Five papers on axiomatic theories of truth

Graham Emil Leigh


Much of the literature on theories of truth can be traced back to Kripke's seminal paper, Outline of a theory of truth. Kripke's aim was to provide an account of self-applicable truth which avoided the inconsistencies highlighted by Tarski's theorem. This involved abandoning the naive approach that truth was a total concept and passing to notions of partial truth where there are formulae that are neither declared true nor false. Let $L_T = L \cup \{T\}$ denote the language of arithmetic augmented by an additional unary predicate symbol $T$. A partial $L_T$-structure of truth over the natural numbers is a triple $\langle N, S_1, S_2 \rangle$ with $S_1, S_2 \subseteq N$. Satisfaction for such a structure is defined as usual for the connectives and atomic formulae with the following additions: $\langle N, S_1, S_2 \rangle \models T(t)$ if $t^{S} \in S_1$, and $\langle N, S_1, S_2 \rangle \models \neg T(t)$ if $t^{S} \in S_2$. The partiality of the structure arises from the fact it need not satisfy $T(t) \lor \neg T(t)$ for every term $t$. Kripke, however, provides a canonical method to close off partial models and obtain classical models; one simply drops the interpretation $S_2$ and sets $\langle N, S_1 \rangle \models \neg T(t)$, if $t^{S} \notin S_1$. A fixed point model is a partial model $\langle N, S_1, S_2 \rangle$ where the pair $S_1, S_2$ is a fixed point of the induced monotone function $\Phi_N : \langle X_1, X_2 \rangle \mapsto \langle \{\langle A^{x} : \langle N, X_1, X_2 \rangle \models A\}, \{\langle A^{x} : \langle N, X_1, X_2 \rangle \models \neg A\} \rangle$. Such models validate a sentence $A$ iff $A \in S_1$, and $\neg A$ iff $A \in S_2$. Moreover, they validate the Tarskian bi-conditional for all sentences that obtain a truth value. These sentences, which Kripke called significant, play a fundamental rôle in Kripke's treatment of truth.

Kripke's approach is purely model-theoretic and there have been attempts to formalise his methods. Feferman, in Reflecting on incompleteness [Journal of Symbolic Logic, vol. 56 no. 1 (1991), pp. 1–49] employs Kripke's techniques to characterise what “ought to be accepted if one has accepted the given basic notions [of a theory]” (p. 44, Feferman's italics) in a manner that harnesses the informal notion of acceptance without making explicit use of ordinals and notation systems. Thus, he formulates a number of operations to obtain new theories from old. The first of these of interest is the ordinary truth theory
associated with a theory $S$, labelled $\text{Tr}(S)$: a new predicate $T_0$ is added to $S$ along with axioms expressing that $T_0$ behaves as a truth predicate for the language of $S$ and all axiom schemata are extended to include the augmented language. Iteration of the underlying operation, $S \mapsto \text{Tr}(S)$, up to an ordinal $\alpha$ yields what has since become known as the ramified hierarchy of truth, $\text{RT}_\alpha$; one has an hierarchy of predicates, $\{T_\beta : \beta < \alpha\}$, each a truth predicate for the language encompassing the predicates of smaller index. The author remarks that $\text{RT}_\alpha$ and the ramified hierarchy at level $\alpha$, $\text{RA}_\alpha$, are inter-translatable, though the proof Feferman refers to is missing from the paper. The reduction of $\text{RA}_\alpha$ to $\text{RT}_\alpha$, however, is relatively straightforward and a proof of the converse reduction can, essentially, be gleaned from the author’s arguments in section 4.3. A direct proof using Feferman’s techniques can be found in section V.18 of Halbach’s book *Axiomatische Wahrheitstheorien* [Akademie Verlag, Berlin, 1996].

By replacing the basic truth predicate of $\text{Tr}(S)$ with self-applicable notions of truth Feferman is able to define stronger theories of truth. The (weak) reflexive closure of $S$, $\text{Ref}(S)$, extends $S$ with predicates $T$ and $F$ for truth and falsity. The predicates are defined to behave compositionally on the language $L_S \cup \{T, F\}$ and hence allow self-applicability. Furthermore all axiom schemata of $S$ are extended to also incorporate the new predicates. The theory $\text{Ref}(S)$ is classical and starting from certain models of $S$, Feferman shows how one may construct closed-off Kripke models of $\text{Ref}(S)$ by interpreting $F(A)$ as $T(\neg A)$. Thus, if $\text{Ref}(S) \vdash T(A)$, $A$ is valid in all fixed point models of $S$ and it is in this sense that truth and falsity in $\text{Ref}(S)$ can be viewed as partial. If $S$ is formulated as a schematic theory with a free predicate symbol $P$ and a rule of the form $P\text{-Subst}$: ‘from $A(P)$ infer $A(B)$ for any formula $B’$, Feferman also considers the strong reflexive closure of $S(P)$, $\text{Ref}^+(S(P))$, in which the truth predicate also applies to $P$ and one has a suitably modified form of $P\text{-Subst}$. The strong reflexive closure forms a significantly stronger theory than the weak version. The main theorem of the paper is that $\text{Ref}(\text{PA}) \equiv \text{RA}_{<\epsilon_0}$ and $\text{Ref}^+(\text{PA}(P)) \equiv \text{RA}_{<\Gamma_0}$, where $\equiv$ denotes the relation of proof-theoretic equivalence and $\Gamma_0$ denotes the so-called Feferman-Schütte ordinal of predicativity. Of particular interest is the embedding of $\text{RA}_{<\epsilon_0}$ into $\text{Ref}(\text{PA})$. Feferman makes use of what he refers to as the determinate sentences in $\text{Ref}(\text{PA})$, those which are either true or false but not both, to enable him to form infinite conjunctions: given a list of sentences $A_0, A_1, \ldots$ and a primitive recursive function $f$ such that $f(n) = \neg A_n$, he constructs (the code of) a sentence $\land f$ such that $T(\land f) \iff \forall n T(A_n)$ is derivable in $\text{Ref}(\text{PA})$ provided every $A_n$ is determinate. In Kripke models where the two predicates are disjoint, the determinate sentences are precisely the significant ones. It is perhaps unsurprising then that the schema $T(A) \iff A$ is derivable in $\text{Ref}(S)$ for all determinate sentences $A$. From this observation it becomes straightforward to show the jump hierarchy and hence also the ramified hierarchy up to $\epsilon_0$ can be interpreted within the reflexive closure of PA.

Key to the motivation behind introducing the notions of reflexive closure is their naturalness. Feferman argues that the axioms added to $S$ to form $\text{Ref}(S)$ “correspond directly to the informal notion of grounded truth and falsity, which is the simplest notion of partial self-applicable truth predicates. Moreover, the use of classical logic and the full schematic principles in $\text{Ref}(S(P))$ accords with informal reasoning” (p. 42). Another important factor for Feferman is the view that the reflexive closure is a general operation and can be applied to a number of differing theories. He suggests applying it to, for example, Howard’s theory of
the countable ordinals, the theory of real numbers and fragments of set theory. In the case of set theory, set-theoretical reflection principles that are expressible with the aid of a truth predicate become candidates for inclusion as further axioms, providing a notion of reflective closure specific to set theory.

Related to Kripke's notion of significance is Gödel's idea that concepts such as 'meaningfully applicable' need not be meaningful when applied to arbitrary arguments. By interpreting ‘A is meaningfully applicable to x’ as ‘A(¯x) is significant’, Reinhardt, in his paper *Some remarks on extending and interpreting theories with a partial predicate for truth* [Journal of Philosophical Logic, vol. 15 no. 2 (1986), pp. 219–251], explores the possibility of Kripke's theory serving as an implementation of Gödel's suggestion. As there are sentences A for which the statement ‘it is significant that A is significant’ does not hold in Kripke's theory, Reinhardt's interpretation accords with Gödel's informal view that the concept 'meaningful' is partial.

Reinhardt's main proposal is to view the Kripke-Feferman theory KF, defined as Ref(PA) plus the axiom stating that truth and falsity do not intersect, as an axiomatisation of Kripke's theory. KF is not itself sound with respect to fixed point models but its internal theory IKF = {A : KF ⊢ T(A)} is; thus the author recommends taking KF ⊢ T(A) as a sufficient condition to declare A to be true. Reinhardt notes that his proposal has a number of desirable properties for an informal theory of truth and significance; for example, its formalisation, ProvKF(⌜T(⌜x⌝)⌝) → T(x), is significant and can be safely admitted as true: adding T(⌜∀x ProvKF(⌜T(⌜x⌝)⌝) → T(x))⌝) to KF results in a stronger, but consistent, theory and a weaker notion of truth. The proposal is explored further by considering variants of the liar sentence and their consequences. His analysis shows that many statements about non-significance are inconsistent with KF, spreading some light on the nature of significance and non-significance within Kripke's theory. These include the naïve 'it is always significant that A is significant' and the more plausible 'if it is provable in KF that A is not significant, it is significant that A is not significant'. With this interpretation, however, KF behaves merely as an instrument for generating the 'real' theorems of interest, which all lie within IKF. Reinhardt is clearly aware of this and suggests embarking on an analogue of Hilbert's programme, the objective being, he writes, “to justify the use of nonsignificant sentences entirely within the framework of the significant sentences” (p. 225).

With this in mind Reinhardt, in *Remarks on significance and meaningful applicability* [in: ed. de Alcantara, *Mathematical logic and formal systems*, Lecture Notes in Pure and Applied Mathematics, vol. 94 (1985), pp. 227–242], asks whether, for every A ∈ IKF there is a proof A0, A1, . . . , An of A, such that A_i ∈ IKF for each i < n. This question is answered negatively by Halbach and Horsten in the third paper under review, *Axiomatizing Kripke's theory of truth* [Journal of Symbolic Logic, vol. 71 no. 2 (2006), pp. 667–712]. It is shown that for any such proof of A ∈ IKF, in fact PA ⊢ A and thus Reinhardt’s goal cannot be obtained. The result does not preclude IKF from having a natural axiomatisation, but does demonstrate the reliance on the ‘ideal’ statements of KF to prove the ‘real’ theorems of IKF. The classical nature of KF provides a number of undesirable features for an axiomatisation of Kripke's theory, not least of which is the discrepancy between its internal and external theories. The authors therefore reject IKF and set themselves the task of determining a natural axiomatisation of Kripke's theory which, they conclude, should be formulated.
directly in partial logic. Halbach and Horsten are not the first to consider formal systems of partial logic; the theory they propose is related to systems of Scott and Blamey and formulated in a Gentzen-style two-sided sequent calculus. This is not the place to explain in detail their system, but it can be recognised as essentially classical logic with, in place of any introduction or elimination rules for negation, a rule of *modus tollens*, sequents for double negation elimination and the curious λ and ¬λ-sequents: λ ⇒ ¬λ and A, ¬A ⇒ λ respectively, where λ denotes the liar sentence λ ↔ ¬T(λ). The idea behind the ¬λ-sequents is as sound replacements for the unsound *ex falso quodlibet* sequents A, ¬A ⇒ ∅. These peculiar sequents are, however, never utilised and appear unnecessary for the analysis. On top of the logic are added the axioms of arithmetic and sequents corresponding to the axioms of compositional truth used in KF, forming the system PKF. Halbach and Horsten show that Kripke fixed point structures form natural models of PKF and, moreover, the soundness argument can be formalised within a sub-theory of KF with restricted induction. By a result due to Cantini this sub-theory has the same proof-theoretic strength as RA<ωω, and thus the authors obtain an upper bound on the strength of PKF.

Although formulated in partial logic, PKF behaves classically on its truth-free sub-language and thus one may query its proof-theoretic strength in relation to classical theories. The authors demonstrate, in fact, that PKF behaves classically on the sub-language of L_T corresponding to RT<ωω, and thus the embedding of RT<ωω into PKF is established with relative ease. This result also acts as a response to an objection made by Feferman that partial logic is unsuitable for everyday reasoning in mathematics: “We do not propose to carry out mathematics in partial logic”, write Halbach and Horsten, “Only when the problematic truth predicate is added, [do] we have to give up classical reasoning” (p. 706). The result also raises an interesting question: Why do KF and PKF differ so much in their content when each appears to provide a natural axiomatisation of Kripke’s theory of truth? The authors suggest that KF instead axiomatises “large parts of its metatheory. If the resources for distinguishing between truth, falsity and indeterminateness are dropped and a thoroughly partial system is adopted, proof-theoretic strength is lost” (p. 707).

Required reading for almost anyone in the subject is Friedman and Sheard’s paper, *An axiomatic approach to self-referential truth* [Annals of Pure and Applied Logic* vol. 33 no. 1 (1987), pp. 1–21]. It stands out in the literature on axiomatic truth in its aim to provide a philosophically neutral analysis of theories of self-applicable truth. This is a delicate undertaking and the authors are careful to avoid passing any personal judgements on the behaviour of truth: “We are not solving a problem in philosophy” they remark, “but rather a problem in logic with a philosophical motivation” (p. 2). The authors isolate twelve axioms, axiom schemata and rules of inference involving a predicate T (together referred to as Optional Axioms) for further analysis. Each Optional Axiom expresses a property one might intuitively (or naively) expect to hold if T is to be interpreted as ‘truth’. These range from direct weakenings of the inconsistent Tarski bi-conditional to principles stating that truth is consistent, complete or commutes with the first-order quantifiers. By adjoining different subsets of the Optional Axioms to a weak base theory of truth conservative over PA, the authors obtain their possible theories. To call each one a ‘theory of truth’, however, could be misleading: “One may reasonably take the attitude that what we are really exploring is the axiomatic properties of concepts which lie some-
where between ‘provability’ (which is well understood and somehow insufficient) and full ‘truth’ (which is mysterious and perhaps inherently unstable)” (p. 3).

The main result of the paper, that over the base theory there are exactly nine maximal consistent subsets of the Optional Axioms, is presented in the form of a sequence of charts illustrating the various sources of inconsistency as well as groups of consistent principles.

The proof of the theorem splits into two parts: constructing a model for each of the nine theories in question and proving that any subset of the Optional Axioms not covered by the first part is inconsistent. Concerning the first half of the proof, it is the theories containing both the rule of inference T-Elim (‘from $T(A)$ infer $A$’) and the axiom U-Inf, stating the predicate commutes with the universal quantifier, that have the most interesting model constructions. The theory named $D$, for example, is further axiomatised by the rule T-Intro (‘from $A$ infer $T(A)$’) and by principles stating that the truth predicate commutes with all connectives and quantifiers. The authors show truth in $D$ can be interpreted as truth in the stable limit of Belnap and Gupta’s revision theory of truth. The theory $H$, however, which contains the schema $T(A) \rightarrow A$, and the axioms T-Rep, $T(x) \rightarrow T(T(x))$, and U-Inf requires the truth predicate to be interpreted by a notion of derivability in $\omega$-logic which fulfils certain ‘self-satisfiability’ criteria. The second half of the proof begins with what can only be described as a thorough analysis of the liar sentence; the inconsistency of sixteen differing subsets of the Optional Axioms is directly computed from the liar sentence over the base theory. This mass of raw computation is enough to answer the question of consistency for all subsets except those containing the quadruple T-Intro, T-Elim, T-Rep and T-Del, $T(T(x)) \rightarrow T(x)$. To deal with such theories the authors turn to Löb’s theorem. They show that T-Intro and T-Rep suffice to derive a schematic form of Löb’s theorem, and by further assuming T-Del and T-Elim one achieves a contradiction.

To conclude the paper two consistent subsets of the Optional Axioms are analysed for their proof-theoretic content. Although the subsets considered are not maximal consistent their analysis is of independent interest. One subset, while expressively rich, is shown to be a conservative extension of Peano Arithmetic, whereas another seemingly simple subset, $H$ without the axiom T-Rep, proves the same arithmetical statements as the theory of Bar Induction, a theory proof-theoretically equivalent to the theory of one inductive definition.

Halbach’s paper, A system of complete and consistent truth [Notre Dame Journal of Formal Logic, vol. 35 no. 3 (1994), pp. 311–327], takes a closer look at the theory $D$ introduced by Friedman and Sheard. Halbach begins by providing a more natural axiomatisation of $D$, labelled $FS$, which eliminates the need for the base theory. Thus $FS$ is described as $PA$ augmented by a self-applicable truth predicate commuting with all logical connectives and quantifiers, and closed under the rules T-Intro and T-Elim. The first result of the paper is that $FS$ essentially axiomatises the finite iterations of Belnap and Gupta’s revision theory of truth: if $FS_0$ denotes $PA$ and $FS_{n+1}$ denotes the set of theorems of $FS$ derivable with at most $n$ applications of each of T-Intro and T-Elim, Halbach proves that the $\omega$-models of $FS_n$ are exactly those standard $L_T$-structures that are the image of $n$ applications of the revision rule for truth. Moreover, each such $\omega$-structure has associated with it a largest natural number $k$ such that it is an $\omega$-model of $FS_k$, thus showing $FS$ is $\omega$-inconsistent. As the author notes, McGee’s theorem can be applied directly to $FS$ to deduce its $\omega$-inconsistency,
but this would not offer the finer grained results Halbach obtained regarding its approximations. The second result is that $FS$ is proof-theoretically equivalent to $RT_{<\omega}$. In his reduction of $RT_{<\omega}$ to $FS$ the author does not make use of the rule T-Elim, and thus determines that the rule is unnecessary for deriving the truth-free theorems of $FS$. A later result by Sheard shows that likewise T-Intro is not required to derive the truth-free consequences of $FS$; it is unknown, however, whether either rule is superfluous in its axiomatisation.