The Review of Symbolic Logic
Volume 15, Number 1, March 2022

THE POWER OF NAIVE TRUTH

HARTRY FIELD
New York University

Abstract. Nonclassical theories of truth that take truth to be transparent have some obvious advantages over any classical theory of truth (which must take it as nontransparent on pain of inconsistency). But several authors have recently argued that there's also a big disadvantage of nonclassical theories as compared to their "external" classical counterparts: proof-theoretic strength. While conceding the relevance of this, the paper argues that there is a natural way to beef up extant internal theories so as to remove their proof-theoretic disadvantage. It is suggested that the resulting internal theories are preferable to their external counterparts.

As is well known, Kripke's (1975) fixed point construction for truth based on the Strong Kleene semantics suggests a number of different axiomatized truth theories. A prominent division among them is between the *external theories* and the *internal theories*. Internal theories are, roughly, those whose only theorems are members of (some or all) fixed points of Kripke's construction; whereas external theories provide a certain kind of external commentary on the fixed points. The internal theories based on the Kleene semantics respect *the naivety of truth*: roughly, the equivalence between any sentence and the attribution of truth to it. (Also known as *transparency*.) The external theories don't respect naivety. Related to this, the internal theories seem more faithful to the intuitions behind the Kripke construction: as Halbach and Horsten put it (2006), the external theories are "not sound with respect to Kripke's semantics in the straightforward sense" (p. 677). *Prima facie*, these seem like strong points in favor of the internal theories.

But as Halbach and Horsten argue in that paper, the standard internal theories are substantially weaker than their external counterparts in their nontruth-theoretic consequences. Halbach (2011) spells this out further, and while conceding the *prima facie* advantages of the internal theories over the external, says that the extra strength of the external theories (as regards nontruth-theoretic consequences) is a factor that far outweighs that. The same conclusion, on the same basis, is reached in Feferman (2012).[1]

I concede to these authors that the extra strength of the external theories is highly desirable. My aim in this paper is to show that this extra strength can be attained

Received: July 19, 2019.
2020 *Mathematics Subject Classification*: 03A05, 03B20, 03B50.
Key words and phrases: truth, fixed point, Saul Kripke, KF, PKF, Volker Halbach, Leon Horsten, Solomon Feferman, proof-theoretic strength, predicativity, schematic theory.

[1] Halbach & Horsten (2006) took a different view (p. 710 top): that while the proof-theoretic advantages of the external theories over the internal are striking, still the external theories are simply not satisfactory as theories of truth.

doi:10.1017/S1755020320000167

(and in a natural way) in an internal framework which still respects the naivety of truth. (That means that its logic must be nonclassical; more on this in a moment.) I first show how to achieve this in connection with the external theory KF^+ ("extended Kripke-Feferman"), which has first order Peano arithmetic as its basis: my strategy is to provide an internal theory INT, also based on first order PA, in which KF^+ can be interpreted. I then turn to Feferman's (1991) schematic extension of KF^+, which I'll call S-KF^+; I provide an analogous schematic internal theory, S-INT, in which S-KF^+ can be interpreted.

Feferman showed that classical predicative analysis can be interpreted in S-KF^+, so when that deep result is combined with the (shallower) result to be presented we get a striking conclusion:

(*) One can do all of classical predicative analysis within a nonclassical internal theory based on arithmetic.

The situation is analogous for theories that add a truth predicate to richer theories such as ZFC (Zermelo-Fraenkel set theory with choice): the method I'll be advocating would yield a rather strong predicative theory of self-applicative properties over the ZFC sets, which I think is a natural version of what Feferman (1991) calls the "reflective closure" of ZFC. But I'll confine my attention here to the arithmetic case, where it's clearer what the target is (viz., predicative analysis).

Feferman also repeatedly made a somewhat separate complaint against internal theories: one cannot carry out "sustained ordinary reasoning" within the logics they employ.[2] The results of this paper also undercut *much of* the force of this. For my conclusion is really stronger than (*): since S-INT allows us to interpret S-KF^+ (and by a very simple translation, as we'll see), and that in turn allows us to interpret predicative analysis, we get

(**) One can carry out what is essentially *the same reasoning as employed in the classical theory S-KF^+* within a nonclassical internal theory.

The trick is that S-INT allows us to carve out a substantial "classical core"; *within that substantial core* it is possible to carry out whatever sustained ordinary reasoning we can carry out classically. Feferman's point about sustained ordinary reasoning *outside the core* remains; but the core is so substantial that I think much of the worry is undercut.

§1. What is to be done? The basic idea will be to extend the usual internal theories by adding an extra predicate "*Scl*," read "strongly classical." This predicate itself is to behave classically (in particular, it is to obey excluded middle).[3] The "truth predicate"

[2] It's easy to ascertain meta-theoretically what inferences they validate, but the need to go meta-theoretic is unattractive.

[3] It is this requirement that leads me to speak of *strong* classicality. In a truth theory within the internal logic I prefer, it's natural to regard a sentence x as "classical" if $True(x) \vee \neg True(x)$. But then we can never prove (or even legitimately assert) of a sentence B that it *isn't* classical: that would amount to asserting $\neg[True(\langle B \rangle) \vee \neg True(\langle B \rangle)]$, which in the logic in question entails the contradiction $\neg True(\langle B \rangle) \wedge True(\langle B \rangle)$, which in turn entails any absurdity one chooses. (We can't even prove or legitimately assert that *there are* nonclassical sentences, on this definition of classicality.) To get around this, we want a notion of strong classicality for which the strong classicality of x entails $True(x) \vee \neg True(x)$, but not conversely; and for which we can prove of many sentences that they are strongly classical and of many others

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

of the external theory is then to be interpreted within the corresponding internal theory, roughly as "strongly classically true," that is, "both true and strongly classical," which I will abbreviate as "*Strue*." The laws governing "*Scl*" will be such that the defined predicate "*Strue*" also behaves classically (in particular, obeys excluded middle). As a result, classical logic applies fully among sentences that only contain "true" in contexts of form "strongly classical and true." Under basically this translation, all the laws of the external theory will be validated. This means that the proof-theoretic strength of the external theory is attained within the internal theory.

I don't claim a lot of pretheoretic clarity for the notion of strong classicality: indeed, it is a notion that can be filled out in various incompatible ways, and I will build into it only those features needed for the purpose of interpreting the external theories. For instance, the theory will be neutral as to whether truth-teller sentences, or sentences asserting that they are either true or not true, are strongly classical. But it is *compatible* with the theory to be presented to read "strongly classical" as "grounded" in the sense of (the Strong Kleene version of) Kripke (1975).[4]

How will the paradoxes be treated? Let's consider both a "genuine Liar sentence" (which I'll abbreviate as) Q, which asserts its own untruth, and an "external Liar sentence" Q^*, which asserts that it itself isn't both true and strongly classical (in whatever way that notion of strong classicality is filled out). Regarding the latter:

(A) The internal theories INT and S-INT will diagnose Q^* as true but not strongly classical.[5]

This is reminiscent of, but seems far more attractive than, the treatment of the Liar sentence in external theories like KF^+. In KF^+ (because "true" amounts to what the internal theory calls "strongly classically true") the Liar sentence is *asserted but simultaneously declared untrue*. (As Halbach and Horsten say, KF^+ "disproves its own soundness" (2006, p. 682).)

The internal theories will treat Q very differently from Q^*. INT and S-INT will employ the Strong Kleene logic K_3, to be described shortly; it restricts excluded middle while keeping modus ponens for $\supset$ (where $\supset$ is defined in terms of $\neg$ and $\vee$ in the usual way). With this logic,

(B) "$True(\langle Q\rangle) \vee \neg True(\langle Q\rangle)$" is not a theorem: indeed it is an *anti-theorem*, in the sense that it entails everything including absurdities like "$0 = 1$." The disjuncts are likewise anti-theorems; so one must reject them.

Whereas the external theories declare a sentence stating its own "untruth" as "not true," but assert it nonetheless, the internal theories reject the sentence and also reject its untruth, since its untruth is equivalent to the sentence itself. (They also reject its truth, which they take to likewise be equivalent.) I think this a much more appealing way to use "true": as many have argued, the equivalence of a sentence to the claim that it is true seems quite central to the uses to which a truth predicate is put.

Be that as it may, the point will be to construct theories INT and S-INT that take such a naive truth predicate as basic, but define (from it together with a predicate "*Scl*") a

that they aren't. And to allow for the interpretation of the external theories, we want strong classicality to be a classical predicate.

[4] Or rather: to read it this way when it is applied to sentences that don't themselves contain 'strongly classical'.

[5] At least, as not strongly classical *in the same sense as used in* Q^*: see note 35.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

notion of strongly classical truth that works entirely in accord with the external theory, and thus preserves the deductive power of the external theory.

The view is somewhat reminiscent of Feferman (2008), which involves a theory of truth together with a separate predicate "Determinate" applicable to sentences for which "things work nicely." The difference is that his truth theory is classical throughout, and "things work nicely for determinate sentences" means that truth is naive for them; whereas for me it's the reverse, we have naive truth everywhere and "things work nicely" means that instances of the formula obey excluded middle and hence are fully classical. I believe this latter view is preferable: on it, we get the full deductive power of the external theory within a philosophically attractive naive theory of truth.[6]

I think the view to be presented is one that should have been appealing to a slightly later time-slice of Feferman. For he says in 2012 (p. 189) that the notion of truth employed in $\mathrm{KF}^{(+)}$ and S-$\mathrm{KF}^{(+)}$ is not the philosophically significant notion: he says that it doesn't really stand for truth, but for *grounded truth* in Kripke's sense.[7] (Obviously he intended a qualification analogous to that in note 4 above.) But he provides no axiomatization of the philosophically significant notion. The present paper can be viewed as filling out his suggestion, by simultaneously axiomatizing the philosophically significant notion of truth and a notion of strong classicality that can but needn't be read as groundedness, so that something akin to grounded truth can be defined in a way that validates his axioms.

That basically is the paper. The next section contains an explanation of the internal/external distinction and some details about the Halbach–Horsten–Feferman argument; the rest of the paper gives the details of the response.

§2. Internal and external theories. There are three main logics based on the Strong Kleene three-valued evaluation rules.[8] (I'll call the three Kleene values 0, $\frac{1}{2}$ and 1, where 1 is the value we give to classical truths and 0 the value we give to classical falsehoods.)

[6] See also Halbach and Fujimoto (in preparation): their theory, like Feferman (2008), is a classical truth theory with a separate determinateness predicate, in whose scope truth behaves naively. Interestingly, their axioms for determinateness are quite similar to mine for strong classicality. (We developed them independently, so I take this as a sign of their intuitive appeal.) Their system strikes me as rather more natural than Feferman (2008), both in its determinateness axioms and in its underlying truth theory. (Theirs, unlike Feferman's but like mine, contains all standard composition principles for truth; and it doesn't have the defect of declaring itself unsound, though it does fail to declare some of its theorems true.) The proof-theoretic strength of their system exceeds that of KF but is still far short of S-KF; presumably a schematic variant matches that of S-KF, and hence of the theory S-INT to be presented below. Seeing their system does not change my verdict that it's best to attain that proof theoretic strength in a fully naive theory.

[7] By contrast, Halbach & Nicolai (2018) insist (p. 241 among other places) that the concept of truth employed in $\mathrm{KF}^{(+)}$ is the same as that employed in the internal theories. Presumably this means that when the internal theory rejects the sentences that the external theory accepts, the dispute is genuine; which is what Feferman appears to have been denying.

[8] I focus on Strong as opposed to Weak Kleene, both because the connectives and quantifiers of the latter seem unnatural, and because as Saul Kripke pointed out to me, Weak Kleene is just a sublogic of Strong Kleene (one with quite limited expressive power). More specifically, its disjunction is definable in Strong Kleene as $(A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge B)$, and its existential quantification (Kripke credits this to Brian Porter) as $\exists x Ax \wedge \forall x(Ax \vee \neg Ax)$.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

My favorite is the logic K_3, that declares an inference valid when in every model where the premises have value 1, so does the conclusion. It is the dual to Priest's "Logic of Paradox" LP (Priest, 1998), which declares an inference valid iff in every model where the premises have nonzero value, so does the conclusion. S_3 is the "symmetric" logic, whose valid inferences are those that are valid both in K_3 and in LP: in other words, those where in every model the value of the conclusion is at least the minimum of the values of the premises. In K_3, the law of excluded middle isn't valid, but the rule of explosion is. (That's the rule that contradictions imply anything; it's equivalent in the current context to the rule of disjunctive syllogism, $A \vee B, \neg A \vdash B$; which in turn is equivalent to Modus Ponens for $\supset$.) In LP it's the other way around: excluded middle is valid, explosion (and Modus Ponens for $\supset$) isn't. In S_3 neither is valid; but their common content, the rule $A \wedge \neg A \vdash B \vee \neg B$, is valid. I favor K_3, but little will turn on this. Still, it's nice not to have to conduct the discussion in a general way, so I'll assume it for purposes of this paper.[9]

Whichever of these three logics one prefers, there are serious issues of expressive inadequacy: the logic does not contain conditionals adequate to our needs, or a satisfactory restricted universal quantifier (for saying "All A are B" when A and B are nonclassical formulas). Much of my work in recent years has been on extending K_3 to address these limitations. In the present paper I want to avoid that: I'll work entirely within K_3. Some things would be smoother with an added conditional, but the overall complexity would bury the basic idea. It's possible here to avoid added conditionals because once the strong classicality predicate is added, the task of recovering the proof-theoretic strength of external theories is done within the strongly classical part, and here the ordinary Kleene $\supset$ suffices. The additional conditional or conditionals are still important to the overall theory, e.g. for the theory of restricted quantification outside the strongly classical realm, but not for the part needed to respond to Halbach, Horsten and Feferman.

My internal theories will include a truth theory over Peano arithmetic, based on the logic K_3. (The arithmetic enables us to develop a syntactic theory for sentences, taken to be the bearers of truth, via a Gödel numbering.) The truth theory will be "naive," which means in part that for each sentence B of the language (including those containing "True"), it will include each instance of these four rules (where the "A" is schematic for sentences, and $\langle A \rangle$ is the Gödel number of A):

T-Elim: $True(\langle A \rangle) \vdash A$
¬T-Introd: $\neg A \vdash \neg True(\langle A \rangle)$
T-Introd: $A \vdash True(\langle A \rangle)$
¬T-Elim: $\neg True(\langle A \rangle) \vdash \neg A$

More generally, naivety means that for any sentence A, and any formula X_A in which A is a subsentence, if $X_{True(\langle A \rangle)}$ results from X_A by replacing one or more occurrences of A by $True(\langle A \rangle)$ then $X_{True(\langle A \rangle)} \vdash X_A$ and $X_A \vdash X_{True(\langle A \rangle)}$. This follows inductively from

[9] The problem of proof-theoretic strength that Halbach, Horsten and Feferman raise for internal theories based on Strong Kleene logic arises also for internal theories based on rather different logics such as Priest's LP, due to the restrictions imposed there on modus ponens (see Picollo, 2018) The solution that I suggest to the problem in the case of Strong Kleene theories can easily be adapted to apply to LP-based theories.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

the four listed rules together with the laws of K_3. I'll refer to it as *the intersubstitutability of* $True(\langle A\rangle)$ *with* A *in inference*.

Many generalizations about truth, e.g. compositional laws, will also be part of the theory; and the arithmetical induction rule will extend to formulas containing "*True*."

By standard techniques one can construct a Liar sentence, which I'll abbreviate as "Q"; its central feature is that it's equivalent to "$\neg True(\langle Q\rangle)$," and so in particular

$$Q \vdash \neg True(\langle Q\rangle), \text{ and}$$
$$\neg True(\langle Q\rangle) \vdash Q.$$

Combining these with the previous, we get

$$Q \vdash \neg Q, \text{ and}$$
$$\neg Q \vdash Q.$$

Classically these are inconsistent, but using the more modest rules of S_3 they lead only to the conclusion

$$Q \vee \neg Q \vdash Q \wedge \neg Q.$$

In the K_3 that I prefer, we get the sharper conclusion

$$Q \vee \neg Q \vdash \bot$$

where $\bot$ is an absurdity (say, $0=1$). Thus using naive truth in K_3, $Q \vee \neg Q$ is an *anti-theorem*: it implies absurdities. Indeed the disjuncts Q and $\neg Q$ are themselves anti-theorems. (If I'd used only the weaker S_3, $Q \vee \neg Q$ would not be an anti-theorem: it would still imply the contradiction $Q \wedge \neg Q$, but whereas in K_3 contradictions imply absurdities, they don't in S_3.) The fact that in K_3 $Q \vee \neg Q$ is an anti-theorem does not mean that its negation is a theorem: $\neg$-Introduction is not a valid rule in K_3. Indeed, not only is $\neg(Q \vee \neg Q)$ not a theorem, it is an anti-theorem of K_3 too, since it is equivalent to $Q \wedge \neg Q$.

What's called the Kripke–Feferman theory (KF) is sort of an "external analog" of the naive truth theory based on S_3: where $A_1, ..., A_n \vdash B$ in S_3, KF proves $True(\langle A_1\rangle) \wedge \cdots \wedge True(\langle A_n\rangle) \supset True(\langle B\rangle)$.[10] For instance, KF proves $True(\langle Q \vee \neg Q\rangle) \supset True(\langle Q \wedge \neg Q\rangle)$. KF^+ adds to KF the principle

(CONSIS) $\neg\exists x[True(x) \wedge True(neg(x))]$

(where when x is the Gödel number of a formula, $neg(x)$ is the Gödel number of its negation; and when x isn't the Gödel number of a formula, $neg(x)$ isn't either). This is an "external analog of" the explosion rule, so KF^+ is an "external analog of" naive truth theory based on K_3. KF^+ thus derives $\neg True(\langle Q \vee \neg Q\rangle)$, from which it derives $\neg True(\langle Q\rangle)$ and $\neg True(\langle \neg Q\rangle)$.

Feferman (1991) showed that KF^+ (and even KF) is powerful enough to interpret ramified analysis up to the ordinal ε_0 but no further. He also suggested a natural way to beef up KF or KF^+ with a schematic variable for the arithmetic induction schema,

[10] The technical result of Halbach & Horsten 2006 shows that the converse fails. They say that KF can be viewed as the theory of the "closed off" Kripke construction that Kripke mentions late in his paper, though this is probably more true of the KF^+ to be mentioned next.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

to yield theories I'll call Schematic-KF and Schematic-KF$^+$[11]; and he showed that these allow the development of ramified analysis up to the much larger ordinal Γ_0 (the Feferman–Schütte ordinal), i.e. what's called "predicative analysis." The usual internal theories based on arithmetic are called PKF and PKF$^+$, and Halbach & Horsten (2006) showed that they yield ramified analysis only up to the much smaller level ω^ω. This has led Halbach (2011) and Halbach & Nicolai (2018) to not only question these particular internal theories but to strongly suggest that *any* theories based on nonclassical logic are inadequate to mathematics (even if they assume excluded middle for standard mathematical predicates and restrict it only for "*True*"): e.g. the latter conclude their paper (p. 251) by saying "We shouldn't expect that the effects of restricting classical logic for use with the truth predicate can be contained."

It would be possible to contest the significance of this argument from proof-theoretic strength. While we certainly want the results of ramified analysis of these higher levels, one might argue that this is just because we accept those results independently of the notion of truth: e.g., because we accept a standard set theory like ZFC, from which ramified analysis at all levels certainly follows. In that case, the real test isn't theories that add a truth theory to PA, but theories that add it to a much more powerful theory such as set theory. Presumably the results about excess strength carry over: adding truth in a KF-like (or KF$^+$-like) way to ZFC will result in more consequences in the language of ZFC than adding it in a way that corresponds to extant internal theories. But it might be less obvious how important that is: if you think that ZFC already contains all the math you need, then even a weak internal extension of it certainly does.

While this response is worth noting, I don't find it satisfying. It does seem to me (along the lines of various papers by Feferman) that there is an attractive project of "reflectively closing" theories by adding predicates for truth and related notions (together with natural principles governing these predicates). The goal of such reflective closures should be to capture "predicative reasoning over those theories"; and that includes far more than one gets in extant internal theories. (Indeed, it includes far more than one gets in the nonschematic theories KF and KF$^+$; but showing how to get up to ε_0 in an internal framework is a clue for how to go farther.) So I think that the challenge to internal theories that these authors raise should be taken seriously.

To this end, I'll begin (§3) by formulating a nonschematic internal theory INT, and show (§4 and §5) that it is sufficient to interpret KF$^+$, thus getting ramified analysis up to ε_0; in §6 I'll give a model-theoretic proof of its consistency.[12] In §7–9 I'll show that Feferman's use of schematic variables is equally available in an internal context, and leads to a consistent theory that interprets Schematic-KF$^+$ and hence full predicative analysis. §10 sketches some extensions, and contains further remarks about the philosophical import of the results.[13]

[11] What I'm calling Schematic-KF is his Ref*(PA(P)).

[12] The explosion rule will be used only to derive the interpretation of **(CONSIS)**, so the analogous internal theory based on S_3 will suffice for KF; and that itself is known to interpret ramified analysis up to the ordinal ε_0. That's why my choice to use K_3 rather than S_3 is inessential. (Consideration of LP would require a longer discussion.)

[13] I should note that there are "cheaper" ways of adding proof-theoretic strength to PKF. One, mentioned in Nicolai (2018), is simply to add induction up to ε_0 (or up to Γ_0) as a primitive principle. Another, considered in Fischer, Nicolai, & Horsten (2018), is to use

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

§3. The nonschematic internal theory INT. I've divided the formalization of INT into four parts: the logic, the arithmetic, the truth theory, and the theory of strong classicality. The first three will probably contain no surprises (together they are very similar to the theory PKF of Halbach & Horsten (2006), though with the addition of the Explosion rule), but it is important to be explicit, and to set the framework within which the fourth part is presented.

3.1. The logic. I start with a rather standard formalization of the logic K_3. (The formalization is similar to the one in Wang, 1961.) It is in the format of a Gentzen sequent system with single-formula consequents. The sequent symbol "$\Rightarrow$" is of course not part of the language. I assume the usual structural rules for $\Rightarrow$. (S_3 is the same system except with (Explosion) weakened to $A, \neg A \Rightarrow B \vee \neg B$; simply dropping (Explosion) gives the four-valued logic FDE.)

($\wedge$-**Ea**): $A \wedge B \Rightarrow A$
($\wedge$-**Eb**): $A \wedge B \Rightarrow B$
($\neg\wedge$-**Ia**): $\neg A \Rightarrow \neg(A \wedge B)$
($\neg\wedge$-**Ib**): $\neg B \Rightarrow \neg(A \wedge B)$
($\wedge$-**I**): $A, B \Rightarrow A \wedge B$

$$(\neg\wedge\text{-}\mathbf{E}): \quad \frac{\Gamma, \neg A \Rightarrow C \quad \Gamma, \neg B \Rightarrow C}{\Gamma, \neg(A \wedge B) \Rightarrow C}$$

($\neg\neg$-**I**): $A \Rightarrow \neg\neg A$
($\neg\neg$-**E**): $\neg\neg A \Rightarrow A$
(Explosion): $A, \neg A \Rightarrow B$
($\forall$-**E**): $\forall x Ax \Rightarrow At$ (when the substitution of t for x is legitimate)
($\neg\forall$-**I**): $\neg At \Rightarrow \neg\forall x Ax$ (when the substitution of t for x is legitimate)

$$(\forall\text{-}\mathbf{I}): \quad \frac{\Gamma \Rightarrow Ax}{\Gamma \Rightarrow \forall x Ax} \quad \text{when } x \text{ not free in any member of } \Gamma$$

$$(\neg\forall\text{-}\mathbf{E}): \quad \frac{\Gamma, \neg Ax \Rightarrow B}{\Gamma, \neg\forall x Ax \Rightarrow B} \quad \text{when } x \text{ not free in } B \text{ or any member of } \Gamma$$

We define $\vee$, $\exists$, $\supset$ and $\equiv$ from the others in the usual way, and we get the expected rules for them. (This includes modus ponens, given that we've included **(Explosion)**.)

It's easy to check that we have restricted conditional proof with side formulas:

progressions of theories based on global reflection principles, rather than single theories: let T_0 be PKF, and if T_α has been defined let $T_{\alpha+1}$ say (roughly) that everything provable in T_α is true (with suitable union at limit ordinals if one goes into the transfinite). (That paper proves that starting from PKF or even something weaker, two stages of iteration gets you to ramified analysis up to ω^{ω^2}, and it conjectures that further iteration would get you all the way up to ε_0. It doesn't discuss the use of schematic induction.) This use of progressions of theories strikes me as "less cheap" than directly building in powerful induction principles, but still cheap in that no stage past the zeroth is a directly motivated theory, but rather results from successively piggybacking on earlier theories. I doubt that either of these procedures successfully answers the challenge that Halbach, Horsten, Nicolai and Feferman raised (and would suspect that the authors of these more recent papers would agree). I hope to do better.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

Restricted conditional proof $\dfrac{\Gamma \Rightarrow A \vee \neg A \quad \Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \supset B}$

By induction on complexity, we also get an intersubstitutivity result: that if X_B results from X_A by substituting some or all occurrences of A in it by B, then from $A \Rightarrow B$, $\neg B \Rightarrow \neg A$, $B \Rightarrow A$ and $\neg A \Rightarrow \neg B$, we can derive $X_A \Rightarrow X_B$ (with the usual restrictions on substituting formulas with free variables into the scope of quantifiers).

Finally for identity we need

(Refl): $\Rightarrow x = x$
(Subst of =): $x = y, A(v/x) \Rightarrow A(v/y)$ when v is a variable and the substitutions of x and y for it are legitimate.

We also need

A0: $\Rightarrow \forall x \forall y (x = y \vee \neg(x = y))$.

Instead of viewing this as a general logical axiom, we might prefer to think of a version restricted to where x and y are natural numbers as an arithmetic axiom. But in the context of theories built on PA, the only objects are natural numbers, so we have the unrestricted law either way (and it will make no difference whether we regard it as logical or arithmetic).

3.2. The arithmetic. I'll employ this logic in connection with the standard language of first-order Peano arithmetic, expanded to include two new one-place predicates "*True*" and "*Scl*." Let L be this expanded language. (For definiteness, suppose that the arithmetic has "=" as its only predicate, and has the constant symbol "0" and the function symbols "*suc*," "+" and "·". If we had more arithmetic predicates then we'd need to include analogs of **A0** for them.)

The axioms (beyond the instances of excluded middle) are the standard Peano axioms, except with the induction schema formulated in rule form:

Induction Rule: $\dfrac{\Gamma, A(x) \Rightarrow A(suc(x))}{\Gamma, A(0) \Rightarrow \forall x A(x)}$ when x not free in members of Γ.

All formulas of L, even those with "*True*" and "*Scl*," are allowed as instances of A in this schema. (It's the possible presence of the nonclassical "*True*" that forces the retreat to rule form.)

It's not hard to see that **A0** together with the (other) logical laws implies universally generalized excluded middle for all formulas of the language of Peano arithmetic. By restricted conditional proof, the usual conditional form of induction is derivable for all formulas in the language of Peano arithmetic.[14] Since the other laws of PA have been built in directly, *we have full classical Peano arithmetic*.

3.3. The compositional truth theory. To state the compositional rules for truth we need some primitive recursive syntactic operations, which can of course be defined in arithmetic as operations on Gödel numbers; so we can conservatively expand arithmetic

[14] From the induction rule as formulated, derive the weaker rule form $A(0) \wedge \forall x(A(x) \supset A(suc(x))) \Rightarrow \forall x A(x)$, by using $\forall x(A(x) \supset A(suc(x)))$ as a side formula (together with (∀-**E**) and modus ponens, the latter of which depends on **Explosion**). Apply restricted conditional proof to that.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

to include function symbols for them. (This is all very standard, but I include it for reference as needed, and because some of the material required in stating it is also required for the compositional theory of strong classicality.) Let # be any specific number that isn't the Gödel number of anything. The function symbols we need are:

$neg(x)$, representing the function that takes the Gödel number of a formula of L to the Gödel number of its negation, and that takes the Gödel number of anything that isn't a formula of L into #;

$conj(x, y)$, representing the function that takes the Gödel numbers of two formulas of L to the Gödel number of their conjunction, and that takes two numbers at least one of which isn't a formula of L into #;

$univ(v, x)$, representing the function that takes the Gödel numbers of a variable of L and a formula of L to the Gödel number of the result of universally quantifying that formula with that variable, and that takes two numbers into # unless the first is the Gödel number of a variable of L and the second the Gödel number of a formula of L;

$subst(x, v, t)$, representing the function that takes the Gödel numbers of a formula, of a variable, and of a term to the Gödel number of the result of substituting the term for the variable in the formula; and that takes other triples of numbers to #;

$eq(x, y)$, representing the function that takes the Gödel numbers of two terms to the Gödel number of the equation between these terms (and that takes other pairs of numbers to #);

$num(x)$, representing the function that takes any number to the Gödel number of its corresponding numeral;

$\mathbf{SC}(x)$, representing the function that takes the Gödel number x of an expression of L to the Gödel number of the corresponding atomic sentence "$Scl(num(x))$" (and that takes other numbers to #);

$\mathbf{TR}(x)$, representing the function that takes the Gödel number x of an expression of L to the Gödel number of the corresponding atomic sentence "$True(num(x))$" (and that takes other numbers to #).[15]

These, together with the usual function symbols of PA, are to be the function symbols of L. There is also a primitive recursive relation that holds between two numbers if the first is the Gödel number of a closed term of L that denotes the second[16]; we can conservatively extend PA to include a predicate "denotes" that represents this. (I also include obvious predicates such as $SENT_L$ and $CTerm$ (closed term) for the various syntactic categories in the full language L; I include the subscript on $SENT$ because we'll later consider sublanguages with some predicates omitted (but with the same closed terms, obviating the need of a subscript on $CTerm$).

[15] It is more common in the literature on KF to use a function symbol Ṭ that represents the function taking the Gödel number of a term t to the Gödel number of the corresponding atomic formula "$True(t)$," so that $Ṭ(num(x))$ is $\mathbf{TR}(x)$. The use of **TR** rather than Ṭ (and analogously for **SC**) simplifies many formulations that follow, especially as regarding the function symbol H of §5.

There is no formal significance to the use of boldface and uppercase in **SC** and **TR**; I simply wanted to make the distinction between these function symbols and the corresponding predicates leap out to the reader.

[16] A referee notes that were we to include *all* primitive recursive function symbols in L, the denotation relation wouldn't be *primitive* recursive but merely recursive (but that this wouldn't affect the rest of the argument).

The compositional truth theory will come in six main parts: three corresponding to the three atomic predicates of L and three corresponding to the three primitive logical operations $\neg$, $\wedge$ and $\forall$. (In the part for $\forall$ I make use of the fact that the language contains a closed term for everything in the intended model.) The parts corresponding to "$=$" and "*Scl*" will be single axioms; the other parts will each consist of four rules (T-Elim, $\neg$T-Introd, T-Introd and $\neg$T-Elim), special cases of the rules presented in §2. It would be possible (and more uniform) to present the parts corresponding to "$=$" and "*Scl*" that way too, and then use **A0** and the **S0** to be introduced later for the forms listed below[17]; but in the interests of ease of comprehension and use I've adopted the simpler formulation below.

($\mathbf{T}_{eq}$): $\Rightarrow \forall s \forall t \forall x \forall y[s$ denotes $x \wedge t$ denotes $y \supset [True(eq(s,t)) \equiv x = y]$
($\mathbf{T_{SC}}$): $\Rightarrow \forall x[True(\mathbf{SC}(x)) \equiv Scl(x)]$
($\mathbf{T_{TR}}$-E): $True(\mathbf{TR}(x)) \Rightarrow True(x)$
($\neg\mathrm{T}_{\mathbf{TR}}$-I): $\neg True(x) \Rightarrow \neg True(\mathbf{TR}(x))$
($\mathbf{T_{TR}}$-I): $True(x) \Rightarrow True(\mathbf{TR}(x))$
($\neg\mathrm{T}_{\mathbf{TR}}$-E): $\neg True(\mathbf{TR}(x)) \Rightarrow \neg True(x)$
($\mathbf{T}_{neg}$-E): $True(neg(x)) \Rightarrow SENT_L(x) \wedge \neg True(x)$
($\neg\mathrm{T}_{neg}$-I): $True(x) \Rightarrow SENT_L(x) \wedge \neg True(neg(x))$
($\mathbf{T}_{neg}$-I): $\neg True(x) \wedge SENT_L(x) \Rightarrow True(neg(x))$
($\neg\mathrm{T}_{neg}$-E): $\neg True(neg(x)) \wedge SENT_L(x) \Rightarrow True(x)$
($\mathbf{T}_{conj}$-E): $True(conj(x,y)) \Rightarrow True(x) \wedge True(y)$
($\neg\mathrm{T}_{conj}$-I): $\neg True(x) \vee \neg True(y) \Rightarrow \neg True(conj(x,y))$
($\mathbf{T}_{conj}$-I): $True(x) \wedge True(y) \Rightarrow True(conj(x,y))$
($\neg\mathrm{T}_{conj}$-E): $\neg True(conj(x,y)) \Rightarrow \neg True(x) \vee \neg True(y)$
($\mathbf{T}_{univ}$-E): $True(univ(v,x) \Rightarrow \forall y[CTerm(y) \supset True(subst(x,v,y))]$
($\neg\mathrm{T}_{univ}$-I): $\neg True(univ(v,x) \Rightarrow \exists y[CTerm(y) \wedge \neg True(subst(x,v,y))]$
($\mathbf{T}_{univ}$-I): $\forall y[CTerm(y) \supset True(subst(x,v,y))] \Rightarrow True(univ(v,x))$
($\neg\mathrm{T}_{univ}$-E): $\exists y[CTerm(y) \wedge \neg True(subst(x,v,y))] \Rightarrow \neg True(univ(v,x))$.

I'll also include the rather trivial

T1: $\Rightarrow \forall x[True(x) \supset SENT_L(x)]$.[18]

(The induction rule for formulas involving "*True*" has already been included.)

Though we can't derive the analog of ($\mathbf{T_{SC}}$) for "**TR**," we can use ($\mathbf{T_{TR}}$-I) and ($\neg\mathbf{T_{TR}}$-I) to derive the weaker

---

[17] That the biconditional formulation I've adopted implies the rule form uses modus ponens for $\supset$. This is generally valid in K_3; but the particular case involves formulas that provably obey excluded middle, so would be unproblematic even in S_3.

[18] In a more general context than arithmetic it would be natural to weaken this, to allow truth to *fully parameterized formulas*, in effect pairs of formulas and assignments of objects to their variables; this would effectively absorb the notion of satisfaction into the notion of truth. In the context of arithmetic this has less obvious point since for every parameterized formula $A(\underline{x}_1, ..., \underline{x}_n)$ there is a corresponding sentence $A(\mathbf{x_1}, ..., , \mathbf{x_n})$, where the $\mathbf{x_i}$ are numerals for the $\underline{x}_i$.

(There's a slight redundancy in my formalization: given **T1**, the "$SENT_L(x)$" conjuncts are unnecessary in the two $\mathbf{T}_{neg}$ rules in which they appear on the right; and given those rules, **T1** could be restricted to atomic sentences and the results derived from the rest. Here and also later, I won't be concerned with eliminating such redundancies.)

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

$(\mathbf{T_{TR}}\textbf{-Cor})$: $True(x) \vee \neg True(x) \Rightarrow \forall x[True(\mathbf{TR}(x)) \equiv True(x)]$

(which however does not suffice to derive the corresponding quadruple of inference rules). Once we have the law **(S-Main)** for "*Scl*," given below, we'll be able to derive from this the generalization

$$\Rightarrow \forall x(Scl(x) \supset [True(\mathbf{TR}(x)) \equiv True(x)]).$$

But this is getting ahead of ourselves.

It is easily shown by induction on complexity that the general schemas of **T-Elim**, **¬T-Introd**, **T-Introd** and **¬T-Elim** given early in §2 all hold; which together with an intersubstitutivity result noted at the end of §3.1 implies intersubstitutivity of $True(\langle A \rangle)$ with A in inference (when A is a sentence). Related to this, we have that on the assumption that $SENT_L(x)$, $True(neg(x))$ is intersubstitutable in inference with $\neg True(x)$ in all contexts. This tells us (in marked contrast to the case of KF) that *for sentences, truth of negation (falsity) is just nontruth*. This intersubstitutivity makes $\neg True(\langle Q \rangle)$ an anti-theorem of the K_3-based theory (whereas it's a theorem of KF^+). It's an anti-theorem because it's equivalent to $True(\langle \neg Q \rangle)$, which in turn is equivalent to $True(\langle True(\langle Q \rangle)\rangle)$, which in turn is equivalent to $True(\langle Q \rangle)$; so it's equivalent to the contradiction $True(\langle Q \rangle) \wedge \neg True(\langle Q \rangle)$, and contradictions are anti-theorems of K_3.

3.4. The compositional theory of strong classicality. I complete the specification of INT by giving the rules for the new predicate "*Scl*." This will basically be a compositional theory that makes sense independent of the truth predicate, though to state it in full generality requires the use of the truth predicate.[19] (In a classical theory, the predicate "*Scl*" is redundant: every sentence satisfies it.) A central assumption will be excluded middle for "*Scl*":

S0: $\Rightarrow \forall x(Scl(x) \vee \neg Scl(x))$.

Less central, I'll restrict the application of "*Scl*" to sentences[20]:

S1: $\Rightarrow Scl(x) \supset SENT_L(x)$.

For atomic sentences I stipulate the following:

$(\mathbf{S}_{eq})$: $CTerm(x) \wedge CTerm(y) \Rightarrow Scl(eq(x, y))$
$(\mathbf{S_{TR}}\textbf{-i})$: $\neg SENT_L(x) \Rightarrow Scl(\mathbf{TR}(x))$
$(\mathbf{S_{TR}}\textbf{-ii})$: $SENT_L(x) \Rightarrow [Scl(\mathbf{TR}(x)) \equiv Scl(x)]$
$(\mathbf{S_{SC}}\textbf{-i})$: $\neg SENT_L(x) \Rightarrow Scl(\mathbf{SC}(x))$
$(\mathbf{S_{SC}}\textbf{-ii})$: $SENT_L(x) \Rightarrow [Scl(\mathbf{SC}(x)) \equiv Scl(x)]$.

---

[19] Without the truth predicate, the laws $(\mathbf{S}_{conj})$, $(\mathbf{S}_{univ})$ and **(S-Main)** must be stated schematically: e.g. **(S-Main)** as $\Rightarrow Scl(\langle A \rangle) \supset (A \vee \neg A)$. Here as elsewhere the role of the truth predicate is as a device of generalization.

[20] From it we could define a notion Scl^* of strong classicality for formulas: a formula is Scl^* if all closed substitution instances are *Scl*. (This would capture the intuitive notion even outside the arithmetic context if we included parameterized instances, as suggested in note 18.) An alternative strategy would take Scl^* as the primitive; that would require adding a new axiom, that if a formula is Scl^* then all substitution instances of it are too. All of this (and the truth theory too) would require complications in a language with a description operator, where the singular terms themselves can generate nonclassicality.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

(Note the contrast between **S1** and ($\mathbf{S_{TR}}$-i). This is no violation of Intersubstitutivity since the discrepancy between $Scl(x)$ and $Scl(\mathbf{TR}(x))$ occurs only when x isn't the Gödel number of a sentence.)

The compositional axioms are as follows:

($\mathbf{S}_{neg}$): $\Rightarrow Scl(neg(x)) \equiv Scl(x)$
($\mathbf{S}_{conj}$): $\Rightarrow Scl(conj(x, y)) \equiv [(Scl(x) \wedge Scl(y)) \vee (Scl(x) \wedge \neg True(x)) \vee (Scl(y) \wedge \neg True(y))]$
($\mathbf{S}_{univ}$): $\Rightarrow Scl(univ(v, x)) \equiv [\forall y(CTerm(y) \supset Scl(subst(x, v, y))) \vee \exists y(Scl(subst(x, v, y)) \wedge \neg True(subst(x, v, y))]$.

One final law:

(S-Main): $\Rightarrow Scl(x) \supset (True(x) \vee \neg True(x))$.

Some comments on these:

($\mathbf{S_{TR}}$**-ii**) is natural given the naivety of truth. There's an obvious parallel between the $\mathbf{S_{SC}}$ axioms and the $\mathbf{S_{TR}}$ axioms.[21] Regarding ($\mathbf{S_{SC}}$**-ii**), the direction "$Scl(x) \supset Scl(\mathbf{SC}(x))$" has an obvious appeal; the converse direction that $\neg Scl(x) \supset \neg Scl(\mathbf{SC}(x))$ (when $SENT_L(x)$) might seem less obviously desirable, but will be essential to the theory. ($\mathbf{S}_{neg}$) encapsulates the idea that "*Scl*" is to be neutral between truth and falsity. The motivating idea behind ($\mathbf{S}_{conj}$) is that if at least one conjunct of a conjunction is both classical and false, the conjunction is classical as well as false. ("False" means "has a true negation" but since truth is naive, this is equivalent to "is an untrue sentence.") ($\mathbf{S}_{univ}$) makes universal quantification analogous to conjunction.

The laws without **(S-Main)** are compatible with *all* sentences being strongly classical; **(S-Main)** excludes that, in the K_3-based system, i.e. given **(Explosion)**. It excludes it because for the usual Liar sentence Q, "$True(\langle Q\rangle) \vee \neg True(\langle Q\rangle)$" is an anti-theorem (it implies everything); so **(S-Main)** makes $Scl(\langle Q\rangle)$ also an anti-theorem; hence, given **S0**, $\neg Scl(\langle Q\rangle)$ is a theorem. A similar result holds (though by a different argument) for the "external Liar sentence" Q^* that asserts that it isn't *both true and strongly classical*. For **(S-Main)** together with **S0** entails that Q^* is either not Sclassical, or Sclassical and true, or Sclassical and not true; but using naivety, the last two disjuncts are ruled out, so $\neg Scl(\langle Q^*\rangle)$ is also a theorem. (The example of the external Liar sentence shows that we cannot consistently have the converse of **(S-Main)**. It will however hold for A that don't contain both "*Scl*" and "*True*.")

§4. Strongly classical truth in INT. In §6 I will prove the consistency of INT. Before then, the main official goal is to interpret KF^+ in INT; this will involve, among

[21] And two obvious disparallels between these and the corresponding T-axioms and T-rules (i.e. $\mathbf{T_{SC}}$ plus the four $\mathbf{T_{TR}}$ rules). One is that for $\mathbf{S_{TR}}$ we can use axioms whereas for $\mathbf{T_{TR}}$ we can't: this is because "strongly classical" will be assumed to obey excluded middle even in application to truth attributions. The other is that for the S-axioms we separately consider the case where x isn't the Gödel number of a sentence, whereas we don't for the T-rules. We don't need to for the T-rules because we want the results $\neg True(\mathbf{TR}(x))$ and $\neg True(\mathbf{SC}(x))$ when $\neg SENT_L(x)$, and these follow from **T1** and **S1** by ($\neg\mathbf{T_{TR}}$ **-I**). By contrast, it's natural to want $Scl(\mathbf{TR}(x))$ and $Scl(\mathbf{SC}(x))$ when $\neg SENT_L(x)$ (since given **T1** and **S1**, excluded middle is guaranteed for attributions of truth and strong classicality to nonsentences). This requires the separation of cases (i) and (ii) in the $\mathbf{S_{SC}}$ and the $\mathbf{S_{TR}}$ rules.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

other things, singling out a certain class of sentences within the language L of INT as (translations of) KF-sentences. (I could say KF^+-sentences, but KF^+ and KF have the same language.) But the notion of strongly classical truth (Struth, where "$Strue(x)$" is defined as "$Scl(x) \wedge True(x)$") is important even for L-sentences that aren't interpretations of KF-sentences: indeed it is probably ultimately more natural to develop ramified analysis up to ε_0 directly in INT, using the notion of Struth, than to go via the KF^+ interpretation. So it's worth studying the notion of Struth directly. And besides, there are some slightly technical issues about the interpretation of KF^+ that I'd rather defer until the main ideas about Struth are on the table. That said, the laws of Struth I'll be demonstrating are closely analogous to standard laws of KF^+, as given e.g. on p. 201 of Halbach (2011). I'll use a numbering for mine that corresponds to the numbers Halbach uses; there are gaps in my numbers since he has laws for $\vee$ and $\exists$ which I won't bother with given that they follow from the others given the definitions of these connectives.

The following will be central:

Lemma 1: $\Rightarrow \forall x[Strue(x) \vee \neg Strue(x)]$.

Proof. By **(S-Main)**, $Scl(x)$ obviously implies $(Scl(x) \wedge True(x)) \vee (Scl(x) \wedge \neg True(x))$, which implies $(Scl(x) \wedge True(x)) \vee \neg(Scl(x) \wedge True(x))$, i.e. $Strue(x) \vee \neg Strue(x)$. That conclusion also follows from $\neg Scl(x)$, so by **S0** it follows without assumptions; and we can then use $\forall$-**I** to universally generalize. □

Using Lemma 1 (and **S0** again) one easily establishes

Lemma 2: $\Rightarrow \forall x[Scl(x) \supset (Strue(x) \equiv True(x))]$.[22]

(So by intersubstitutivity of "*True*," we have $\Rightarrow Scl(\langle A \rangle) \supset [Strue(\langle A \rangle \equiv A]$.)

I next turn to the one result in this section whose proof requires **Explosion**; the proofs of the numbered Str-laws that follow would go through in the S_3-based theory.[23] This is analogous to the axiom **(CONSIS)** of KF^+ (the only axiom included in KF^+ but not in KF).

Str-CONSIS: $\Rightarrow \neg\exists x[Strue(x) \wedge Strue(neg(x))]$.

Proof. $Strue(x) \wedge Strue(neg(x))$ amounts to $Scl(x) \wedge True(x) \wedge True(neg(x))$, which implies $0 = 1$ by ($\mathbf{T}_{neg}$-**E**) and Explosion. But $Strue(x) \wedge Strue(neg(x))$ obeys excluded middle (by Lemma 1 and the fact that the conjunction of two things that obey excluded middle itself obeys excluded middle); so we can apply restricted conditional proof. With $\neg(0=1)$ we get $\Rightarrow \neg(Strue(x) \wedge Strue(neg(x)))$. Universally generalize (and re-express using $\exists$). □

[22] Proof: $Strue(x) \Rightarrow True(x)$ by definition of *Strue*, so by Lemma 1 and restricted conditional proof, $\Rightarrow Strue(x) \supset True(x)$; so certainly (i) $Scl(x) \Rightarrow Strue(x) \supset True(x)$. Also $Scl(x), \neg Strue(x) \Rightarrow \neg True(x)$ by definition of *Strue*, so by Lemma 1 and restricted conditional proof again, (ii) $Scl(x) \Rightarrow \neg Strue(x) \supset \neg True(x)$. By (i) and (ii), $Scl(x) \Rightarrow Strue(x) \equiv True(x)$; so by **S0** and another conditional proof, followed by $\forall$-**I**, we get the result.

[23] This fact would be useful in extending the results of this paper to LP-based theories, since LP extends S_3 but doesn't extend K_3.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

Next, analogs of the KF axioms. The reader who doesn't want to wade through all this should probably still look at the proofs of Str5 and Str13, since the key ideas arise there. (An alternative to looking at Str5 is to look at the slightly simpler Str9.)

Str1: $\Rightarrow \forall s \forall t \forall x \forall y [s \text{ denotes } x \wedge t \text{ denotes } y \supset (Strue(eq(s,t)) \equiv x = y)]$.

Proof. $Strue(eq(s,t))$ implies $True(eq(s,t))$ by definition, and that together with "s denotes $x \wedge t$ denotes y" implies $x = y$ by $(\mathbf{T}_{eq})$; so

s denotes $x \wedge t$ denotes $y, Strue(eq(s,t)) \Rightarrow x = y$.

So using Lemma 1 and restricted conditional proof,

(i) s denotes x, t denotes $y \Rightarrow Strue(eq(s,t)) \supset x = y$.

For the converse, use $(\mathbf{S}_{eq})$ and the definition of "*Strue*" to get $\neg Strue(eq(s,t)) \Rightarrow \neg True(eq(s,t))$. Then by analogous reasoning to that in (i), we get

(ii) s denotes $x \wedge t$ denotes $y \Rightarrow \neg Strue(eq(s,t)) \supset \neg(x = y)$.

So the denotation premise implies the biconditional, and so by another restricted conditional proof and universal generalization we get the result. □

Str2: $\Rightarrow \forall s \forall t \forall x \forall y (s \text{ denotes } x \wedge t \text{ denotes } y \supset [Strue(neg(eq(s,t))) \equiv \neg(x = y)])$.

Proof. $Strue(neg(eq(s,t)))$ implies $True(neg(eq(s,t)))$ by definition, and that together with "s denotes $x \wedge t$ denotes y" implies $\neg(x = y)$ by $(\mathbf{T}_{eq})$; so as with Str1 we get

(i) s denotes $x \wedge t$ denotes $y \Rightarrow Strue(neg(eq(s,t))) \supset \neg(x = y)$.

Conversely, $\neg(x = y)$ plus the denotation assumptions implies $True(neg(eq(s,t))$ by $(\mathbf{T}_{neg})$ and $(\mathbf{T}_{eq})$, and they imply $Scl(neg(eq(s,t)))$ by $(\mathbf{S}_{eq})$ and $(\mathbf{S}_{neg})$. So

(ii) s denotes $x \wedge$ t denotes $y \Rightarrow \neg(x = y) \supset Strue(neg(eq(s,t)))$.

Apply restricted conditional proof and universal generalization to the conjunction of (i) and (ii). □

Str3: $\Rightarrow \forall x [Strue(neg(neg(x))) \equiv Strue(x)]$

Proof. (i) $Strue(x) \Rightarrow Scl(x) \wedge True(x)$, so using two applications of $(\mathbf{S}_{neg})$ plus $(\mathbf{T}_{neg}\text{-}\mathbf{I})$ and $(\neg\mathbf{T}_{neg}\text{-}\mathbf{I})$, $Strue(x) \Rightarrow Scl(neg(neg(x))) \wedge True(neg(neg(x)))$, i.e. $Strue(x) \Rightarrow Strue(neg(neg(x)))$. By Lemma 1 and restricted conditional proof, $\Rightarrow Strue(x) \supset Strue(neg(neg(x)))$.

(ii) $Strue(neg(neg(x))) \Rightarrow Scl(neg(neg(x))) \wedge True(neg(neg(x)))$, so using two applications of $(\mathbf{S}_{neg})$ together with and $(\mathbf{T}_{neg}\text{-}\mathbf{E})$ and $(\neg\mathbf{T}_{neg}\text{-}\mathbf{E})$, $Strue(neg(neg(x))) \Rightarrow Scl(x) \wedge True(x)$, i.e. $Strue(neg(neg(x))) \Rightarrow Strue(x)$. By Lemma 1 and restricted conditional proof, $\Rightarrow Strue(neg(neg(x))) \supset Strue(x)$. □

Str4: $\Rightarrow \forall x \forall y [Strue(conj(x,y)) \equiv Strue(x) \wedge Strue(y)]$

Proof. Again, we derive rule forms, then use restricted conditional proof and universal generalization.

R to L rule: RHS implies $Scl(x) \wedge Scl(y)$, so $Scl(conj(x,y))$ by $(\mathbf{S}_{conj})$. RHS also implies $True(x) \wedge True(y)$, so $True(conj(x,y))$ by truth rules. So $Strue(conj(x,y))$.

L to R: LHS implies $Scl(conj(x,y))$, so by $(\mathbf{S}_{conj})$, either $Scl(x) \wedge Scl(y)$, or $Scl(x) \wedge \neg True(x)$, or $Scl(y) \wedge \neg True(y)$. But LHS also implies $True(conj(x,y))$, which implies $True(x)$ and $True(y)$, knocking out second and third disjuncts above. So $Scl(x) \wedge Scl(y) \wedge True(x) \wedge True(y)$; i.e. $Strue(x) \wedge Strue(y)$. □

Str5: $SENT_L(x) \wedge SENT_L(y) \Rightarrow Strue(neg(conj(x,y))) \equiv Strue(neg(x)) \vee Strue(neg(y))$.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

Proof. The restriction to sentences is unnecessary for the left to right of the biconditional, but is needed in the other direction.

First I derive the following claims:

(1a) $\Rightarrow Strue(neg(x)) \supset Scl(neg(conj(x, y)))$

(1b) $SENT_L(y) \Rightarrow Strue(neg(x)) \supset True(neg(conj(x, y)))$

For (1a), note that $Strue(neg(x))$ implies $Scl(neg(x))$, which implies $Scl(x)$; and that $Strue(neg(x))$ also implies $True(neg(x))$, which implies $\neg True(x)$; so by ($\mathbf{S}_{conj}$), $Scl(conj(x, y))$, and hence $Scl(neg(conj(x, y)))$. Now use Lemma 1 and restricted conditional proof to get (1a).

For (1b), note that $Strue(neg(x))$ implies $True(neg(x))$, hence $SENT_L(x) \wedge \neg True(x)$, hence $SENT_L(x) \wedge \neg True(conj(x, y))$. But with $SENT_L(y)$ we get $SENT_L(conj(x, y))$, so $True(neg(conj(x, y)))$. Use Lemma 1 and restricted conditional proof to get (1b).

Putting these together, we have $SENT_L(y) \Rightarrow Strue(neg(x)) \supset Strue(neg(conj(x, y)))$. By a similar argument we get $SENT_L(x) \Rightarrow Strue(neg(y)) \supset Strue(neg(conj(x, y)))$. From these we derive

(1) $SENT_L(x) \wedge SENT_L(y) \Rightarrow Strue(neg(x)) \vee Strue(neg(y)) \supset Strue(neg(conj(x, y)))$.

For the converse, I first derive

(2) $\Rightarrow Strue(neg(conj(x, y))) \supset Strue(neg(x)) \vee Strue(neg(y))$.

To do this, note first that the antecedent requires that $SENT_L(neg(conj(x, y)))$, which obviously requires both $SENT_L(x)$ and $SENT_L(y)$. Next, given **S0**, we have that either (i) $\neg Scl(x) \wedge \neg Scl(y)$, or (ii) $Scl(x) \wedge \neg Scl(y)$, or (iii) $\neg Scl(x) \wedge Scl(y)$, or (iv) $Scl(x) \wedge Scl(y)$.

If we suppose $Strue(neg(conj(x, y)))$ we have $Scl(conj(x, y))$ (using ($\mathbf{S}_{neg}$)). Using ($\mathbf{S}_{conj}$), this rules out case (i), and allows us to expand cases (ii) and (iii) as follows: (ii*) $Scl(x) \wedge \neg Scl(y) \wedge \neg True(x)$; (iii*) $Scl(y) \wedge \neg Scl(x) \wedge \neg True(y)$. But since $SENT_L(x)$, case (ii*) yields $Scl(neg(x)) \wedge True(neg(x))$, so $Strue(neg(x))$; analogously in (iii*), $Strue(neg(y))$. So in both these cases, $Strue(neg(x)) \vee Strue(neg(y))$.

In case (iv) we also get $Strue(neg(x)) \vee Strue(neg(y))$, again under the assumption that $Strue(neg(conj(x, y)))$. For that assumption entails $True(neg(conj(x, y)))$, which by the truth rules yields $True(neg(x)) \vee True(neg(y))$, and by case (iv) assumptions this yields $Strue(neg(x)) \vee Strue(neg(y))$.

So by the four cases together, we have $Strue(neg(conj(x, y))) \Rightarrow Strue(neg(x)) \vee Strue(neg(y))$; so Lemma 1 and restricted conditional proof yield (2).

Another restricted proof followed by universal generalization give the desired claim. □

Str8: $\Rightarrow \forall x[Strue(univ(v, x)) \equiv \forall y[CTerm(y) \supset Strue(subst(x, v, y))]]$

Proof. Again, I derive rule forms, then use restricted conditional proof and universal generalization.

R to L rule: RHS implies both $\forall y[CTerm(y) \supset Scl(subst(x, v, y))]$ and $\forall y[CTerm(y) \supset True(subst(x, v, y))]$. The first implies $Scl(univ(v, x))$ by ($\mathbf{S}_{univ}$), and the second implies $True(univ(v, x))$ by ($\mathbf{T}_{univ}$-**I**), and so the two together imply $Strue(univ(v, x))$.

L to R: LHS implies both (i) $Scl(univ(v, x))$ and (ii) $True(univ(v, x))$. (ii) implies that $\forall y[CTerm(y) \supset True(subst(x, v, y))]$. By ($\mathbf{S}_{univ}$), (i) implies that either $\forall y[CTerm \supset Scl(subst(x, v, y))]$ or $\exists y[Scl(subst(x, v, y)) \wedge \neg True(subst(x, v, y))]$;

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

but the second disjunct contradicts the conclusion from (ii). Using (ii) again with the remaining first disjunct, we have $\forall y[CTerm(y) \supset Strue(subst(x, v, y))]$. □

Str9: $\Rightarrow \forall x[Strue(neg(univ(v, x))) \equiv \exists y[CTerm(y) \wedge Strue(neg(subst(x, v, y)))]]$

Proof. Again, I derive rule forms, then use restricted conditional proof and universal generalization.

R to L rule: Suppose that for some closed term y, $Strue(neg(subst(x, v, y))$. Then $Scl(neg(subst(x, v, y))$ and $True(neg(subst(x, v, y))$. So $Scl(subst(x, v, y)$ and $\neg True(subst(x, v, y))$; so by $(\mathbf{S}_{univ})$, $Scl(univ(v, x))$ and hence

(i) $Scl(neg(univ(v, x)))$.

$Scl(subst(x, v, y))$also entails $SENT_L(subst(x, v, y))$, so with $\neg True(subst(x, v, y))$ it entails $True(neg(subst(x, v, y)))$; which then entails

(ii) $True(neg(univ(x, v, y)))$.

By (i) and (ii) together, $Strue(neg(subst(x, v, y)))$ entails $Strue(neg(univ(v, x)))$. By ∃-**Elim** (equivalently, ¬∀-**Elim**), $\exists y[CTerm(y) \wedge Strue(neg(subst(x, v, y)))]$ entails $Strue(neg(univ(v, x)))$.

L to R rule: $Strue(neg(univ(v, x)))$ implies both (i) $Scl(neg(univ(v, x)))$ and (ii) $True(neg(univ(v, x)))$. (i) implies $Scl(univ(v, x))$, which by $(\mathbf{S}_{univ})$ implies that either (a) $\forall y[CTerm(y) \supset Scl(subst(x, v, y))]$ or else (b) $\exists y[CTerm(y) \wedge [Scl(subst(x, v, y)) \wedge \neg True(subst(x, v, y)]]$. (ii) implies that $\exists y[CTerm(y) \wedge \neg(True(subst(x, v, y)))]$; given this, (a) above implies (b), so we've proved (b). But (given that everything Scl is a sentence, and the closure of Scl under negation), (b) implies that $\exists y[CTerm(y) \wedge Scl(neg(subst(x, v, y))) \wedge True(neg(subst(x, v, y)))]$; i.e. that $\exists y[CTerm(y) \wedge Strue(neg(subst(x, v, y)))]$. □

I now turn to laws of Struth that are analogs of Halbach's axiom KF12 and to some extent KF13; but I'll divide up KF13 and its analogs into three parts. (As we'll see, the task of interpreting KF12 and KF13 involves a major issue that doesn't arise for these analogs, and because of this, the analogy in 13c may be only partial.)

Let **STR** represent the function taking the Gödel number of an expression x to the Gödel number of $Scl(num(x)) \wedge True(num(x))$ (and taking other numbers to #).[24] (So $\mathbf{STR}(y)$ is equivalent to $conj(\mathbf{SC}(y), \mathbf{TR}(y))$.) Then

Str12: $\Rightarrow \forall x[Strue(\mathbf{STR}(x)) \equiv Strue(x)]$

Proof. Since both sides of biconditional are bivalent by Lemma 1, it suffices to establish the rule forms

(1) $Strue(\mathbf{STR}(x)) \Rightarrow Strue(x)$

(2) $Strue(x) \Rightarrow Strue(\mathbf{STR}(x))$.[25]

(1): If $Strue(\mathbf{STR}(x))$ then $True(\mathbf{STR}(x)))$, so $Strue(x)$ by $(\mathbf{T_{TR}}\text{-}\mathbf{E})$.

(2): The premise implies $True(\mathbf{STR}(x))$ by $(\mathbf{T_{TR}}\text{-}\mathbf{I})$, so to get the conclusion we only need that $Scl(\mathbf{STR}(x))$; which is equivalent to $Scl(conj(\mathbf{SC}(x), \mathbf{TR}(x)))$. But the premise also implies $Scl(x)$, so by L to R of $(\mathbf{S_{SC}}\text{-}\mathbf{ii})$ and $(\mathbf{S_{TR}}\text{-}\mathbf{ii})$ it implies both

[24] If x is the Gödel number of a nonsentence, then this function takes it not to # but rather to the sentence that falsely attributes Struth to that nonsentence.

[25] Both these claims, and Str12 itself, would hold if "**TR**" were substituted for "**STR**." Also note that (1) would hold with the weaker premise $True(\mathbf{STR}(x))$; but the stronger premise is needed for going on to invoke conditional proof.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

$Scl(\mathbf{SC}(x))$ and $Scl(\mathbf{TR}(x))$. By $(\mathbf{S}_{conj})$, this suffices for $Scl(conj(\mathbf{SCL}(x), \mathbf{TR}(x)))$; which is equivalent to $Scl(\mathbf{STR}(x))$. □

Str13a: $\Rightarrow \forall x[Strue(neg(\mathbf{STR}(x))) \wedge SENT_L(x) \supset Strue(neg(x))]$

Proof. By Lemma 1, restricted conditional proof, and universal generalization, and the definition of *Strue*, it suffices to prove these:

(i) $Strue(neg(\mathbf{STR}(x))) \Rightarrow Scl(neg(x))$

(ii) $SENT_L(x), Strue(neg(\mathbf{STR}(x))) \Rightarrow True(neg(x))$.

The premise of (i) implies $Scl(neg(\mathbf{STR}(x)))$, which implies $Scl(\mathbf{STR}(x))$ by $(\mathbf{S}_{neg})$, which amounts to $Scl(conj(\mathbf{SC}(x)), \mathbf{TR}(x))$. So by $(\mathbf{S}_{conj})$, either (I) $Scl(\mathbf{SC}(x))$ or (II) $Scl(\mathbf{TR}(x))$. ($(\mathbf{S}_{conj})$ gives more detailed information, but this is all that's needed.) But each of (I) and (II) imply $Scl(x)$ (by $(\mathbf{S_{SC}}\text{-}\mathbf{ii})$ and $(\mathbf{S_{TR}}\text{-}\mathbf{ii})$ respectively). So by $(\mathbf{S}_{neg})$ again, $Scl(neg(x))$.

As for (ii), $Strue(neg(\mathbf{STR}(x)))$ implies $True(neg(\mathbf{STR}(x)))$, which implies $\neg Strue(x)$, i.e. $\neg Scl(x) \vee \neg True(x)$; and we've already established $Scl(x)$, so $\neg True(x)$. The truth rules plus $SENT_L(x)$ give $True(neg(x))$. □

Str13b: $\Rightarrow \forall x[Strue(neg(x)) \supset Strue(neg(\mathbf{STR}(x)))]$.

Proof. (Again I prove the rule form, and conditionalize on the basis of Lemma 1.) $Strue(neg(x))$ implies both (A) $True(neg(x))$ and (B) $Scl(neg(x))$. (A) implies $\neg True(x)$ by $(\neg\mathbf{T}_{neg}\text{-}\mathbf{E})$; and hence $\neg Strue(x)$ by definition of "*Strue*." It also implies $SENT_L(neg(x))$ and hence $SENT_L(x)$ and hence $EXPRESSION(x)$, so $SENT_L(\mathbf{STR}(x))$; so $(\mathbf{T}_{neg}\text{-}\mathbf{I})$ yields $True(neg(\mathbf{STR}(x)))$.

(B) implies $Scl(x)$ by $(\mathbf{S}_{neg})$, hence $Scl(\mathbf{SC}(x))$ by $(\mathbf{S_{SC}}\text{-}\mathbf{ii})$ and $Scl(\mathbf{TR}(x))$ by $(\mathbf{S_{TR}}\text{-}\mathbf{ii})$. By $(\mathbf{S}_{conj})$ these entail $Scl(conj(\mathbf{SCL}(x), \mathbf{TR}(x)))$, which amounts to $Scl(\mathbf{STR}(x))$. So by $(\mathbf{S}_{neg})$ again, $Scl(neg(\mathbf{STR}(x)))$. This with the result of (A) yields $Strue(neg(\mathbf{STR}(x)))$. □

Str13c: $\forall x[\neg SENT_L(x) \supset Strue(neg(\mathbf{STR}(x)))]$.

Proof. If $\neg SENT_L(x)$ then $\neg Strue(x)$, so $\neg True(\mathbf{STR}(x))$. But $Sent_L(\mathbf{STR}(x))$, so $True(neg(\mathbf{STR}(x)))$.

Also $\neg SENT_L(x)$ implies $Scl(\mathbf{TR}(x))$ by $(\mathbf{S_{TR}}\text{-}\mathbf{i})$, and it implies $\neg True(x)$ by **T1** and hence $\neg True(\mathbf{TR}(x))$ by $(\neg\mathbf{T}_{neg}\text{-}\mathbf{I})$. That is, $\mathbf{TR}(x)$ is both strongly classical and untrue; so $(\mathbf{S}_{conj})$ says that its conjunction with anything is strongly classical. In particular, $Scl(conj(\mathbf{SC}(x), \mathbf{TR}(x)))$, i.e. $Scl(\mathbf{STR}(x)))$. So by $(\mathbf{S}_{neg})$, $Scl(neg(\mathbf{STR}(x)))$. □

Finally a useful pair of lemmas:

Lemma 3A: $\Rightarrow \forall x[Strue(\mathbf{STR}(x)) \equiv Strue(\mathbf{TR}(x))]$

Proof. $Strue(\mathbf{TR}(x))$ means $Scl(\mathbf{TR}(x)) \wedge True(\mathbf{TR}(x))$; by $(\mathbf{S_{TR}}\text{-}\mathbf{ii})$ and the truth rules, that's equivalent to $Scl(x) \wedge True(x)$, i.e. to $Strue(x)$. And by Str12, $Strue(\mathbf{STR}(x))$ is also equivalent to $Strue(x)$. This establishes the biconditional. □

Lemma 3B: $\Rightarrow \forall x[Strue(neg(\mathbf{STR}(x))) \equiv Strue(neg(\mathbf{TR}(x)))]$

Proof. (i) If x isn't the Gödel number of an expression, then $\mathbf{STR}(x)$, $\mathbf{TR}(x)$, $neg(\mathbf{STR}(x))$ and $neg(\mathbf{TR}(x))$ each denote #; this implies the negations of each side of the bicondional, so the biconditional holds.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

(ii) If x is the Gödel number of a nonsentence, then $\mathbf{STR}(x)$ and $\mathbf{TR}(x)$ each denote strongly classical falsehoods, so $neg(\mathbf{STR}(x))$ and $neg(\mathbf{TR}(x))$ each denote strongly classical truths. So the biconditional holds.

(iii) Suppose x is the Gödel number of a sentence. $Strue(neg(\mathbf{TR}(x)))$ means $Scl(neg(\mathbf{TR}(x))) \wedge True(neg(\mathbf{TR}(x)))$; by ($\mathbf{S}_{neg}$) and ($\mathbf{S_{TR}}$-**ii**) and ($\mathbf{S}_{neg}$) again, and the truth rules, that's equivalent to $Scl(neg(x)) \wedge True(neg(x))$, i.e. to $Strue(neg(x))$. By Str13a and b, $Strue(neg(\mathbf{STR}(x)))$ is also equivalent to $Strue(neg(x))$. This establishes the biconditional. □

§5. Interpreting KF$^+$ in INT. Here I show that INT interprets KF$^+$, and thus by previous results (Feferman (1991)) interprets ramified analysis up to ε_0. (The interpretation will leave the arithmetic and logical vocabulary fixed, including the range of quantifiers; only the truth predicate will be reinterpreted.) Most of the work for this was done in §4.

There is an issue here about the goal. Halbach's axiomatization of KF includes what amounts to

KF13c: $\forall x[\neg SENT_k(x) \supset True_k(neg_k(\mathbf{TR}_k(x)))]$.

(The reader can ignore the subscript "k," but I think it helpful to use it for nonlogical vocabulary in the language of KF *prior to the reinterpretation that validates KF$^+$ in INT*. I'll soon introduce subscripts "KF" for the reinterpretations that will validate KF$^+$.) In KF$^+$ (though not KF) the consequent of this implies $\neg True_k(\mathbf{TR}_k(x))$, which together with KF12 implies $\neg True_k(x)$; so

Corollary to KF13c: $\forall x[\neg SENT_k(x) \supset \neg True_k(x)]$.

If one accepts the Corollary (a KF analog of T1) then KF13c seems intuitively plausible. But as Halbach suggests (p. 199), neither KF13c or its corollary are needed for the proof-theoretic power of KF or KF$^+$. Since the main aim here is to show that INT has all the proof-theoretic power of KF$^+$, I will start out with a simple interpretation of the language of KF in INT that delivers all of KF$^+$ except for KF13c; it doesn't deliver the corollary either. (It does deliver the analogs with $SENT_k$ replaced by $SENT_L$.[26]) Then I will briefly sketch a more complicated interpretation that should yield the full KF$^+$.

The simple interpretation interprets "$True_k(x)$" as "$Strue(x)$," as long as x isn't the Gödel number of a formula of the language of KF that contains the function symbol $\mathbf{TR}_k$. It would be inappropriate to interpret "$True_k(x)$" as "$Strue(x)$" for other x, because a correct interpretation also needs to shift the denotation of terms in (the formula whose Gödel number is) x that contain $\mathbf{TR}_k$. So we need to introduce a function symbol H for a function that shifts the denotation of these terms. (That function is to be the identity except on formulas that contain $\mathbf{TR}_k$.) "$True_k(x)$" will then be interpreted as "$Strue(H(x))$." H is only applied to occurrences within the scope of $True_k$.

The intuitive idea of H is that it replaces any formula containing $\mathbf{TR}_k$ by the corresponding formula containing **STR**, "to arbitrary depths of embedding"; so that e.g. if x is a sentence not containing $\mathbf{TR}_k$, H maps $\mathbf{TR}_k(\mathbf{TR}_k(x))$ into $\mathbf{STR}(\mathbf{STR}(x))$. But as discussed in Halbach (2011, pp. 36–38) in a more general setting, it's tricky to

[26] This formulation uses a Gödel numbering of L in KF, but that's unproblematic.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

formalize this. (Because of self-referential uses of $True_k$, we can't do it by a simple induction on the depth of embedding.) To see the correct strategy, note that we want not only

(i) "$True_k(t)$" is interpreted as "$Strue(H(t))$,"

but also

(ii) the arithmetic function h that H stands for is the interpretation function induced by (i);

where this means that (a) h is the identity on any number that isn't the Gödel number of a formula of the language of KF that contains $True_k$; (b) if n is the Gödel number of a formula of form $True_k(t)$ then $h(n)$ will be the Gödel number of the corresponding $Strue(H(t))$; and (c) h preserves the logical operations on formulas, e.g. if n is the Gödel number of a formula then $h(neg(n)) = neg(h(n))$. Given (b), there's an apparent circularity in demanding both (i) and (ii), but as Halbach observes, we can get an H satisfying these constraints via the version of Kleene's second recursion theorem for primitive recursive functions given in Hinman (1978, p. 41). See Halbach for the details. One can easily check that since H satisfies (i) and (ii), we can derive

(\$): $H(\mathbf{TR}_k(x)) = \mathbf{STR}(H(x)))$.

(Both sides denote $h(True_k(x))$.) This guarantees that H does indeed "translate the function symbol $\mathbf{TR}_k$ as **STR** to arbitrary depths," in the way it ought to.

Given the interpretation of "$True_k(x)$," there is an obvious strategy for interpreting "$SENT_k(x)$" in a way that might validate KF^+: we define in the arithmetic language a predicate $AtFORM_{KF}$ for the Gödel numbers of L-formulas that are either identities or of form of "$Scl(H(t)) \wedge True(H(t))$"; we then define $FORMULA_{KF}$ and $SENT_{KF}$ from these in the standard way.

By an easy induction on complexity, we then establish

Lemma 4: $\Rightarrow \forall x[SENT_{KF}(x) \supset True(disj(x, neg(x))) \wedge \neg True(conj(x, neg(x)))]$,

where of course $disj$ represents disjunction. By the truth rules, this gives the schema $SENT_{KF}(\langle A \rangle) \supset A \vee \neg A$. *So INT validates the application of classical logic to KF-sentences, and also validates the* KF^+ *axiom CONSIS.*

The arithmetic axioms of KF^+, with the possible exception of the induction axioms, are immediate since these are axioms of INT and the interpretation leaves arithmetic sentences invariant. And induction is validated too: I observed at the end of §3.2 that the induction axioms in the language of arithmetic are consequences of the induction rule of INT, and the argument for this turned only on the classical nature of the induction formula. Since the interpretation of $True_k$ in INT is classical, the interpretation of the induction formulas of KF that involve it are also classical, and so the interpretations of these too are theorems of INT.

So we need only that the interpretation validates the truth axioms (other than KF13c) under this simple interpretation. (By validating B under this interpretation, I mean that INT proves $\Rightarrow B^*$ where B^* is the interpretation of B.)

And that the KF truth axioms other than 13c are validated is almost immediate given the corresponding Str theorems. For instance,

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

- KF1 says that if s and t denote$_k$ x and y respectively, $True_k(eq(s,t)) \equiv x = y$). Under the interpretation, this says that if s and t denote $H(x)$ and $H(y)$ respectively, $Strue(eq(s,t)) \equiv H(x) = H(y)$); and this is an instance of Str1.
- KF3 is that for any sentence x of KF, $True_k(neg_k(neg_k(x))) \equiv True(x)$. The proper interpretation of neg_k is just neg restricted to things that satisfy $FORMULA_{KF}$, so the interpretation of this is in effect that if $SENT_{KF}(x)$ then $Strue[H(neg(neg(x)))] \equiv Strue[H(x)]$. But since H commutes with negation, this amounts to $Strue[neg(neg(H(x)))] \equiv Strue[H(x)]$, which is an instance of Str3.
- KF12 is $\forall x[True_k(\mathbf{TR}_k(x)) \equiv True_k(x))]$,[27] so the interpretation of KF12 is $\forall x[Strue(H(\mathbf{TR}_k(x))) \equiv Strue(H(x))]$. By (\$) this is equivalent to $\forall x[Strue(\mathbf{STR}(H(x))) \equiv Strue(H(x))]$, and that is an instance of Str12.

The others are similar.

In the case of KF13c and its corollary (given at the start of the section), the translation (and the fact that H preserves sentencehood and nonsentence-hood) yields only the weak versions where $\neg SENT_k$ is replaced by $\neg SENT_L$. For instance, if x is (the Gödel number of) the sentence "$True(\langle 0 = 0 \rangle)$," then $\neg SENT_{KF}(x)$, since '*True*' occurs in x unconjoined with "*Scl*"; nonetheless $True_{KF}(x)$, i.e. $Strue(H(x))$, the H here being vacuous.

As I've said, Halbach's KF13c isn't needed for the proof-theoretic power of KF$^+$. So what I've established, in conjunction with Feferman (1991), shows that INT suffices to interpret ramified analysis up to ε_0.

If we want to interpret the full KF$^+$, we can probably complicate the foregoing to achieve this. The basic idea is to define $AtFORM_{KF}$ in a more complicated way than before, define $FORMULA_{KF}$ and $SENT_{KF}$ from it in the usual way, and then interpret "$True_{KF}(x)$" as "$SENT_{KF}(H(x)) \wedge Scl(H(x)) \wedge True(H(x))$." (The occurrence of H in the first conjunct is redundant.) Because of the first conjunct, the interpretation of "All truths$_k$ are sentences$_k$" come out correct, and KF13c could easily be argued as well.

The complication is in defining $AtFORM_{KF}$. The atomic formulas of KF that aren't equations now need to be equivalent to L-formulas of form "$FORMULA_{KF}(x) \wedge Scl(H(x)) \wedge True(H(x))$," but the first conjunct is defined in terms of $AtFORM_{KF}$, so the procedure looks circular. So we need to recast the definition of $AtFORM_{KF}$ as a more complicated kind of inductive definition. Again this will go by the recursion theorem.

If it weren't for the H, a fairly simple use of the recursion theorem would allow for the definition of $AtFORM_{KF}$ (and hence $FORMULA_{KF}$ and $SENT_{KF}$), independent of the interpretation of "$True_k$". We could then introduce the H needed for interpreting "$True_k$" by a separate use of the recursion theorem just as before. But given the need for H even in the interpretation of the predicate $AtForm_{KF}$, things are more difficult: instead of two separate uses of the recursion theorem we need a single but more complicated one. I have little doubt that it can be provided, but I will leave that to those more expert in these matters than myself. I can take this attitude because the

[27] Recall from note 15 that I've simplified the Halbach formulation by using **TR** instead of his Ṭ.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

simple interpretation, though it doesn't yield KF-13c, suffices for the proof-theoretic strength I've claimed.

Indeed, if I were working from scratch rather than appealing to the results of Feferman, I'd have skipped the interpretation of KF entirely, and directly interpreted $RA_{<\varepsilon_0}$ in INT: mimicking Feferman's proof rather than simply appealing to the result. (The analog of KF13c in this setting is Str13c, which though not needed would be available if convenient.) The Str-theorems are more general than the corresponding KF-axioms, in that they apply to L-sentences that aren't the interpretations of KF-sentences, i.e. that have "*True*" in contexts other than "$Scl(t) \wedge True(t)$"; going by way of KF would be quite unnatural if proceeding from scratch.

§6. Consistency of the nonschematic internal theory. We need to prove INT consistent.[28] Let INT_0 be the part without the extra predicate "*Scl*." Such a theory was in effect shown consistent by Kripke's well known fixed point construction on a three-valued model theory using Strong Kleene semantics, based on the standard model of arithmetic. Label the three values 0, $\frac{1}{2}$ and 1, with 1 "best", i.e. the value assigned to theorems, and 0 "worst," i.e. assigned to negations of theorems. Call a sequent GOOD if it preserves value 1 at the minimal fixed point (or at all fixed points, it won't matter) of Kripke's construction, for all instantiations of the variables, in this model. The transition rules between sequents are (i) the structural rules (which I haven't bothered to list); (ii) the logical rules ($\neg\wedge$-E), ($\forall$-I), and ($\neg\forall$-E); and (iii) the induction rule. These preserve GOODness. (The induction rule includes instances with "*True*": that it preserves GOODness is guaranteed since we've taken the model to be standard.) And the sequents that are axioms of the logic, arithmetic and truth theory are easily seen to be GOOD on the Kripke construction. So any sequent that is a theorem of INT_0 is GOOD.

A sequent $A_1, ..., A_n \Rightarrow B$ is naturally viewed as encoding the inference from the A_i to B (and a sequent of the special form $\Rightarrow B$ as encoding the endorsement of B as a theorem). So another way to put this is: Kripke's construction shows that any inference that INT_0 endorses preserves value 1 on all instantiations of the variables (at the minimal fixed point, or at all of them); and in particular, any formula it endorses as a theorem has value 1 on all instantiations of the variables.[29] Since some sentences don't get value 1 in all fixed points (indeed in any nontrivial one), this shows that INT_0 is Post-consistent; and indeed it is negation-consistent since no sentence of form $A \wedge \neg A$ gets value 1 in any nontrivial fixed point.

But what about the full theory INT that includes the axioms for "*Scl*"? I will now provide a beefed up Kripke fixed point construction, which proves the negation-consistency of the full INT in exactly the same way that the ordinary Kripke construction proves the consistency of its subtheory INT_0.

---

[28] Obviously this needs to be done in a stronger theory; a fairly weak classical set theory (with no additional truth predicate) suffices.

[29] This is in marked contrast to external theories like KF^+ prior to their reinterpretation in terms of strong classicality. In KF^+ prior to reinterpretation, there are theorems such as $\neg True(\langle Q \rangle)$ where Q is an ordinary Liar sentences, and this gets value $\frac{1}{2}$ in all Kripke fixed points. There are also other sentences that get value $\frac{1}{2}$ in all fixed points that are anti-theorems of KF^+: for instance, $True(\langle Q \rangle)$. The fixed points thus don't adequately distinguish between sentences that KF^+ takes to be good and sentences it takes to be bad.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

The trick is to run the Kripke construction twice over.[30] In the first run, we treat "$Scl(x)$" as equivalent to "$SENT_L(x) \wedge (True(x) \vee \neg True(x))$"; so in this run, sentences of form "$Scl(t)$" get only values $\frac{1}{2}$ or 1 unless the denotation of t isn't a sentence. We continue to a fixed point Ω in the usual way. In the second run, we start off by "closing off" "$Scl(x)$" (but not "$True(x)$"): at each stage we give atomic sentences containing it value 1 if they got value 1 at the fixed point in the first run, and 0 if they got value $\frac{1}{2}$ or 0 at the fixed point in the first run (keeping the values fixed throughout the second run). But the values of sentences of form "$True(t)$" vary in the second run, in the usual Kripkean way, until we reach a second fixed point Ω^*. Every sentence that gets value 1 or 0 in the first run gets the same value in the second, but some sentences that get value $\frac{1}{2}$ in the first run get another value in the second. Because of this, there will be sentences A such that "$A \vee \neg A$" gets value 1 in the second run even though "$Scl(\langle A \rangle)$" gets value 0 in the second run.[31] But the reverse can't happen: If "$Scl(\langle A \rangle)$" gets value 1 in the second run, "$A \vee \neg A$" does too.

For those who prefer a more mathematical statement, here goes. For notational simplicity I initially state the construction for sentences only, rather than for formulas relative to a variable-assignment, exploiting the fact that the language contains a closed term for every object in the domain of the standard model.

The construction will proceed from the standard model of arithmetic, by assigning to each (Gödel number of a) sentence of the language L and each ordinal $\sigma \leq \omega_1 \cdot 2$ a value in $\{0, \frac{1}{2}, 1\}$; where ω_1 is the first uncountable ordinal. (ω_1 is big enough to serve as the first fixed point Ω in the sketch above, and $\omega_1 \cdot 2$ the second one Ω^*.[32] If we generalized the procedure for other languages and other models, we'd go to $c^+ \cdot 2$, where c is the maximum of the cardinalities of the domain of the model and the vocabulary of the language, and c^+ is the first ordinal of cardinality greater than c.) At every stage σ, $|eq(s,t)|_\sigma$ is 1 iff s and t are Gödel numbers of closed terms for the same number; it's 0 otherwise. At every stage σ, $|neg(x)|_\sigma$ (when x is the Gödel number of a sentence) is $1 - |x|_\sigma$; $|conj(x,y)|_\sigma$ (with a similar restriction) is $min\{|x|_\sigma, |y|_\sigma\}$; and $|univ(v,x)|_\sigma$ is $min\{|x(v/\mathbf{m})|_\sigma : \mathbf{m}$ a numeral$\}$. The interesting thing is the values of the other atomic sentences.

Following Kripke, we let $|True(t)|_\sigma$ for closed terms t be

- 1 if t denotes (the Gödel number of) a sentence A for which $(\exists \rho < \sigma)(\forall \tau \in [\rho, \sigma))(|A|_\tau = 1)$
- 0 if either t denotes (the Gödel number of) a sentence A for which $(\exists \rho < \sigma)(\forall \tau \in [\rho, \sigma))(|A|_\tau = 0)$, or else t doesn't denote (the Gödel number of) a sentence of L
- $\frac{1}{2}$ otherwise.

[30] A referee has pointed out to me that this idea was used in Gupta & Martin (1984) in the context of adding a nonclassicality predicate to truth theory in *Weak* Kleene logic (on which, see note 8).

[31] A simple example is the external Liar Q^*. It and $Scl(\langle Q^* \rangle)$ get value $\frac{1}{2}$ throughout the first run. At the start of the second run, $|Scl(\langle Q^* \rangle)|$ is set at 0, so Q^* and hence $Q^* \vee \neg Q^*$ get value 1.

[32] Actually ω_1^{CK} and $\omega_1^{CK} \cdot 2$ would suffice for the present construction, but not for the generalized version in §9.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

I formulate the clauses this way to make the consistency of the 1 and 0 clauses obvious, but once we've proved monotonicity as below, it will follow that the $(\exists\rho < \sigma)(\forall\tau \in [\rho, \sigma))$ in them can be simplified to $(\exists\tau < \sigma)$.

In the same spirit as Kripke, we let $|Scl(t)|_\sigma$ be:

- 1 if t denotes (the Gödel number of) a sentence A for which $(\exists\tau < min(\sigma, \omega_1))(|A|_\tau \in \{0, 1\})$
- 0 if either $\sigma > \omega_1$ and $\neg(\exists\tau < \omega_1)(|A|_\tau \in \{0, 1\})$, or else t doesn't denote (the Gödel number of) a sentence of L
- $\frac{1}{2}$ otherwise.

In this case there is no initial worry of conflict between the 1 and 0 clauses, so there's no need to resort to the $(\exists\rho < \sigma)(\forall\tau \in [\rho, \sigma))$ formulation.

Generalizing Kripke, all sentences of form $True(t)$ or $Scl(t)$ for which t denotes the Gödel number of a sentence get value $\frac{1}{2}$ at stage 0. (For other t they get value 0.) Also, and crucially, we have monotonicity: letting $u \leq_K v$ (for u and v in $\{0, \frac{1}{2}, 1\}$) mean that either $u = \frac{1}{2}$, or $u = v = 0$, or $u = v = 1$, then we can easily argue that for any sentence x of L, if $\tau < \sigma$ then $|x|_\tau \leq_K |x|_\sigma$.[33]

We can now extend Kripke's fixed point argument to show the existence of "double fixed points." First, there can be only countably many changes prior to ω_1, so there must be a $\sigma < \omega_1$ where for every sentence x, $|x|_{\sigma+1} = |x|_\sigma$. So for any sentence of form $True(s)$ where s denotes a sentence y, $|True(s)|_\sigma = |y|_\sigma$. Thus at a fixed point value, truth is naive. Any later σ up through ω_1 will give the same values as this σ gives, so for convenience we can choose ω_1 as the ordinal Ω for the "first run" fixed point. At $\omega_1 + 1$ the rule for "*Scl*" jolts the construction, but from there on it's basically just another Kripke construction with a different starting point, and thus produces another fixed point prior to $\omega_1 \cdot 2$; anything after, including $\omega_1 \cdot 2$, is another fixed point, so for convenience we can take $\omega_1 \cdot 2$ as the ordinal Ω^* for the "second run" fixed point.

From now on, only the values at Ω $(=\omega_1)$ and Ω^* $(=\omega_1 \cdot 2)$ will matter. Truth behaves naively at both. At Ω, sentences of form $Scl(t)$ where t denotes the Gödel number of a sentence get only values $\frac{1}{2}$ and 1 at Ω; since there are some $\frac{1}{2}$s, we do not have excluded middle for the formula $Scl(v)$. At Ω^*, sentences of this form get only values 0 and 1 (1 if they got 1 at Ω, 0 if they got $\frac{1}{2}$ at Ω); so at Ω^*, $\forall v(Scl(v) \vee \neg Scl(v))$ does get value 1.

It remains to check in detail that all provable sequents of the internal theory are GOOD in a sense analogous to the one explained before: they preserve value 1 at Ω^* for any instantiation of the variables, in this model. I'll write an instantiation of a formula $A(u_1, ..., u_n)$ by members $x_1, ..., x_n$ of the domain as $A(\underline{x}_1, ..., \underline{x}_n)$ (a notation

[33] Suppose not; then there is a smallest σ, call it σ_0, for which there is a sentence x that is a "failure at σ" in the sense that for some $\tau < \sigma$, $|x|_\tau$ is 0 or 1 and $|x|_\sigma \neq |x|_\tau$. But x can't be of form $True(t)$, since that would require that t denotes a sentence y that is a failure prior to σ_0. And it can't be of form $Scl(t)$, since if $\sigma_0 \leq \omega_1$ this would likewise require a failure prior to σ_0, and since the only changes in valuation for $Scl(t)$ when $\sigma > \omega_1$ are when σ is $\omega_1 + 1$ and go from value $\frac{1}{2}$ to value 0. x also can't be an equality, since they never change in value as σ increases, so it can't be any atomic sentence. And the valuation rules for the Kleene connectives are such that no failure for atomic sentences implies no failure for any sentences.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

which will only be used in the context where the instantiated formula is being evaluated in the model).[34]

The proof that the transition rules preserve GOODness, and that the sequent rules of logic, arithmetic, and truth theory are GOOD, is the same as before: here the shift to the "second fixed point" changes nothing. Thus the only axiom-sequents that need checking are those governing "*Scl*."

Since there are no transition rules special to "*Scl*," I'll adopt the simplified formulation of inference rules and theorems introduced in the second paragraph of this section: so the task is to prove that the inference rules preserve value 1 in all instantiations, and that the theorems get value 1 in all instantiations.

- We've already verified **S0**, excluded middle for "*Scl*." (And induction for "*Scl*" is included in the arithmetic part.)
- **S1:** If $|\neg SENT_L(\underline{x})|_{\Omega^*} = 1$, then obviously x is not (the Gödel number of) a sentence of L. So by the valuation rule for *Scl*, $|Scl(\underline{x})|_\sigma$ is 0 at every stage of the construction and hence at Ω^*.
- ($\mathbf{S}_{eq}$), stating the strong classicality of closed equations, is likewise validated: the rules give such equations value 0 or 1 at every stage of the construction, so the claim of their Sclassicality gets value 1 for each $\sigma > 0$ and hence at Ω^*.
- ($\mathbf{S_{TR}}$**-i**) and ($\mathbf{S_{SC}}$**-i**): Suppose that $|\neg SENT_L(\underline{x})|_{\Omega^*}$ is 1, so that x is not (the Gödel number of) a sentence of an L-sentence. By the valuation rules for "*True*" and "*Scl*," $|True(\underline{x})|_\sigma$ and $|Scl(\underline{x})|_\sigma$ are 0 for all σ; so by the valuation rules for "*Scl*," $|Scl(\mathbf{TR}(\underline{x}))|_\sigma$ and $|Scl(\mathbf{SC}(\underline{x}))|_\sigma$ are 1 for all $\sigma > 0$, and hence for Ω^*.
- ($\mathbf{S_{TR}}$**-ii**): Suppose that $|SENT_L(\underline{x})|_{\Omega^*}$ is 1, so that x is an L-sentence. By the fixed point property of Ω, $|True(\underline{x})|_\Omega = |\underline{x}|_\Omega$, and indeed the same is true for all sufficiently large σ prior to Ω. So by the valuation rules for "*Scl*" and the fixed point property, $|Scl(\mathbf{TR}(\underline{x}))|_\Omega$ is the same as $|Scl(\underline{x})|_\Omega$ (either 1 or $\frac{1}{2}$). So $|Scl(\mathbf{TR}(\underline{x}))|_{\Omega^*}$ is the same as $|Scl(\underline{x})|_{\Omega^*}$, and in this case either 1 or 0. So $|Scl(\mathbf{TR}(\underline{x})) \equiv Scl(\underline{x})|_{\Omega^*} = 1$. (Indeed, the biconditional has value 1 for any σ strictly greater than Ω.)
- ($\mathbf{S_{SC}}$**-ii**): Again, from $|SENT_L(\underline{x})|_{\Omega^*} = 1$ we get that x is an L-sentence. If $|\underline{x}|_\Omega \in \{0, 1\}$ then $|Scl(\underline{x})|_\Omega$ is 1 and hence $|Scl(\mathbf{SC}(\underline{x}))|_\Omega$ is also 1; and if $|\underline{x}|_\Omega$ is $\frac{1}{2}$ then $|Scl(\underline{x})|_\Omega$ is $\frac{1}{2}$ and hence $|Scl(\mathbf{SC}(\underline{x}))|_\Omega$ is also $\frac{1}{2}$. Moving to Ω^*, we get that if $|\underline{x}|_\Omega \in \{0, 1\}$ then $|Scl(\underline{x})|_{\Omega^*}$ and $|Scl(\mathbf{SC}(\underline{x}))|_{\Omega^*}$ are both 1 and otherwise they are both 0. So the biconditional $Scl(\mathbf{SC}(x)) \equiv Scl(x)$ has value 1 at Ω^*. (Indeed, it does at any $\sigma > \Omega$.)

[34] In the current context, where every object in the domain gets a name, the value $|A(\underline{x}_1, ..., \underline{x}_n)|_\sigma$ of a parameterized formula at stage σ of the construction can be identified with the value $|A(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)|_\sigma$ at σ of the sentence involving the corresponding numerals. In the more general case there is no such correspondence. (In that case, a proper treatment of quantification requires that the Kripke construction itself be done not in terms of sentences but of parameterized formulas—or what is essentially the same, ordinary formulas relative to an assignment function for the variables. In the parameterized formula formulation, we need to take the Gödel number of a parameterized formula as a finite sequence of the ordinary Gödel number of the formula from which it was composed and the parameter values.) While what follows could be done in terms of the evaluation of the sentences $A(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$, I find it less confusing to think in terms of the values of parameterized formulas $A(\underline{x}_1, ..., \underline{x}_n)$, as would be required in the more general case.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

- ($\mathbf{S_{neg}}$): Obvious.
- ($\mathbf{S_{conj}}$): We show that at Ω^*, the formulas "$Scl(conj(x, y))$" and "$(Scl(x) \wedge Scl(y)) \vee (Scl(x) \wedge \neg True(x)) \vee (Scl(y) \wedge \neg True(y))$" have the same value, and it's never $\frac{1}{2}$. That suffices for the biconditional between them to have value 1 at Ω^*.

  (i) $|Scl(conj(\underline{x}, \underline{y}))|_{\Omega^*}$ is 1 if $|conj(\underline{x}, \underline{y})|_{\Omega} \in \{0, 1\}$, and 0 otherwise. (Note shifts here and in rest of proof from Ω^* to Ω.) Hence it's 1 iff either $|\underline{x}|_{\Omega}$ and $|\underline{y}|_{\Omega}$ are both in $\{0, 1\}$ or one of them is 0; i.e. if one of them is 0 or both are 1. And it's 0 otherwise.

  (ii) $|Scl(\underline{x}) \wedge Scl(\underline{y})|_{\Omega^*}$ is 1 if $|\underline{x}|_{\Omega}$ and $|\underline{y}|_{\Omega}$ are both in $\{0, 1\}$; and 0 otherwise. And $|Scl(\underline{x}) \wedge \neg True(\underline{x})|_{\Omega^*}$ is 1 iff $|\underline{x}|_{\Omega} = 0$, and otherwise is 0; similarly for y in place of x. So at Ω^*, the three-disjunct disjunction has value 1 iff either $|\underline{x}|_{\Omega}$ and $|\underline{y}|_{\Omega}$ are both in $\{0,1\}$, or one of them is 0; and 0 otherwise. This is the same as the 1-condition in (i). And here too the disjunction has value 0 in all other cases.
- ($\mathbf{S_{univ}}$): Unless v is the Gödel number of a variable and x the Gödel number of a formula with only that variable free then obviously (the instantiations of) both sides have value 0 (at Ω^* and everywhere else), so the biconditional is trivial when the condition isn't met. Assuming it is met, $|\forall y(CTerm \supset Scl(subst(\underline{x}, \underline{v}, y)))|_{\Omega^*}$ is 1 iff for every closed term t, the value of $subst(x, v, t)$ at Ω is in $\{0, 1\}$. And $|\exists y(Scl(subst(\underline{x}, \underline{v}, y)) \wedge \neg True(subst(\underline{x}, \underline{v}, y))|_{\Omega^*}$ is 1 iff for some closed term t, the value of $subst(x, v, t)$ at Ω is 0. So the value of the disjunction that's the RHS of ($\mathbf{S_{univ}}$) is 1 iff either for some closed term the value of $subst(x, v, t)$ at Ω is 0, or else for every closed term the value of $subst(x, v, t)$ at Ω is 1. And neither term of the disjunction can have value $\frac{1}{2}$, so the disjunction can't either. Clearly these are exactly the values one gets for $|Scl(univ(\underline{v}, \underline{x}))|_{\Omega^*}$. Since the values at Ω^* are the same and never $\frac{1}{2}$, the biconditional has value 1 at Ω^*.
- **(S-Main):** $|Scl(\underline{x})|_{\Omega^*}$ is 0 or 1, and if 0 then $|Scl(\underline{x}) \supset (True(\underline{x}) \vee \neg True(\underline{x}))|_{\Omega^*}$ is obviously 1, so we need only consider the case where $|Scl(\underline{x})|_{\Omega^*}$ is 1. But in that case $Scl(x)$ gets value 1 already at Ω, so $True(x) \vee \neg True(x)$ does too, and it retains this value at Ω^*.

This completes the model-theoretic proof of the consistency of INT.

§7. The schematic expansions of KF$^+$ and INT. Since it's known that KF$^+$ (indeed, KF) can interpret $RA_{<\varepsilon_0}$ (ramified analysis up to ε_0), the above results show that the internal theory we've been considering can consistently do so too.

It would be nice to go further: to $RA_{<\alpha}$ for larger α. An obvious way to do so, though not so attractive, is to introduce weaker and weaker strong classicality predicates. Pick an ordinal α (say Γ_0) such that for each predicate $\beta < \alpha$ we have a satisfactory ordinal notation, and introduce a separate predicate Scl_β for each such β. Then iterate the Kripke construction through $\omega_1 \cdot \alpha$: whenever $1 \leq \beta < \alpha$ we close off Scl_β at $\omega_1 \cdot \beta$.[35] But it isn't altogether attractive to use a hierarchy of primitive predicates. (This would be an analog of Kripke's "ghost of the Tarski hierarchy," though with strong classicality

[35] There is then a hierarchy of external Liar sentences: for each $\beta < \alpha$, a sentence declaring itself not both true and Scl_β. The theory would prove it true and $Scl_{\beta+1}$ though not Scl_β.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

stratified instead of KF-like truth. Presumably each level of the "ghost hierarchy", with an array of KF-like predicates $True_\beta$, would be interpretable in a theory with a corresponding array of predicates Scl_β.)

There's a much better approach, which is an internal analog of the classical approach of Feferman (1991). Feferman noted that the standard presentation of induction in Peano arithmetic (or expansions of it like KF), as simply consisting of the set of first order instances of the schema, doesn't really capture the intended strength. The intended strength is better represented as a *schematic axiom*: something like a second order axiom, but in a context where there are no second order quantifiers. We have just a single unquantifiable schematic variable P, and formulate induction in a classical setting using this variable as

(Classical Schematic Induction): $[P(0) \wedge \forall x(P(x) \supset P(suc(x)))] \supset \forall x P(x)$.

When we introduce the schematic variable P in the context of KF or KF$^+$, it is natural to allow it to occur in instances of the truth rules. Feferman does this (see **[Ext2]** and **[Ext3]** below), and also introduces a substitution rule for schematic variables in which the legitimate substituends of a schematic formula include those that themselves contain the schematic variable "P." (A restriction on the schematic formula in which the substitution is made is also required; see **[Ext4]** below.) He shows that when the theory S-KF (or S-KF$^+$) is formulated in this way, it is highly nonconservative over KF (or KF$^+$): instead of merely interpreting $RA_{<\varepsilon_0}$, it gives full predicative analysis ($RA_{<\Gamma_0}$).

There's no need to go through how it does this.[36] As in the case of KF$^+$, my strategy will be to show that whatever can be done in this theory S-KF$^+$ can be done in an internal analog of it, which I'll call S-INT. S-INT will simply be a schematic version of the theory INT already presented; and I'll extend the strategy used with unschematic KF$^+$, by showing that S-INT is consistent and can interpret S-KF$^+$ within it.[37]

First I sketch Feferman's theory S-KF$^+$. Its language L(P) is the language of PA expanded to include both "*True*" and a unary schematic variable (i.e. unquantifiable second order variable) "P." Its formalization involves

> **Ext 1**: An arithmetic theory like that of KF, except with the Classical Schematic Induction given above in place of the usual first order induction schema;
> **Ext 2**: All the axioms and rules of both first order logic and KF$^+$, understood to apply to formulas that contain "P" as well as ones that don't. In the case of Axioms 5, 13a and 13c of KF$^+$ we understand "$SENT_k$" as "formula of KF(P)

[36] Feferman's argument is complicated, but its basic strategy is to show via schematic reasoning that if ramified analysis holds up to α then it holds up to $\Phi(\alpha, 0)$, where Φ is the Veblen function; since Γ_0 is the first fixed point of the function $\lambda\alpha\Phi(\alpha, 0)$, this allows us to bootstrap our way up to $RA_{<\Gamma_0}$.

[37] There's another extension of KF$^+$ that yields full predicative analysis and beyond: Burgess's KFμ (2014), which incorporates within it a minimality axiom, expressing that the "truths" are just the sentences that get value 1 in the *minimal* fixed point, i.e. what Kripke (1975) calls the "grounded" truths. I doubt that there's a way to find a natural internal theory in which one can interpret this minimality axiom: that axiom seems to have a rather impredicative character beyond the reach of my sort of internal theory. (In my consistency proof I've used minimal fixed points for convenience, but others would do: the theory of strong classicality in this paper is not commited one way or the other on the identification of the strongly classical with the grounded.)

with no free variables *other than* 'P"' (*P-sentence*, for short). [Feferman actually uses KF instead of KF$^+$, but the difference won't matter.]

Ext 3: A new pair of truth axioms: Let **P** represent the function taking the Gödel number of a term t to the Gödel number of the corresponding schematic formula "$P(t)$" (and taking other numbers to #); then

($\mathbf{T}_{schem}$**-i**): $\Rightarrow \forall s[\neg CTerm(s) \supset \neg True(\mathbf{P}(s)]$

($\mathbf{T}_{schem}$**-ii**): $\Rightarrow \forall s \forall x[s \text{ denotes } x \supset [True(\mathbf{P}(s)) \equiv P(x)]$.[38]

Ext 4: An enhanced substitution rule. To formulate it, we need a notion of substituting a formula $B(v)$ for "P" in a schematic formula $A(P)$. ($B(v)$ may contain free variables in addition to "v"; these may include "P."") What this means is basically just that whenever a formula of form $P(t)$ appears in $A(P)$ it is replaced by the corresponding $B(v/t)$. (For the usual reasons, some replacement of bound variables in $A(P)$ may be required if they occur free in $B(v)$; I won't bother to spell this out.) We can write the result of such a substitution as $A(^vB(v))$. Feferman's enhanced rule is then:

Feferman Rule: $\dfrac{\vdash A(P)}{\vdash A(^vB(v))}$ if $A(P)$ doesn't contain "$True$."

$B(v)$ is allowed to contain both "$True$" and "P"; as remarked, the allowance of "P" in B is crucial to the proof-theoretic strength of the theory.

I now turn to S-INT, the internal analog of S-KF$^+$. Its language will be the language of INT expanded to include the new schematic variable; in other words, the language of S-KF expanded to include "Scl" (though with "$True$" understood differently than in S-KF).

Int 0: In line with the interpretation of schematic variables just mentioned, we include the new axiom

(P-LEM): $\Rightarrow \forall x(Px \vee \neg Px)$.

(This will of course mean that we need new restrictions in the substitution rule.)

Int 1: We keep the Induction Rule of INT, but *now allow the schematic variable* "P" *in instances of the schema*. From this together with **(P-LEM)**, we easily derive Classical Schematic Induction. (Instantiate $A(x)$ as $P(x)$, and reason as in note 14.)[39]

Int 2: All the axioms and rules of both K$_3$ and INT are understood to apply to formulas that contain "P" as well as ones that don't. In the case of **T1** and some of the S-rules, we understand "$SENT_L$" as "P-sentence," i.e. "formula of L(P) with no free variables *other than* 'P"'.

[38] One might be surprised at the inclusion of this in a classical theory where naive truth is impossible, but Feferman interprets it as effectively saying that for subsets X of the natural numbers, "$P(x)$" is true relative to the assignment of X to "P" iff $x \in X$; on that interpretation, restrictions on ($\mathbf{T}_{schem}$**-ii**) would be unwarranted. The restriction on $A(P)$ in the substitution rule below makes clear that there is no threat of paradox.

[39] We must extend the INT induction rule in this way rather than simply *replacing* it by Classical Schematic Induction: for the modification to be proposed below in the Feferman substitution rule would prevent Classical Schematic Induction from entailing instances of the unschematic induction rule with nonclassical "$A(x)$." (Similarly, a formulation of schematic induction in S-INT that simply replaced the "$A(x)$" in the Induction Rule of INT by "$P(x)$" would not have worked: substitution won't deliver the nonclassical instances not involving "P."")

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

Int 3a: We add the truth axioms ($\mathbf{T}_{schem}$**-i**) and ($\mathbf{T}_{schem}$**-ii**) for the atomic "Pt." (Instead of the latter we could use four rules of T-E, ¬T-I, T-I and¬T-E for P, but given (P-LEM) this is equivalent to adding ($\mathbf{T}_{schem}$**-ii**).)
Int 3b: We also need a corresponding pair of axioms for "Scl":
($\mathbf{S}_{schematicvbl}$**-i**): $\Rightarrow \forall s[\neg CTerm(s) \supset \neg Scl(\mathbf{P}(s)]$
($\mathbf{S}_{schematicvbl}$**-ii**): $\Rightarrow \forall s[CTerm(s) \supset Scl(\mathbf{P}(s)]$.
Int 4: For substitution rule I propose

Modified Feferman Rule: $\dfrac{\Rightarrow A(P)}{\forall v(B(v) \vee \neg B(v)) \Rightarrow A(^{v}B(v))}$ if $A(P)$ doesn't contain '*True*' or "*Scl*."

Again, there is no restriction on the substituting formula $B(v)$ (as long as we fiddle with the bound variables of $A(P)$ in case of clashes): it can contain "*True*" and "*Scl*," and also "*P*." (A weaker version of this replaces the result below the line by $\Rightarrow \forall s[CTerm(s) \supset Scl(B(v/s))] \supset A(^{v}B(v))$; that would suffice for interpreting S-KF^{+}, but as we'll see, the stronger version can be validated.)

§8. Interpreting S-KF^{+} in S-INT. The interpretation of S-KF^{+} in S-INT will be just like the interpretation of KF^{+} in INT: interpret the "*True*" of S-KF^{+} as strong truth, i.e. the conjunction of strong classicality and truth, for sentences not containing the function symbol "**TRUE**" of KF; and generalize this to sentences that do contain "**TRUE**" by the same trick as before. (And leave the interpretation of arithmetic and logical vocabulary fixed).[40] Of course, this now applies not just to ordinary sentences but to *P*-sentences: i.e., formulas with no free variables other than "*P*."

We need to verify that this does indeed validate the axioms and rules of S-KF^{+}, including its use of excluded middle across the board. But this involves little new: by and large the discussion of §4 and §5 carries over without change when the formulas are schematic. Besides this there are only three things that need to be checked, and one of them, Classical Schematic Induction, has already been discussed. This leaves the ($\mathbf{T}_{schem}$) axioms (**Ext 3**) and the Feferman Substitution Rule (**Ext 4**).

Regarding the former, we've included corresponding ($\mathbf{T}_{schem}$) axioms in S-INT, but we need the analogs with "*Strue*" for "*True*." But given also the ($\mathbf{S}_{schem}$) axioms of S-INT, this is a trivial consequence.

As for the Feferman Substitution Rule, this involves $A(P)$ where A is in the schematic arithmetical language only (no additional predicates). By the analogous rule in S-INT, $\vdash \forall v(B(v) \vee \neg B(v)) \Rightarrow A(^{v}B(v))$. But for $B(v)$ in the restricted language L* built up from identities and formulas of form "$Scl(t) \wedge True(t)$," $\vdash \forall v(B(v) \vee \neg B(v))$; so for such formulas, $\vdash A(^{v}B(v))$.

To summarize where we are, this result together with the powerful results of Feferman (1991) shows that all of predicative analysis can be carried out in the internal theory S-INT. That's the power of naive truth (or to be more accurate, of naive truth plus, or given the background of, a strong classicality predicate).

[40] As we've seen there are really two interpretations, a simple one that doesn't deliver the inessential KF-13c and a more sophisticated one that does. For simplicity I've built my remarks here around the simple one, but with more complicated wording the point would extend to the other.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

§9. Consistency of the schematic internal theory. I turn finally to the consistency of S-INT. Here too little needs to be changed from the "double Kripke construction" of §5. We need to discuss how the model-theory applies to the schematic theory S-INT. This is analogous to how Feferman (1991) applied model theory in the case of S-KF$^+$, but for clarity I spell this out.

Let a *standard schematic model* $M_{Y,Z}$ of the language of S-INT consist of the standard model M of arithmetic together with two functions Y and Z, each of which takes an arbitrary subset X of the domain of M (the natural numbers) to functions on the domain of M with values in $\{0, \frac{1}{2}, 1\}$. For any $X \subseteq M$, let $M_{Y,Z}(X)$ (equivalently, $M_{Y(X),Z(X)}$) be the three-valued model of the language of S-INT that treats "P" as an ordinary $\{0,1\}$-valued predicate with extension X (i.e. it's evaluated by the characteristic function of X) and that evaluates "*Scl*" by $Y(X)$ and "*True*" by $Z(X)$. Then in any such $M_{Y,Z}$, and for any subset X of the domain of M, every sentence in $\mathrm{L}(P)$ gets a value in this three-valued model. Of course this won't be a very useful assignment of values unless Y and Z are chosen properly.

To choose a good Y and Z, we run the double Kripke construction of §5 in this general setting: we construct sequences Y_σ and Z_σ for ordinals $\sigma \leq \omega_1 \cdot 2$, by relativizing to X the rules for "*True*" and "*Scl*" given before. (The X is held fixed in the construction.) That is: for each such σ, and subset X of the domain, and natural number n,

$[Y_\sigma(X)](n)$ is

- 1 iff n is the Gödel number of a sentence A such that for some τ that precedes both σ and ω_1, A gets value 0 or 1 in $M_{Y_\tau(X),Z_\tau(X)}$,
- 0 iff either $\sigma > \omega_1$ and n is the Gödel number of a sentence A that gets only value $\frac{1}{2}$ in any $M_{Y_\tau(X),Z_\tau(X)}$ for which τ precedes both σ and ω_1, or else n isn't the Gödel number of a sentence;
- $\frac{1}{2}$ otherwise.

$[Z_\sigma(X)](n)$ is

- 1 iff n is the Gödel number of a sentence A such that for some interval just prior to σ, A gets value 1 relative to $M_{Y_\tau(X),Z_\tau(X)}$, for each τ in the interval;
- 0 iff either n is the Gödel number of a sentence A such that for some interval just prior to σ, A gets value 0 relative to $M_{Y_\tau(X),Z_\tau(X)}$, for each τ in the interval; or else n isn't the Gödel number of a sentence;
- $\frac{1}{2}$ otherwise.

The argument from before then tells us that for each X, we get a three-valued fixed point model by assigning the extension X to "P" and evaluating "*Scl*" and "*True*" by $Y_\Omega(X)$ and $Z_\Omega(X)$, where Ω is ω_1; similarly if we evaluate them by $Y_{\Omega^*}(X)$ and $Z_{\Omega^*}(X)$, where Ω^* is $\omega_1 \cdot 2$. (The ordinals at which the construction first reach the two fixed points will depend on the X; but since, for any X, the first fixed point is reached prior to ω_1 and remains through ω_1, and analogously for the second, we get the common fixed points as claimed.)

From now on the only two ordinals that will matter are Ω and Ω^*; for each X, we will evaluate all sentences, relative to any X, at both Ω and Ω^*. As before, it's the values at Ω^* that are important for evaluating inferences, but those at Ω are needed for determining the extension of "*Scl*" at Ω^* relative to X.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

We must state the good-making property that we want sequents of the schematic theory to have. Call a sequent UNIFORMLY GOOD if for every set X of natural numbers, the inference it encodes preserves the property of having value 1 relative to X at Ω^*, for all instantiations of the variables, in this model.

The transition rules between sequents now include the Modified Feferman Rule, so a main task is to verify that this rule preserves UNIFORM-GOODness. (The only other transition rules of the system are the structural rules that I haven't bothered to list, and the logical rules ($\neg\wedge$-**E**), ($\forall$-**I**), and ($\neg\forall$-**E**); that these ones preserve UNIFORM-GOODness is evident.)

As for the sequent axioms themselves, all the ones whose P-free versions were shown GOOD for INT are obviously UNIFORMLY GOOD, since for sequents themselves (as opposed to transition rules between them) the schematic variable simply functions as a new predicate. (This handles what I earlier called **Int 1** and **Int 2** in my comparison to Feferman's schematic theory.) So besides the Modified Feferman substitution rule (**Int 4**), we need only consider the new sequent axioms of this theory, which are **(P-LEM)**, ($\mathbf{S}_{schem}$-**i**), ($\mathbf{S}_{schem}$-**ii**), ($\mathbf{T}_{schem}$-**i**) and ($\mathbf{T}_{schem}$-**ii**) (which were **Int 0**, **Int 3a** and **Int 3b**).

- That **(P-LEM)** is UNIFORMLY GOOD is trivial, since for any σ, $|P(y)|^X_\sigma$ is 1 if $y \in X$ and 0 otherwise.
- ($\mathbf{T}_{schem}$-**i**) and ($\mathbf{S}_{schem}$-**i**): If s isn't (the Gödel number of) a closed term then $\mathbf{P}(s)$ isn't a P-sentence, so for any X and any $\sigma > 0$, $|True(\mathbf{P}(\underline{s}))|^X_\sigma = 0$ and $|Scl(\mathbf{P}(\underline{s}))|^X_\sigma = 0$. So for any X and any $\sigma > 0$, $|\forall s[\neg CTerm(s) \supset \neg True(\mathbf{P}(s))]|^X_\sigma = 1$, and analogously with $Scl(\mathbf{P}(s))$ for $\neg True(\mathbf{P}(s))$, establishing that ($\mathbf{T}_{schem}$-**i**) and ($\mathbf{S}_{schem}$-**i**) are UNIFORMLY GOOD.
- ($\mathbf{S}_{schem}$-**ii**): If s is (the Gödel number) of a closed term, say for x, then $P(s)$ is a sentence whose value for any X and σ is 1 if $x \in X$ and 0 otherwise. In that case, for any X and any $\sigma > 0$, $|Scl(\mathbf{P}(\underline{s}))|^X_\sigma = 1$. So for any X and any $\sigma > 0$, $|\forall s[CTerm(s) \supset Scl(\mathbf{P}(s))]|^X_\sigma = 1$. So ($\mathbf{S}_{schem}$-**ii**) is UNIFORMLY GOOD.
- ($\mathbf{T}_{schem}$-**ii**): If s denotes x then $\mathbf{P}(s)$ is a sentence; so for any X and any $\sigma > 0$, $|True(\mathbf{P}(\underline{s}))|^X_\sigma = |P(\underline{x})|^X_\sigma$. This is 1 or 0 (depending on whether $x \in X$), so $|True(\mathbf{P}(\underline{s})) \equiv \mathrm{P}(\underline{x})|^X_\sigma = 1$. So for any X and any $\sigma > 0$, $|\forall s \forall x[s \text{ denotes } x \supset [True(\mathbf{P}(s)) \equiv P(x)]]|^X_\sigma = 1$, establishing that ($\mathbf{T}_{schem}$-**ii**) is UNIFORMLY GOOD.
- Finally the **Modified Feferman Rule**. Suppose (1) that $\Rightarrow A(P)$ is UNIFORMLY GOOD, with $A(P)$ in the arithmetical language, i.e. not containing "*True*" or "*Scl*." We need (2) that for any given $B(v)$, $\forall v(B(v) \vee \neg B(v)) \Rightarrow A(^vB(v))$ is UNIFORMLY GOOD; where $B(v)$ may contain 'P'. Fix $B(v)$, and for any set Y, let X_Y be $\{x : |B(x)|^Y_{\Omega^*} = 1\}$. (If $B(v)$ doesn't contain 'P' then this is independent of Y.) Then (1) gives that for any Y, $|A(P)|^{X_Y}_{\Omega^*} = 1$.

  The only occurrences of "P" in $A(^vB(v))$ are within B (since all occurrences of "P" in $A(P)$ are replaced in going to $A(^vB(v))$). And since $A(P)$ is in the arithmetical language, its instances $A(^vB(v))$ are classical whenever (3) $|\forall v(B(v) \vee \neg B(v))|^Y_{\Omega^*} = 1$. So when (3) holds, $|A(^vB(v))|^Y_{\Omega^*}$ is $|A(P)|^{X_Y}_{\Omega^*}$, which we've seen is 1. That gives the desired (2).

That completes the proof of the consistency of the internal theory S-INT.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

§10. Concluding remarks.

10.1. Extensions. I mentioned early in the paper that for reasons independent of the proof-theoretic considerations that have been the topic of this paper, the logic K_3 seems expressively inadequate: it lacks a conditional that is well-behaved in nonclassical contexts, and relatedly, it lacks a well-behaved way of expressing restricted universal generalizations in nonclassical contexts. Those defects didn't raise their heads in this paper because the idea of this paper was to add a new predicate that allowed us to maximize the classical contexts. But the work in the present paper does nothing to make the problems go away when we do deal with nonclassical contexts.

There thus remains the task of combining the added strong classicality predicate with the conditionals and the restricted quantifier. I suspect that there is no difficulty in doing this, though there are choices to be made and there might be some issue about which is best.

Typical approaches to adding new conditionals (not only my own, as given for instance in Field, 2016, but also earlier approaches such as Brady, 1983) work by a macroconstruction consisting of a series of Kripke constructions, leading up to a privileged Kripke construction (in the case of my approach, at what are called reflection ordinals).[41] The simplest way to incorporate a strong classicality predicate would be to simply perform the kind of double Kripke construction considered here *at the privileged stage*. An alternative approach would be to substitute the double Kripke construction for the single *at each stage* of the macroconstruction, not just the privileged stage, using the second half as the basis for later stages of the macroconstruction. (If one does that, then to avoid problems about the treatment of "*Scl*" at limit stages, it's probably best to make it behave in a Brady-like way despite the overall construction being revision-theoretic, in the manner I advocated for property-identity in my 2020b. This would make fewer sentences involving the new conditionals "strongly classical", but should improve the laws for the strong classicality predicate in this setting.)

I should mention that on my currently preferred approach (2020a), which works from a semantics based on the unit interval [0,1] rather than on the three-valued semantics, it is almost certainly possible to significantly expand the present approach by adding not only a strong classicality predicate but also a weaker "strong regularity" predicate. Call a formula B *regular* if $(\top \rightarrow B) \leftrightarrow B$ holds of it (where $\rightarrow$ is the conditional related to restricted quantification). A formula that obeys excluded middle is regular, but the converse is far from the case: indeed, in the theory of my 2020a, nearly all of the standard paradoxical sentences are regular. (The ones that can be handled with a naive truth predicate in Łukasiewicz continuum-valued semantics are all regular.) *Strong regularity* is to be a bivalent property that guarantees regularity, in the same way that strong classicality is a bivalent property that guarantees excluded middle. Adding a strong regularity predicate in addition to a strong classicality predicate would not only greatly expand the domain in which classical logic holds (as done in this paper), but also supply a wider expanded domain in which Łukasiewicz continuum-valued logic holds. But that is a matter for another time (and, I hope, another person).

[41] In earlier work I'd called them "acceptable ordinals," but that was before Anil Gupta pointed out to me that my work proving their existence duplicated previous work establishing them under another name.

https://doi.org/10.1017/S1755020320000167 Published online by Cambridge University Press

10.2. Other morals. Many people, including myself in my 2008, have viewed internal theories based on nonclassical logics and external theories like KF or KF^+ or S-KF^+ as competitors: they've assumed that choosing one of them involves rejecting the other. But of course this needn't be so: one could take one to be basic, and interpret the other within it. The task of interpreting the internal within the external is not promising, given the classicality of the latter: one could simply piggyback on the external, by taking $A_1, ..., A_n \vdash^* B$ to mean that $\vdash_{KF} True(\langle A_1 \rangle) \wedge \cdots \wedge True(\langle A_n \rangle) \supset True(\langle B \rangle)$, but any attempt to do something similar in a more autonomous way results in a very weak internal theory. Any such approach results in internal theories that are impoverished and unnatural restrictions of the external theories that interpret them.

My approach in this paper has been the reverse: to interpret theories like KF^+ or S-KF^+ within internal theories. This approach seems to me better on two grounds. First, it accords with the idea (compellingly argued by many philosophers) that the most generally useful notion of truth (or perhaps "the philosophically basic" one) is naive. Second, the theory that this approach leads to is strictly stronger than the classical theory, in that it contains the latter within it. (I don't claim that it has greater proof-theoretic strength, i.e. that it proves more arithmetic sentences: I'm quite sure that it doesn't.) This is very much in contrast to the strong suggestion in Halbach (2011) and Halbach & Nicolai (2018) that nonclassical theories are inherently impoverished.

The approach here also shows that my rhetoric in my 2008 against external theories was misplaced: they are perfectly good theories. They aren't theories of truth in the philosophically most important sense, but they are good theories of a notion of strongly classical truth, which (though perhaps containing some arbitrary elements) is perfectly intelligible.

Acknowledgments. Thanks to Leon Horsten, Lavinia Picollo, Matteo Zicchetti and two anonymous referees for corrections and useful suggestions.

BIBLIOGRAPHY

Brady, R. (1983). The simple consistency of a set theory based on the logic CSQ. *Notre Dame Journal of Formal Logic*, **24**, 431–149.

Burgess, J. (2014). Friedman and the axiomatization of Kripke's theory of truth. In Tennant, N., editor. *Foundational Adventures*: *Essays in Honor of Harvey M. Friedman*. London: College Publications, pp. 125–148.

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

———. (2008). Axioms for determinateness and truth. *Review of Symbolic Logic*, **1**, 204–217.

Feferman, S. (2012). Axiomatizing truth: Why and how? In U. Berger, H. Diener, P. Schuster, and M. Seisenberger editors. *Logic*, *Construction*, *Computation*. Frankfurt: Ontos Verlag, pp. 185–200.

Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.

———. (2016). Indicative conditionals, restricted quantification and naive truth. *Review of Symbolic Logic*, **9**, 181–208.

Field, H. (2020a) Properties, propositions and conditionals. *Australasian Philosophical Review*, **4**.

Field, H. (2020b) Reply to Zach Weber. *Australasian Philosophical Review*, **4**.

Fischer, M., Nicolai, C., & Horsten, L. (2018). Iterated reflection over full disquotational truth. *Journal of Logic and Computation*, **27**, 2631–2651.
Gupta, A. & Martin, R. (1984). A fixed point theorem for the weak kleene valuation scheme. *Journal of Philosophical Logic*, **13**, 131–135.
Halbach, V. (2011). *Axiomatic Theories of Truth*. Cambridge: Cambridge University Press.
Halbach, V. & Fujimoto, K. (n.d.) Classical Determinate Truth I, in preparation.
Halbach, V. & Horsten, L. (2006). Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic*, **71**, 677–712.
Halbach, V. & Nicolai, C. (2018). On the costs of nonclassical logic. *Journal of Philosophical Logic*, **47**, 227–257.
Hinman, P. (1978). *Recursion-Theoretic Hierarchies*. Berlin: Springer.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690–716.
Nicolai, C. (2018). Provably true sentences across axiomatizations of Kripke's theory of truth. *Studia Logica*, **106**, 101–130.
Picollo, L. (2018). Truth in a logic of formal inconsistency: How classical can it get? *Logic Journal of the IGPL*, jzy059. doi.org/10.1093/jigpal/jzy059.
Priest, G. (1998). What's so bad about contradictions? *Journal of Philosophy*, **95**, 410–426.
Wang, H. (1961). The calculus of partial predicates and its extension to set theory. *Mathematical Logic Quarterly*, **7**, 283–288.

PHILOSOPHY DEPARTMENT
NEW YORK UNIVERSITY
5 WASHINGTON PLACE
NEW YORK, NY 10003, USA
E-mail: hf18@nyu.edu