A Note on Typed Truth and Consistency Strength: Abstract

Word Count: 890 words

In the talk we will be concerned with the characterisation of the logical strength of the so-called compositional truth axioms over a minimal theory of syntax. More specifically, since there is no definite sense of what counts as a compositional axiom, we focus on a collection of principles inspired by the inductive clauses of Tarski's celebrated definition, that is axioms such as

(1) a conjunction is true if and only if both conjuncts are true.

We claim that any axiomatisation of the truth predicate that counts as compositional should contain this collection: a precise description of the strength of these principles will thus amount to an identification of a *lower bound* for the logical strength of any class of compositional axioms.

Proof-theoretic analyses of systems such as

$$\text{Peano Arithmetic} + \text{'there is a satisfaction class'}$$

(a.k.a $\mathsf{CT}\!\upharpoonright$ in [3]) seem to support the thesis that compositional principles have no strength at all: $\mathsf{CT}\!\upharpoonright$ is in fact known to be a conservative extension of PA and also relatively interpretable in it.

In the contribution a *typed* (Tarskian) axiomatisation of the truth predicate will be characterised as an *abstract consistency statement*, answering some conjectures by Richard Heck. More precisely, but still vaguely, starting with an arbitrary theory U formulated, for simplicity, in the language of arithmetic, we let $Tr[.]$ denote the application of a Tarski-style theory of truth to the object theory U and Con_{U} a canonical consistency statement for U constructed intensionally in a suitable theory of syntax interpretable in Robinson's arithmetic Q. We show that Con_{U} can be seen as the unique Π_1-sentence σ—unique in the sense of $I\Delta_0(\mathsf{exp})$-provable equivalence—such that

$$Tr[\mathsf{U}] \ \text{ is mutually interpretable with } \ \mathsf{Q} + \sigma.$$

By Pudlák's fascinating strengthening of Gödel's Second Incompleteness Theorem we know that U is not interpretable in $\mathsf{Q} + Con_{\mathsf{U}}$.[1] Therefore any theory containing our version of a Tarski-style axiomatisation of the truth predicate will be logically stronger—at least in the sense of relative

[1] Cfr. [2, III.].

interpretability—than the theory U. Crucially, our characterisation will tell us exactly *how much* stronger.

The compositional axiomatisation of typed truth that we will employ—and that we provisionally labelled $Tr[.]$—is based upon an unconventional way of formulating axiomatic theories of truth, in which syntactic and logical notions concerning the object theory U are formalised in a language whose quantifiers range over a domain which is disjoint from the domain over which quantifiers of $\mathcal{L}_{\mathsf{U}}$ range. Semantic machinery is then applied to codes of formulas that belong to the disjoint 'syntactic' domain. By employing a suggestive notation we call the resulting theory of truth $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$: the syntax for U will be formalised in a weak theory of truth interpretable in Robinson's arithmetic; moreover, $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$ will contain a weak theory of sequences that connects the 'syntactic' and the 'mathematical' domain and that it is essential to formulate the axioms for the satisfaction predicate.[2] Our theory of syntax will contain no schemata extended to the entire vocabulary of the language of $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$. [4] represents a starting point for the investigation of these theories.

To obtain an interpretation of $\mathsf{Q} + Con_{\mathsf{U}}$ in $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$, we will compensate the lack of semantic induction by employing the method of definable cuts introduced by Robert Solovay in an unpublished work. In other words, we will prove in $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$ the consistency of U on a cut.[3] Some care is required in choosing a suitable $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$-definable cut in which some essential syntactic principles hold and in which all logical axioms of U are true and all its rules of inference preserve truth. The domain of the interpretation will thus be represented by a shortening of this cut.

By contrast, to see that $\mathsf{Q} + Con_{\mathsf{U}}$ interprets $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$, the arithmetisation of Henkin construction will be employed. Feferman's original formalisation ([1]) of the construction of the completion S^* of a set of sentences S required Δ_2-induction, as S^* was known to be representable by a Δ_2-formula. The result can be improved significantly by employing the method of cuts (cfr. [5]). In particular, for any recursive U, it is possible construct an arithmetised model for it in $I\Delta_0 + \Omega_1 + Con_{\mathsf{U}}$ equipped with a truth predicate. The truth predicate of $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}}$, unlike the one of $\mathsf{CT}\restriction$, can be interpreted in $I\Delta_0 + \Omega_1 + Con_{\mathsf{U}}$—and thus in $\mathsf{Q} + Con_{\mathsf{U}}$—as truth in the sense of the arithmetised model.

The main result, that is the desired characterisation Con_{U}, will follow by employing a result of Paris and Wilkie that identifies the Π_1-sentences provable in $I\Delta_0(\mathsf{exp})$ as the Π_1-sentences that are interpretable in Q.

In the concluding remarks, we will focus on the one hand on the close relationships occurring between the proposed description of the compositional axioms and predicative comprehension; on the other, we consider the impact of our results on the debate around the explanatory role of the truth predicate.

References

[1] Feferman, S. (1960), 'Arithmetization of metamathematics in a general setting', *Fundamenta Mathematicae*, 49, 35-91.

[2] Given the need of connecting the two domains in the theory of truth, a binary satisfaction predicate is axiomatised instead of a unary truth predicate.

[3] If U is infinitely axiomatised, we will instead need to focus on the theory $\mathsf{T}[\mathcal{L}_{\mathsf{U}}]_{\mathsf{Q}} + AxT_{\mathsf{U}}$, where AxT_{U} is the formalisation in our 'disentangled' syntax of the statement 'all axioms of U are true'.

[2] Hájek, P., P. Pudlák (1993), *Metamathematics of First-Order Arithmetic*, Springer, Berlin.

[3] Halbach, V. (2011), *Axiomatic Theories of Truth*, Cambridge University Press.

[4] X and Y (2013), 'Axiomatic Truth, Syntax and Metatheoretic Reasoning', *The Review of Symbolic Logic* 6(4), (2013), pp. 613–636.

[5] Visser, A. (1991), 'The Formalization of Interpretability', *Studia Logica* 50(1): 81-106, 1991.