A Theory of Truth, I. Preliminary Report, and

A Theory of Truth, II

Saul A. Kripke

These are published abstracts from Baldwin, J., Martin, D. A., Soare, R. I., and Tait, W. W. (1976), Meeting of the Association for Symbolic Logic, Chicago 1975. *Journal of Symbolic Logic*, 41(2): 551-560, which can be obtained from the publisher at https://doi.org/10.1017/S0022481200051665. They are reproduced here by permission of the Association for Symbolic Logic which holds the copyright.

© Association for Symbolic Logic
given operations on $\mathcal{F}$ are monotonic with respect to $\leq$, (ii) $\forall \varphi \exists \psi (\forall x (\varphi x \leq \psi x) \rightarrow \varphi \leq \psi)$ and (iii) each chain $\mathbb{M}$ in $\mathcal{F}$ has an upper bound $\psi$ satisfying the condition

$$\forall \varphi \forall \chi (\exists \theta (\theta \in \mathbb{M} \rightarrow \varphi \theta x \leq \chi) \rightarrow \varphi \psi x \leq \chi).$$

The iteration of $\varphi$ controlled by $\chi$ is by definition the least solution $\xi$ of the equation $\xi = (\chi \circ I, \theta, \varphi)$. A “function” $\psi$ is called recursive in some “functions” $\psi_1, \ldots, \psi_n$ iff $\psi$ can be obtained from $I, L, R, T, F, \psi_1, \ldots, \psi_n$ by means of the three given operations on $\mathcal{F}$ and iteration. For this notion of recursiveness we prove a normal form theorem, an enumeration theorem and the first and second recursion theorems.

Consider a set $M$ together with a pairing mechanism $J, L, R$ on it and let $c_i \in M_i \subseteq M$ ($i = 0, 1), M_0 \cap M_1 = \emptyset$ (using the notations from [1], we can for example take $M$ to be the set $B^*$ corresponding to an arbitrary set $B$ and take $J = \text{ast.}(s, t), L = \text{cr}, R = \text{8}, M_0 = B^0, M_1 = B^* - B^0, c_0 = 0, c_1 = 1$). We obtain a model for the given system of axioms taking $\mathcal{F}$ to be the set of all partial multiple-valued mappings of $M$ into $M$ with the natural rule of composition and the natural partial ordering and taking $\vartheta = (\lambda c : c \in M), (\varphi, \psi) = \lambda s.J(\varphi(s), \psi(s)), (\chi \circ \varphi, \psi) = \lambda s.(\chi(s) \cap M_0 \neq \emptyset \land t \in \varphi(s)) \lor (\chi(s) \cap M_1 \neq \emptyset \land t \in \psi(s))); T = \lambda s.c_0, F = \lambda s.c_1; \text{in the special case which corresponds to [1] our notion of recursiveness will be equivalent to absolute prime computability. Other models can be obtained by taking } \mathcal{F} \text{ to be the set of the partial mappings of } M \text{ into } M \text{ or by considering fuzzy or probabilistic mappings of } M \text{ into } M. \text{ We can also take } \mathcal{F} \text{ to be the set of all pairs } <D, f>, \text{ where } f \text{ is a partial multiple-valued mapping of } M \text{ into } M \text{ and } D \subseteq \text{Dom } f, \text{ and define multiplication by } <D_0, f_0> - <D_1, f_1> = <D, f_0 f_1>, \text{ where } D = \{t: t \in D_1 \land f_0(t) \subseteq D_0\}. \text{ Then for a suitable interpretation of the rest of the primitive notions we again obtain a model for the considered system.}

REFERENCE


Let $L_0$ be an interpreted first-order language whose domain $D$ includes the natural numbers. ('First-order' is only for definiteness; generalized quantifiers, even infinitary connectives, etc., could be included without damaging the construction.) Extend it to a language $L$ with one additional (uninterpreted) monadic predicate $T(x)$. A partial subset of $D, S = (S_1, S_2)$ is a pair of disjoint subsets of $D; S_1 (S_2)$ contains the members (nonmembers) of $S$, $S$ is 'undefined' elsewhere. $S \leq S'$ means $S_1 \subseteq S'_1$ and $S_2 \subseteq S'_2$. When $T(x)$ is interpreted by a partial subset of $L$, formulae of $L$ can be evaluated as true, false, or undefined by an 'appropriate' 3-valued scheme. Appropriate schemes include (the natural quantificational extension of) Kleene's 3-valued logic and van Fraassen's supervaluations. (There are others; the 'weak' 3-valued logic is 'appropriate' but is an unhappy choice unless $L_0$ is enriched by a restricted universal quantifier.) One such 3-valued scheme is assumed to be chosen throughout.

If $T(x)$ is interpreted by $S = (S_1, S_2)$, let $\phi(S) = S' = (S'_1, S'_2)$, where $S'_1 (S'_2)$ is the set of Gödel numbers of true (false) sentences of $L$ under the suggested interpretation. Then if $S_1 \leq S_2, \phi(S_1) \leq \phi(S_2)$.

An easy theorem: $\phi$ has a (least) fixed point. Indeed, if we define $S_0 =$ the completely undefined partial set, $S_{a+1} = \phi(S_a)$ for $\beta$ a limit ordinal $\phi(S_\beta) =$ l.u.b. $S_\alpha (\alpha < \beta)$, then the least $\alpha$ (‘$\alpha_0$’) such that $S_\alpha = S_{\alpha+1}$ gives the least fixed point of $\phi$. Every f.p. is $\leq$ a maximal f.p. Usually, there is no greatest f.p., but there is a greatest ‘intrinsic’ f.p. (f.p. compatible with every other f.p.). The intrinsic f.p.'s form a complete lattice under $\leq$.

If $T(x)$ is interpreted by a fixed point of $\phi$, $L$ becomes a language with its own truth predicate.

SAUL KRIPEKE, A theory of truth. II. We continue the preceding abstract.

It is suggested that interpretations of $L$ using fixed points of $\phi$ form an approximate model for the intuitive concept of (expressing a) truth in natural language. The least fixed point probably is the most natural, but the others are useful to make certain intuitive distinctions
precise. (‘I am false’ lacks truth-value in \textit{all} f.p.’s; ‘I am true’ has no truth-value in any intrinsic f.p., but has one in every maximal f.p.; etc.)

In the usual Tarski ‘hierarchy’ approach, if Smith says, (1) “Everything Jones says is true”, he must choose a ‘level’ for “true”. If (unbeknownst to Smith) some of Jones utterances are on too high a ‘level’, (1) may not ‘cover’ everything Jones says. The present proposal assigns no ‘level’ to “true”. If (1) has a truth-value in the least f.p. (as it will in ‘normal’ cases), in a sense it has a ‘level’ (the least \( \beta \) such that (1) \( \in S_\beta \)), but the ‘level’ depends on the facts about what Jones says, rather than the ‘truth-predicate’ of (1).

Nevertheless if \( L_0 \) contains arithmetic, a Tarski hierarchy of truth-predicates and meta-languages for \( L_0 \) can be constructed within \( L \). The idea: if \( A(x) \) is true of the formulae of \( L_0 \), \( T_0(x) = T(x) \land A_0(x), L_1 \) is the sublanguage formed by adjoining \( T_0(x) \) to \( L_0 \), etc. The construction can be continued through high ordinals \(< a_0 \); we omit details, which are related to the hyperarithmic hierarchy, and can be used to construct it. Formulae such as (1), however, appear on no level of the Tarski hierarchy.

One can make a fixed point two-valued by declaring \( T(x) \) false wherever it was undefined. Tarski’s convention \( T \) then becomes: \( T(\neg A') \lor T(\neg \neg A') \Rightarrow (A \equiv T(\neg A') \land (\neg A \equiv T'(\neg \neg A')) \). This leads to a simple axiomatic theory of truth.

If \( D \) contains (codes of) all finite sequences of elements of \( D \), an analogous construction allows \( L_0 \) to be extended to a language with its own satisfaction predicate. (Here the ordinal of the induction may be uncountable.) Extensions to intensional languages are also possible. Interesting technical questions and theorems arise in the investigation.

**REFERENCES**


1 This research was partially supported by NSF P038823.