Complexity for Kripke's theory of truth

Stanislav Speranski

Steklov Mathematical Institute of RAS

(SIMC)

WDCM 2022

Kripke's theory of truth

Consider the signature of Peano arithmetic and its expansion obtained by adding an extra unary predicate symbol T, viz.

$$\sigma \;:=\; \{0, \mathrm{s}, +, \times, =\} \quad \text{and} \quad \sigma_T \;:=\; \sigma \cup \{T\}.$$

Throughout this presentation the following assumptions are in force:

- the connective symbols are $\neg$, $\wedge$ and $\vee$;
- the quantifier symbols are $\forall$ and $\exists$.

We abbreviate $\neg\varphi \vee \psi$ to $\varphi \to \psi$, $(\varphi \to \psi) \wedge (\psi \to \varphi)$ to $\varphi \leftrightarrow \psi$, etc. Let $\mathcal{L}$ and $\mathcal{L}_T$ be the first-order languages of σ and σ_T respectively.

Here is some related notation:

$$\begin{aligned}
For &:= \text{the collection of all } \mathcal{L}\text{-formulas;}\\
Sen &:= \text{the collection of all } \mathcal{L}\text{-sentences;}\\
For_T &:= \text{the collection of all } \mathcal{L}_T\text{-formulas;}\\
Sen_T &:= \text{the collection of all } \mathcal{L}_T\text{-sentences.}
\end{aligned}$$

Assume some Gödel numbering $\#$ of $\mathcal{L}_T$ has been chosen. Then we call $A \subseteq \mathbb{N}$ consistent iff there is no $\phi \in Sen_T$ s.t. both $\#\phi$ and $\#\neg\phi$ are in A. If $A \subseteq \mathbb{N}$, we write $\langle \mathfrak{N}, A \rangle$ for the expansion of the standard model $\mathfrak{N}$ of Peano arithmetic to σ_T in which T is interpreted as A.

In his 'Outline of a theory of truth', Kripke used partial interpretations of T, i.e. pairs of the form $S = \langle S^+, S^- \rangle$ where S^+ and S^- are disjoint subsets of $\mathbb{N}$, resp. called the extension of S and the anti-extension of S. Henceforth we limit ourselves to partial interpretations of T with consistent extensions. A partial valuation for σ_T is a mapping from Sen_T to a superset of $\left\{0, \frac{1}{2}, 1\right\}$.

By a valuation scheme we mean a function from partial interpretations to partial valuations. Among the most interesting such schemes are the schemes based on Kleene's strong and weak three-valued logics, sK and wK, which treat $\frac{1}{2}$ as 'undefined' and 'meaningless' respectively.

Define the strong Kleene valuation scheme V_{sK} inductively as follows:

- for any closed $\mathcal{L}$-terms t_1 and t_2,

$$V_{\mathsf{sK}}(S)(t_1 = t_2) := \begin{cases} 1 & \text{if } \mathfrak{N} \models t_1 = t_2, \\ 0 & \text{if } \mathfrak{N} \models t_1 \neq t_2; \end{cases}$$

- for every closed $\mathcal{L}$-term t,

$$V_{\mathsf{sK}}(S)(T(t)) := \begin{cases} 1 & \text{if } t^{\mathfrak{N}} \in S^{+}, \\ 0 & \text{if } t^{\mathfrak{N}} \in S^{-}, \\ \frac{1}{2} & \text{otherwise}; \end{cases}$$

- $V_{\mathsf{sK}}(S)(\varphi \wedge \psi) := \bigwedge_{\mathsf{sK}} \{V_{\mathsf{sK}}(S)(\varphi), V_{\mathsf{sK}}(S)(\psi)\}$;
- $V_{\mathsf{sK}}(S)(\forall x\, \varphi(x)) := \bigwedge_{\mathsf{sK}} \{V_{\mathsf{sK}}(S)(\varphi(t)) \mid t \text{ is a closed } \mathcal{L}\text{-term}\}$;

- $V_{\mathrm{sK}}(S)(\neg\varphi) := 1 - V_{\mathrm{sK}}(S)(\varphi)$;
- $V_{\mathrm{sK}}(S)(\varphi \vee \psi) := V_{\mathrm{sK}}(S)(\neg(\neg\varphi \wedge \neg\psi))$;
- $V_{\mathrm{sK}}(S)(\exists x\, \varphi(x)) := V_{\mathrm{sK}}(S)(\neg\forall x\, \neg\varphi(x))$.

To get the weak Kleene valuation scheme V_{wK}, we replace sK by wK.

Notice that each valuation scheme V induces a function $\mathfrak{J}_V$ from partial interpretations to partial interpretations, called the Kripke-jump operator for V, as follows:

$$\begin{aligned}
\mathfrak{J}_V(S)^+ &:= \{\#\varphi \mid \varphi \in Sen_T \text{ and } V(S)(\varphi) = 1\},\\
\mathfrak{J}_V(S)^- &:= \{\#\varphi \mid \varphi \in Sen_T \text{ and } V(S)(\varphi) = 0\}.
\end{aligned}$$

In turn $\mathcal{J}_V$ generates a transfinite sequence indexed by ordinals:

$$\mathcal{J}_V^{\alpha}(S) \;:=\; \begin{cases} S & \text{if } \alpha = 0, \\ \mathcal{J}_V\left(\mathcal{J}_V^{\beta}(S)\right) & \text{if } \alpha = \beta + 1, \\ \langle \bigcup_{\beta<\alpha} \mathcal{J}_V^{\beta}(S)^{+}, \bigcup_{\beta<\alpha} \mathcal{J}_V^{\beta}(S)^{-} \rangle & \text{if } \alpha \in \text{L-Ord}. \end{cases}$$

We shall often write T_V^{α} instead of $\mathcal{J}_V^{\alpha}(\varnothing, \varnothing)^{+}$ — these sets constitute the truth hierarchy for V.

Moreover Kripke dealt with monotone schemes, i.e. those which satisfy the condition that for any partial interpretations S_1 and S_2,

$$S_1^{+} \subseteq S_2^{+} \;\;\&\;\; S_1^{-} \subseteq S_2^{-} \quad \Longrightarrow$$
$$\Longrightarrow \quad \mathcal{J}_V(S_1)^{+} \subseteq \mathcal{J}_V(S_2)^{+} \;\;\&\;\; \mathcal{J}_V(S)^{-} \subseteq \mathcal{J}_V(S)^{-}.$$

Observation (Kripke)

For every monotone valuation scheme V there exists an ordinal α s.t. $\mathfrak{J}_V^{\alpha+1}(\varnothing,\varnothing) = \mathfrak{J}_V^{\alpha}(\varnothing,\varnothing)$ — yielding the least fixed point of $\mathfrak{J}_V$.

It is easy to verify that practically every valuation scheme considered in the literature is monotone and also satisfies:

- if $\mathfrak{J}_V(S) = S$, then $V(S)(T(\ulcorner\varphi\urcorner)) = V(S)(\varphi)$;
- $\#\varphi \in \mathfrak{J}_V^{\alpha}(S)^{-}$ iff $\#\neg\varphi \in \mathfrak{J}_V^{\alpha}(S)^{+}$;
- $\#\varphi \in \mathfrak{J}_V^{\alpha}(S)^{+}$ iff $\#\neg\varphi \in \mathfrak{J}_V^{\alpha}(S)^{-}$;
- $\mathfrak{J}_V$ turns out to be a Π^1_1-operator — so, by a well-known theorem, we have $\mathsf{T}_V^{\alpha} = \mathsf{T}_V^{\alpha+1}$ already for some $\alpha \in \text{C-Ord} \cup \{\omega_1^{\mathsf{CK}}\}$.

Kleene's $\mathcal{O}$

Remember that Kleene's system of notation for C-Ord consists of:

- a special partial function $\nu_{\mathcal{O}}$ from $\mathbb{N}$ onto C-Ord;
- an appropriate ordering relation $<_{\mathcal{O}}$ on $\operatorname{dom}(\nu_{\mathcal{O}})$ — which mimics the usual ordering relation on C-Ord.

Call $n \in \mathbb{N}$ a notation for $\alpha \in$ C-Ord iff $\nu_{\mathcal{O}}(n) = \alpha$. To simplify the statements I often write $n \in \mathcal{O}$ instead of $n \in \operatorname{dom}(\nu_{\mathcal{O}})$.

Folklore

$\operatorname{dom}(\nu_{\mathcal{O}})$ *is* Π^1_1*-complete.*

Fix one's favorite universal partial computable (two-place) function U.

Folklore

There exists a computable function f such that for every $n \in \mathcal{O}$,

$$\{k \in \mathbb{N} \mid k <_{\mathcal{O}} n\} \;=\; \mathrm{dom}\left(U_{f(n)}\right).$$

Folklore (Effective Transfinite Recursion)

Suppose f is a computable function such that for any $e \in \mathbb{N}$ and $n \in \mathcal{O}$,

$$\{k \in \mathbb{N} \mid k <_{\mathcal{O}} n\} \;\subseteq\; \mathrm{dom}\left(U_e\right) \quad\Longrightarrow\quad n \;\in\; \mathrm{dom}\left(U_{f(e)}\right).$$

Then there is a $c \in \mathbb{N}$ for which $U_{f(c)} = U_c$, and $\mathrm{dom}\left(\nu_{\mathcal{O}}\right) \subseteq \mathrm{dom}\left(U_c\right)$.

About least fixed-points

Let us call a valuation scheme V ordinary iff for any $\alpha \in \mathrm{Ord}$, $\chi \in Sen$, $\psi \in Sen_T$ and $\varphi(x) \in For_T$ the following conditions hold:

1. $\mathsf{T}_V^\alpha \subseteq \mathsf{T}_V^{\alpha+1}$;
2. $\chi \in \mathsf{T}_V^\alpha$ iff $\alpha \neq 0$ and $\mathfrak{N} \models \chi$;
3. $\psi \in \mathsf{T}_V^\alpha$ iff $T(\ulcorner \psi \urcorner) \in \mathsf{T}_V^{\alpha+1}$;
4. $\forall x\, \varphi(x) \in \mathsf{T}_V^{\alpha+1}$ iff $\{\varphi(\underline{n}) \mid n \in \mathbb{N}\} \subseteq \mathsf{T}_V^{\alpha+1}$;
5. $\chi \wedge \psi \in \mathsf{T}_V^\alpha$ iff $\mathfrak{N} \models \chi$ and $\psi \in \mathsf{T}_V^\alpha$;
6. if $\chi \vee \psi \in \mathsf{T}_V^\alpha$ and $\mathfrak{N} \models \neg\chi$, then $\psi \in \mathsf{T}_V^\alpha$;
7. if $\mathfrak{N} \models \chi$ and $\alpha \neq 0$, then $\chi \vee \psi \in \mathsf{T}_V^\alpha$.

In effect, V_{sK} and many other schemes considered in the literature turn out to be ordinary. But V_{wK} is not ordinary.

Given V, by the rank of $\psi \in Sen_T$ — denoted by $rank_V(\psi)$ — I mean the least ordinal α for which $\psi \in \mathsf{T}_V^{\alpha+1}$.

Proposition

Let V be a valuation scheme satisfying (3–4). Then for every $\psi \in Sen_T$ and every $\varphi(x) \in For_T$,

$$\begin{aligned} rank_V(T(\ulcorner\psi\urcorner)) &= rank_V(\psi) + 1 \quad \textit{and} \\ rank_V(\forall x\, \varphi(x)) &= \sup\{rank_V(\varphi(\underline{n})) \mid n \in \mathbb{N}\}. \end{aligned}$$

Theorem $\sharp$

For each ordinary scheme V there exists a computable function ρ_V such that for every $n \in \mathcal{O}$, $rank_V(\rho_V(n)) = \nu_{\mathcal{O}}(n) + 1$.

Corollary ♯

Each ordinary scheme V has the following property: for every ordinal α, if $\mathsf{T}_V^\alpha = \mathsf{T}_V^{\alpha+1}$, then $\alpha \geqslant \omega_1^{\mathsf{CK}}$ and T_V^α is Π_1^1-hard.

The technique used in the proofs of these facts can be applied in various other situations as well. Let us see how it works e.g. for V_{wK}. Still, as it was shown by Cain and Damnjanovic, one should be warned:

> *Actually certain complexity results for the weak Kleene scheme depend on the Gödel numbering and the language of the "standard" model of arithmetic we choose.*

I am aiming at a deeper understanding of this intensionality phenomenon.

In their article from 1991, Cain and Damnjanovic suggested expanding σ to avoid the conflict. More precisely, assuming an appropriate coding $\mathsf{M}_0, \mathsf{M}_1, \ldots$ of all Turing machines, they added a new function symbol π of arity 4, whose interpretation is given by

$$\pi(e, i, j, k) \;:=\; \begin{cases} n & \text{if } \mathsf{M}_e \text{ halts on input } i \text{ at step } j \text{ with output } n, \\ k & \text{if } \mathsf{M}_e \text{ does not halt on input } i \text{ at step } j. \end{cases}$$

Clearly this function is primitive recursive. *So what can we do with π?*

Observation A

If we include π in σ, then both Theorem $\sharp$ and Corollary $\sharp$ generalise to arbitrary valuation schemes satisfying (1–5).

As an alternative to Cain–Damnjanovic' suggestion, I propose to add a symbol $\dot{-}$ for cut-off subtraction, i.e. $i \dot{-} j := \max\{0, i - j\}$:

Observation B

Similar to Observation A, but with $\dot{-}$ instead of π.

Another modification, with $\mathcal{L}$ unchanged, deals with the following condition (for any $\alpha \in \mathrm{Ord}$ and $\theta(x) \in \mathit{For}$):

8. $\exists x\,(\theta(x) \wedge T(x)) \in \mathrm{T}_V^{\alpha}$ iff $\mathfrak{N} \models \theta(\underline{n})$ and $T(\underline{n}) \in \mathrm{T}_V^{\alpha}$ for some $n \in \mathbb{N}$.

Observation C

The analogues of Theorem $\sharp$ and Corollary $\sharp$ hold for all schemes satisfying (1–5) and (8).

In fact, although (8) fails for the weak Kleene scheme, the customary treatment of $\exists$ in the case of V_{wK} does not seem to be well motivated. Alternatively, we can define V^*_{wK} exactly as V_{wK} except that

$$V^*_{\mathrm{wK}}(S)(\exists x\, \varphi(x)) := \begin{cases} 1 & \text{if } V^*_{\mathrm{wK}}(S)(\varphi(t)) = 1 \text{ for some closed } \mathcal{L}\text{-term } t, \\ 0 & \text{if } V^*_{\mathrm{wK}}(S)(\varphi(t)) = 0 \text{ for all closed } \mathcal{L}\text{-terms } t, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

(treating $\exists$ like in the strong Kleene scheme V_{sK}). Then V^*_{wK} satisfies (1–5) and (8), so Observation C applies.

Earlier we took $\to$ as an abbreviation. However, interpreting $\varphi \to \psi$ as $\neg\varphi \vee \psi$ is not always the right choice. To avoid confusion, I add a new connective symbol $\twoheadrightarrow$ to the original three (viz. $\neg$, $\wedge$ and $\vee$). Of course *For*, For_T, *Sen* and Sen_T are easily modified to accommodate $\twoheadrightarrow$. Now consider the following variation on (6–7):

6' if $\chi \twoheadrightarrow \psi \in \mathsf{T}^{\alpha}_{V}$ and $\mathfrak{N} \models \chi$, then $\psi \in \mathsf{T}^{\alpha}_{V}$;

7' if $\mathfrak{N} \models \neg\chi$, then $\chi \twoheadrightarrow \psi \in \mathsf{T}^{\alpha}_{V}$

— where χ and ψ range over the modified versions of *Sen* and Sen_T resp. Evidently, even when we treat $\twoheadrightarrow$ as the material conditional on $\{0, 1\}$, the meanings of $\varphi \twoheadrightarrow \psi$ and $\neg\varphi \vee \psi$ may differ on $\left\{0, \frac{1}{2}, 1\right\}$.

Observation D

If we expand $\mathcal{L}$ and $\mathcal{L}_T$ by adding $\twoheadrightarrow$, then the analogues of Theorem $\sharp$ and Corollary $\sharp$ hold for all schemes satisfying (1–5) and (6'–7').

This is closely related to a three-valued scheme from Feferman's article 'Axioms for determinateness and truth'. It can be obtained by extending V_{wK} to formulas containing $\twoheadrightarrow$ by setting

$$V'_{\mathsf{wK}}(S)(\varphi \twoheadrightarrow \psi) := \begin{cases} 1 & \text{if } V'_{\mathsf{wK}}(S)(\varphi) = 0 \text{ or } V'_{\mathsf{wK}}(S)(\psi) = V'_{\mathsf{wK}}(S)(\varphi) = 1, \\ 0 & \text{if } V'_{\mathsf{wK}}(S)(\varphi) = 1 \text{ and } V'_{\mathsf{wK}}(S)(\psi) = 0, \\ \frac{1}{2} & \text{otherwise;} \end{cases}$$

(the other clauses are the same as in the definition of V_{wK}). Now V'_{wK} satisfies (1–5) and (6'–7'), so Observation D applies.

Some strengthenings

One curious scheme emerges from Leitgeb's 'What truth depends on' — although the definition presented below was stated explicitly by Thomas Schindler. Say that $\varphi \in Sen_T$ depends on $A \subseteq \mathbb{N}$ iff for every $B \subseteq \mathbb{N}$,

$$\langle \mathfrak{N}, B \rangle \models \varphi \quad \Longleftrightarrow \quad \langle \mathfrak{N}, B \cap A \rangle \models \varphi.$$

Now define Leitgeb's valuation scheme V_{L} by

$$V_{\mathsf{L}}(S)(\varphi) := \begin{cases} 1 & \text{if } \varphi \text{ depends on } S^+ \cup S^- \text{ and } \langle \mathfrak{N}, S^+ \rangle \models \varphi, \\ 0 & \text{if } \varphi \text{ depends on } S^+ \cup S^- \text{ and } \langle \mathfrak{N}, S^+ \rangle \models \neg\varphi, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Note that every $\mathcal{L}_T$-formula can be viewed as an arithmetical monadic second-order $\sigma_{\mathbb{N}}$-formula whose only set variable is T, and vice versa. Given an $\mathcal{L}_T$-sentence ψ and an $\mathcal{L}$-formula $\chi(x)$, we construct

$$\psi_\chi \ := \ \text{the result of replacing each } T(t) \text{ in } \psi \text{ by } \chi(t) \wedge T(t).$$

Observation E

Let $\chi(x)$ be an $\mathcal{L}$-formula defining an infinite computable subset of $\mathbb{N}$ in $\mathfrak{N}$. Then $\{\psi_\chi \mid \psi \in Sen_T \text{ and } \mathfrak{N} \models \forall T\, \psi_\chi(T)\}$ is Π^1_1-complete.

It gives an alternative and probably the shortest proof for the following.

Theorem ♣ (Welch, Hjorth, Meadows)

Let V be V_{L}. Then for every $\alpha \in \mathrm{Ord}^{+}$, T_V^{α} is Π^1_1-hard. The same holds if V is a reasonable supervaluation scheme.

Proof.

Assume $V = V_{\mathsf{L}}$. Take A to be $\#\{\mu, T(\ulcorner\mu\urcorner), T(\ulcorner T(\ulcorner\mu\urcorner)\urcorner), \ldots\}$ with μ denoting some fixed 'truthteller', and let χ be an $\mathcal{L}$-formula defining A in $\mathfrak{N}$. Since $A \cap \mathsf{G} = \varnothing$, we obtain

$$\begin{aligned}
\#\psi_\chi \in \mathsf{T}_V^{\beta+1} \quad &\Longleftrightarrow \quad \#\psi_\chi \in \mathsf{G}_{\beta+1} \text{ and } \langle \mathfrak{N}, \mathsf{T}_V^{\beta} \rangle \models \psi_\chi \\
&\Longleftrightarrow \quad \mathfrak{N} \models \forall T\,(\psi_\chi(T \cap \mathsf{G}_\beta) \leftrightarrow \psi_\chi(T)) \wedge \psi_\chi(\mathsf{T}_V^{\beta}) \\
&\Longleftrightarrow \quad \mathfrak{N} \models \forall T\,(\psi_\chi(\varnothing) \leftrightarrow \psi_\chi(T)) \wedge \psi_\chi(\varnothing) \\
&\Longleftrightarrow \quad \mathfrak{N} \models \forall T\,\psi_\chi(T).
\end{aligned}$$

Clearly $\mathsf{T}_V^{\alpha} = \bigcup_{\beta<\alpha} \mathsf{T}_V^{\beta+1}$, so the Π^1_1-hardness of T_V^{α} follows by Observation E. Perfectly analogous arguments apply to the other schemes. □

Some references

1975: S. Kripke. Outline of a theory of truth. *Journal of Philosophy* 72:19, 690–716.

1986: J. P. Burgess. The truth is never simple. *J. Symbolic Logic* 51:4, 663–681.

1991: J. Cain and Z. Damnjanovic. On the weak Kleene scheme in Kripke's theory of truth. *J. Symbolic Logic* 56:4, 1452–1468.

2005: H. Leitgeb. What truth depends on. *J. Philosophical Logic* 34:2, 155–192.

2008: S. Feferman. Axioms for determinateness and truth. *Review of Symbolic Logic* 1:2, 204–217.

2014: P. D. Welch. The complexity of the dependence operator. *J. Philosophical Logic* 44:3, 337–340.

2015: T. Schindler. *Type-Free Truth* (Ph.D. Thesis). Ludwig-Maximilians-Universität München.

2017: S. Speranski. Notes on the computational aspects of Kripke's theory of truth. *Studia Logica* 105:2, 407–429.