A proof-theoretic account of classical principles of truth

Graham E. Leigh
University of Oxford

In the analysis of truth it is often logic that is blamed for the truth-theoretic paradoxes. It is well known that compositionality and self-referentiality do not mix well in a purely classical environment, so logics weaker than classical logic have been suggested for the truth predicate. These typically involve moving to partial or paraconsistent logics (see for example [3, 6]). Compared with classical logic, such logics are not well understood, and as Feferman observes, it is questionable whether "[anything] like sustained ordinary reasoning can be carried [out in them]" [2, p. 95]. In this paper we seek to explore more deeply the interface between principles of self-applicable truth and classical logic, in particular the rôle classical principles play in restricting the free use of truth principles. We propose to do this by analysing various axiomatisations of truth over intuitionistic logic.

Intuitionistic logic has not received much attention to date from the truth theory community, but it does have a number of virtues. As a logic upon which to study the effect of classical reasoning, it is better suited than other weakenings of classical logic because of its mature model theory and proof theory. This means that consistency and conservativeness results for truth can be easily interpreted outside the field of theories of truth (for example in set theory or second order analysis) as they can with a classical base theory. It also provides a firm base on which to study constructive interpretations of truth. Of course few authors, if any, believe that truth is inherently constructive. But since an axiomatic theory of truth can provide at best only an approximation for "real" truth, it is natural at least not to rule out constructive formulations.

The analysis we present is inspired by the work of Friedman and Sheard [4]. There the authors determine the consistency of all subsets of a collection of twelve principles of truth over a classical base theory. The outcome is nine maximal consistent theories of truth. Subsequent work by Cantini [1], Halbach [5] and Leigh and Rathjen [8] establish the proof-theoretic strength of each of these theories showing they range from conservative extensions of Peano arithmetic to the strength of one inductive definition, ID_1. These results are summarised in table 2 below. Beginning with [7] the work has been transferred to intuitionistic logic. It is shown that some of the inconsistencies noted in [4] can be attributed to classical principles inherent in the base theory, the law of excluded middle or the statement that the logic under the truth predicate is classical. Moreover, these sets of principles turn out to be consistent over a purely intuitionistic base theory. For example, over the classical base theory utilised in [4], the principles $\mathtt{T}(A \vee B) \rightarrow \mathtt{T}A \vee \mathtt{T}B$ and $\mathtt{T}A \vee \mathtt{T}(\neg A)$ (stated for arbitrary sentences A, B) are equivalent; nevertheless, there are models of the intuitionistic base theory that satisfy the former but refute the latter principle. In fact, the former axiom is consistent with all consistent sets of truth principles and the latter is inconsistent with roughly half. The upshot is that although there are still exactly nine maximal consistent sets of principles over the new base theory, reverting to intuitionistic logic provides more freedom to express principles of truth while avoiding the pitfalls of inconsistency.

In the present paper we discuss the proof-theoretic analysis of the induced maximal consistent theories. Four of the theories remain sub-theories of their classical counterparts. The other theories are inconsistent with either the law of excluded middle or the principle of a classical truth predicate (or both). Upper bounds on their proof-theoretic strength do not, therefore, immediately result from the analysis of their classical cousins and a more refined analysis is required.

We denote by $\mathcal{L}_{\mathtt{T}}$ the language of arithmetic, $\mathcal{L}$, extended to include a unary predicate $\mathtt{T}$. $\mathsf{HA}_{\mathtt{T}}$ is Heyting arithmetic formulated in $\mathcal{L}_{\mathtt{T}}$, that is with the schema of induction extended to formulæ of $\mathcal{L}_{\mathtt{T}}$, likewise $\mathsf{PA}_{\mathtt{T}}$. If S is a classical theory we denote by S^i the appropriate theory based on intuitionistic logic. Let $\ulcorner . \urcorner$ denote some fixed primitive recursive Gödel coding of $\mathcal{L}_{\mathtt{T}}$. Though not essential for the present work, it is convenient to assume that $\ulcorner . \urcorner$ is total, i.e. every number is the code of some formula of $\mathcal{L}_{\mathtt{T}}$. Usual conventions apply to the arithmetisation of syntax that should be clear from the context in which they occur.

The base theory of truth we consider, called $\mathsf{Base}^i_{\mathtt{T}}$, extends $\mathsf{HA}_{\mathtt{T}}$ by the three axioms

1. $\forall x \forall y (\mathtt{T}x \wedge \mathtt{T}(x \dot{\rightarrow} y) \rightarrow \mathtt{T}y)$,
2. $\forall x (\mathtt{val}^i(x) \rightarrow \mathtt{T}x)$,
3. $\forall x (\mathtt{Ax}_{\mathsf{PRA}}(x) \rightarrow \mathtt{T}x)$.

Here $\mathtt{val}^i(x)$ expresses that x is the Gödel number of an intuitionistically valid first-order $\mathcal{L}_{\mathtt{T}}$-sentence and $\mathtt{Ax}_{\mathsf{PRA}}(x)$ expresses that x is the Gödel number of the universal closure of a non-logical axiom of primitive recursive arithmetic. The classical base theory of [4] and [8], called $\mathsf{Base}_{\mathtt{T}}$, is obtained from $\mathsf{Base}^i_{\mathtt{T}}$ by adding the law of excluded middle and the principles $\forall \ulcorner A \urcorner \mathtt{T}(\ulcorner A \vee \neg A \urcorner)$ stating that the underlying logic of the predicate $\mathtt{T}$ is classical. Thus $\mathsf{Base}^i_{\mathtt{T}}$ should be viewed as a natural base theory for the study of theories of truth over intuitionistic logic; it is a conservative extension of Heyting arithmetic and its only truth-theoretic commitments are that the collection of statements provably true is closed under modus ponens and comprises intuitionistic $\mathsf{I}\Sigma_1$.

Two theories are said to be *proof-theoretically equivalent* if they share the same theorems of $\mathcal{L}$ and, moreover, that this fact can be established within HA. We say a theory has *proof-theoretic ordinal* α if it is proof-theoretically equivalent to either $\mathsf{HA} + \mathtt{TI}(<\alpha)$ or $\mathsf{PA} + \mathtt{TI}(<\alpha)$ where $\mathtt{TI}(<\alpha)$ represents the schema of transfinite induction on initial segments of (a natural representation of) α.

Name	Axiom	Name	Axiom
Rep	$\mathtt{T}A \rightarrow \mathtt{TT}A$	Cons	$\neg(\mathtt{T}A \wedge \mathtt{T}\neg A)$
Del	$\mathtt{TT}A \rightarrow \mathtt{T}A$	Comp	$\mathtt{T}A \vee \mathtt{T}\neg A$
$\forall$-Inf	$\forall x \mathtt{T}A(\dot{x}) \rightarrow \mathtt{T}(\forall x A)$	$\vee$-Inf	$\mathtt{T}(A \vee B) \rightarrow \mathtt{T}A \vee \mathtt{T}B$
$\exists$-Inf	$\mathtt{T}(\exists x A) \rightarrow \exists x\, \mathtt{T}A(\dot{x})$	$\rightarrow$-Inf	$(\mathtt{T}A \rightarrow \mathtt{T}B) \rightarrow \mathtt{T}(A \rightarrow B)$

Name	Axiom Schema	Name	Rule of inference
In	$A \rightarrow \mathtt{T}A$	Intro	From A infer $\mathtt{T}A$
Out	$\mathtt{T}A \rightarrow A$	Elim	From $\mathtt{T}A$ infer A
		$\neg$Intro	From $\neg A$ infer $\neg\mathtt{T}A$
		$\neg$Elim	From $\neg\mathtt{T}A$ infer $\neg A$.

Table 1: Table of truth-theoretic principles

Table 1 shows the fifteen principles of truth considered in [4] and [7]. For presentation purposes these principles are given in shorthand and should be read in the usual more general formulation universally quantified with parameters (where applicable). For instance, the axiom schema $A \rightarrow \mathtt{T}A$ should be read as the schema $A(x_0, x_1, \ldots, x_n) \rightarrow \mathtt{T}\ulcorner A(\dot{x_0}, \dot{x_1}, \ldots, \dot{x_n})\urcorner$ for each formula A of $\mathcal{L}_{\mathtt{T}}$ with at most $x_0, \ldots, x_n$ free.

The following theorem outlines the known results regarding sets of principles over $\mathsf{Base}_{\mathtt{T}}$.

Theorem 1. Table 2 presents the complete list of maximal consistent sets of principles from table 1 over $\mathsf{Base_T}$ with theories that are proof-theoretically equivalent alongside. The axioms Comp(w), $\vee$-Inf and $\rightarrow$-Inf are omitted from the list as they are all equivalent to Comp over $\mathsf{Base_T}$.

Maximal consistent set	Equivalent theories
$\forall$. In, Intro, Rep, Del, Comp, $\neg$Elim, $\forall$-Inf, $\exists$-Inf.	PA
B. Rep, Cons, Comp, $\forall$-Inf, $\exists$-Inf.	ACA, $\mathsf{PA} + \mathtt{TI}(<\epsilon_{\epsilon_0})$
C. Del, Cons, Comp, $\forall$-Inf, $\exists$-Inf.	ACA, $\mathsf{PA} + \mathtt{TI}(<\epsilon_{\epsilon_0})$
D. Intro, Elim, Cons, Comp, $\neg$Intro, $\neg$Elim, $\forall$-Inf, $\exists$-Inf.	ACA_0^+, $\mathsf{PA} + \mathtt{TI}(<\varphi 20)$, $\mathsf{RA}_{<\omega}$
$\exists$. Intro, Elim, Del, Cons, $\neg$Intro, $\forall$-Inf.	Σ^1_1-DC_0, ID_1^*, $\mathsf{PA} + \mathtt{TI}(<\varphi\omega 0)$
F. Intro, Elim, Del, $\neg$Elim, $\forall$-Inf.	Σ^1_1-DC_0, ID_1^*, $\mathsf{PA} + \mathtt{TI}(<\varphi\omega 0)$
G. Intro, Elim, Rep, $\neg$Elim, $\forall$-Inf.	ACA_0^+, $\mathsf{PA} + \mathtt{TI}(<\varphi 20)$
H. Out, Elim, Del, Rep, Cons, $\neg$Intro, $\forall$-Inf.	ID_1, $\mathsf{KP}\omega$, $\mathsf{PA} + \mathtt{TI}(<\vartheta\epsilon_{\Omega+1})$
I. Rep, Del, Elim, $\neg$Elim, $\forall$-Inf.	ACA_0^+, $\mathsf{PA} + \mathtt{TI}(<\varphi 20)$

Table 2: Known results over classical logic

Isolation of the maximal consistent theories is due to Friedman and Sheard [4]. The proof-theoretic analysis of theories D and H is provided by Halbach [5] and Cantini [1] respectively. A lower bound on the strength of H is also present in [4]. The strength of the remaining seven theories is due to Leigh and Rathjen [8].

The consistency of sets of truth principles over $\mathsf{Base_T^{\mathit{i}}}$ has also been explored. Leigh and Rathjen, in [7], classify all sets of truth principles as either consistent or inconsistent over $\mathsf{Base_T^{\mathit{i}}}$. The present work completes the picture by determining the proof-theoretic strength of each resulting theory, outlined by the next theorem.

Theorem 2. Table 3 below lists all maximal consistent collections of the fifteen principles from table 1 over $\mathsf{Base_T^{\mathit{i}}}$ together with proof-theoretically equivalent theories. For space considerations obvious redundancies in listing the axioms of each theory have been omitted.

Maximal consistent set	Equivalent theories
$\forall^i$. In, Intro, Rep, Del, Comp, $\neg$Elim, $\forall$-Inf, $\exists$-Inf.	HA
B^i. Rep, Cons, Comp, $\forall$-Inf, $\exists$-Inf.	ACA, $\mathsf{PA} + \mathtt{TI}(<\epsilon_{\epsilon_0})$
C^i. Del, Cons, Comp, $\forall$-Inf, $\exists$-Inf.	ACA, $\mathsf{PA} + \mathtt{TI}(<\epsilon_{\epsilon_0})$
D^i. Intro, Elim, Cons, Comp, $\neg$Intro, $\neg$Elim, $\forall$-Inf, $\exists$-Inf.	ACA_0^+, $\mathsf{PA} + \mathtt{TI}(<\varphi 20)$, $\mathsf{RA}_{<\omega}$
$\exists^i$. Intro, Elim, Del, Cons, $\neg$Intro, $\forall$-Inf, $\exists$-Inf, $\vee$-Inf.	Σ^1_1-DC_0^i, $\mathsf{HA} + \mathtt{TI}(<\varphi\omega 0)$
F^i. Intro, Elim, Del, $\neg$Elim, $\forall$-Inf, $\exists$-Inf, Comp(w), $\vee$-Inf.	Σ^1_1-DC_0^i, $\mathsf{HA} + \mathtt{TI}(<\varphi\omega 0)$
G^i. In, Elim, $\forall$-Inf, $\exists$-Inf, Comp(w), $\vee$-Inf, $\rightarrow$-Inf.	ACA_0^{i+}, $\mathsf{HA} + \mathtt{TI}(<\varphi 20)$
H^i. Out, Rep, Cons, $\forall$-Inf, $\exists$-Inf, $\vee$-Inf.	ID_1^i, $\mathsf{KP}\omega^i$, $\mathsf{HA} + \mathtt{TI}(<\vartheta\epsilon_{\Omega+1})$
I^i. Rep, Del, Elim, $\neg$Elim, $\forall$-Inf, $\exists$-Inf, Comp(w), $\vee$-Inf.	ACA_0^{i+}, $\mathsf{HA} + \mathtt{TI}(<\varphi 20)$

Table 3: New results over intuitionistic logic

At first glance it may not seem surprising that each of the theories $\forall^i$–I^i has the same proof-theoretic ordinal as its classical cousin. After all, the only axioms lacking compared to the classical theories is the law of excluded middle. The model construction used to verify consistency of the intuitionistic theories, however, have much in common with the fixed point construction utilised by Feferman in the analysis of KF, a theory whose intuitionistic formulation is a conservative extension of Heyting arithmetic. In the hope of providing a clearer understanding as to why, in contrast with KF, the theories do not collapse, we compare the ordinal analysis of F^i with that of F.

In [8] a classical sequent calculus $\mathsf{F}_\infty\big|\!\frac{m}{\alpha}\ \Delta$ is defined into which F without Elim embeds and such that the following soundness result holds: If $\mathsf{F}_\infty\big|\!\frac{m+1}{\alpha}\ \Delta$ is derivable and Δ consists of only formulæ in which the truth predicate occurs positively then some element of Δ becomes a true sentence if the truth predicate is interpreted as the set of sentences B for which $\mathsf{F}_\infty\big|\!\frac{m}{\varphi(m+1)\alpha}\ B$ is derivable, where φ represents the two-placed Veblen function. It is crucial for the soundness result above that the calculus enjoys cut elimination, since then if $\mathsf{F}_\infty\big|\!\frac{m}{\alpha}\ \mathtt{T}s$ is derivable, there is a derivation in which any sub-derivation with the same rank m has the form $\mathsf{F}_\infty\big|\!\frac{m}{\gamma}\ \mathtt{T}s_1, \mathtt{T}s_2, \ldots, \mathtt{T}s_n$. It follows that the consistency proof for F can be formalised in ID_1^*, the sub-theory of ID_1 in which proof by induction and the schema of induction is only stated for formulæ that do not contain negative instances of other fixed point predicates. This is because the properties required to establish that F is interpretable in the calculus $\mathsf{F}_\infty\vdash$, specifically cut elimination and closure under Elim, are purely positive in their expression, being of the form "if X is derivable in the sequent calculus, then so is Y". It is known, however, that F cannot be embedded into ID_1^* by interpreting the truth predicate simply as membership of some positive inductive definition because the latter theory does not contain the full schema of induction, whereas the former theory does. One would expect the study of F to offer some explanation as to why this is the case, but it is surprisingly silent on the issue. Indeed it is only from the analysis of F^i, which is more involved, that we begin to find some answers.

As with the analysis of F, the first step towards achieving an upper bound for F^i is to embed the theory without the rule Elim into an infinitary sequent calculus. We thus define $\Gamma \Rightarrow^m_\alpha A$ according to the usual rules for an infinitary version of HA together with rules representing each axiom of F^i and the rule Intro, so that if A is a sentence derivable in F^i without using Elim and with at most m applications of Intro then $\emptyset \Rightarrow^m_\omega A$. To embed all theorems of F^i one needs to establish that the calculus admits the rule Elim. The rule cannot be added directly to $\Rightarrow$ because of its impact on cut elimination, a key factor in obtaining the final interpretation into arithmetic.

The second step is to prove a soundness result similar to the one mentioned above. In this case we cannot make do with analysing only $\mathtt{T}$-positive derivations. The rules corresponding to Comp(w) and the axiom $\mathtt{T}(A \to B) \wedge \mathtt{T}A \to \mathtt{T}B$ require negative occurrences of the truth predicate in their premises, so unlike with F, it is not obvious that $\Rightarrow$ permits the elimination of cuts. As such, to deduce the admissibility of Elim in the context of F^i, a more general soundness result is required, one that applies to derivations involving sequents with both positive and negative occurrences of the truth predicate. The solution is to make use of asymmetric interpretations for truth to obtain an analogous soundness result. These are formalised models in which positive and negative occurrences of the truth predicate may have different interpretations. In this setting they take the following form.

Lemma. There is a hierarchy of intuitionistic Kripke models $(\mathfrak{F}^m_\alpha \mid \alpha < \Gamma_0, m < \omega)$ and normal functions $\{f^m_\alpha \mid m < \omega\}$ such that i) if $\mathfrak{F}^m_\alpha \models \mathtt{T}\ulcorner A \urcorner$ then $\emptyset \Rightarrow^n_{f^m_\alpha(0)} A$ is derivable for some $n < m$; and ii) if $\Gamma \Rightarrow^m_\alpha A$ is derivable and $\Gamma \cup \{A\}$ consists solely of atomic formulæ, then whenever $\mathfrak{F}^m_\gamma$ satisfies all elements in Γ, $\mathfrak{F}^m_{f^m_\alpha(\gamma)}$ satisfies A.

For f^m_α we pick the function $\gamma \mapsto \varphi m(\gamma + \omega^\alpha)$; the model $\mathfrak{F}^m_\alpha$ is then defined to be a linear Kripke ω-structure with two worlds. In the top world every sentence is considered true, while the second world is defined so that (i) in the above lemma holds. Therefore the rule Elim is admissible in $\Rightarrow$ and if $\mathsf{F}^i \vdash A$ then the universal closure of A is derivable with finite m and height bounded by $\varphi\omega 0$. In order to deduce that F^i is interpretable in $\mathsf{HA} + \mathtt{TI}(<\varphi\omega 0)$ it is necessary to establish that the resources required to prove lemma do not go beyond what is expressible in HA itself. Thus reference to the models $\mathfrak{F}^m_\alpha$ must be removed in favour of arithmetically definable notions. This is possible because $\Rightarrow$ enjoys partial cut elimination, so any derivation of $\Gamma \Rightarrow^m_\alpha A$ satisfying the assumption of (ii) in the lemma has a derivation involving atomic formulæ only.

The use of asymmetric interpretations appears essential. For if an interpretation is available in which the truth predicate is directly replaced by a notion of derivability then it would be expected that F^i would be interpretable in the intuitionistic version of ID_1^*. Since the latter theory is a conservative extension of arithmetic this is impossible. One direction for future research is to determine an interpretation of F into ID_1^*. We expect more success in this matter by looking at variants of the above asymmetric interpretation rather than at direct interpretations motivated by the classical analysis.

A classical theory with proof-theoretic ordinal $\varphi\omega 0$ that retains its strength when transferred to intuitionistic logic is the theory of Σ_1^1 dependent choice (Σ_1^1-DC_0). It is reasonable to suppose there is a correlation between this axiom schema and the theories F and F^i. Thus another avenue of interest is to explore further connections between axiomatisations of truth and formulations of choice in second-order analysis.

Bibliography

[1] Cantini, A. A Theory of Formal Truth Arithmetically Equivalent to ID1. *Journal of Symbolic Logic 55*, 1 (1990), 244–259.

[2] Feferman, S. Toward useful type-free theories. I. *The Journal of Symbolic Logic 49*, 1 (1984), 75–111.

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

[4] Friedman, H., and Sheard, M. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic 33* (1987), 1–21.

[5] Halbach, V. A System of Complete and Consistent Truth. *Notre Dame Journal of Formal Logic 35*, 3 (1994), 311–327.

[6] Halbach, V., and Horsten, L. Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic 71*, 2 (2006), 677–712.

[7] Leigh, G. E., and Rathjen, M. The Friedman-Sheard programme in intuitionistic logic. *Journal of Symbolic Logic 77*, 3 (2012), 777–806.

[8] Leigh, G. E., and Rathjen, M. An ordinal analysis for theories of self-referential truth. *Archive for Mathematical Logic 49*, 2 (Jan. 2010), 213–247.