

## *Can a many-valued language functionally represent its own semantics?*

JEFFREY KETLAND

Intuitively, Tarski's Indefinability Theorem (Tarski 1935/6, Theorem I) says that no sufficiently rich language can represent its own semantics. The semantics of such a language must be described in an 'essentially richer' meta-language. Concerning the philosophical significance of Tarski's Theorem, Bell and Machover comment,

An argument like the one used in the proof of Thm. 4.7 can be applied in a wide variety of cases. ... This suggests that there cannot exist a formal language which – under some 'standard' interpretation – could adequately serve as its own metalanguage; for the syntax and semantics of a formal language cannot be adequately expressed within the language itself. In particular, the dream of certain philosophers, that some day a precise formal language will be constructed in which *all* scientific notions and theories would be expressible, is most probably unrealizable. (Bell and Machover 1974: 330–31).

The proof of Tarski's Theorem is based on the construction of a self-referential sentence, akin to the traditional Liar sentence.<sup>1</sup> Although there is still no consensus concerning the resolution of the Liar and related semantic paradoxes, one group of approaches involves various systems of Many-Valued Logic (MVL). In particular, one may construct fixed-point languages, as pioneered by Kripke 1975 and Martin & Woodruff 1975.<sup>2</sup> For such a fixed point language, where  $T$  is the truth predicate symbol, the *extension* (say,  $E$ ) of  $T$  is exactly the *set of truths* in the language *containing*  $T$ . This guarantees, for example, that where  $\phi$  is a sentence, then

<sup>1</sup> Suppose that the language contains a function symbol  $diag(\cdot)$  representing the diagonal function, where the *diagonalization* of a formula  $\phi(x)$  with just the variable  $x$  free is the result of substituting the quotation of  $\phi$  for any free occurrence of  $x$  in  $\phi(x)$ . The diagonalization of  $\phi(x)$  is thus  $\phi(\ulcorner\phi(x)\urcorner)$ . The Liar sentence  $\lambda$  is constructed by diagonalizing the formula  $\neg T(diag(x))$ . (Compare this with the Gödel formula  $G$  for an axiomatic mathematical theory  $S$ . This formula  $G$  is constructed by diagonalizing the formula  $\neg Prov_S(diag(x))$ , where  $Prov_S(x)$  is a formula representing provability in  $S$ .) So,  $\lambda$  has the explicit form  $\neg T(diag(\ulcorner\neg T(diag(x))\urcorner))$ .

<sup>2</sup> For more recent work, see Visser 1984 for the extension to 4-valued semantics, and Leitgeb 1999 for a general treatment in terms of de Morgan-valued models. For a valuable discussion of the main properties of fixed point theories of truth, see Gupta and Belnap 1993.

$\varphi \in E$  iff  $T(\ulcorner \varphi \urcorner) \in E$ . In this sense, the resulting interpreted language ‘represents its own semantics’.<sup>3</sup>

Here, I wish to argue that, in another quite reasonable sense, such languages do *not* represent their own semantics. The idea is that the semantics for a many-valued language is defined (in the informal, ‘external’, mathematical metalanguage) using a certain semantic *valuation function*, written  $\|\cdot\|$ , which assigns ‘truth degrees’ to formulae of the language. For such a many-valued language to represent its own semantics properly, I propose that the language should contain not merely a truth *predicate*, but a *function symbol*, say  $v$ , which denotes the semantic valuation function  $\|\cdot\|$ . Then, by a modification of Tarski’s Theorem, we show that this condition cannot be satisfied (modulo a certain caveat: the language must be able to discriminate the truth degrees).

MVL introduces a domain of possible ‘truth values’ or ‘truth degrees’, beyond the familiar classical truth values  $\{\top, \perp\}$ . An interpreted MV-language  $L$  can be modelled as a pair  $(L, M)$  where  $L$  is a (usually, first-order) language and  $M$  is an MV-model for  $L$ .<sup>4</sup> Let’s use the notation  $\|\varphi\|_M$  to mean the truth degree of an  $L$ -sentence  $\varphi$  in the interpreted language  $(L, M)$ . For a classical language, the semantic value  $\|\varphi\|_M$  is either  $\top$  or  $\perp$ . In particular, the MV-model  $M$  determines:

- (a) A *term valuation function*,  $(\cdot)^{M,s}: \text{Term}(L) \rightarrow D$ , which assigns to each  $L$ -term a value in the domain  $D$  (relative to some assignment  $s$ ).
- (b) A *formula valuation function*,  $\|\cdot\|_{M,s}: \text{Form}(L) \rightarrow D_\Omega$ , which assigns to each  $L$ -formula a truth degree from  $D_\Omega$  (relative to some assignment  $s$ ).

As is well known, many-valued approaches to the semantic paradoxes are prone to the Revenge Problem (see, e.g. Haack 1978: 147–48). For example, consider the strengthened Liar sentence  $\lambda$ , equivalent to ‘ $\lambda$  is not true’. On a many-valued approach, assuming a primitive truth predicate  $T$ , such a sentence  $\lambda$  may be constructed, and is assigned some non-standard truth degree  $\mathbf{d}$ , perhaps  $\mathbf{u}$  (meaning ‘neither-true-nor-false’) or  $\mathbf{b}$  (meaning ‘both-true-and-false’). In general, the Liar sentence receives some truth

<sup>3</sup> As Martin & Woodruff (1975) put it, we have represented ‘true-in- $L$ ’ in  $L$ .

<sup>4</sup> An MV-model  $M$  for  $L$  can be represented as a quadruple  $(\Omega, TF, D, \mathfrak{S})$  where  $\Omega$  is an algebraic structure of the form  $(D_\Omega, \top, \perp, \neg, \wedge, \vee)$  satisfying certain conditions (the precise conditions are not too important for what follows),  $TF$  is a filter in  $\Omega$  (also known as the set of designated values) and  $\mathfrak{S}$  is an  $L$ -interpretation (with domain  $D$ ) assigning characteristic functions to the primitive predicate and function symbols of  $L$ . See Gupta and Belnap 1993 or Leitgeb 1999 for more details about the metatheory for many-valued semantics.

degree  $\mathbf{d}$  such that  $\sim\mathbf{d} = \mathbf{d}$ .<sup>5</sup> In any case, it seems that what  $\lambda$  'says' is the case. For if the truth degree of  $\lambda$  is  $\mathbf{d}$ , then surely  $\lambda$  is *not* true simpliciter: its degree certainly isn't  $\top$ . So, the reasoning goes, since  $\lambda$  expresses a true proposition, the degree of  $\lambda$  should be  $\top$ . But this leads to contradiction. One might, and indeed I think should, conclude from this that the relevant *proposition* (that the truth degree of  $\lambda$  is not  $\top$ ) is not expressed by this sentence  $\lambda$ . And the underlying reason for  $\lambda$ 's failure to express this proposition is that the *semantic* concept 'the truth degree [semantic value] of ...' is not *definable* in the language.

The point is general. No matter how we squeeze and tug the carpet, certain *semantic concepts* which are well defined in the informal metalanguage description of the fixed-point MV-language are *inexpressible* within the very MV-language in question. This point is not original. Kripke remarked, having described his fixed-point theory of truth, that

... the necessity to ascend to a metalanguage may be one of the weaknesses of the present theory. The ghost of the Tarski hierarchy is still with us. (Kripke 1975: 74)

The concepts we *use* to describe fixed-point languages are expressed in some informal metalanguage, and there is no guarantee that they can be defined *inside* the language under discussion. And in general the Revenge Problem (i.e. the appropriate use of Tarski's Indefinability Theorem) shows that they *can't*. The metalanguage is richer than the object language. Gupta and Belnap make a similar point:

Irrespective of whether we consider classical or non-classical schemes, then, only logically incomplete schemes have the fixed point property ... It can be shown that there are related semantical notions that these [fixed-point] languages cannot contain. ... All these claims of indefinability, and others like them, can be established by means of variants of the Liar paradox. (Gupta and Belnap 1993: 95)

Here, I wish to try and analyse more carefully how Tarski's Indefinability Theorem reappears in connection with MVL approaches to the semantic paradoxes. The simple result below does not turn on details of the MVL in question. However, it does turn crucially on the bivalence of *identity claims* about the truth degrees themselves. Suppose that we treat the identity predicate = as bivalent. If  $t$  and  $u$  are terms denoting the same object,

<sup>5</sup> The fixed point condition forces  $\lambda$  to have the same truth degree as its negation. For the Liar construction entails that  $\lambda$  has the same truth degree as  $\neg T(\lambda)$ . And the fixed point condition requires that  $\lambda$  has the same truth degree as  $T(\lambda)$ . Thus,  $\lambda$  has the same value as  $\neg\lambda$ . If  $\mathbf{d}$  is the truth degree of  $\lambda$  and  $\sim$  is the semantic operation which interprets the negation symbol  $\neg$ , then  $\mathbf{d} = \sim\mathbf{d}$ . Classical bivalent semantics obviously lacks such a truth degree.

the equation  $t = u$  is true ( $\top$ ); if they denote distinct objects, then the equation is false ( $\perp$ ). In symbols,  $\|t = u\|_M = \top$  if  $t^M = u^M$ , and  $\|t = u\|_M = \perp$  if  $t^M \neq u^M$ .

For the following argument, our interpreted language must be ‘sufficiently rich’, which involves three conditions. First, it must represent its own syntax; second, there must be some mechanism of self-reference; and third, it must represent its own truth domain. We shall say first that an interpreted MV-language  $L = (L, M)$  *represents its own syntax* just in case (a) every expression of  $L$  is an element of the domain  $D$ , and (b) for each element of  $L$ , there is a term  $t$  (written  $\ulcorner \varepsilon \urcorner$ ) such that  $t^M = \varepsilon$ , and (c) the sets of variables, constants, terms, atomic formulae, formulae, etc., in  $L$  are definable in  $L$  (there exist  $L$ -formulae whose extensions in the appropriate MV sense are precisely these sets). Second, we require the presence of self-reference. We say that an interpreted language  $L$  *has self-reference* just in case for any  $L$ -formula  $\varphi$  with just  $x$  free, there is an  $L$ -term  $t$  such that  $t^M = \varphi(t)$ .

Third, a suitably rich MV-language can also talk about the *truth degrees* themselves. It is prima facie clear that fixed-point languages can be constructed with the property that the underlying structure  $\Omega$  of truth degrees exists as a *substructure* of the model  $(D, \mathfrak{S})$ . For example, the language may contain constants for each truth degree of the structure  $\Omega$ , and a relation symbol  $<$  expressing its ordering properties (or function symbols defining the operations  $\sim$ ,  $\wedge$  and  $\vee$  on the truth degrees). The reason one can do this consistently is rather trivial. Unless one *explicitly says so*, the language doesn’t ‘know’ that certain elements of its domain (i.e. the elements of the truth domain) are also being used as truth degrees. It simply treats them as elements of its domain  $D$ . So, we say that an MV-language  $L$  *represents its own truth domain* just in case  $D_\Omega \subseteq D$  and for each truth degree  $a \in D_\Omega$ , there is a unique constant  $\underline{a}$  such that  $\underline{a}^M = a$ .<sup>6</sup>

Using the bivalence property of identity, we can now show that it is impossible to define within a sufficiently rich MV-language  $L = (L, M)$  a *function symbol*  $v$  which denotes in  $L$  the ‘external’ formula valuation function  $\|\cdot\|_M$ .

**Definition:** A 1-place function symbol  $v$  in  $L$  *represents* the valuation function  $\|\cdot\|_M$  in an MV-language  $L = (L, M)$  just in case, for every  $L$ -sentence  $\varphi$ , we have

<sup>6</sup> E.g. the truth degree  $\top$  is denoted by the constant  $\underline{\top}$ . The condition that all truth degrees have names is obviously rather strong if the truth domain is the compact interval  $[0, 1]$ , as used in fuzzy logic. But we do not appeal to any such strength below. Also, we do not require that the operations on truth degree are represented. That is, we do not demand that there exist in  $L$  function symbols  $\underline{\sim}$ ,  $\underline{\wedge}$ ,  $\underline{\vee}$  which represent the corresponding truth-functional operations on the truth degrees.

296 JEFFREY KETLAND

$$(*) \quad v(\ulcorner \varphi \urcorner)^M = \|\varphi\|_M.$$

If such a function symbol  $v$  exists, we say that  $L$  *represents its own formula valuation function*.

This constraint is strong. For with such a function symbol, one can *define* a specific semantic concept for each truth degree  $a$  that is named in the language: put  $S_a(x)$  for  $v(x) = a$ . So,  $T(x)$  is just  $v(x) = \top$  and  $F(x)$  is just  $v(x) = \perp$ . Furthermore, as long as the truth degrees are distinguishable (identity is bivalent), no sufficiently rich MV-language can represent its own formula valuation function:

**Theorem:** Let  $L$  be a MV-language with truth degree  $\top$ , and which represents its own syntax, has self-reference and in which identity is bivalent. Suppose also that  $L$  represents its own truth domain. Then  $L$  does not represent its own formula valuation function  $\|\cdot\|$ .

**Proof:** This is modification of Tarski's Theorem. For a *reductio*, assume that  $L$  represents its own formula valuation function. That is, there is a function symbol  $v$  which represents  $\|\cdot\|$ . By self-reference, we construct a formula  $\lambda$  of the form  $(v(t) \neq \top)$ , where the term  $t$  denotes  $\lambda$ . In particular,  $\|\lambda\| = \|v(\ulcorner \lambda \urcorner)\| \neq \top$ . First, suppose that  $\|\lambda\| = \top$ . Then,  $\|v(\ulcorner \lambda \urcorner)\| \neq \top = \top$ . So,  $v(\ulcorner \lambda \urcorner)^M \neq \top$ . So, using (\*),  $\|\lambda\| \neq \top$ . Contradiction. Instead, suppose that  $\|\lambda\| \neq \top$ . By (\*),  $v(\ulcorner \lambda \urcorner)^M \neq \top$ . Hence, because the truth domain is represented,  $\|v(\ulcorner \lambda \urcorner)\| \neq \top = \top$ . Hence,  $\|\lambda\| = \top$ . Contradiction. **QED.**

Of course, we *can* coherently define the semantics of many valued languages. Working in our informal mathematical metalanguage, we may coherently define an interpreted MV-language  $L$ , and define the formula valuation function  $\|\cdot\|$ , which assigns a truth degree to each formula. Also, because of the various fixed point theorems, we are also able to construct MV-languages which do represent their own semantics in the sense that the extension of some *predicate* (the truth predicate) coincides with the sentences assigned the value  $\top$  by  $\|\cdot\|$ . Many-valued logic is interesting, valuable and so on. But many-valued logic doesn't resolve the Liar paradox. In our informal metalanguage, we assume that we are able to reason concerning the possible truth degrees. In particular, the truth degrees are (ontologically) distinct. For example, our description assumes that the degree  $\top$  is different from the degrees  $\mathbf{u}$ ,  $\perp$ , etc. If  $L$  is such a language, satisfying conditions of 'sufficient richness', and can properly discriminate amongst its own truth degrees, then the semantic notion 'the truth degree of  $\varphi$  in  $L$ ' is *not* definable *inside*  $L$ .<sup>7</sup>

<sup>7</sup> Hannes Leitgeb (personal communication) has pointed out to me that if the bivalence condition on the identity predicate is relaxed, then fixed-point languages *can* be constructed again. One needs to treat certain identities amongst the truth degrees as not having a standard truth degree. For example, consider a 3-valued semantics with truth degrees  $\{\top, \perp, \mathbf{u}\}$  with the strong Kleene truth tables. Suppose that an identity

University of Leeds  
Leeds LS2 9JT, UK  
j.j.ketland@leeds.ac.uk

### References

- Bell, J. L. and M. Machover. 1974. *A Course in Mathematical Logic*. Amsterdam: North-Holland.
- Gupta, A. and N. Belnap. 1993. *The Revision Theory of Truth*. Cambridge, MA: MIT Press.
- Haack, S. 1978. *Philosophy of Logics*. Cambridge: Cambridge University Press.
- Kripke, S. 1975. Outline of a theory of truth. *Journal of Philosophy* 72: 690–716.
- Leitgeb, H. 1999. Truth and the Liar in de Morgan-valued models. *Notre Dame Journal of Formal Logic* 40: 496–514.
- Martin, R. M. and P. Woodruff. 1975. On representing ‘true-in- $L$ ’ in  $L$ . *Philosophia* 5: 217–21.
- Tarski, A. 1935/6. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica* I: 261–405. English translation by J. H. Woodger, ‘The Concept of Truth in Formalized Languages’ in A. Tarski 1956: *Logic, Semantics and Metamathematics: Papers by Alfred Tarski 1923–1938*. Oxford: Clarendon Press.
- Visser, A. 1984. Four-valued semantics and the liar. *Journal of Philosophical Logic* 13: 181–212.

---

$t = u$  is assigned the value  $\mathbf{u}$  (‘undefined’) when either  $t^M$  or  $u^M$  is  $\mathbf{u}$  and assigned either  $\top$  or  $\perp$ , as the case may be, when the terms denote objects distinct from  $\mathbf{u}$ . Then the resulting Kripke jump operator  $J$  (which in our case maps the function  $f$  which the function symbol  $v$  denotes to a new function,  $J(f)$ ) is monotonic, and it follows that there are fixed points.

## Why the Converse Consequence Condition cannot be accepted

LUCA MORETTI

Three general confirmation principles discussed by Hempel (1965) are the following:

*Converse Consequence Condition* (CCC) If an observation statement  $E$  confirms a hypothesis  $H$  and if another hypothesis  $H^*$  entails  $H$ , then  $E$  confirms  $H^*$ .

*Special Consequence Condition* (SCC) If an observation statement  $E$  confirms a hypothesis  $H$ , then  $E$  confirms any of  $H$ ’s logical consequences.