suggest that the chances of success in this context, where the interpreted or significant part of the language includes such powerful notions as truth, are somewhat better than in Hilbert's context, where the contentual part was very restricted." [36, p.225].

With this program in the background, Reinhardt [35, p. 239] posed:

Reinhardt's Problem. *Is there for every* KF*-theorem of the form* $T\ulcorner\varphi\urcorner$*, a* KF*-proof*

$$\varphi_1, \ldots, \varphi_n, \varphi$$

such that for each $1 \leq i \leq n$, $\mathsf{KF} \vdash T\ulcorner\varphi_i\urcorner$?

Reinhardt's program and Hilbert's program differ in some respects essentially: The ideal and the real statements are syntactically distinguished in the latter case, while the inner and the outer logic of KF are formulated in exactly the same language. Also we do not want to go into the discussion whether a purely instrumentalist understanding of Hilbert's program is adequate. At any rate it is not hard to see, however, that Reinhardt's analogue of Hilbert's program suffers the same fate as that of Hilbert's program.[8] Theorem 8 shows that in many cases, the detour via 'ideal' statements to prove 'real' theorems of KF is essential. In order to prove Theorem 8, we need two lemmata.

Lemma 6. $\mathsf{KF} \vdash T\ulcorner\varphi\urcorner \leftrightarrow \varphi$ *holds for all* $\varphi \in \mathscr{L}_{\mathrm{PA}}$.

Proof. This standard lemma is proved by a meta-induction on the buildup of φ. Actually we prove a 'uniform' version of the lemma, i.e., a version with free variables in φ:

$$\forall x_1 \ldots \forall x_n (T\ulcorner\psi(\dot{x}_1, \ldots, \dot{x}_n)\urcorner \leftrightarrow \psi(x_1, \ldots, x_n)). \tag{1}$$

The uniform version is useful for proving the induction step. Assume as induction hypothesis that (1) has been established. We aim to show the claim for $\forall x_n \psi(x_1, \ldots, x_n)$ and proceed as follows:

$$\begin{aligned} \mathsf{KF} \vdash T\ulcorner\forall x_n \psi(\dot{x}_1, \ldots, \dot{x}_{n-1}, x_n)\urcorner &\leftrightarrow \forall x_n\, T\ulcorner\psi(\dot{x}_1, \ldots, \dot{x}_n)\urcorner && \mathsf{KF8} \\ &\leftrightarrow \forall x_n \psi(x_1, \ldots, x_n) && (1) \end{aligned}$$

The cases of atomic arithmetical formulas and other complex formulas are trivial. ⊣

Lemma 7. *Assume that* $\varphi_1, \ldots \varphi_n, \varphi$ *is a* KF*-proof of* $\varphi \in \mathscr{L}_{\mathrm{PA}}$ *such that for each* $1 \leq i \leq n$, $\mathsf{KF} \vdash T\ulcorner\varphi_i\urcorner$. *Then* $\mathsf{PA} \vdash \varphi$.

Proof. It is not hard to see that none of the axioms KF3–KF13 and CONS holds in a partial fixed point model. For instance, the liar sentence λ lacks a truth value in all fixed point models, as does $T\ulcorner\lambda\urcorner$ and $T\ulcorner\neg\neg\lambda\urcorner$. Therefore

$$T\ulcorner\neg\neg\lambda\urcorner \leftrightarrow T\ulcorner\lambda\urcorner$$

does not receive a truth value in any fixed point model. Consequently KF3 does not hold in any such model by the soundness of KF, that is, Proposition 4. Thus φ_i $(1 \leq i \leq n)$ is not KF3. Similarly all other axioms KF3–KF13 and CONS

[8] Although it took awhile for *us* to see this.

cannot occur in the proof $\varphi_1, \dots, \varphi_n, \varphi$. The only axioms in the proof that involve the truth predicate may be KF1, KF2, certain induction axioms and KF14. Given such a proof, replace any occurrence of T by Tr_0 in this proof, where Tr_0 is the usual truth definition in $\mathscr{L}_{\mathrm{PA}}$ of atomic sentences of $\mathscr{L}_{\mathrm{PA}}$. It is not hard to check that KF1, KF2, all induction axioms and KF14 become theorems of PA. Thus the resulting structure of sentences can easily be turned into a proof in PA by adding some further subproofs. For instance, the translations of KF1 and KF2 are not themselves axioms of PA but merely theorems of PA; thus their proofs must be added in order to convert the translation of a KF-proof into a proof in PA. $\dashv$

This lemma solves Reinhardt's Problem:

THEOREM 8. *For some* $\varphi \in$ IKF, *there exists no proof* $\varphi_1, \dots, \varphi_n, \varphi$ *of* KF *such that for all* $i \leq n$, $\varphi_i \in$ IKF.

PROOF. KF is not conservative over PA. Feferman [12] has determined the proof-theoretic strength of KF: it is arithmetically equivalent to the system $\mathsf{RA}_{<\varepsilon_0}$ of ramified analysis up to ε_0. For our purposes a mundane example will do: $\mathsf{KF} \vdash \mathsf{Con}_{\mathsf{PA}}$. From Lemma 6 we obtain $\mathsf{KF} \vdash T\ulcorner\mathsf{Con}_{\mathsf{PA}}\urcorner$. Thus $\mathsf{Con}_{\mathsf{PA}} \in$ IKF holds, but since $\mathsf{PA} \nvdash \mathsf{Con}_{\mathsf{PA}}$ by Gödel's theorem, there is no proof $\varphi_1, \dots, \varphi_n, \mathsf{Con}_{\mathsf{PA}}$ of KF such that for all $i \leq n$, $\varphi_i \in$ IKF because of the preceding lemma. $\dashv$

In fact, this theorem can be strengthened somewhat.[9] Let us define *strict* KF (SKF) to be just like KF, except that the truth predicate is not allowed to appear in the induction axioms. The inner logic of SKF (strict IKF) is denoted by SIKF. It is argued in McGee [29] that SKF is an interesting theory, particularly for disquotationalists.

SKF is a conservative extension of PA (see Cantini [5]). Thus $\mathsf{Con}_{\mathsf{PA}}$ is not a theorem of SKF. So if we want to prove Theorem 8 for SKF instead of for KF, we cannot use $\mathsf{Con}_{\mathsf{PA}}$. Nevertheless, another example is not hard to find:

THEOREM 9. *For some* $\varphi \in$ SIKF, *there exists no proof* $\varphi_1, \dots, \varphi_n, \varphi$ *of* SKF *such that for all* $i \leq n$, $\varphi_i \in$ SIKF.

PROOF. We can use the same interpretation as in Lemma 7. That is, we replace in any such proof the truth predicate by Tr_0. This shows also that $T\ulcorner\forall x(x = x)\urcorner$ is not provable in KF by a proof consisting entirely of members of IKF, even though $\forall x(x = x)$ is in SIKF. $\dashv$

Yet another theory that one might consider is the version of KF where the truth axioms are formulated as *axiom schemes*. If for this theory one considers the corresponding inner logic and poses Reinhardt's question, the answer is not so straightforward. In comparison to full KF, schematic KF is again quite weak: it is not hard to see that schematic KF is arithmetically conservative over PA. The proof of Theorem 9, however, carries over to schematic KF as well.

§2. The consistency axiom. The consistency axiom CONS enabled us to prove the liar sentence in KF and, consequently, to prove in KF that the liar sentence is not true. This asymmetry between inner and outer logic provided a strong reason for rejecting KF itself as a plausible theory of truth and for concentrating on the internal logic of KF. Now one might surmise that we too hastily rejected

[9]This strengthening was suggested to us by Vann McGee. McGee has proved Theorem 8 independently (but he has not published it).

classical logic and that dropping CONS from the list of axioms might already dissolve our worries about KF. Moreover, CONS differs in several respects from the axioms KF1–KF13 and has a peculiar status among the other truth-theoretic axioms of KF. Consequently, some authors have formulated KF without CONS. In particular McGee [29] advocates KF without CONS as a theory of truth that should be useful to the disquotationalist.

We shall first elaborate on the special status of CONS before arguing that it should be retained and that dropping CONS does not make KF more attractive as a theory of truth.

One of the special features of CONS is that does not form part of a positive inductive 'definition' of truth (or rather the axiomatic counterpart of such a definition) like the other axioms KF1–KF13. For the following discussion and future reference we provide alternative formulations of CONS. In particular, we show that CONS is equivalent to 'uniform T-reflection', that is, to the scheme $\forall x\big(T\ulcorner\varphi(\dot{x})\urcorner \rightarrow \varphi(x)\big)$:

Lemma 10. *Over* KF–CONS, *that is, over* KF *without axiom* CONS, *the following are equivalent*:

(i) CONS, *that is,* $\forall x\big(\mathrm{Sent}(x) \rightarrow \neg(T\,x \wedge T\,\dot{\neg}x)\big)$,
(ii) $\forall x\big(\mathrm{Sent}(x) \rightarrow (T\,\dot{\neg}x \rightarrow \neg\,T\,x)\big)$,
(iii) *the schema* $\forall\vec{x}(T\ulcorner\varphi(\dot{\vec{x}})\urcorner \rightarrow \varphi(\vec{x}))$ *for all formulas* $\varphi(\vec{x})$ *of* $\mathscr{L}_{\mathrm{T}}$; $\vec{x}$ *stands here for a string* $x_1, \dots, x_n$ *of variables,*
(iv) *the schema* $\forall x\big(T\ulcorner\varphi(\dot{x})\urcorner \rightarrow \varphi(x)\big)$ *for all formulas* $\varphi(x)$ *of* $\mathscr{L}_{\mathrm{T}}$; *this schema allows only one free variable in the respective instantiating formula,*
(v) $\forall x(T\ulcorner\neg\,T\,\dot{x}\urcorner \rightarrow \neg\,T\,x)$.

Of course, finite subsets of the set of all instances of (iii) and (iv), respectively, will suffice for deriving (i), (ii) and (v) in KF–CONS.

Proof. (i) logically implies (ii).

It is known that KF yields the schema (iii). Cantini [7, p. 54, Theorem 8.8(i)] proved this implication for a closely related system. The following argument for (iii) from (ii) follows this proof.

The implication from (ii) to (iii) is proved simultaneously for $\varphi(\vec{x})$ and $\neg\varphi(\vec{x})$ by meta-induction on the complexity of $\varphi(\vec{x})$.

If $\varphi(\vec{x})$ is an atomic arithmetical formula, KF1 and KF2 yield the claim. In the case where φ is $\psi \wedge \chi$ (we suppress free variables), the claim is established in the following way:

$$\begin{aligned} \text{KF-CONS} + \text{(ii)} \vdash T\ulcorner\psi \wedge \chi\urcorner &\rightarrow T\ulcorner\psi\urcorner \wedge T\ulcorner\chi\urcorner && \text{KF4} \\ &\rightarrow \psi \wedge \chi && \text{induction hypothesis} \end{aligned}$$

The case $\neg\varphi$ is treated in the following way:

$$\begin{aligned} \text{KF-CONS} + \text{(ii)} \vdash T\ulcorner\neg(\psi \wedge \chi)\urcorner &\rightarrow T\ulcorner\neg\psi\urcorner \vee T\ulcorner\neg\chi\urcorner && \text{KF5} \\ &\rightarrow \neg\psi \wedge \neg\chi && \text{induction hypothesis} \end{aligned}$$

If φ is of the form $T\,t$ where t is a term (again we suppress free variables in t), the claim follows directly by instantiating axiom KF14:

$$\text{KF-CONS} + \text{(ii)} \vdash T\ulcorner T\,t\urcorner \rightarrow T\,t$$

Only in order to prove the claim for $\neg T t$, we need (ii) (or axiom CONS):

$$\begin{aligned} \mathsf{KF\text{-}CONS} + \text{(ii)} \vdash T\ulcorner \neg T t \urcorner &\rightarrow T \underset{\cdot}{\neg} t \vee \neg \mathrm{Sent}(t) && \mathsf{KF13} \\ &\rightarrow \neg T t && \text{(ii) and } \mathsf{KF14} \end{aligned}$$

All instances of the schema in (iv) are also instances of the schema in (iii). Therefore (iv) follows trivially from (iii).

Similarly, (v) is an instance of the schema (iv).

In order to show that (i) can be derived from (v), one can reason as follows in KF with CONS replaced by the schema (v).

$$\begin{aligned} \exists x(\mathrm{Sent}(x) \wedge T x \wedge T \underset{\cdot}{\neg} x) &\rightarrow \exists x(T x \wedge T \underset{\cdot}{\neg} \underset{\cdot}{T} \dot{x}) && \mathsf{KF13} \\ &\rightarrow \exists x(T x \wedge \neg T x) && \text{(v)} \end{aligned}$$

Since the succedent in the last line is a logical contradiction, the negation of $\exists x(\mathrm{Sent}(x) \wedge T x \wedge T \underset{\cdot}{\neg} x)$ follows, that is, CONS. ⊣

The axioms KF4, KF6, KF8 and KF10 claim that T commutes with conjunction, disjunction, the universal and the existential quantifier. (ii) of the lemma, that is,

$$\mathsf{KF} \vdash \forall x\big(\mathrm{Sent}(x) \rightarrow (T \underset{\cdot}{\neg} x \rightarrow \neg T x)\big)$$

shows that CONS yields one half of the equivalence that expresses that T commutes also with negation. Adding the full equivalence, that is, the claim that T commutes with negation, would render KF inconsistent. Thus, while all axioms KF1–KF13 are in a sense compositional, CONS is not a compositional axiom (cf. Halbach [18] and below).

Part (iii) and (iv) of the Lemma 10 show that CONS forces one direction of the T-sentences: KF proves $T\ulcorner\varphi\urcorner \rightarrow \varphi$ for all sentences φ of $\mathscr{L}_{\mathrm{T}}$. Therefore we cannot consistently also have $\varphi \rightarrow T\ulcorner\varphi\urcorner$ in any consistent extension of KF. As a side effect of this asymmetry KF proves the liar sentence, as has been noted in Remark 5. Thus CONS forces the inner and the outer logic of KF to be different. Even adding to KF a rule that implies that the inner logic contains the outer logic renders KF inconsistent; such a rule of 'necessitation' would allow us to conclude $T\ulcorner\varphi\urcorner$ from φ. If this rule were added, CONS would allow us to prove $T\ulcorner\varphi\urcorner \rightarrow \varphi$ for all sentences by Lemma 10(iii); and this reflection scheme is inconsistent with the above mentioned rule of necessitation by Montague's theorem (see Montague [30]).

All this seems to support the view that CONS should be excluded from the list of the KF-axioms. When CONS is abandoned, truth value gluts, that is, sentences that are both true and false are admitted. CONS excludes truth value gluts by saying that no sentence is true together with its negation. Dropping this postulate no longer rules out such gluts. Several authors—most prominently Visser [43]—have expanded Kripke's theory to a four-valued logic, more precisely, they allow the extension and antiextension of the truth predicate to overlap. Cantini [5] has investigated KF without CONS and its relation with fixed point models with gluts *and* gaps.

Although several annoying pathologies are removed from KF by dropping CONS, some desirable features of KF are also lost. Therefore we prefer to formulate KF with axiom CONS.

One reason for keeping CONS is that the truth predicate no longer distributes over the material conditional in the absence of CONS. That is, the use of CONS is essential in establishing

$$\mathsf{KF} \vdash T\ulcorner \varphi \to \psi \urcorner \to (T\ulcorner \varphi \urcorner \to T\ulcorner \psi \urcorner)$$

for any two sentences φ and ψ of $\mathscr{L}_{\mathrm{T}}$. Material implication is defined here in the usual way: $\varphi \to \psi$ stands for $\neg\varphi \vee \psi$. Without CONS we could not even prove the corresponding rule.

There is another, more important reason for keeping CONS: We do not think that CONS is the real source of the above mentioned asymmetries; the real source is rather the axiomatization of a partial notion of truth in classical logic. Dropping CONS from the list of the KF-axioms does not render the inner and outer logic identical; the inner and the outer logic will be still different. In particular, the inner logic still will not comprise all classical tautologies. But KF does not only fall short of proving the same theorems in its inner and outer logic: one cannot even consistently postulate in the context of KF that the inner and outer logic are identical. For in order to force the identity of inner and outer logic one might try to add the following two rules for all sentences $\varphi \in \mathscr{L}_{\mathrm{T}}$:

$$\frac{\varphi}{T\ulcorner \varphi \urcorner}\mathsf{NEC} \qquad \frac{T\ulcorner \varphi \urcorner}{\varphi}\mathsf{CONEC}$$

Adding these two rules to KF yields an inconsistency even if CONS is dropped from the list of axioms.

Lemma 11. KF *without axiom* CONS *but with the two rules* NEC *and* CONEC *is inconsistent*.

Proof. If λ is the liar sentence with $\mathsf{PA} \vdash \lambda \leftrightarrow \neg T\ulcorner \lambda \urcorner$, the following two sentences are theorems of PA:

$$\lambda \vee T\ulcorner \lambda \urcorner,$$
$$\neg\lambda \vee \neg T\ulcorner \lambda \urcorner.$$

By applying NEC to these two sentences and distributing the truth predicate over $\vee$ with KF6, respectively, we obtain the following sentences:

(2) $$T\ulcorner \lambda \urcorner \vee T\ulcorner T\ulcorner \lambda \urcorner \urcorner,$$

(3) $$T\ulcorner \neg\lambda \urcorner \vee T\ulcorner \neg T\ulcorner \lambda \urcorner \urcorner.$$

KF12 can be used for simplifying (2) to $T\ulcorner \lambda \urcorner$, while (3) yields $T\ulcorner \neg\lambda \urcorner$ with KF13. From

$$T\ulcorner \lambda \urcorner \wedge T\ulcorner \neg\lambda \urcorner$$

we obtain using KF4 the sentence

$$T\ulcorner \lambda \wedge \neg\lambda \urcorner.$$

An application of CONEC yields an inconsistency. ⊣

In sum, not only is KF without CONS not closed under NEC and CONEC: it *cannot* be consistently closed under these rules. So CONS is not the only source of the mentioned asymmetries. CONS might highlight them, but the KF is essentially asymmetric because it is formulated in classical logic, while its internal logic is partial. So removing CONS from KF does not really solve the problems for KF that Reinhardt and others have pointed out.

§3. **Reinhardt's Challenge.** Following Reinhardt, we shall focus on deductive theories that are valid in partial fixed point models. KF is not sound with respect to partial fixed point models; only Reinhardt's instrumentalist interpretation could render it useful for our purposes. Concentrating on IKF, the inner logic of KF, initially sounds like a promising idea for obtaining a suitable deductive theory. From the definition of IKF it follows that IKF is a recursively enumerable set of sentences, so it is recursively axiomatizable by Craig's method [9].

The acceptability of this theory, however, relies on the acceptability of the classical system KF. One could draw a dramatic parallel to reductive programs in the foundations of mathematics again: One could take the set of arithmetical sentences provable in Zermelo-Fraenkel set theory as one's theory of arithmetic (using the usual embedding of arithmetic in set theory); this could hardly be considered as a satisfactory axiomatization of arithmetic, because such an axiomatization ought to be *direct* and not dependent on a theory of sets.

In a similar way IKF is not a satisfactory axiomatization of Kripke's theory of truth, because it relies on a system that is not directly sound with respect to Kripke's theory. IKF in combination with Craig's trick does not produce a list of sentences that are even *candidates* for being basic principles of truth.[10] Therefore Reinhardt [35, p. 239] put it as a *challenge* to find a *natural* axiomatization of IKF which is not obviously parasitic on the untrustworthy system KF.

In this paper, we provide a *direct* axiomatization of Kripke's theory. This implies that our system is formulated in *partial logic*, and not in classical logic as KF.[11] The central but necessarily tentative claim of the present paper is that this is *the* natural way to axiomatize Kripke's theory of truth with the Strong Kleene scheme over Peano arithmetic. We take this as evidence that Reinhardt's Challenge cannot be met.

The uniqueness claim that our system is actually the *only* natural formalization of Kripke's theory is to be understood up to reaxiomatization. That is, any natural axiomatization of Kripke's theory in partial logic will yield the same theorems. Actually for our purposes a weaker claim would suffice. We are interested in a comparison of our system with the classical alternative KF and with commonly considered subsystems of Second-Order arithmetic. Since we are interested in the proof-theoretic strength of the natural axiomatization of Kripke's theory, we only need to claim uniqueness up to proof-theoretic equivalence. Which notion of proof-theoretic equivalence is applicable will become evident below.

Of course we cannot mathematically prove our uniqueness claim, because the notion of a *natural* axiomatization has not been mathematically captured. We hope that our system will make our claim plausible, but we shall not tire the reader by providing alternative approaches and subsequently proving that the systems yield actually the same theorems.

By axiomatizing Kripke's theory of truth in partial logic, we will ensure that, in accordance with Reinhardt's strictures, the inner logic and the outer logic of the resulting system coincide.

[10]The artificiality of Craigian theories is pointed out in Hempel [21, section 9].

[11]Maudlin [27] too argues on philosophical grounds, similar to the ones that are advanced here, that Kripke's theory of truth ought to be expressed in partial (Strong Kleene) logic. But he does not carry out this project in all formal details. In particular, he does not express the theory in an arithmetical setting.

In the following sections we describe our formal system for generating sentences that hold in all fixed point models in Strong Kleene logic.

§4. Strong Kleene logic and arithmetic in the sequent calculus. We employ a sequent calculus for Strong Kleene logic. Our system is Scott's [38] with slight modifications; Blamey [4] presents a very similar system. For an alternative approach see Aoyama's [1]. Later we shall add specific rules for arithmetic and truth.

The system can be rewritten as a natural deduction system. The formulas preceding the sequent arrow $\Rightarrow$ are then treated as assumptions, and the sentences in the succedent are disjunctively joined into one conclusion. The natural deduction version may be more natural, because it can dispense with the sequent arrow $\Rightarrow$, which is much different from the conditional $\rightarrow$ of the object language, but for technical purposes the sequent calculus is preferable. We will return to the comparison with natural deduction below.

4.1. Logic. Sequents are conceived as given by a pair Γ and Δ of finite sets of formulas. The sequent is written as $\Gamma \Rightarrow \Delta$. If Γ and Δ are sets of sentences and $\Gamma \Rightarrow \Delta$ is derivable, our system PKF is sound in the sense that if all sentences in Γ are true in a partial model, then at least one sentence in Δ is true in that model, and if all sentences in Δ are false in a partial model, then at least one sentence in Γ is false in the model. Truth and falsity in a partial model are determined by Strong Kleene logic. The initial sequents and rules are also complete with respect to Strong Kleene logic. We shall return to this issue at the end of this section.

4.1.1. *Structural rules and initial sequents.* All sequents of one of the following forms are initial sequents:

(IN) $\quad \Gamma \Rightarrow \Delta$, where $\Gamma \cap \Delta \neq \emptyset$

(weakening 1) $$\frac{\Gamma \Rightarrow \Delta}{\Gamma, \varphi \Rightarrow \Delta}$$

(weakening 2) $$\frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \varphi, \Delta}$$

(cut) $$\frac{\Gamma \Rightarrow \varphi, \Delta \qquad \Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

4.1.2. *Laws of truth values.* Scott uses constants for the truth values. For our purposes it is more convenient not to expand the language. $\top$ is the sentence $0{=}0$, $\bot$ the sentence $0 = 1$ and λ is the liar sentence (that is a sentence that is "gappy" under the intended interpretation). The following sequents are then initial sequents:

($\top$-sequent) $\quad \Rightarrow \top$

($\bot$-sequent) $\quad \bot \Rightarrow$

(λ-sequents) $\quad \lambda \Leftrightarrow \neg\lambda$

In the last line and in the following the double arrow indicates that both, $\lambda \Rightarrow \neg\lambda$ and $\neg\lambda \Rightarrow \lambda$ are initial sequents. This convention will also be applied below.

4.1.3. *Laws of negation.* If Γ is a set of sentences, $\neg\Gamma$ designates the set of all negations of sentences in Γ.

(¬-rule) $$\frac{\Gamma \Rightarrow \Delta}{\neg\Delta \Rightarrow \neg\Gamma}$$

(¬¬-sequents) $$\varphi \Leftrightarrow \neg\neg\varphi$$

(¬λ-sequents) $$\varphi, \neg\varphi \Rightarrow \lambda$$

Not all *ex falso quodlibet*-sequents $\varphi, \neg\varphi \Rightarrow$ are sound, because if φ lacks a truth value then all formulas in the consequent are false (there are none), but at the same time none of the (two) formulas in the antecedent is false. The $\neg\lambda$-sequents $\varphi, \neg\varphi \Rightarrow \lambda$, in contrast, are sound: since λ lacks a truth value, not all formulas in the consequent are false.

The usual rules for ¬-introduction are not sound, that is, for instance, one cannot bring a single formula from the antecedent to the succedent by affixing a negation symbol to the formula. If φ is a sentence lacking a truth value, $\varphi \Rightarrow \varphi$ is sound in the sense explained above, but $\Rightarrow \varphi, \neg\varphi$ is not sound.

4.1.4. *Laws of* ∨ *and* ∧. Scott employs here an additional connective ⨳, which is dispensable here.

(∧1) $$\varphi, \psi \Rightarrow \varphi \wedge \psi$$

(∧2) $$\varphi \wedge \psi \Rightarrow \varphi$$

(∧3) $$\varphi \wedge \psi \Rightarrow \psi$$

(∨1) $$\varphi \vee \psi \Rightarrow \varphi, \psi$$

(∨2) $$\varphi \Rightarrow \varphi \vee \psi$$

(∨3) $$\psi \Rightarrow \varphi \vee \psi$$

4.1.5. *Laws of quantifiers.* For the quantifiers we have the following initial sequents and rules:

(∀1) $$\forall x\varphi \Rightarrow \varphi(t/x)$$

(∃1) $$\varphi(t/x) \Rightarrow \exists x\varphi$$

(∀2) $$\frac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma \Rightarrow \forall x\varphi, \Delta}$$ x not free in lower sequent

(∃2) $$\frac{\Delta, \varphi \Rightarrow \Gamma}{\Delta, \exists x\varphi \Rightarrow \Gamma}$$ x not free in lower sequent

4.1.6. *Laws of Identity.* We also add initial sequents for identity for arbitrary terms s and t:

(=1) $$\Rightarrow t = t$$

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

The transitivity and symmetry follow from these initial sequents in the usual way.

The following theorem states that these initial sequents and rules are sound with respect to the intended notion of logical consequence in Strong Kleene logic. It is proved by a routine inductive argument.

THEOREM 12. *If* $\Gamma \Rightarrow \Delta$ *is derivable, then the following holds for all partial models* $\mathscr{M}$ *and variable assignments* $\mathfrak{b}$:

(i) *If* $\mathscr{M} \models_{\mathsf{SK}} \psi[\mathfrak{b}]$ *for all formulas* $\psi \in \Gamma$, *then there is a formula* $\varphi \in \Delta$ *with* $\mathscr{M} \models \varphi[\mathfrak{b}]$

(ii) *If* $\mathscr{M} \models_{\mathsf{SK}} \neg\psi[\mathfrak{b}]$ *for all formulas* $\psi \in \Delta$, *then there is a formula* $\varphi \in \Gamma$ *with* $\mathscr{M} \models \neg\varphi[\mathfrak{b}]$

A partial model is a model of $\mathscr{L}_{\mathrm{T}}$ where all predicate symbols ($=$ and T in the present case) are allowed to be partially interpreted. It should be clear how to define validity in such models according to the Strong Kleene scheme (see also the references below). Completeness proofs exist for several systems designed for various concepts of logical consequence in Strong Kleene logic; see Aoyama [1], Blamey [4], Cleave [8], Kearns [22] and Wang [44]. We do not give a completeness proof for our system here, but Blamey's [4] proof can easily be adapted. Thus the set of rules and initial sequents listed so far suffices for proving all correct sequents of Strong Kleene logic. Of course, this has motivated the above system.

4.2. Arithmetic. Since PKF will contain PA we add the additional sequents $\Rightarrow \varphi$ where φ is an axiom of PA except for the induction axioms.

We could also allow all induction axioms of $\mathscr{L}_{\mathrm{PA}}$ as initial sequents. But we also want to extend induction to the language with the truth predicate. Therefore we postulate the stronger:

$$\text{(IND)} \qquad \frac{\Gamma, \varphi(x) \Rightarrow \varphi(x+1), \Delta}{\Gamma, \varphi(\overline{0}), \Rightarrow \varphi(t), \Delta}$$

t is here an arbitrary term, φ any formula of $\mathscr{L}_{\mathrm{T}}$. x must not occur freely in $\varphi(\overline{0})$, Γ or Δ, but the term t is allowed to contain x.

4.3. Truth. There have been various attempts to set up a formal system for Kripke's fixed point semantics. Michael Kremer [23] has presented a system in the sequent calculus with aims that differ from ours. He adopts a derivability relation of the kind mentioned above. But his theory does not comprise arithmetic or a comparable system for expressing syntactical facts.

We now add additional initial sequents corresponding to the axioms of KF. For an explanation of the notation the reader is invited to consult the note preceding the list of KF axioms in Section 1.

PKF1 (i) $\mathrm{ClTerm}(x), \mathrm{ClTerm}(y), \mathrm{val}(x) = \mathrm{val}(y) \Rightarrow T\, x \dot{=} y$,
(ii) $\mathrm{ClTerm}(x), \mathrm{ClTerm}(y), T\, x \dot{=} y \Rightarrow \mathrm{val}(x) = \mathrm{val}(y)$,
PKF2 (i) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\, x \wedge T\, y \Rightarrow T(x \dot{\wedge} y)$,
(ii) $\mathrm{Sent}(x), \mathrm{Sent}(y), T(x \dot{\wedge} y) \Rightarrow T\, x \wedge T\, y$,
PKF3 (i) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\, x \vee T\, y \Rightarrow T(x \dot{\vee} y)$,
(ii) $\mathrm{Sent}(x), \mathrm{Sent}(y), T(x \dot{\vee} y) \Rightarrow T\, x \vee T\, y$,
PKF4 (i) $\mathrm{Var}(v), \mathrm{For}(x, v), \forall y\, T\, x(\dot{y}/v)) \Rightarrow T\, \dot{\forall} v x$,
(ii) $\mathrm{Var}(v), \mathrm{For}(x, v), T\, \dot{\forall} v x \Rightarrow \forall y\, T\, x(\dot{y}/v))$,
PKF5 (i) $\mathrm{Var}(v), \mathrm{For}(x, v), \exists y\, T\, x(\dot{y}/v)) \Rightarrow T\, \dot{\exists} v x$,
(ii) $\mathrm{Var}(v), \mathrm{For}(x, v), T\, \dot{\exists} v x \Rightarrow \exists y\, T\, x(\dot{y}/v))$,
PKF6 (i) $\mathrm{ClTerm}(x), T\, \mathrm{val}(x) \Rightarrow T\, \dot{T} x$,
(ii) $\mathrm{ClTerm}(x), T\, \dot{T} x \Rightarrow T\, \mathrm{val}(x)$,

PKF7 (i) $\mathrm{Sent}(x), \neg\, T\, x \Rightarrow T\, \dot{\neg} x$,
(ii) $\mathrm{Sent}(x), T\, \dot{\neg} x \Rightarrow \neg\, T\, x$,

PKF8 $T\, x \Rightarrow \mathrm{Sent}(x)$.

This concludes the description of PKF.

§**5. Some basic facts on PKF.** At first we shall prove that two rules of inference concerning $\vee$ are admissible in PKF, because we will need them in later proofs.

LEMMA 13. *The following rule is admissible in* PKF:

$$\frac{\varphi \Rightarrow \chi \qquad \psi \Rightarrow \chi}{\varphi \vee \psi \Rightarrow \chi}$$

PROOF.

$$\cfrac{\text{cut}\ \cfrac{(\vee 1)\quad \varphi \vee \psi \Rightarrow \varphi, \psi \qquad \varphi \Rightarrow \chi}{\varphi \vee \psi \Rightarrow \chi, \psi} \qquad \psi \Rightarrow \chi}{\varphi \vee \psi \Rightarrow \chi}\ \text{cut}$$

⊣

LEMMA 14. *The following rule is admissible in* PKF:

$$\frac{\Rightarrow \varphi, \psi}{\Rightarrow \varphi \vee \psi}$$

PROOF.

$$\cfrac{\cfrac{\Rightarrow \varphi, \psi \qquad \varphi \Rightarrow \varphi \vee \psi \quad (\vee 2)}{\Rightarrow \varphi \vee \psi, \psi} \qquad \psi \Rightarrow \varphi \vee \psi \quad (\vee 3)}{\Rightarrow \varphi \vee \psi}$$

⊣

PKF should behave classically on the arithmetical formulas. In PKF we have all rules and initial sequents of classical logic except for the rules that allow us to shift a formula from one side of the sequent arrow to the other side by affixing a negation symbol, while all other formulas are left in their place. These rules of the classical sequent calculus are also admissible for a formula φ in PKF, if PKF proves $\Rightarrow \varphi, \neg\varphi$.

LEMMA 15. *If* $\Rightarrow \varphi, \neg\varphi$ *is derivable in* PKF, *then the following two rules are derived rules of* PKF:

$$\frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \neg\varphi, \Delta}$$

$$\frac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma, \neg\varphi \Rightarrow \Delta}$$

PROOF. The first rule can be established by an application of the cut rule:

$$\frac{\Rightarrow \varphi, \neg\varphi \qquad \Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \neg\varphi, \Delta}$$

For the second rule we employ the $\neg\neg$-sequent and the $\neg$-rule:

$$\cfrac{\Gamma \Rightarrow \varphi, \Delta \qquad \cfrac{\varphi \Rightarrow \neg\neg\varphi \qquad \cfrac{\Rightarrow \varphi, \neg\varphi}{\neg\varphi, \neg\neg\varphi \Rightarrow}\ \neg\text{-rule}}{\neg\varphi, \varphi \Rightarrow}\ \text{cut}}{\Gamma, \neg\varphi \Rightarrow \Delta}\ \text{cut}$$

⊣

Thus we need only prove $\Rightarrow \varphi, \neg\varphi$ in order to show that arithmetical formulas behave classically in PKF

LEMMA 16. *$\Rightarrow \varphi, \neg\varphi$ is derivable in* PKF *for arithmetical φ.*

PROOF. The claim is established by an induction on the complexity of φ. First we prove the claim for atomic formulas (cf. Scott [38, p. 19]), that is, for formulas $s = t$, where s and t are terms. The leftmost line in the following proof is an initial sequent by =2; the rightmost is a law of negation.

$$\text{cut}\ \dfrac{\dfrac{\dfrac{s = t, \neg s = t \Rightarrow \neg t = t \qquad \dfrac{\Rightarrow t = t}{\neg t = t \Rightarrow}}{s = t, \neg s = t \Rightarrow}}{\Rightarrow \neg s = t, \neg\neg s = t}\ \neg\text{-rule} \qquad \neg\neg s = t \Rightarrow s = t}{\Rightarrow s = t, \neg s = t}$$

Thus the claim is proved for all atomic $\varphi \in \mathscr{L}_{\mathrm{PA}}$. As an example we show the claim for conjunctions. Thus $\Rightarrow \varphi, \neg\varphi$ and $\Rightarrow \psi, \neg\psi$ hold by induction hypothesis. The first line is $(\wedge 2)$.

$$\dfrac{\Rightarrow \varphi, \neg\varphi \qquad \dfrac{\varphi \wedge \psi \Rightarrow \varphi}{\neg\varphi \Rightarrow \neg(\varphi \wedge \psi)}}{\Rightarrow \varphi, \neg(\varphi \wedge \psi)}$$

Similarly, we derive $\Rightarrow \neg(\varphi \wedge \psi), \psi$ and proceed as follows:

$$\dfrac{\Rightarrow \neg(\varphi \wedge \psi), \psi \qquad \dfrac{\Rightarrow \varphi, \neg(\varphi \wedge \psi) \qquad \varphi, \psi \Rightarrow \varphi \wedge \psi \quad (\wedge 1)}{\psi \Rightarrow \varphi \wedge \psi, \neg(\varphi \wedge \psi)}}{\Rightarrow \varphi \wedge \psi, \neg(\varphi \wedge \psi)}$$

We skip the other cases. ⊣

For future reference, we state following obvious consequence of Lemmata 16 and 15:

COROLLARY 17. *If φ is arithmetical, then the following two rules are derived rules of* PKF:

$$\dfrac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \neg\varphi, \Delta}$$

$$\dfrac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma, \neg\varphi \Rightarrow \Delta}$$

Since these two rules are the only rules that are missing from classical logic, we obtain the following corollary:

COROLLARY 18. *Classical Peano arithmetic* PA *restricted to the language $\mathscr{L}_{\mathrm{PA}}$ is a subsystem of* PKF.

PROOF. This follows from Corollary 17. In the presence of classical logic the unrestricted induction rule IND yields all induction axioms. ⊣

Each axiom of KF has been split up into two initial sequents of PKF, because the sequent arrow $\Rightarrow$ has only one direction. Apart from this, however, the axioms for PKF are easier to formulate than the axioms for KF because separate rules for negated conjunctions, disjunctions etc. are not required. The sequents corresponding to the axioms of KF for negated connectives and quantifiers can be proved in PKF.

Lemma 19. *The following sequents are derivable in* PKF.

1. (i) $\mathrm{ClTerm}(x), \mathrm{ClTerm}(y), \mathrm{val}(x) \neq \mathrm{val}(y) \Rightarrow T\, \underset{\cdot}{\neg} x \underset{\cdot}{=} y$,
   (ii) $\mathrm{ClTerm}(x), \mathrm{ClTerm}(y), T\, \underset{\cdot}{\neg} x \underset{\cdot}{=} y \Rightarrow \mathrm{val}(x) \neq \mathrm{val}(y)$,
2. (i) $\mathrm{Sent}(x), T\, x \Rightarrow T\, \underset{\cdot}{\neg}\underset{\cdot}{\neg} x$,
   (ii) $\mathrm{Sent}(x), T\, \underset{\cdot}{\neg}\underset{\cdot}{\neg} x \Rightarrow T\, x$,
3. (i) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\, \underset{\cdot}{\neg} x \vee T\, \underset{\cdot}{\neg} y \Rightarrow T\, \underset{\cdot}{\neg}(x \underset{\cdot}{\wedge} y)$,
   (ii) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\, \underset{\cdot}{\neg}(x \underset{\cdot}{\wedge} y) \Rightarrow T\, \underset{\cdot}{\neg} x \vee T\, \underset{\cdot}{\neg} y$,
4. (i) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\, \underset{\cdot}{\neg} x \wedge T\, \underset{\cdot}{\neg} y \Rightarrow T\, \underset{\cdot}{\neg}(x \underset{\cdot}{\vee} y)$,
   (ii) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\, \underset{\cdot}{\neg}(x \underset{\cdot}{\vee} y) \Rightarrow T\, \underset{\cdot}{\neg} x \wedge T\, \underset{\cdot}{\neg} y$,
5. (i) $\mathrm{Var}(v), \mathrm{For}(x, v), \exists y\, T\, \underset{\cdot}{\neg} x(\dot{y}/v)) \Rightarrow T\, \underset{\cdot}{\neg}\underset{\cdot}{\forall} v x$,
   (ii) $\mathrm{Var}(v), \mathrm{For}(x, v), T\, \underset{\cdot}{\neg}\underset{\cdot}{\forall} v x \Rightarrow \exists y\, T\, \underset{\cdot}{\neg} x(\dot{y}/v))$,
6. (i) $\mathrm{Var}(v), \mathrm{For}(x, v), \forall y\, T\, \underset{\cdot}{\neg} x(\dot{y}/v)) \Rightarrow T\, \underset{\cdot}{\neg}\underset{\cdot}{\exists} v x$,
   (ii) $\mathrm{Var}(v), \mathrm{For}(x, v), T\, \underset{\cdot}{\neg}\underset{\cdot}{\exists} v x \Rightarrow \forall y\, T\, \underset{\cdot}{\neg} x(\dot{y}/v))$,
7. (i) $\mathrm{ClTerm}(x), T\, \underset{\cdot}{\neg}\, \mathrm{val}(x) \Rightarrow T\, \underset{\cdot}{\neg} \underset{\cdot}{T} x$,
   (ii) $\mathrm{ClTerm}(x), T\, \underset{\cdot}{\neg} \underset{\cdot}{T} x \Rightarrow T\, \underset{\cdot}{\neg}\, \mathrm{val}(x)$.

PKF7(ii) corresponds to CONS if CONS is formulated as $\forall x (T \ulcorner \neg\, T\, \dot{x} \urcorner \rightarrow \neg\, T\, x)$ (see Lemma 10(v)). In PKF we have, in addition, the converse 'completeness' direction. This makes PKF a more symmetrical system than KF. PKF7 plays the crucial rôle in the proof of Lemma 19, which we skip.

In the following we write PKF $\vdash \varphi$ if and only if the sequent $\Rightarrow \varphi$ is derivable in PKF.

Lemma 20 (soundness). *If* PKF $\vdash \varphi$, *then φ holds in all fixed point models.*

Outline of Proof. In order to prove the lemma one shows by induction on the length of derivations the following stronger claim:

> If $\Gamma \Rightarrow \Delta$ is derivable, then the two following conditions obtain for all fixed point models $\mathscr{M}$ and assignments $\mathfrak{b}$:
> - If all formulas in Γ are true in $\mathscr{M}$ at $\mathfrak{b}$, then at least one formula in Δ is true in $\mathscr{M}$ at $\mathfrak{b}$.
> - If all formulas in Δ are false in $\mathscr{M}$ at $\mathfrak{b}$, then at least one formula in Δ is false in $\mathscr{M}$ at $\mathfrak{b}$.

We do not go through the numerous initial sequents and the rules. The proof will later be formalized in a subtheory of KF in Theorem 27. ⊣

By the previous lemma, in PKF the T-sentences are not derivable for arbitrary formulas of $\mathscr{L}_{\mathrm{T}}$. For arithmetical sentences, however, they are provable.

Lemma 21. PKF $\vdash \forall \vec{x}(T \ulcorner \varphi(\dot{\vec{x}}) \urcorner \leftrightarrow \varphi(\vec{x}))$ *for all arithmetical formulas $\varphi(\vec{x})$ with the indicated free variables.*

This can be proved by a meta-induction on the buildup of $\varphi(\vec{x})$ using the initial sequents of PKF.

For arbitrary sentences we do not get the T-sentences, but only a corresponding rule:

Theorem 22. *For all sentences $\varphi \in \mathscr{L}_{\mathrm{T}}$: φ is provable in* PKF *if and only if $T \ulcorner \varphi \urcorner$ is provable in* PKF.

Proof. The result that is actually shown is somewhat stronger. One shows that $\Gamma \Rightarrow \varphi(\vec{x}), \Delta$ is provable if and only if $\Gamma \Rightarrow T \ulcorner \varphi(\dot{\vec{x}}) \urcorner, \Delta$ is provable in PKF. $\vec{x}$ is the string of variables free in φ.

The proof proceeds then on the buildup of φ and its negation. We suppress the inactive formulas. Let $s(\vec{x})$ and $t(\vec{y})$ be terms with the indicated free variables. Then the claim is established for the atomic case in the following way:

$$\dfrac{\dfrac{\Rightarrow s(\vec{x}) = t(\vec{x})}{\Rightarrow \mathrm{val}\ulcorner s(\dot{\vec{x}})\urcorner = \mathrm{val}\ulcorner t(\dot{\vec{x}})\urcorner}\ \text{arithmetic}}{\Rightarrow T\ulcorner s(\dot{\vec{x}}) = t(\dot{\vec{x}})\urcorner}\ \mathsf{PKF1}$$

We have been somewhat sloppy: The first step involves a more lengthy reasoning in PA. The second step is also abbreviated and needs some additional steps involving identity axioms. We allow ourselves such abbreviations also in the following. The derivation also can be inverted, and it obviously works for negated identity statements as well.

If φ is an atomic formula of the form $T\,t$ or a negation of such a formula, PKF6 can be used. For the negated formulas of the form $T\,t$ Lemma 19.7 can be employed.

If φ is a doubly negated formula Lemma 19.2 is employed. Next we consider as an example the case that φ is a negated conjunction $\neg(\psi(\vec{x}) \wedge \chi(\vec{y}))$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\psi(\vec{x}) \wedge \chi(\vec{y}) \Rightarrow \neg\neg(\psi(\vec{x}) \wedge \chi(\vec{y})) \qquad \dfrac{\Rightarrow \neg(\psi(\vec{x}) \wedge \chi(\vec{y}))}{\neg\neg(\psi(\vec{x}) \wedge \chi(\vec{y})) \Rightarrow}\ \neg\text{-rule}}{\psi(\vec{x}) \wedge \chi(\vec{y}) \Rightarrow}}{\psi(\vec{x}), \chi(\vec{y}) \Rightarrow}\ \wedge 1}{\Rightarrow \neg\psi(\vec{x}), \neg\chi(\vec{y})}\ \neg\text{-rule}}{\Rightarrow T\ulcorner\neg\psi(\dot{\vec{x}})\urcorner, T\ulcorner\neg\chi(\dot{\vec{y}})\urcorner}\ \text{induction hypothesis}}{\Rightarrow T\ulcorner\neg\psi(\dot{\vec{x}})\urcorner \vee T\ulcorner\neg\chi(\dot{\vec{y}})\urcorner}\ \text{Lemma 14}$$

Then Lemma 19.3(i) plus some additional arithmetical steps yield the desired conclusion

$$\Rightarrow T\ulcorner\neg(\psi(\dot{\vec{x}}) \wedge \chi(\dot{\vec{y}}))\urcorner.$$

Again, it is not hard to reverse the proof and to derive the sequent $\Rightarrow \neg(\psi(\vec{x}) \wedge \chi(\vec{y}))$ from this last sequent.

The remaining cases with connectives and quantifiers are treated in a similar way. For the quantifier cases we need in the induction hypothesis that φ may contain free variables. $\dashv$

The closure under NEC and CONEC, that is, the identity of inner an outer logic was of course the main point for considering the system PKF in the first place. For Theorem 22 shows that the inner and outer logic of PKF—in contrast to KF—coincide. In fact, the argument proves a stronger statement. It shows that *all extensions* of PKF by additional axioms have the property of being closed under the Necessitation and the Conecessitation rule.

In §2 we have considered the special rôle of the consistency axiom CONS in KF. There we argued that CONS allows us to distribute the truth predicate over the conditional. PKF also allows the truth predicate to be distributed over the conditional, although this holds —like every law in partial logic— only as a rule of inference.

LEMMA 23. PKF *proves the sequents*

$$T\ulcorner\varphi \to \psi\urcorner \Leftrightarrow T\ulcorner\varphi\urcorner \to T\ulcorner\psi\urcorner$$

for all sentences φ *and* ψ *of* $\mathscr{L}_{\mathrm{T}}$. *Therefore if* $\mathsf{PKF} \vdash T\ulcorner\varphi \to \psi\urcorner$ *obtains, we have also* $\mathsf{PKF} \vdash T\ulcorner\varphi\urcorner \to T\ulcorner\psi\urcorner$.

PROOF. $\varphi \to \psi$ is defined as $\neg\varphi \vee \psi$. PKF3 yields

$$T\ulcorner\neg\varphi \vee \psi\urcorner \Rightarrow T\ulcorner\neg\varphi\urcorner \vee T\ulcorner\psi\urcorner.$$

Applying PKF7 we obtain the following:

$$T\ulcorner\neg\varphi \vee \psi\urcorner \Rightarrow \neg\, T\ulcorner\varphi\urcorner \vee T\ulcorner\psi\urcorner,$$

which yields one half the lemma. The other direction is proved by reading the above proof from bottom to top. ⊣

Friedman and Sheard [14] have investigated the following analogue of the disjunction property and numerical existence property for intuitionistic systems:

DEFINITION 24.

(DP) A theory S formulated in $\mathscr{L}_{\mathrm{T}}$ has the Disjunction Property (DP) if and only if for all sentences $\varphi, \psi \in \mathscr{L}_{\mathrm{T}}$: if $S \vdash T\ulcorner\varphi\urcorner \vee T\ulcorner\psi\urcorner$, then $S \vdash T\ulcorner\varphi\urcorner$ or $S \vdash T\ulcorner\psi\urcorner$.

(NEP) A theory S formulated in $\mathscr{L}_{\mathrm{T}}$ has the Numerical Existence Property (NEP) if and only if for all $\varphi(x) \in \mathscr{L}_{\mathrm{T}}$: if $S \vdash \exists x\, T\ulcorner\varphi(\dot{x})\urcorner$, then $S \vdash \varphi(\overline{n})$ for some (standard) numeral $\overline{n}$.

Like KF, PKF does not have either of these properties:

PROPOSITION 25. PKF *does not have* DP *and* NEP.

PROOF. We only consider DP; the argument for NEP is similar. By Lemma 20, PKF is consistent. $\mathsf{Con}_{\mathsf{PKF}}$ does not contain the truth predicate.

$$\mathsf{PKF} \vdash \mathsf{Con}_{\mathsf{PKF}} \vee \neg\mathsf{Con}_{\mathsf{PKF}},$$
$$\mathsf{PKF} \vdash T\ulcorner\mathsf{Con}_{\mathsf{PKF}} \vee \neg\mathsf{Con}_{\mathsf{PKF}}\urcorner,$$
$$\mathsf{PKF} \vdash T\ulcorner\mathsf{Con}_{\mathsf{PKF}}\urcorner \vee T\ulcorner\neg\mathsf{Con}_{\mathsf{PKF}}\urcorner.$$

On the one hand, if $\mathsf{PKF} \vdash T\ulcorner\mathsf{Con}_{\mathsf{PKF}}\urcorner$ we have by Theorem 22 also $\mathsf{PKF} \vdash \mathsf{Con}_{\mathsf{PKF}}$, which is impossible by Gödel's second theorem. On the other hand,

$$\mathsf{PKF} \vdash T\ulcorner\neg\mathsf{Con}_{\mathsf{PKF}}\urcorner$$

implies $\mathsf{PKF} \vdash \neg\mathsf{Con}_{\mathsf{PKF}}$ contradicting the soundness of PKF. ⊣

§6. The proof-theoretic strength of PKF.

6.1. The upper bound. In this section we shall determine an upper bound for the proof-theoretic strength of PKF. We employ a result by Cantini [5]. In his paper Cantini studies KF with restricted induction schemes. Cantini calls one of his systems KF, but it differs from what we call KF. Therefore we introduce a new label for this system and call it $\mathsf{KF}_{\mathsf{int}}$. $\mathsf{KF}_{\mathsf{int}}$ is given by the axioms of "our" KF with the induction scheme replaced by the following single axiom of "internal" induction (whence the subscript int).

$$\forall v \forall x \big(\mathrm{Var}(v) \wedge \mathrm{For}(x, v) \wedge T\, x(\overline{0}/v) \wedge \forall y (T\, x(\dot{y}/v) \to T\, x(\dot{y} + 1/v)) \to \forall y\, T\, x(\dot{y}/v)\big).$$

We shall show that PKF can be embedded in $\mathsf{KF}_{\mathsf{int}}$ in a sense to be specified below. In combination with Cantini's proof-theoretic analysis of $\mathsf{KF}_{\mathsf{int}}$, this will yield an upper bound for the proof-theoretic strength of PKF.

For the embedding we need the following lemma:[12]

LEMMA 26. $\mathsf{KF}_{\mathsf{int}} \vdash \forall v \forall x \big(\mathrm{Var}(v) \wedge \mathrm{For}(x, v) \wedge \neg T\, x(\overline{0}/v) \wedge \forall y(\neg T\, x(\dot{y}/v) \rightarrow \neg T\, x(\dot{y}+1/v)) \rightarrow \forall y \neg T\, x(\dot{y}/v)\big)$.

PROOF. We apply Parsons's [33] well known trick that is also used for showing that Π_n- and Σ_n-induction are equivalent. Subtraction $(-)$ is defined in PA in the usual way; if $n < k$ the difference $n - k$ is stipulated to be 0.

We reason in $\mathsf{KF}_{\mathsf{int}}$ as follows. For a *reductio as absurdum* assume

(4) $$\neg T\, x(\overline{0}/v),$$

(5) $$\forall y(\neg T\, x(\dot{y}/v) \rightarrow \neg T\, x(\dot{y}+1/v)),$$

(6) $$T\, x(\dot{z}/v).$$

From the induction axiom of $\mathsf{KF}_{\mathsf{int}}$ we obtain:

(7) $$\forall v \forall x \big(\mathrm{Var}(v) \wedge \mathrm{For}(x, v) \wedge T\, x(z - \overline{0}/v) \wedge \forall y(T\, x(z - \dot{y}/v) \rightarrow T\, x(z - \dot{y} - 1/v)) \rightarrow \forall y\, T\, x(z - \dot{y}/v)\big).$$

(5) implies $\forall y(T\, x(z - \dot{y}/v) \rightarrow T\, x(z - \dot{y} - 1/v))$. This, (6) and (7) then imply $\forall y\, T\, x(z - \dot{y}/v)$ and thus $T\, x(\overline{0}/v)$ contradicting (4). ⊣

For finite non-empty sets Γ of sentences $\bigwedge\!\!\!\bigwedge\Gamma$ is the conjunction of all elements of Γ, $\bigvee\!\!\!\bigvee\Gamma$ is their disjunction. If Γ is empty, $\bigwedge\!\!\!\bigwedge\Gamma$ is $\overline{0} = \overline{0}$ and $\bigvee\!\!\!\bigvee\Gamma$ is $\neg\overline{0} = \overline{0}$.

We now formalize the soundness theorem 12 in $\mathsf{KF}_{\mathsf{int}}$.

THEOREM 27. *If the sequent* $\Gamma \Rightarrow \Delta$ *is derivable in* PKF, *then*

(i) $\mathsf{KF}_{\mathsf{int}} \vdash \forall \vec{x}(T \ulcorner \bigwedge\!\!\!\bigwedge \Gamma(\dot{\vec{x}}) \urcorner \rightarrow T \ulcorner \bigvee\!\!\!\bigvee \Delta(\dot{\vec{x}}) \urcorner)$,

(ii) $\mathsf{KF}_{\mathsf{int}} \vdash \forall \vec{x}(T \ulcorner \neg \bigvee\!\!\!\bigvee \Delta(\dot{\vec{x}}) \urcorner \rightarrow T \ulcorner \neg \bigwedge\!\!\!\bigwedge \Gamma(\dot{\vec{x}}) \urcorner)$.

The variables $\vec{x}$ *are the free variables in* Γ *and* Δ.

PROOF. The proof proceeds by induction on the length of the proof in PKF. We give only some examples.

In order to handle the $\neg\lambda$-sequents $\varphi, \neg\varphi \Rightarrow \lambda$, we use CONS.[13] In order to prove part (i) of the claim, we proceed as follows suppressing free variables again:

$\mathsf{KF}_{\mathsf{int}} \vdash \forall x \big(\mathrm{Sent}(x) \rightarrow \neg(T\, x \wedge T\, \underset{\cdot}{\neg} x)\big)$	CONS
$\mathsf{KF}_{\mathsf{int}} \vdash \neg\, T \ulcorner \varphi \wedge \neg\varphi \urcorner$	for all $\varphi \in \mathscr{L}_{\mathrm{T}}$
$\mathsf{KF}_{\mathsf{int}} \vdash T \ulcorner \varphi \wedge \neg\varphi \urcorner \rightarrow T \ulcorner \lambda \urcorner$	for all $\varphi \in \mathscr{L}_{\mathrm{T}}$

For part (ii) of the theorem we need to show

(8) $$T \ulcorner \neg\lambda \urcorner \rightarrow T \ulcorner \neg(\varphi \wedge \neg\varphi) \urcorner.$$

[12] This observation is due to Cantini (personal communication). We thank a referee for bringing to our attention that the technique used for the proof is due to Charles Parsons.

[13] We thank a referee who spotted a gap in our original proof and made us aware of the need to use CONS.

It is easy to see from the proof of Lemma 10 that it also holds for $\mathsf{KF_{int}}$. So from Lemma 10(iii) we have:

(9) $$\mathsf{KF_{int}} \vdash T\ulcorner\neg\lambda\urcorner \rightarrow \neg\lambda.$$

Then we reason as follows:

(10) $$\begin{aligned} \mathsf{KF_{int}} \vdash T\ulcorner\neg\lambda\urcorner &\rightarrow \neg\, T\ulcorner\lambda\urcorner && \text{Lemma 10(ii)} \\ &\rightarrow \lambda && \text{definition of } \lambda \end{aligned}$$

(9) and (10) imply $\mathsf{KF_{int}} \vdash \neg\, T\ulcorner\neg\lambda\urcorner$ and therefore also (8).

As an example for a truth-theoretic initial sequent we choose PKF7(ii), that is, $\mathrm{Sent}(x), \neg\, T\,x \Rightarrow T\,\underset{\cdot}{\neg} x$. We have to show two claims corresponding to parts (i) and (ii) of the theorem respectively:

(11) $$\mathsf{KF_{int}} \vdash \forall x (T\ulcorner\mathrm{Sent}(\dot{x}) \wedge \neg\, T\,\dot{x}\urcorner \rightarrow T\ulcorner T\,\underset{\cdot}{\neg}\dot{x}\urcorner),$$

(12) $$\mathsf{KF_{int}} \vdash \forall x (T\ulcorner\neg\, T\,\neg\dot{x}\urcorner \rightarrow T\ulcorner\neg(\mathrm{Sent}(\dot{x}) \wedge \neg\, T\,\dot{x})\urcorner).$$

To prove (11) we proceed as follows:

$\mathsf{KF_{int}} \vdash \forall x \big(\mathrm{ClTerm}(x) \wedge \mathrm{Sent}(\mathrm{val}(x)) \wedge T\,\underset{\cdot}{\neg}\,\mathrm{val}(x) \rightarrow T\,\underset{\cdot}{\neg}\,\mathrm{val}(x)\big)$	logic
$\mathsf{KF_{int}} \vdash \forall x \big(\mathrm{ClTerm}(x) \wedge \mathrm{Sent}(\mathrm{val}(x)) \wedge T\,\underset{\cdot}{\neg}\underset{\cdot}{T} x \rightarrow T\,\underset{\cdot}{T}\underset{\cdot}{\neg} x\big)$	KF12, KF13
$\mathsf{KF_{int}} \vdash \forall x \big(\mathrm{ClTerm}(x) \wedge \mathrm{Sent}(\mathrm{val}(x)) \wedge T\ulcorner\underset{\cdot}{\neg}\, T\,\dot{x}\urcorner \rightarrow T\ulcorner T\,\underset{\cdot}{\neg}\dot{x}\urcorner\big)$	arithmetic
$\mathsf{KF_{int}} \vdash \forall x \big(\mathrm{Sent}(\mathrm{val}(x)) \rightarrow \mathrm{ClTerm}(x)\big)$	assumption on val
$\mathsf{KF_{int}} \vdash \forall x \big(\mathrm{Sent}(\mathrm{val}(x)) \wedge T\ulcorner\neg\, T\,\dot{x}\urcorner \rightarrow T\ulcorner T\,\underset{\cdot}{\neg}\dot{x}\urcorner\big)$	two preceding lines
$\mathsf{KF_{int}} \vdash \forall x \big(T\ulcorner\mathrm{Sent}(\dot{x})\urcorner \wedge T\ulcorner\neg\, T\,\dot{x}\urcorner \rightarrow T\ulcorner T\,\underset{\cdot}{\neg}\dot{x}\urcorner\big)$	Lemma 6
$\mathsf{KF_{int}} \vdash \forall x \big(T\ulcorner\mathrm{Sent}(\dot{x}) \wedge \neg\, T\,\dot{x}\urcorner \rightarrow T\ulcorner T\,\underset{\cdot}{\neg}\dot{x}\urcorner\big)$	KF4

We skip the proof of (12), which is similar.

For PKF14 we need to establish the following two claims:

(13) $$\mathsf{KF_{int}} \vdash \forall x \big(T\ulcorner T\,\dot{x}\urcorner \rightarrow T\ulcorner\mathrm{Sent}(\dot{x})\urcorner\big),$$

(14) $$\mathsf{KF_{int}} \vdash \forall x \big(T\ulcorner\neg\,\mathrm{Sent}(\dot{x})\urcorner \rightarrow T\ulcorner\neg\, T\,\dot{x}\urcorner\big).$$

We prove only (13). We observe first the following:

(15) $$\mathsf{KF_{int}} \vdash \forall x (T\,\underset{\cdot}{T}\dot{x} \rightarrow \mathrm{ClTerm}(x)).$$

This follows from $\mathsf{KF_{int}} \vdash \forall x (T\,\underset{\cdot}{T} x \rightarrow \mathrm{Sent}(\underset{\cdot}{T} x))$, which is an instance of KF14. Then we prove (13) using mainly KF12:

$\mathsf{KF_{int}} \vdash \forall x \big(\mathrm{ClTerm}(x) \rightarrow (T\,\underset{\cdot}{T} x \leftrightarrow T\,\mathrm{val}(x))\big)$	
$\mathsf{KF_{int}} \vdash \forall x \big(T\,\underset{\cdot}{T} x \rightarrow T\,\mathrm{val}(x)\big)$	(15)
$\mathsf{KF_{int}} \vdash \forall x \big(T\,\underset{\cdot}{T} x \rightarrow \mathrm{Sent}(\mathrm{val}(x))\big)$	KF14
$\mathsf{KF_{int}} \vdash \forall x \big(T\,\underset{\cdot}{T} x \rightarrow T\ulcorner\mathrm{Sent}(\dot{x})\urcorner\big)$	Lemma 6

Lemma 6 was used in its form with free variables. As an example of a rule we consider the cut-rule of PKF (suppressing free variables):

$$\frac{\Gamma \Rightarrow \varphi, \Delta \qquad \Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

By induction hypothesis we have

$$\mathsf{KF}_{\mathsf{int}} \vdash T\ulcorner \bigwedge\Gamma \urcorner \rightarrow T\ulcorner \varphi \vee \bigvee\Delta\urcorner,$$
$$\mathsf{KF}_{\mathsf{int}} \vdash T\ulcorner \varphi \wedge \bigwedge\Gamma \urcorner \rightarrow T\ulcorner \bigvee\Delta\urcorner.$$

Applying KF4 and KF6 we obtain from these two lines respectively:

$$\mathsf{KF}_{\mathsf{int}} \vdash T\ulcorner \bigwedge\Gamma \urcorner \rightarrow T\ulcorner \varphi\urcorner \vee T\ulcorner \bigvee\Delta\urcorner,$$
$$\mathsf{KF}_{\mathsf{int}} \vdash T\ulcorner \varphi\urcorner \wedge T\ulcorner \bigwedge\Gamma \urcorner \rightarrow T\ulcorner \bigvee\Delta\urcorner.$$

This implies by propositional logic

$$\mathsf{KF}_{\mathsf{int}} \vdash T\ulcorner \bigwedge\Gamma \urcorner \rightarrow T\ulcorner \bigvee\Delta\urcorner.$$

The claim corresponding to (ii) in the theorem is proved in a similar way.

The most interesting rule is induction:

(IND) $$\frac{\Gamma, \varphi(x) \Rightarrow \varphi(x+1), \Delta}{\Gamma, \varphi(\overline{0}), \Rightarrow \varphi(t), \Delta}$$

In the following we do not mention the additional free variables of Γ, Δ and $\varphi(x)$ in order to keep the presentation more transparent. The induction hypothesis is:

(16) $$\mathsf{KF}_{\mathsf{int}} \vdash \forall x (T\ulcorner \varphi(\dot{x}) \wedge \bigwedge\Gamma\urcorner \rightarrow T\ulcorner \varphi(\dot{x}+1) \vee \bigvee\Delta\urcorner),$$

(17) $$\mathsf{KF}_{\mathsf{int}} \vdash \forall x (T\ulcorner \neg(\varphi(\dot{x}+1) \vee \bigvee\Delta)\urcorner \rightarrow T\ulcorner \neg(\varphi(\dot{x}) \wedge \bigwedge\Gamma)\urcorner).$$

We do not show that the translation corresponding to (i) of the theorem is derivable, but proceed directly to (ii).

We start with (17):

$T\ulcorner \neg(\varphi(\dot{x}+1) \vee \bigvee\Delta)\urcorner \rightarrow T\ulcorner \neg(\varphi(\dot{x}) \wedge \bigwedge\Gamma)\urcorner$	
$(T\ulcorner \neg\varphi(\dot{x}+1)\urcorner \wedge T\ulcorner \neg\bigvee\Delta\urcorner) \rightarrow (T\ulcorner \neg\varphi(\dot{x})\urcorner \vee T\ulcorner \neg\bigwedge\Gamma\urcorner)$	KF7, KF5
$(T\ulcorner \neg\bigvee\Delta\urcorner \rightarrow T\ulcorner \neg\bigwedge\Gamma\urcorner) \vee (T\ulcorner \neg\varphi(\dot{x}+1)\urcorner \rightarrow T\ulcorner \neg\varphi(\dot{x})\urcorner)$	prop. logic
$(T\ulcorner \neg\bigvee\Delta\urcorner \rightarrow T\ulcorner \neg\bigwedge\Gamma\urcorner) \vee (\neg T\ulcorner \neg\varphi(\dot{x})\urcorner \rightarrow \neg T\ulcorner \neg\varphi(\dot{x}+1)\urcorner)$	prop. logic
$(T\ulcorner \neg\bigvee\Delta\urcorner \rightarrow T\ulcorner \neg\bigwedge\Gamma\urcorner) \vee (\neg T\ulcorner \neg\varphi(\overline{0})\urcorner \rightarrow \neg T\ulcorner \neg\varphi(\dot{x})\urcorner)$	Lemma 26
$(T\ulcorner \neg\bigvee\Delta\urcorner \rightarrow T\ulcorner \neg\bigwedge\Gamma\urcorner) \vee (T\ulcorner \neg\varphi(\dot{x})\urcorner \rightarrow T\ulcorner \neg\varphi(\overline{0})\urcorner)$	prop. logic
$(T\ulcorner \neg\bigvee\Delta\urcorner \wedge T\ulcorner \neg\varphi(\dot{x})\urcorner) \rightarrow (T\ulcorner \neg\bigwedge\Gamma\urcorner \vee T\ulcorner \neg\varphi(\overline{0})\urcorner)$	prop. logic
$T\ulcorner \neg(\bigvee\Delta \vee \neg\varphi(\dot{x}))\urcorner \rightarrow T\ulcorner \neg(\bigwedge\Gamma \wedge \varphi(\overline{0}))\urcorner$	KF7, KF5

The last line shows that the translation of $\Gamma, \varphi(\overline{0}) \Rightarrow \varphi(t), \Delta$ corresponding to claim (ii) of the Theorem is derivable. ⊣

COROLLARY 28. *Every sentence provable in* PKF *is also provable in* $\mathsf{KF}_{\mathsf{int}}$.

PROOF. If a sentence $\Rightarrow \varphi$ is provable PKF, then $\mathsf{KF}_{\mathsf{int}}$ proves $T\ulcorner\varphi\urcorner$ by Theorem 27(i). But by Lemma 10(iv) $\mathsf{KF}_{\mathsf{int}}$ proves $T\ulcorner\varphi\urcorner \rightarrow \varphi$ for every sentence φ. Therefore $\mathsf{KF}_{\mathsf{int}}$ proves φ. ⊣

This concludes the reduction of PKF to $\mathsf{KF_{int}}$. In particular, PKF proves all *arithmetical* sentences that are in $\mathsf{KF_{int}}$ provable.
This raises the question whether PKF is identical with the inner logic of $\mathsf{KF_{int}}$. We leave this question open.

Theorem 27 allows us to relate PKF to standard systems, because Cantini determined the strength of the system $\mathsf{KF_{int}}$, which is called KF in his article:

THEOREM 29 (Cantini [5], §9). $\mathsf{KF_{int}}$ *is proof-theoretically equivalent to the system* $\mathsf{RA}_{<\omega^\omega}$ *of ramified analysis up to* ω^ω.

From this we obtain an upper bound for PKF:

THEOREM 30. *Every arithmetical sentence provable in* PKF *is also provable in* $\mathsf{RA}_{<\omega^\omega}$.

6.2. Lower bound. Interpreting classical systems in PKF is not straightforward because PKF is formulated in partial logic. However, PKF behaves classically on a sublanguage of $\mathscr{L}_T$. In this section we shall show that the classical part of PKF is sufficient for interpreting $\mathsf{RA}_{<\omega^\omega}$. Together with Theorem 30 this shows that PKF is equivalent to $\mathsf{RA}_{<\omega^\omega}$: both theories prove the same arithmetical sentences.

We shall interpret $\mathsf{RA}_{<\omega^\omega}$ in PKF by showing that ramified truth predicates up to level ω^ω can be defined in PKF and that all sentences involving these ramified truth predicates satisfy classical logic in PKF. We need to show that the truth predicates T_α are 'Tarskian' truth predicates for the language that contains only truth predicates T_β with $\beta < \alpha$.

To this end we need a notation system for the ordinals $\alpha < \omega^\omega$. We do not give the details of such a notation system and use the ordinals freely in PKF (see, for instance, Troelstra and Schwichtenberg [42], Schwichtenberg [37], Pohlers [34] or Takeuti [41]). We define sublanguages of $\mathscr{L}_T$ with ramified truth predicates recursively. The language $\mathscr{L}_0$ is the language $\mathscr{L}_{PA}$ of arithmetic. $\mathscr{L}_{\alpha+1}$ is the language $\mathscr{L}_\alpha$ expanded by the predicate $T_\alpha\, x := (T\,x \wedge x \in \mathscr{L}_\alpha)$, and at limit levels $\mathscr{L}_\lambda$ is the union of all languages $\mathscr{L}_\alpha$ with $\alpha < \lambda$.

In T_α the level ordinal α appears as an index. However, it follows from the definition of T_α that the level index can be quantified over as in $\forall\alpha < \omega^\omega\; T_\alpha \ulcorner 0 = 0 \urcorner$, for instance.

The formula $\mathrm{Sent}_\alpha(x)$ expresses in $\mathscr{L}_{PA}$ that x is a sentence of the language $\mathscr{L}_\alpha$.

The main challenge for recovering a classical hierarchy of truth theories up to any level ω^ω consists in proving that PKF behaves classically on the languages $\mathscr{L}_\alpha$ for $\alpha < \omega^\omega$, that is, on the language $\mathscr{L}_{\omega^\omega} = \bigcup_{\alpha<\omega^\omega} \mathscr{L}_\alpha$.

As pointed out in §5, the logic of PKF is separated from classical logic only by the absence of a rule that allows one to shift formulas back and forth from the antecedent to the succedent and vice versa by affixing a negation symbol to the formula. We shall show that shifting formulas in this way is permissible for all formulas in $\mathscr{L}_{\omega^\omega}$. Actually we shall establish a slightly stronger claim in PKF, namely that PKF proves for every sentence of $\mathscr{L}_\alpha$ ($\alpha < \omega^\omega$) that either the sentence itself or its negation is true, that is, we shall prove:

$$\forall x(\mathrm{Sent}_\alpha(x) \to T\,x \vee \neg\, T\,x) \tag{18}$$

formally in PKF for all $\alpha < \omega^\omega$ by transfinite induction on α and side induction on the complexity of x. Once this is proved, it is easy to show that $\Rightarrow \varphi, \neg\varphi$ is derivable for all $\varphi \in \mathscr{L}_{\omega^\omega}$, which in turn implies by Lemma 15 that φ can be shifted from the antecedent to the succedent by affixing a negation symbol and vice versa.

The side induction for the proof of (18) is covered by the following lemma:

LEMMA 31. *The following sequents are provable in* PKF:

(i) $\mathrm{Sent}(x), T\,x \vee \neg\, T\,x \Rightarrow T\,\underset{\cdot}{\neg}x \vee \neg\, T\,\underset{\cdot}{\neg}x$
(ii) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\,x \vee \neg\, T\,x, T\,y \vee \neg\, T\,y \Rightarrow T(x\underset{\cdot}{\wedge}y) \vee \neg\, T(x\underset{\cdot}{\wedge}y)$
(iii) $\mathrm{Sent}(x), \mathrm{Sent}(y), T\,x \vee \neg\, T\,x, T\,y \vee \neg\, T\,y \Rightarrow T(x\underset{\cdot}{\vee}y) \vee \neg\, T(x\underset{\cdot}{\vee}y)$
(iv) $\mathrm{Var}(v), \mathrm{For}(x,v), \forall y\; T\,x(\dot{y}/v) \vee \neg\forall y\; T\,x(\dot{y}/v) \Rightarrow T\,\underset{\cdot}{\forall}vx \vee \neg\, T\,\underset{\cdot}{\forall}vx$
(v) $\mathrm{Var}(v), \mathrm{For}(x,v), \exists y\; T\,x(\dot{y}/v) \vee \neg\exists y\; T\,x(\dot{y}/v) \Rightarrow T\,\underset{\cdot}{\exists}vx \vee \neg\, T\,\underset{\cdot}{\exists}vx$

PROOF. As an example we prove (iv).

We abbreviate $\mathrm{Var}(v), \mathrm{For}(x,v)$ as $\Gamma(x,v)$ and start with PKF4(i).

$$\dfrac{\dfrac{\Gamma(x,v), \forall y\; T\,x(\dot{y}/v) \Rightarrow T\,\underset{\cdot}{\forall}vx}{\Gamma(x,v), \forall y\; T\,x(\dot{y}/v) \Rightarrow T\,\underset{\cdot}{\forall}vx, \neg\, T\,\underset{\cdot}{\forall}vx}\;\text{weakening 2}}{\Gamma(x,v), \forall y\; T\,x(\dot{y}/v) \Rightarrow T\,\underset{\cdot}{\forall}vx \vee \neg\, T\,\underset{\cdot}{\forall}vx}\;\text{Lemma 14}$$

$\Gamma(x,v)$ is purely arithmetical and therefore Corollary 17 applies to it, and we can leave it in its place in the antecedent in the second line of the following proof. The first line is PKF4(ii).

$$\dfrac{\dfrac{\dfrac{\Gamma(x,v), T\,\underset{\cdot}{\forall}vx \Rightarrow \forall y\; T\,x(\dot{y}/v))}{\Gamma(x,v), \neg\forall y\; T\,x(\dot{y}/v)) \Rightarrow \neg\, T\,\underset{\cdot}{\forall}vx}\;\neg\text{-rule}}{\Gamma(x,v), \neg\forall y\; T\,x(\dot{y}/v)) \Rightarrow T\,\underset{\cdot}{\forall}vx, \neg\, T\,\underset{\cdot}{\forall}vx}\;\text{weakening 2}}{\Gamma(x,v), \neg\forall y\; T\,x(\dot{y}/v)) \Rightarrow T\,\underset{\cdot}{\forall}vx \vee \neg\, T\,\underset{\cdot}{\forall}vx}\;\text{Lemma 14}$$

Now we apply Lemma 13 to the last lines of the two preceding proofs, respectively, to obtain (iv):

$$\Gamma(x,v), \forall y\; T\,x(\dot{y}/v) \vee \neg\forall y\; T\,x(\dot{y}/v) \Rightarrow T\,\underset{\cdot}{\forall}vx \vee \neg\, T\,\underset{\cdot}{\forall}vx.$$

The other cases can be dealt with in a similar manner. ⊣

According to the conventions introduced above, $\mathrm{Sent}_0(x)$ expresses that x is a sentence of the language $\mathscr{L}_0 = \mathscr{L}_{\mathrm{PA}}$. We now prove claim (18) for $\alpha = 0$.

LEMMA 32. $\mathsf{PKF} \vdash \forall x(\mathrm{Sent}_0(x) \rightarrow T\,x \vee \neg\, T\,x)$

PROOF. The lemma is proved by a formal induction on the buildup of x. The induction step is covered by Lemma 31; so we only need to prove the claim for the atomic sentences of $\mathscr{L}_{\mathrm{PA}}$. They are all of the form $s = t$ for closed terms s and t. So we can employ PKF1(i):

$$\dfrac{\dfrac{\mathrm{ClTerm}(x), \mathrm{ClTerm}(y), \mathrm{val}(x) = \mathrm{val}(y) \Rightarrow T\,x\underset{\cdot}{=}y}{\mathrm{ClTerm}(x), \mathrm{ClTerm}(y), \mathrm{val}(x) = \mathrm{val}(y) \Rightarrow T\,x\underset{\cdot}{=}y, \neg\, T\,x\underset{\cdot}{=}y}\;\text{weakening 2}}{\mathrm{ClTerm}(x), \mathrm{ClTerm}(y), \mathrm{val}(x) = \mathrm{val}(y) \Rightarrow T\,x\underset{\cdot}{=}y \vee \neg\, T\,x\underset{\cdot}{=}y}\;\text{Lemma 14}$$

Similarly we can use PKF1(ii):

$$
\begin{array}{c l}
\text{ClTerm}(x), \text{ClTerm}(y), T\, x \underset{\cdot}{=} y \Rightarrow \text{val}(x) = \text{val}(y) & \\
\hline
\neg\,\text{val}(x) = \text{val}(y) \Rightarrow \neg\, T\, x \underset{\cdot}{=} y, \neg\,\text{ClTerm}(x), \neg\,\text{ClTerm}(y) & \neg\text{-rule} \\
\hline
\neg\neg\,\text{ClTerm}(x), \neg\neg\,\text{ClTerm}(y), \neg\,\text{val}(x) = \text{val}(y) \Rightarrow \neg\, T\, x \underset{\cdot}{=} y & \text{Corollary 17} \\
\hline
\text{ClTerm}(x), \text{ClTerm}(y), \neg\,\text{val}(x) = \text{val}(y) \Rightarrow \neg\, T\, x \underset{\cdot}{=} y & \neg\neg\text{-sequents and cuts} \\
\hline
\text{ClTerm}(x), \text{ClTerm}(y), \neg\,\text{val}(x) = \text{val}(y) \Rightarrow T\, x \underset{\cdot}{=} y, \neg\, T\, x \underset{\cdot}{=} y & \text{weakening 2} \\
\hline
\text{ClTerm}(x), \text{ClTerm}(y), \neg\,\text{val}(x) = \text{val}(y) \Rightarrow T\, x \underset{\cdot}{=} y \vee \neg\, T\, x \underset{\cdot}{=} y & \text{Lemma 14}
\end{array}
$$

Combining the last lines of both derivations by Lemma 13 yields the following:

$$
\begin{aligned}
&\text{ClTerm}(x), \text{ClTerm}(y), \text{val}(x) = \text{val}(y) \vee \neg\,\text{val}(x) = \text{val}(y) \Rightarrow \\
&\qquad\qquad T\, x \underset{\cdot}{=} y \vee \neg\, T\, x \underset{\cdot}{=} y.
\end{aligned} \tag{19}
$$

According to Lemma 16, the sequent

$$\Rightarrow \text{val}(x) = \text{val}(y) \vee \neg\,\text{val}(x) = \text{val}(y)$$

is derivable. An application of the cut rule to (19) yields the following sequent:

$$\text{ClTerm}(x), \text{ClTerm}(y) \Rightarrow T\, x \underset{\cdot}{=} y \vee \neg\, T\, x \underset{\cdot}{=} y.$$

Then Lemma 31 and IND yield the claim. ⊣

We turn to the induction step.

LEMMA 33. *The following sequent is provable in* PKF:

$$\forall x(\text{Sent}_\alpha(x) \rightarrow T\, x \vee \neg\, T\, x) \Rightarrow \forall x(\text{Sent}_{\alpha+1}(x) \rightarrow T\, x \vee \neg\, T\, x).$$

PROOF. We start with PKF6(i):

$$
\begin{array}{c l}
\text{ClTerm}(x), T\,\text{val}(x) \Rightarrow T\, \underset{\cdot}{T} x & \\
\hline
\text{ClTerm}(x), T\,\text{val}(x) \Rightarrow T\, \underset{\cdot}{T} x \vee \neg\, T\, \underset{\cdot}{T} x & \text{Lemma 14}
\end{array}
$$

Similarly we employ PKF6(ii):

$$
\begin{array}{c l}
\text{ClTerm}(x), T\, T\, x \Rightarrow T\,\text{val}(x) & \\
\hline
\text{ClTerm}(x), \neg\, T\,\text{val}(x) \Rightarrow \neg\, T\, \underset{\cdot}{T} x & \neg\text{-rule} \\
\hline
\text{ClTerm}(x), \neg\, T\,\text{val}(x) \Rightarrow T\, \underset{\cdot}{T} x \vee \neg\, T\, \underset{\cdot}{T} x & \text{Lemma 14}
\end{array}
$$

Combining both yields by Lemma 13 the sequent

$$\text{ClTerm}(x), T\,\text{val}(x) \vee \neg\, T\,\text{val}(x) \Rightarrow T\, \underset{\cdot}{T} x \vee \neg\, T\, \underset{\cdot}{T} x.$$

From this one easily obtains the following:

$$\text{Sent}_\alpha(\text{val}(x)), T\, x \vee \neg\, T\, x \Rightarrow T(\underset{\cdot}{T} x \wedge \text{Sent}_\alpha(x)) \vee \neg\, T(\underset{\cdot}{T} x \wedge \text{Sent}_\alpha(x)).$$

Finally we employ Lemma 31 to establish the claim by formal induction on the buildup of x. ⊣

This implies also the following lemma:

LEMMA 34. $\forall \alpha < \beta(\text{Sent}_\alpha(x) \rightarrow T\, x \vee \neg\, T\, x) \Rightarrow (\text{Sent}_\beta(x) \rightarrow T\, x \vee \neg\, T\, x)$ *is provable in* PKF.

PROOF. If β is a successor ordinal this follows from the previous lemma. If β is a limit ordinal, the claim can be proved in a straightforward way because $\mathscr{L}_\beta$ is the union of all $\mathscr{L}_\alpha$ with $\alpha < \beta$. ⊣

In order to prove (18) from the previous lemma, we need to show transfinite induction up to any ordinal $\alpha < \omega^\omega$. The general scheme of transfinite induction is not provable in PKF. The corresponding rule, however, is provable.

LEMMA 35. *The following rule is a derived rule in* PKF *for any given natural number* k:

$$\frac{\Gamma, \forall\alpha < \beta\, \varphi(\alpha) \Rightarrow \varphi(\beta), \Delta}{\Gamma \Rightarrow \forall\alpha < \omega^k\, \varphi(\alpha), \Delta}$$

PROOF. The following proof is hardly innovative. However, we present the proof of transfinite induction in some detail, because it has to be carried out in the non-classical system PKF: care is needed when carrying out arguments familiar from classical logic in partial logic.

In the following we suppress Γ and Δ. In order to distinguish addition and multiplication on ordinals in the language $\mathscr{L}_{\mathrm{T}}$ from addition and multiplication on natural numbers we circle the former; thus $\oplus$ and $\odot$ are (formalizations of) functions pertaining to (codes of) ordinal numbers.

There is a function that sends any given natural number to the code of the corresponding ordinal. This function is expressed in $\mathscr{L}_{\mathrm{T}}$ by [...]

We assume that the following sequent is provable in PKF

$$\forall\alpha < \beta\, \varphi(\alpha) \Rightarrow \varphi(\beta) \tag{20}$$

and show from this by meta-induction on k that for every natural number k the following is provable in PKF:

$$\forall\alpha < \beta\, \varphi(\alpha) \Rightarrow \forall\alpha < \beta \oplus \omega^k\, \varphi(\alpha). \tag{21}$$

ω^k is fixed here, it is a numeral for a code of the respective ordinal. Therefore we do not have to apply the operation [...] and we do not have to overline k.

The case $k = 0$ is trivial: from (20) and

$$\forall\alpha < \beta\, \varphi(\alpha) \Rightarrow \forall\alpha < \beta\, \varphi(\alpha)$$

using some properties of ordinals in PKF we obtain

$$\forall\alpha < \beta\, \varphi(\alpha) \Rightarrow \forall\alpha < \beta \oplus \overline{[1]}\, \varphi(\alpha).$$

This covers the case $k = 0$ because $\omega^0 = 1$.

The induction step is established in the following way:

$$\forall\alpha < \beta\, \varphi(\alpha) \Rightarrow \forall\alpha < \beta \oplus \omega^k\, \varphi(\alpha) \qquad \text{ind. hyp.}$$

$$\forall\alpha < \beta \oplus (\omega^k \odot [n])\, \varphi(\alpha) \Rightarrow \forall\alpha < \beta \oplus (\omega^k \odot [n]) \oplus \omega^k\, \varphi(\alpha) \qquad \text{univ. inst.}$$

$$\forall\alpha < \beta \oplus (\omega^k \odot [n])\, \varphi(\alpha) \Rightarrow \forall\alpha < \beta \oplus (\omega^k \odot [n + \overline{1}])\, \varphi(\alpha)$$

$$\forall\alpha < \beta \oplus \omega^k \odot [\overline{0}]\, \varphi(\alpha) \Rightarrow \forall n \forall\alpha < \beta \oplus (\omega^k \odot [n])\, \varphi(\alpha) \qquad \mathsf{IND}$$

$$\forall\alpha < \beta\, \varphi(\alpha) \Rightarrow \forall\alpha < \beta \oplus \omega^{k+1}\, \varphi(\alpha)$$

This concludes the proof of the induction step and (21) is established. Substituting 0 for β in (21) yields the claim. ⊣

Now we can prove (18) by applying Lemma 35 to Lemma 34:

THEOREM 36. PKF *proves* $\forall x(\mathrm{Sent}_\alpha(x) \to T\,x \vee \neg\, T\,x)$ *for all* $\alpha < \omega^\omega$.

COROLLARY 37. *The set of sentences in* $\mathscr{L}_{\omega^\omega}$ *provable in* PKF *is closed under classical logic.*

PROOF. This follows from Theorem 36 and Lemma 15. ⊣

Corollary 37 shows that we can relatively interpret classical systems in PKF as long as the range of the interpretation does not not exceed the language $\mathscr{L}_{\omega^\omega}$.

In the next theorem we shall show that the truth predicates T_α behave like ramified truth predicates. Above $T_\alpha\, x$ has been defined as $T\, x \wedge x \in \mathscr{L}_\alpha$.

THEOREM 38. PKF *proves the following theorems for all* $\alpha < \omega^\omega$:

(i) $\forall x, y\big(\,\mathrm{ClTerm}(x) \wedge \mathrm{ClTerm}(y) \rightarrow (T_\alpha\, x \mathbin{\dot{=}} y \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$,

(ii) $\forall x\big(\mathrm{Sent}_\alpha(x) \rightarrow (T_\alpha\, \dot{\neg} x \leftrightarrow \neg\, T_\alpha\, x)\big)$,

(iii) $\forall x \forall y\big(\mathrm{Sent}_\alpha(x) \wedge \mathrm{Sent}_\alpha(y) \rightarrow (T_\alpha(x \dot{\wedge} y) \leftrightarrow T_\alpha\, x \wedge T_\alpha\, y)\big)$,

(iv) $\forall x \forall y\big(\mathrm{Sent}_\alpha(x) \wedge \mathrm{Sent}_\alpha(y) \rightarrow (T_\alpha(x \dot{\vee} y) \leftrightarrow T_\alpha\, a \vee T_\alpha\, y)\big)$,

(v) $\forall v \forall x\big(\mathrm{Var}(v) \wedge \mathrm{For}_\alpha(x, v) \rightarrow (T_\alpha\, \dot{\forall} v x \leftrightarrow \forall y\, T_\alpha\, x(\dot{y}/v))\big)$,
$\mathrm{Var}(v)$ *expresses that* v *is a variable*; $\mathrm{For}_\alpha(x, v)$ *expresses that* x *is a formula of* $\mathscr{L}_\alpha$ *with only* v *free*,

(vi) $\forall v \forall x\big(\mathrm{Var}(v) \wedge \mathrm{For}_\alpha(x, v) \rightarrow (T_\alpha\, \dot{\exists} v x \leftrightarrow \exists y\, T_\alpha\, x(\dot{y}/v))\big)$,

(vii) $\forall x \forall \beta < \alpha\big(\,\mathrm{ClTerm}(x) \rightarrow (T_\alpha\, \dot{T}_\beta x \leftrightarrow T_\beta\, \mathrm{val}(x))\big)$,

(viii) $\forall x \forall \beta < \alpha\big(\mathrm{Sent}_\beta(\mathrm{val}(x)) \wedge \mathrm{ClTerm}(x) \rightarrow (T_\alpha\, \dot{T}_\beta x \leftrightarrow T_\alpha\, \mathrm{val}(x))\big)$.

PROOF. By Corollary 37 and Lemma 15 we can freely shift formulas in $\mathscr{L}_{\omega^\omega}$ between the antecedent and the succedent. Thus the above clauses can be obtained from the rules PKF1–PKF8. ⊣

Theorem 38 shows that PKF interprets a system of ramified truth up to any ordinal below ω^ω. Systems of ramified truth were mentioned by Feferman [12] and studied by Halbach [17]. It is known that the system $\mathsf{RA}_{<\alpha}$ of ramified analysis $(\alpha \le \varepsilon_0))$ is equivalent to a system of ramified truth $\mathsf{RT}_{<\alpha}$ as described in Halbach [17]. This just generalizes the well known result that the theory of arithmetical comprehension ACA is equivalent to 'Tarskian' truth (see Feferman [12] or Halbach [17]) to transfinite iterations of these theories. Finally Theorem 38 proves that $\mathsf{RT}_{<\omega^\omega}$ is a subtheory of PKF. We skip the details.

THEOREM 39. *Ramified truth* $\mathsf{RT}_{<\omega^\omega}$ *and ramified analysis* $\mathsf{RA}_{<\omega^\omega}$ *up to any level below* ω^ω *and* PKF *prove the same arithmetical sentences.*

§7. PKF and classical logic. The philosophical significance of the proof-theoretic analysis of PKF depends on whether PKF is actually equivalent to all natural systems for Kripke's theory. The claim could fail in two ways: First, there could be natural axiomatizations of Kripke's theory that are stronger than PKF; second, PKF itself might not be natural and exceed the strength of all really natural axiomatizations.

Concerning the first possibility, we are fairly confident that PKF cannot be naturally enriched by further initial sequents and rules such that a stronger system is obtained. This is not a mathematically provable claim because the notion of a *natural* extension is not mathematically defined. It seems at least that PKF naturally captures the compositional nature of truth in Kripke's theory and allows us to iterate truth in a straightforward way. Thus it seems that the essential truth-theoretic intuitions behind Kripke's theory are captured in PKF. It therefore seems implausible to assume the existence of formalizations of Kripke's theory that properly extend PKF.

In favor of the second possibility—that PKF is actually too strong and not natural— one might argue that PKF achieves its proof-theoretic strength only through a connective that cannot be iterated and that does not obey the rules of Strong Kleene logic. For instance, $\varphi \Rightarrow \varphi$ is derivable for *all* formulas φ of $\mathscr{L}_{\mathrm{T}}$, even for the liar sentence. Thus $\Rightarrow$ is a connective exceeding pure partial logic.

However, this objection is rooted in a misunderstanding of the interpretation of the sequent arrow, which is not a logical connective in the object language but a derivation sign of the metalanguage. This is shown by the fact, mentioned above, that we could have presented PKF in a natural deduction system. Such a system would differ from classical logic in not allowing an unrestricted rule for introducing material implication.

Another much more general objection against PKF might be raised: Feferman rejected the use of partial logic for seemingly good reasons. Strong Kleene logic is simply impractical. Replacing classical logic by partial logic seems to be a price that is much too high for a solution of the liar paradox.

Our reply essentially follows Kripke's defense of partial logic. For the language $\mathscr{L}_{\mathrm{PA}}$ we retain classical logic, as was shown in Corollary 18. We do not propose to carry out mathematics in partial logic. Only when the problematic truth predicate is added, we have to give up classical reasoning. Our proof-theoretic analysis in the previous section has added another aspect that supports this line of defense: The fragment of PKF that contributes to its proof-theoretic strength is purely classical. As has been shown in the previous section, one could completely dispense with partial logic by employing truth predicates with type indices. Feferman's KF is a further alternative: he regains classical logic without typing the truth predicate; but his KF describes a partial truth predicate that is very different from the actual classical semantics of KF.

§8. Truth and the reflective closure of PA. In this section we speculate on some consequences of our results for Feferman's proposal to employ KF as a system for capturing the reflective closure of PA. We start with a very brief sketch of the main motivation for considering the reflective closure of theories.

It has been argued that the acceptance of Peano arithmetic commits us implicitly also to the acceptance of a stronger system. For instance, the acceptance of PA commits us also to the acceptance of PA plus the consistency statement of PA or the uniform reflection principle for PA, which in turn commits us to the consistency (or uniform reflection principle) of this theory and so on. This addition of consistency statements and reflection principles can be iterated into the transfinite along those ordinals for which transfinite induction can be proved. Since the resulting systems will prove stronger instances of transfinite induction, the process can be extended further through so called autonomous progressions (see Feferman [10]) until the process closes off.

A canonical system containing all assumptions implicit in the acceptance of PA has been called the *reflective closure of* PA. In order to bypass the intricacies of the autonomous progressions and to provide a more perspicuous characterization of the reflective closure, Feferman has proposed several alternative techniques.

A particular method for describing the reflective closure consists in iterating the usual 'Tarskian' theory of truth. This leads to the hierarchy RT_α of truth theories. Like other theories of this kind (e.g., systems of ramified analysis), the Tarskian theory of truth can be iterated up to ε_0 (or further to Γ_0 in an autonomous reflection process).

Iterating reflection or comprehension principles or truth theories and the like usually relies on some notation system of the ordinals along which these principles are iterated (see Feferman [10], [11] and many further articles). These ordinals and the corresponding notations are explicitly used in the presentation of the reflective closure.

Feferman has tried to describe the reflective closure of PA (and other system) by systems that do not make explicit use of ordinals and notation systems. KF is such a system. Feferman [12, p. 1] explains the motivation behind his work on KF in the following way:

> [Gödel's incompleteness theorems] point to the possibility of systematically generating larger and larger systems whose acceptability is implicit in acceptance of the starting theory. The engines for that purpose are what have come to be called *reflection principles*. These may be iterated into the constructive transfinite, leading to what are called *recursive progressions of theories*. [...] [F]or some years I had hoped to give a more realistic and perspicuous finite generation procedure. [...] What is presented here [i.e., Feferman's work on KF and its variants] is a new and simple notion of the *reflective closure of a schematic theory* which can be applied quite generally.

In contrast to the systems RT_α of ramified truth, the presentation of KF does not require truth predicates indexed by ordinal notations; rather KF has only one single truth predicate. Thus it is tempting to see KF as a system that captures the content of the reflective closure of PA in one fell swoop instead of approaching it by iterations and ramifications.

The proof-theoretic analysis of KF shows that the ordinal levels are still present in KF; but they are not explicit, rather they are only revealed by the analysis of the system. Therefore the system can be presented in a smooth way that does not require a notation system for ordinals, ramified languages or the like.

On the semantical side, Kripke's theory of truth might be seen as a type-free version of Tarski's hierarchy of truth predicates up to the first non-recursive ordinal. If KF is an axiomatization of Kripke's theory it should achieve the same as the axiomatization of the of ramified theory of truth $\mathsf{RT}_{<\varepsilon_0}$ up to ε_0. Of course in terms of proof-theoretic strength, KF is successful because it matches $\mathsf{RT}_{<\varepsilon_0}$.

However, our results above make it less plausible that KF only sums up what can also be achieved in many steps by ramified systems of truth. For KF is not an axiomatization of Kripke's theory of truth but rather an axiomatization of the 'closed off' version of Kripke's theory. In proving certain arithmetical sentences KF makes essential detours through truth-valueless territory, because it has the resources to distinguish between truth, falsity and indeterminateness, whereas on Kripke's original partial approach the lack of a truth value cannot be expressed within the language. KF is, so to speak, not an axiomatization of Kripke's theory of truth, but rather of large parts of its metatheory. If the resources for distinguishing between truth, falsity and indeterminateness are dropped and a thoroughly partial system is adopted, proof-theoretic strength is lost. PKF no longer captures the mathematical content of $\mathsf{RT}_{<\varepsilon_0}$.

Theorem 39 shows that PKF can only reflect the iteration of Tarskian truth up to any level below ω^{ω}. Only the transition to classical logic, that is, to KF yields a system that exhausts the strength of the reflective closure of PA as characterized by the iteration $\mathsf{RT}_{<\varepsilon_0}$ of Tarskian truth (or even of $\mathsf{RT}_{<\Gamma_0}$ in the case of Feferman's [12] Strong Reflective Closure). But KF is, as we have argued, neither a natural axiomatization of Kripke's theory nor can it be reduced to any such axiomatization. So we do not see how an axiomatic theory of truth should form a useful tool in the analysis of the reflective closure. PKF is intuitively a sound way to make explicit some implicit assumptions of PA. But PKF is weaker than, for instance, systems of autonomously iterated truth theories or other accepted characterizations of the reflective closure. PKF is therefore too weak to be useful in a characterization of the reflective closure of PA. KF (or one of its variants) is strong enough, but it imports commitments that are not implicit in the acceptance of PA.

§9. Comparison with other axiomatic theories of truth. This article was mainly devoted to a comparison of the classical system KF and the system PKF formulated in strong Kleene logic. In this final section we compare PKF with other truth systems. The alternative axiomatic theories of truth that we consider in this section fall into one of the following three categories: The first category is constituted by systems for the same semantics: Kripke's theory of truth with the Strong Kleene scheme. Second, there are systems for Kripke's theory with the Strong Kleene scheme substituted by some other evaluation scheme. Finally we shall consider axiomatic systems describing altogether different semantical constructions.

First we shall consider strengthenings of KF and PKF whilst retaining the same underlying semantical construction. As mentioned above, Feferman [12] considers also a variant of KF, the *Strong* Reflective Closure of PA, which has the strength of $\mathsf{RA}_{<\Gamma_0}$ or iterated truth $\mathsf{RT}_{<\Gamma_0}$ up to Γ_0. One could try to define a system that relates to PKF in the way the Strong Reflective Closure Ref*(PA) of PA relates to the Weak Reflective Closure Ref(PA) in [12]. One could expect that thereby the proof-theoretic strength of PKF can be increased, but partial logic makes the formulation of a suitable system already difficult. However, we cannot exclude that a natural strengthening of PKF might be obtained in this manner.

Both KF and PKF are supposed to axiomatize Kripke's theory of truth formulated with the Strong Kleene evaluation scheme. Kripke and others have considered formulations of the theory with alternative evaluation schemes (see Belnap and Gupta [3], Kripke [24] and McGee [28]). We think that the Strong Kleene scheme is superior to other schemes in many aspects, but we do not attempt to justify this claim: a comparison of the different schemes in the context of Kripke's theory is beyond the scope of this paper.

A minor departure from the Strong Kleene scheme is the use of four-valued logic. This variant of the theory admits truth-value 'gluts' along with truth-value gaps. If truth-value gluts are allowed, the consistency axiom CONS of KF will have to be dropped, because sentences can be both true and false; this does not affect the proof-theoretic strength of KF. In the case of PKF slight modifications of the underlying logic will be required in order to allow gluts besides the gaps. In particular, the $\neg\lambda$-sequents will have to be dismissed as unsound. These sequents, however, are not needed for the interpretation of $\mathsf{RA}_{<\omega^{\omega}}$ in PKF. Moreover, CONS is only required in the proof of Theorem 27 because of the $\neg\lambda$-sequents. Therefore our proof-theoretic results would equally apply if CONS and the $\neg\lambda$-sequents were dropped in order to obtain systems for Kripke's theory with gaps and gluts.

It is also possible to retain gluts but to rule out gaps. Cantini [5] has proved several results on the variant of KF and subsystems of KF that are sound with respect to such an interpretation. Our main results would still not be affected by this modification.

The Weak Kleene evaluation scheme is an alternative evaluation scheme for partial logic, that is, non-classical logic with truth-value gaps only. According to this scheme a sentence is neither true nor false if and only if some component lacks a truth value or–for quantified sentences–a least one instance does not have a truth value. Feferman [12] mentions a variant of his theory that is based on this evaluation scheme and claims that it is as strong as the system based on the strong Kleene scheme. We could also modify our system PKF accordingly. Since the sentences without truth value are irrelevant for the interpretation of $\mathsf{RA}_{<\omega^\omega}$ we would expect that this variant of PKF is as strong as the original version. Again, our main result would apply to this modification.

The Weak Kleene scheme, however, has not found many supporters. The supervaluation scheme is regarded as more attractive. Cantini [6] proposed a system VF that stands in the same relation to Kripke's theory with supervaluations as KF stands to Kripke's theory with the Strong Kleene scheme. In particular, VF is sound with respect to 'closed off' supervaluational fixed point models. If we tried to formulate a system for Kripke's theory with supervaluations in analogy to PKF, we would have to provide a system for supervaluational logic in analogy to 4.1.4–4.1.5. But this is not possible. Woodruff [45] proved that the compactness theorem fails for the consequence relation in supervaluations; hence the consequence relation cannot be fully described by a formal system. Thus it is at least unclear what a natural axiomatization of Kripke's theory with supervaluations would look like or whether such an axiomatization exists at all. Consequently, we do not have anything to say about the proof-theoretic strength of such an axiomatization in comparison with Cantini's classical system VF.

Kripke's theory with the Strong Kleene scheme has a distinctive compositional flavor in contrast to the supervaluation version, which clearly fails to be compositional. Under supervaluational semantics the truth of a complex sentence does not merely depend on the semantical value of its constituents, at least if their semantical value is defined in a standard way. For instance, in supervaluations the liar sentence λ still lacks a truth value as does its negation; $\lambda \vee \neg\lambda$, however, will be declared true by supervaluations, while other disjunctions of sentences without truth value will fail to have a truth value.

Therefore KF has been advocated as a *compositional* theory of type-free truth. It has been conjectured that KF and its variants including Feferman's Strong Reflective Closure of PA might demarcate the limits of compositional truth. A straightforward axiomatization of Kripke's theory with supervaluations exceeds these limits; it is as strong as the theory ID_1 of non-iterated inductive definitions, as Cantini's [6] work on VF has shown. Halbach [18] has conjectured that the limits of compositionality thus coincide with the limits of predicativity. For no compositional theory exceeds the strength of $\mathsf{RA}_{<\Gamma_0}$, while Feferman's Strong Reflective Closure of PA exhausts the strength of predicative analysis $\mathsf{RA}_{<\Gamma_0}$.

The results of this paper shed some doubt on this picture: we deny that KF and its variants are satisfactory theories of truth. Moreover the presence of CONS renders KF a non-compositional theory, as we have argued in the light of Remark 4, which shows that KF decides the liar sentence and that KF proves a sentence whose truth does not rely on the semantic status of its components in KF.

Moreover, we have seen that even if CONS is dropped from KF, one is left with a theory of truth in which the outer logic necessarily differs from the inner logic. Therefore PKF assumes the rôle of the strongest type-free theory of truth currently on the market that is still compositional. Thus the proof-theoretic limits of compositionality would still be unclear. As we have shown, PKF, which is in our view the natural axiomatization of Kripke's theory with the Strong Kleene scheme, is much weaker than certain predicative theories like $\mathsf{RA}_{<\Gamma_0}$, because PKF has only the strength of the system $\mathsf{RA}_{<\omega^\omega}$ of ramified analysis up to any level smaller than ω^ω. There still might be unexplored axiomatic systems of compositional truth that exceed the strength of PKF; but we are not aware of such systems.

All the systems of type-free truth we have discussed so far rely on non-classicality either in the inner or their outer logic. Systems that rely entirely on classical logic have been proposed as well. The strongest thoroughly classical theory of truth to date is the Friedman-Sheard system FS.[14] The arithmetical strength of FS is given by the theory $\mathsf{RA}_{<\omega}$ (that is, finitely iterated elementary comprehension) or the ramified theory of truth of all finite levels. Thus, FS is mathematically much weaker than PKF, because the latter proves also transfinite iterations. Halbach and Horsten [20] argued that there is little hope to obtain a consistent, natural and thoroughly classical theory of truth that allows us to prove transfinite iterations of truth.

Although we have been extolling the virtues of PKF, there is a price to pay: PKF is a decidedly *nonclassical* system. This seems to be the price for having transfinite iterations of truth. KF obscures this fact because it is formulated in classical logic and yields iterations of truth up to ε_0, but this is only achieved by 'closing off'; internally KF relies on partial logic, and this creates the unacceptable asymmetry between inner and outer logic.

REFERENCES

[1] Hiroshi Aoyama, *The strong completeness of a system based on Kleene's strong three-valued logic*, ***Notre Dame Journal of Formal Logic***, vol. 35 (1994), pp. 355–368.

[2] JC Beall and Brad Armour-Garb, ***Deflationism and paradox***, Clarendon Press, Oxford, 2005.

[3] Nuel Belnap and Anil Gupta, ***The revision theory of truth***, MIT Press, Cambridge, 1993.

[4] Stephen Blamey, *Partial logic*, ***Handbook of philosophical logic*** (Dov M. Gabbay and Franz Guenthner, editors), vol. 5, Kluwer, Dordrecht, second ed., 2002, pp. 261–353.

[5] Andrea Cantini, *Notes on formal theories of truth*, ***Zeitschrift für mathematische Logik und Grundlagen der Mathematik***, vol. 35 (1989), pp. 97–130.

[6] ———, *A theory of formal truth arithmetically equivalent to ID_1*, this Journal, vol. 55 (1990), pp. 244–259.

[7] ———, ***Logical frameworks for truth and abstraction. An axiomatic study***, Studies in Logic and the Foundations of Mathematics, vol. 135, Elsevier, Amsterdam, 1996.

[14] The system FS was first presented by Friedman and Sheard [13] (under a different name). It was further analyzed by Halbach [16] and defended by Halbach and Horsten [20].

[8] JOHN CLEAVE, *The notion of logical consequence in the logic of inexact predicates*, ***Zeitschrift für mathematische Logik und Grundlagen der Mathematik***, vol. 20 (1974), pp. 307–324.

[9] WILLIAM CRAIG, *Replacement of auxiliary expressions*, ***Philosophical Review***, vol. 65 (1956), pp. 38–55.

[10] SOLOMON FEFERMAN, *Transfinite recursive progressions of axiomatic theories*, this JOURNAL, vol. 27 (1962), pp. 259–316.

[11] ———, *Systems of predicative analysis*, this JOURNAL, vol. 29 (1964), pp. 1–30.

[12] ———, *Reflecting on incompleteness*, this JOURNAL, vol. 56 (1991), pp. 1–49.

[13] HARVEY FRIEDMAN and MICHAEL SHEARD, *An axiomatic approach to self-referential truth*, ***Annals of Pure and Applied Logic***, vol. 33 (1987), pp. 1–21.

[14] ———, *The disjunction and existence properties for axiomatic systems of truth*, ***Annals of Pure and Applied Logic***, vol. 40 (1988), pp. 1–10.

[15] MICHAEL GLANZBERG, *Minimalism, deflationism, and paradoxes*, in [2, 107–132], 2005.

[16] VOLKER HALBACH, *A system of complete and consistent truth*, ***Notre Dame Journal of Formal Logic***, vol. 35 (1994), pp. 311–327.

[17] ———, ***Axiomatische Wahrheitstheorien***, Akademie Verlag, Berlin, 1996.

[18] ———, *Truth and reduction*, ***Erkenntnis***, vol. 53 (2000), pp. 97–126.

[19] ———, *Axiomatic theories of truth*, ***Stanford Encyclopedia of Philosophy*** (Edward N. Zalta, editor), Spring 2006, forthcoming, `http://plato.stanford.edu/archives/spr2006/entries/truth-axiomatic/`.

[20] VOLKER HALBACH and LEON HORSTEN, *The deflationist's axioms for truth*, in [2, 203–217], 2005.

[21] CARL HEMPEL, *The theoretician's dilemma*, ***Minnesota studies in the philosophy of science, II*** (Herbert Feigl et al., editor), University of Minnesota press, 1958, pp. 37–98.

[22] JOHN KEARNS, *The strong completeness of a system for Kleene's three-valued logic*, ***Zeitschrift für mathematische Logik und Grundlagen der Mathematik***, vol. 25 (1979), pp. 61–68.

[23] MICHAEL KREMER, *Kripke and the logic of truth*, ***Journal of Philosophical Logic***, vol. 17 (1988), pp. 225–278.

[24] SAUL KRIPKE, *Outline of a theory of truth*, ***Journal of Philosophy***, vol. 72 (1975), pp. 690–712, reprinted in [25].

[25] Robert L. Martin (editor), ***Recent essays on truth and the Liar paradox***, Clarendon Press and Oxford University Press, Oxford and New York, 1984.

[26] ROBERT L. MARTIN and PETER W. WOODRUFF, *On representing 'true-in-L' in L*, ***Philosophia***, vol. 5 (1975), pp. 213–217, Reprinted in [25].

[27] TIM MAUDLIN, ***Truth and paradox. Solving the riddles***, Clarendon Press, Oxford, 2004.

[28] VANN MCGEE, ***Truth, vagueness, and paradox: An essay on the logic of truth***, Hackett Publishing, Indianapolis and Cambridge, 1991.

[29] ———, *In praise of the free lunch: Why disquotationalists should embrace compositional semantics*, manuscript, 2003.

[30] RICHARD MONTAGUE, *Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability*, ***Acta Philosophica Fennica***, vol. 16 (1963), pp. 153–67, Reprinted in [31, 286–302].

[31] ———, ***Formal philosophy: Selected papers of Richard Montague***, Yale University Press, New Haven and London, 1974, Edited and with an introduction by Richmond H. Thomason.

[32] YIANNIS N. MOSCHOVAKIS, ***Elementary induction on abstract structures***, Studies in Logic and the Foundations of Mathematics, no. 77, North-Holland and Elsevier, Amsterdam, London and New York, 1974.

[33] CHARLES PARSONS, *On n-quantifier induction*, this JOURNAL, vol. 37 (1972), pp. 466–482.

[34] WOLFRAM POHLERS, ***Proof theory. An introduction***, Springer, Berlin, 1989.

[35] WILLIAM REINHARDT, *Remarks on significance and meaningful applicability*, ***Mathematical logic and formal systems. A collection of papers in honor of Professor Newton C.A. Da Costa*** (Luiz Paulo de Alcantara, editor), Lecture Notes in Pure and Applied Mathematics, vol. 94, Marcel Dekker inc., 1985, pp. 227–242.

[36] ———, *Some remarks on extending and interpreting theories with a partial predicate for truth*, ***Journal of Philosophical Logic***, vol. 15 (1986), pp. 219–251.

[37] HELMUT SCHWICHTENBERG, *Proof theory: Some applications of cut-elimination*, ***Handbook of mathematical logic*** (Jon Barwise, editor), Studies in Logic and the Foundations of Mathematics, no. 90,

North-Holland, Amsterdam, London, New York and Tokyo, 1977, pp. 867–895.

[38] DANA SCOTT, *Combinators and classes*, ***λ-calculus and computer science*** (Corrado Böhm, editor), Lecture Notes in Computer Science, Springer, Berlin, 1975, pp. 1–26.

[39] MICHAEL SHEARD, *Truth, probability, and naive criteria*, ***Principles of truth*** (Volker Halbach and Leon Horsten, editors), Dr. Hänsel-Hohenhausen, Frankfurt a.M., 2002, pp. 169–181.

[40] SCOTT SOAMES, ***Understanding truth***, Oxford University Press, New York and Oxford, 1999.

[41] GAISI TAKEUTI, ***Proof theory***, second ed., North Holland, Amsterdam, 1987.

[42] ANNE S. TROELSTRA and HELMUT SCHWICHTENBERG, ***Basic proof theory***, Cambridge Tracts in Theoretical Computer Science, no. 43, Cambridge University Press, Cambridge, 1997.

[43] ALBERT VISSER, *Semantics and the liar paradox*, ***Handbook of philosophical logic*** (Dov Gabbay and Franz Günthner, editors), vol. 4, Reidel, Dordrecht, 1989, pp. 617–706.

[44] HAO WANG, *The calculus of partial predicates and its extension to set theory*, ***Zeitschrift für mathematische Logik und Grundlagen der Mathematik***, vol. 7 (1961), pp. 283–288.

[45] PETER W. WOODRUFF, *On supervaluations in free logic*, this JOURNAL, vol. 49 (1984), pp. 943–950.

NEW COLLEGE
OXFORD OX1 3BN
UNITED KINGDOM
E-mail: volker.halbach@philosophy.ox.ac.uk

UNIVERSITY OF LEUVEN
INSTITUTE OF PHILOSOPHY
KARDINAAL MERCIERPLEIN 2
B-3000 LEUVEN, BELGIUM
E-mail: Leon.Horsten@hiw.kuleuven.be