Adding a conditional to Kripke's theory of truth

Lorenzo Rossi

University of Oxford

lorenzo.rossi@philosophy.ox.ac.uk

Several authors put great effort into developing a theory of truth consistent with (some form of) the naive conception of truth by giving up classical semantics.

Kripke's construction ([3]) is a successful starting point in this project. Take some theory (usually Peano arithmetic) that can "describe" its own syntax, and let $\mathcal{L}$ be its language. Let $\mathcal{L}_T := \mathcal{L} \cup \{T\}$ for a fresh predicate T for "... is true". T applies to names of sentences, where $\ulcorner\varphi\urcorner$ is the name of the sentence φ in some coding system. In Kripke's models, for every $\varphi \in \mathcal{L}_T$, φ and $T\ulcorner\varphi\urcorner$ have the same truth-value. Moreover, Kripke's semantics gives an appealing account of long-debated paradoxes: if λ denotes the liar sentence, then λ and $\neg\lambda$ have the same truth-value in Kripke's construction. Regrettably, the logic associated to Kripke's construction is weak and not very expressive: in particular, there is no way to represent the above fact about φ and $T\ulcorner\varphi\urcorner$, or λ and $\neg\lambda$, inside the theory. Some writers (notably Field [1]) identify this expressive problem of Kripke's theory in its lack of a non-trivial conditional connective. The proposed solution is to add to Kripke's theory a new conditional ($\rightarrow$), different from the material one, and a new biconditional ($\leftrightarrow$) s.t. $\varphi \leftrightarrow \psi := (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$, providing a new semantics that validates sentences such as $\varphi \leftrightarrow T\ulcorner\varphi\urcorner$ and $\lambda \leftrightarrow \neg\lambda$. To this end, Field gives a revision-theoretic construction; unfortunately, his model is exceedingly complex from the recursion-theoretic standpoint, it is not entirely free from revenge paradoxes (see Welch [4]), and it is not clear which logic can be associated to it (if any). One could wonder whether it is possible to improve on Kripke's theory, equipping it with an interesting conditional whose semantics validates (many of) the equivalences we would like to express, in a model of low complexity. To address this question, I stick to inductive constructions, that are simple and treatable. Let $\mathcal{L}_T^{\rightarrow} := \mathcal{L}_T \cup \{\rightarrow\}$ (for a new $\rightarrow$). Kripke's construction for $\mathcal{L}_T^{\rightarrow}$ is enriched by adding the "monotone fragment" of the semantics of $\rightarrow$ as given in Łukasiewicz's 3-valued logic (Ł3). The result is a partial semantics for $\mathcal{L}_T$, which approximates Ł3 and is obtained via an inductive construction. Such construction is formalized by a monotone operator Υ that incorporates the Kripkean jump and acts on triples of sets of sentences A, B, C . A (B, C) represents the sentences we suppose to be true-in-the-model (false-, gappy-in-the-model) at the beginning. Given $\langle A, B, C\rangle$ as input, Υ yields a new triple $\langle A', B', C'\rangle$, and in each of A', B', and C' also sentences of the form $\varphi \rightarrow \psi$ are introduced, according to the previous evaluation of φ and ψ. The process grows monotonically, until it reaches a fixed point $\langle A_\infty, B_\infty, C_\infty\rangle$. As in [3], every fixed point depends largely on the sets we start with, but Υ exhibits a nice behaviour on some standard choices. In particular, consistent fixed points exist that contain interesting equivalences expressed via $\rightarrow$ such as $\lambda \leftrightarrow \neg\lambda$ and $\varphi \leftrightarrow T\ulcorner\varphi\urcorner$, for a large class of sentences φ of $\mathcal{L}_T^{\rightarrow}$ (including "gappy" ones). This semantics is

3-valued and partial: sentences have value 1, 0, $1/2$, or no value. Unlike in Kripke's models, if φ is valued $1/2$, this is a positive semantic information, it doesn't follow because φ has not value 1 nor 0. Some "not 1- nor 0-valued" sentences can be safely valued $1/2$ and used in a monotone evaluation for a Ł3-like conditional, while some others must remain valueless for a consistent fixed point. Now the question is how to devise a simple way to get (many) safe sentences in the starting choice for C. Such safe sentences may result from some simple model-theoretic method or from some simple search process (say, recursively enumerable). Let's consider an example of the latter case. Consider Halbach and Horsten's [2] theory PKF. It is well-known that, for example, $\lambda \vdash_{\mathsf{PKF}} \neg\lambda$ and $\neg\lambda \vdash_{\mathsf{PKF}} \lambda$ ($\vdash_{\mathsf{PKF}}$ denotes provability in PKF). 1- and 0-valued sentences cannot consistently be closed under addition or removal of an odd number of negations, but this is permitted for gappy sentences, unless we adopt an unconventional notion of gappiness. Therefore, PKF proves the liar sentence to be gappy. In this way, the gappiness of the liar sentence is given a "positive" account: it is effectively shown, it's not merely a by-product of the fact that the liar sentence has not value 1 nor 0 (where both values 1 and 0 are independently built). Thus, a positive and precise notion of gappiness exists: gappiness provable in PKF. In the case of PKF, it is shown that the resulting semantic construction (via Υ) is consistent, expressive and purely inductive. Of course, other theories of "provable gappiness" can be considered. A simple variant of PKF yields particularly uniform fixed points that could admit a partial axiomatization.

References

[1] Field, Hartry (2008), *Saving truth from paradox*, OUP, Oxford.

[2] Halbach, Volker and Leon Horsten (2006), "Axiomatizing Kripke's Theory of Truth", *The Journal of Symbolic Logic*, 71, 2, pp. 677-712.

[3] Kripke, Saul (1975), "Outline of a theory of truth", *Journal of Philosophy*, 72, 19: 690-750.

[4] Welch, Philip (2014), "Some Observations on Truth Hierarchies", in *The Review of Symbolic Logic*, 7, 01, pp. 1-30.