Fuzzy Semantics for the Language of Precise Truth*

Theodor Nenu

Hertford College, University of Oxford
theodor.nenu@hertford.ox.ac.uk

Abstract

This short paper investigates the prospects of designing semantically satisfactory fuzzy models for the formal language of precise truth. We start by showing that this language fails to admit fuzzy models based on Kronecker-Delta semantics for sharp truth-predications, and then we explore some alternative semantic possibilities.

1 Species of Truth Predicates

In his work on the topic of vagueness, Smith [Smi08] made an important logico-philosophical distinction between two kinds of truth predicates:

1. The global truth-predicate T with the property that the semantic value of any truth predication $\mathsf{T}(\overline{\ulcorner\varphi\urcorner})$ matches the semantic value of the underlying sentence φ, i.e. $[\![\mathsf{T}(\overline{\ulcorner\varphi\urcorner})]\!] = [\![\varphi]\!]$.

2. The family $\{\, \mathsf{T}_\mathsf{i} \mid i \in [0,1] \,\}$ of *indexed* truth-predicates that we use in order to say that a sentence has a *specific* degree of truth—e.g. that it is true to degree 0.54, written as $\mathsf{T}_{0.54}(\overline{\ulcorner\varphi\urcorner})$.

The formal semantics literature contains many non-classical ways in which one can successfully add the basic symbol T to the (object-)language of arithmetic, viz. $\mathcal{L}_{\mathsf{PA}}$ (e.g. [Kri75]). We shall now add more symbols to $\mathcal{L}_{\mathsf{PA}}$ in order to enhance its expressive powers, so that precise truth-predications can be articulated. Let $\mathcal{L}_{\mathsf{T}}^{\infty} := \mathcal{L}_{\mathsf{PA}} \cup \{\, \mathsf{T}_\mathsf{i} \mid i \in [0,1] \,\}$.[1] Ideally, the semantics for precise truth-predications should be governed by the Kronecker-Delta function $\delta : [0,1]^2 \to \{\, 0, 1 \,\}$ given by:

$$\delta(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

More explicitly, if $\mathsf{T}_\mathsf{r}(\overline{\ulcorner\varphi\urcorner})$ is an atomic precise truth-predication for some $r \in [0,1]$, we want that:

$$[\![\mathsf{T}_\mathsf{r}(\overline{\ulcorner\varphi\urcorner})]\!] = \begin{cases} 1 & \text{if } r = [\![\varphi]\!] \\ 0 & \text{otherwise} \end{cases}$$

*I would like to thank Johannes Stern for many insightful discussions. This paper is based on Section 6.5 of my PhD thesis, and my research was supported by an ERC Starting Grant (TRUST, Grant No. 803684).

[1] Of course, we can avoid making our language uncountable. We can restrict our indexes to range over $\mathbb{Q} \cap [0,1]$ and use a two-place predicate T with the property that $\mathsf{T}(\overline{\mathsf{n}}, \overline{\ulcorner\varphi\urcorner})$ is (perfectly) true iff φ has the n^{th} rational number—in the canonical enumeration of the countable set $\mathbb{Q}$—as its truth-degree. In other words, when we're writing $\mathsf{T}_\mathsf{i}(\overline{\ulcorner\varphi\urcorner})$, this can be seen as shorthand for $\mathsf{T}(\overline{\#\mathsf{i}}, \overline{\ulcorner\varphi\urcorner})$, where $\#i \in \mathbb{N}$ is i's code. For simplicity, in this paper we'll carry out our formal investigation as if everything were real-valued.

2 δ-Semantics for $\mathcal{L}_{\mathsf{T}}^{\infty}$

Under this picture, there are *prima facie* no obvious obstacles in providing semantic values for some sentences of interest. For example, consider the following lemma, that we shall prove, which concerns the semantic value of the sentence that denies its bivalence:

Lemma 2.1 (The Bivalence Denier)**.** The $\mathcal{L}_{\mathsf{T}}^{\infty}$-sentence which says about itself that it is a counterexample to the principle of bivalence can only be perfectly false (regardless of the truth-structure underpinning the fuzzy semantics).

Proof. The bivalence denier, τ, asserts that its semantic value is neither 0 nor 1. Let τ be a fixed point of the open formula $\neg(\mathsf{T}_0(\mathsf{x}) \vee \mathsf{T}_1(\mathsf{x}))$. By the semantic definition of a fixed point, it follows that $[\![\tau]\!] = [\![\neg(\mathsf{T}_0(\overline{\ulcorner\tau\urcorner}) \vee \mathsf{T}_1(\overline{\ulcorner\tau\urcorner}))]\!]$. Suppose $[\![\tau]\!] \in (0,1)$. Then $[\![\neg(\mathsf{T}_0(\overline{\ulcorner\tau\urcorner}) \vee \mathsf{T}_1(\overline{\ulcorner\tau\urcorner}))]\!] = f_{\neg}(f_{\vee}(\delta([\![\tau]\!],0),\delta([\![\tau]\!],1))) = f_{\neg}(f_{\vee}(0,0)) = f_{\neg}(0) = 1$. Hence we cannot assign a truth degree strictly between 0 and 1 to τ. Now suppose that $[\![\tau]\!] \in \{0,1\}$. In this case we have that $[\![\neg(\mathsf{T}_0(\overline{\ulcorner\tau\urcorner}) \vee \mathsf{T}_1(\overline{\ulcorner\tau\urcorner}))]\!] = f_{\neg}(f_{\vee}(\delta([\![\tau]\!],0),\delta([\![\tau]\!],1))) = f_{\neg}(1) = 0$. Thus, truth-value 1 is discounted and 0 is the only possibility. □

That being said, some unfortunate news are due. Even though the Kronecker-Delta semantics for $\mathcal{L}_{\mathsf{T}}^{\infty}$ seems promising with respect to a multitude of sentences, we can mathematically prove the negative result that $\mathcal{L}_{\mathsf{T}}^{\infty}$ has no models at all. The result resembles in many respects Tarski's [Tar56] classical argument:

Theorem 2.2 (The Undefinability of Precise Truth)**.** There are no fuzzy models $\mathcal{M}$ of language $\mathcal{L}_{\mathsf{T}}^{\infty}$.[2]

Proof. Suppose there is a model $\mathcal{M}$ of our language, where the semantics of the indexed truth-predicates is guided by the Kronecker-Delta proposal. For any $y \in [0,1]$, let λ_{y} be $\neg\mathsf{T}_{\mathsf{y}}(\mathsf{x})$'s liar sentence. To show that $\mathcal{M}$ cannot exist, we just need to show that there is at least one number r in the unit interval such that there's no truth-degree that can be assigned to λ_{r}.

Let's start by checking what happens to λ_1. Given the semantics of the indexed truth-predicates (and the workings of the generalised negation function), it follows that $[\![\neg\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})]\!] \in \{0,1\}$, which means that λ_1 itself can only be interpreted as 0 or 1. Now, if $[\![\lambda_1]\!] = 1$, then $\delta([\![\lambda_1]\!],1) = 1$, so $[\![\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})]\!] = 1$, which in turn means that $[\![\neg\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})]\!] = 0$. Since $\neg\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})$ and λ_1 must have matching semantic values, this is impossible.

On the other hand, if $[\![\lambda_1]\!] = 0$, then $\delta([\![\lambda_1]\!],1) = 0$, so $[\![\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})]\!] = 0$, which means that $[\![\neg\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})]\!] = 1$. Just as in the last case, this cannot obtain. In conclusion, there cannot be any fuzzy models of the entire language $\mathcal{L}_{\mathsf{T}}^{\infty}$ because there is at least one uninterpretable symbol of $\mathcal{L}_{\mathsf{T}}^{\infty}$—and T_1 serves as an explicit example. □

3 Alternative Fuzzy Semantics for $\mathcal{L}_{\mathsf{T}}^{\infty}$

Perhaps the Kronecker-Delta semantics that we relied on is overly punishing of close mismatches of values. Under this brand of semantics, if some sentence φ has semantic value r,

[2] Hájek, Paris and Shepherdson [HPS00] prove a result in this vicinity, but theirs is slightly different than ours. In their paper, they show that the standard model of arithmetic, $\mathcal{N}$, cannot be extended to a model of $\mathsf{PAT}_{\text{Ł}\forall}$. There are no indexed truths in their framework—it's only about a global, disquotational truth-predicate T. They have also shown that theory $\mathsf{PAT}_{\text{Ł}\forall}$ is actually consistent, but it immediately becomes inconsistent if one attempts to extend it with truth-theoretic axioms which say that T commutes with connectives.

then for any small $\varepsilon > 0$ and $s \in (r - \varepsilon, r + \varepsilon) \setminus \{ r \}$, the precise truth-predication $\mathsf{T}_\mathsf{s}(\overline{\ulcorner \varphi \urcorner})$ will have semantic value 0. This seems too harsh. For a concrete example, suppose that:

$$[\![\varphi]\!] = 0.67583 \text{ for some } \varphi \in \mathsf{Sent}_{\mathcal{L}_\mathsf{T}^\infty}$$

Then, using our semantics, the following assignment will obtain:

$$[\![\mathsf{T}_{0.67584}(\overline{\ulcorner \varphi \urcorner})]\!] = 0$$

even though, roughly speaking, the indexed-predicate $\mathsf{T}_{0.67584}$ "got it right"—the error is just 0.00001. It seems reasonable to suggest that the proper semantic value of $\mathsf{T}_{0.67584}(\overline{\ulcorner \varphi \urcorner})$ ought to be some number $s \in (1 - \varepsilon, 1)$ for some very small $\varepsilon > 0$. One way of accomplishing this might be to suggest that, if $\varphi_\mathsf{r} \in \mathsf{Sent}_{\mathcal{L}_\mathsf{T}^\infty}$ is a sentence with semantic value $r \in [0, 1]$, then:

$$[\![\mathsf{T}_\mathsf{s}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})]\!] = \gamma^{d(s,r)}$$

where $\gamma > 0$ is some tiny, epsilonic number, e.g. Liouville's constant (or any other small quantity), and $d : \mathbb{R}^2 \to R_{\geq 0}$ is the ordinary distance function on the reals, defined as follows:

$$d(x, y) = \begin{cases} y - x & \text{if } x \leq y \\ x - y & \text{otherwise} \end{cases}$$

This semantic framework has the following upshots:

- When $s = r$, then $[\![\mathsf{T}_\mathsf{s}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})]\!] = \gamma^0 = 1$.
- When $s \approx r$, then $[\![\mathsf{T}_\mathsf{s}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})]\!] \approx 1$.

Thus, unlike the Kronecker-Delta semantics, where indexed truth predications can only take Boolean values, we now allow for fuzzy semantic values for precise statements such as $\mathsf{T}_\mathsf{s}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})$. That being said, this fuzzified framework makes some odd predictions of its own. For instance, this framework makes it impossible for any precise truth-predication to be perfectly false, since 0 is not in the range of function $f(x, y) := \gamma^{d(x,y)}$. This means that even statements that attribute perfect truth to outright falsities, e.g. $\mathsf{T}_1(\overline{\ulcorner 2 + 2 = 5 \urcorner})$, will turn out *partially true*. This seems seriously problematic.

What if, instead of imposing $[\![\mathsf{T}_\mathsf{s}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})]\!] = \gamma^{d(s,r)}$, we designed our semantics to assign the product $s \times r$ as the semantic value of $\mathsf{T}_\mathsf{s}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})$? This sounds like a natural suggestion, but it comes with some other problematic predictions. For example, the sentence $\mathsf{T}_\mathsf{r}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})$ should be a paradigmatic example of a perfectly true sentence, since it says that sentence φ_r has semantic value r, which it does.[3] However, the product-semantics gets this wrong, since for any $r \in (0, 1)$, we have that $[\![\mathsf{T}_\mathsf{r}(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})]\!] = r^2$, which is strictly less than 1. Another problem arises when we consider positive truth-predications of perfectly false sentences, e.g. $\mathsf{T}_\mathsf{r}(\overline{\ulcorner \varphi_0 \urcorner})$, or perfectly false predications of partially true sentences, e.g. $\mathsf{T}_0(\overline{\ulcorner \varphi_\mathsf{r} \urcorner})$. With respect to the former case: if $r \approx 0$, the product-semantics makes the wrong assignment $[\![\mathsf{T}_\mathsf{r}(\overline{\ulcorner \varphi_0 \urcorner})]\!] = 0$, when in fact it should be the case that $[\![\mathsf{T}_\mathsf{r}(\overline{\ulcorner \varphi_0 \urcorner})]\!] \approx 1$.[4]

[3] In particular, $[\![\mathsf{T}_{\frac{1}{2}}(\overline{\ulcorner \lambda \urcorner})]\!]$ should arguably be 1, where λ is the liar sentence with $[\![\lambda]\!] = \frac{1}{2}$.

[4] The truth-predication *correctly* states that the perfectly false sentence φ_0 has a truth-value that is extremely

4 Modulus Semantics for $\mathcal{L}_{\mathsf{T}}^{\infty}$

Hence, we must look for a new binary function f to underpin our semantics. In light of the discussion above, it seems reasonable to impose that the function $f : [0,1]^2 \to [0,1]$ such that $[\![\mathsf{T}_{\mathsf{x}}(\overline{\ulcorner\varphi_{\mathsf{y}}\urcorner})]\!] = f(x,y)$ should obey the following desiderata:

- f ought to be a continuous function.
- $f(1,0) = f(0,1) = 0$.
- $f(x,x) = 1$ for all $x \in [0,1]$.
- If $x \approx y$, then $f(x,y) \approx 1$.
- If $d(x,y) \approx 1$, then $f(x,y) \approx 0$.

We will denote the distance between x and y, viz. $d(x,y)$ via the usual modulus notation, i.e. $|x-y|$. The cleanest function $f : [0,1]^2 \to [0,1]$ which obeys all of these properties is:

$$f(x,y) = 1 - |x-y|$$

Under the modulus semantics for $\mathcal{L}_{\mathsf{T}}^{\infty}$, we do not have the same obstacle with respect to the truth-value of the fixed point of $\neg\mathsf{T}_1$.

Theorem 4.1 (Perfect Truth and Modulus Semantics)**.** The $\mathcal{L}_{\mathsf{T}}^{\infty}$ sentence which says about itself that it is not perfectly true can only have fuzzy equilibriums as semantic values, i.e. fixed points of the truth-function for negation.

Proof. Let λ_1 be the fixed point of $\neg\mathsf{T}_1$. Then $[\![\lambda_1]\!] = [\![\neg\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})]\!] = f_{\neg}([\![\mathsf{T}_1(\overline{\ulcorner\lambda_1\urcorner})]\!]) = f_{\neg}(1 - |1 - [\![\lambda_1]\!]|) = f_{\neg}([\![\lambda_1]\!])$. Thus, depending on the negation truth-function that one chooses, the semantic value of λ_1 will need to be a fuzzy equilibrium.

□

There are a handful of choices for the truth-functions of our usual connectives. With respect to the foregoing theorem, the choice of the negation function will directly impact the fuzzy equilibriums that can serve as the semantic value of λ_1, which should be a value in $[0,1]$ such that $f_{\neg}([\![\lambda_1]\!]) = [\![\lambda_1]\!]$.

This immediately discounts the Gödel and Product semantics for $\mathcal{L}_{\mathsf{T}}^{\infty}$, because it is impossible for λ_1 to have a semantic value in $[0,1]$ such that $[\![\lambda_1]\!] = f_{\neg}^{G}([\![\lambda_1]\!])$ or $[\![\lambda_1]\!] = f_{\neg}^{P}([\![\lambda_1]\!])$. The proof is straightforward. Both $f_{\neg}^{G}$ and $f_{\neg}^{P}$ are identical to the function g which returns 1 on argument 0 and returns 0 on any other positive argument in the unit interval. There's no value in $[0,1]$ that $[\![\lambda_1]\!]$ can have such that $g([\![\lambda_1]\!]) = [\![\lambda_1]\!]$, because if $[\![\lambda_1]\!] = 0$, then $g([\![\lambda_1]\!]) = 1$ and if $1 \geq [\![\lambda_1]\!] > 0$, then $g([\![\lambda_1]\!]) = 0$.

The only contender left amongst the canonical fuzzy systems is $\text{Ł}_{\aleph_1}$ semantics, because the function $f_{\neg}^{\text{Ł}} : [0,1] \to [0,1]$ demonstrably admits a unique fixed point, since it is a decreasing continuous function from a real interval to itself. This fixed point happens to be $\frac{1}{2}$ and our semantics for $\mathcal{L}_{\mathsf{T}}^{\infty}$ ought to be designed such that $[\![\lambda_1]\!]$ gets assigned this value.

close to 0. Thus, its overall value should be extremely close to 1, and yet it actually happens to be as far as possible from 1.

References

[Tar56] Alfred Tarski. "The Concept of Truth in Formalized Languages". In: *Logic, semantics, metamathematics*. Ed. by Alfred Tarski. Clarendon Press, 1956, pp. 152–278.

[Kri75] Saul Kripke. "Outline of a Theory of Truth". In: *Journal of Philosophy* 72.19 (1975), pp. 690–716. DOI: 10.2307/2024634.

[HPS00] Petr Hájek, Jeff Paris, and John Shepherdson. "The Liar Paradox and Fuzzy Logic". In: *Journal of Symbolic Logic* 65.1 (2000), pp. 339–346. DOI: 10.2307/2586541.

[Smi08] Nicholas J. J. Smith. *Vagueness and Degrees of Truth*. Oxford, England: Oxford University Press, 2008.