Meaning, Self-Reference & Verifiability

Luc Batty

Abstract

The Verifiability Criterion of Meaning judges a sentence to be meaningful if and only if it is either verifiable empirically or is a logical truth or falsehood. This paper points out an issue with the verificationist project of employing the criterion to distinguish between meaningful sentences and nonsense. In particular, it is argued that the criterion is only successful if it can partition the set of all sentences into the meaningless and the meaningful, which is an impossible task. As such, the criterion is rejected as an inadequate arbiter of meaningfulness.

Introduction

What does it mean for a sentence to be meaningful? The doctrine of verificationism, championed in the first half of the twentieth century by the logical positivists, holds that for a sentence to be judged as meaningful it must bear some verifiable connection to the real world accessed either through sensory perception or logic. More precisely, verificationism proposes that a sentence is meaningful if and only if it is empirically verifiable or is a matter of logical necessity.

The aim of this short paper is to show that the project of separating the meaningful from the meaningless, where meaningfulness is defined according to this verifiability criterion, is futile, because the definition does not permit

a clear enough separation. Specifically, the set of all sentences of a language cannot be partitioned into the meaningful and the meaningless under these circumstances, simply because the criterion allows that at least one sentence be included in both cells, or neither, which is impossible by definition.

In what follows, (§1) I will briefly introduce the relevant concepts, before (§2) deriving the failure of the verificationist program from four primary assumptions: first, that there exists a set S of all sentences; second, that the property of meaningfulness is bivalent and has an excluded middle; third, that a sentence is meaningful if and only if it is empirically verifiable, or is a logical truth or falsehood (the verifiability criterion); and fourth, that, given the verifiability criterion, there is at least one well-formed sentence ψ in S that cannot be consistently judged as either meaningful or meaningless. This failure comes in the form of a contradiction: it will be shown that the sentence ψ, given the assumptions, is meaningful if and only if it is meaningless. The next section (§3) will discuss the consequences of this contradiction, and I will argue that the only deniable premise is the third. I will conclude (§4) that the verifiability criterion must be abandoned if the verificationist wishes to pursue the project of demarcating fact from nonsense.

I. Meaningfulness and Verifiability

It would not, I suspect, strike anyone as remarkable if a group of persons, well enough acquainted with the English language, agreed that a sentence such as 'the goose flies south' is meaningful, whereas 'flies goose south the' is not. Presumably, this is because acquaintance with language is accompanied by familiarity with the rules governing its grammar, and such familiarity informs judgments of meaningfulness. This presumption, however, is incomplete; the rules of sentence construction, as are known (at least to some degree) by the users of a language, are not enough to provide an infallible method of detecting meaningfulness. While a well-formed sentence may more obviously present itself as meaningful than an apparently random or chaotic string of words, such as 'apple apple apple jazz', it is not the well-forming of that sen-

tence in itself that makes this so. Without prior understanding of the meaning of the terms used, and of which predicates may be applied to which objects, a sentence may appear to be well-formed grammatically, and may thereby mistakenly judged to be meaningful. 'The fantasy of unsaturated knowledge is indistinguishable from the actualisation of the real' is a well-formed sentence, but it is meaningless; it is nonsense, as is 'the Nothing is prior to the Not and the Negation.'[1] The difference between these two sentences, however, is that the former is a randomly generated string of words,[2] and the latter is a quote from Heidegger's *Was ist Metaphysik?*

The separation of the meaningless from the meaningful was, and remains, a significant philosophical project. Historically, this project was a criticism of metaphysics, and in particular a way to identify and eradicate meaningless sentences from otherwise fruitful fields of inquiry. A sentence that is false, such as 'the speed of light through a vacuum is 30 kilometres per hour' may still be meaningful, whereas a sentence devoid of meaning is neither true nor false, but simply nonsensical. The metaphysician, according to the logical positivists, was guilty not merely of advocating a falsehood, but of producing as results sentences that 'fail to conform to the conditions under which alone a sentence can be literally significant.'[3] Consequently, A. J. Ayer proposed a criterion by which a sentence may be judged as meaningful or meaningless:[4]

> **Verifiability Criterion of Meaning (VCM):** For any sentence χ, χ is meaningful if and only if it is (i) verifiable empirically or (ii) is a logical truth or falsehood.

For the purposes of this paper, the varying forms of the criterion will be

---

[1] As quoted by Rudolf Carnap in 'The Elimination of Metaphysics Through Logical Analysis of Language' 69.

[2] The sentence was randomly generated using LSE's Dr Bryan Roberts' *Nonsense Generator*, available at http://personal.lse.ac.uk/ROBERT49/teaching/ph201/lectures/lecture2.html>.

[3] Ayer, 'The Elimination of Metaphysics' 15.

[4] Ibid.

largely subsumed under a single umbrella of *weak*, *in principle* verification-ism,[5] which holds that a sentence is verifiable empirically if it is at least theoretically possible for an empirical test to increase the probability that it is true (or false, as the case may be).

While the VCM arguably succeeds in labelling most of metaphysics as nonsense, it is not immune to criticism. On the one hand, it may be said to discard too much as nonsense, for instance sentences of the form 'all *X*s are *Y*s' that employ a universal quantifier, or certain sentences about the past or future, which may not be verifiable even in principle. On the other hand, the criterion may admit too much as meaningful. If 'the Nothing is prior to the Not and the Negation' is a target for VCM, then I suspect 'the Nothing is prior to the Not and the Negation, or 2+2=4' is also a target, however the latter sentence cannot be dismissed as meaningless according to the VCM, since it is a logical truth.

These shortcomings, and responses thereto, have been extensively covered throughout the literature,[6] and no more will be said about them here. The purpose of this paper is not to argue that some sentence χ has been unjustly deemed meaningless or meaningful, and therefore that VCM is an inadequate criterion for judging the meaningfulness of a sentence. Rather, the argument presented herein against this form of the VCM, and indeed against verificationism in general, rests on the fact that the project of neatly separating the meaningful from the meaningless, under this definition, is simply impossible.

II. Meaningful Self-Reference

The VCM is designed to apply specifically to sentences. If we begin with the finite set of words *W*, and take its closure under the rule for constructing well-formed sentences (the correct statement of which is not required here, so long as we include the condition that a well-formed sentence cannot contain infinitely many words), we are left with the countably infinite set of sentences

---

[5] For more on the distinctions see Ayer, 'The Elimination of Metaphysics' 16-23.

[6] The curious reader should consult Creath, 'Logical Empiricism'.

S. The goal of the VCM, therefore, can be said to be the partitioning of *S* into a pair of disjoint, non-empty, and exhaustive subsets of *S*: the set *M* of meaningful sentences, and the set *N* of meaningless sentences. Let us assume the following two principles: (1) that there exists a countably infinite set of well-formed sentences *S*, and (2) that the property of meaningfulness, as defined above, is bivalent (in the sense that any given sentence χ is either meaningful or meaningless), and has an excluded middle (χ cannot be both meaningful and meaningless). The first principle, I suspect, is uncontroversial. The second, on the other hand, may need further justification.

Given the VCM, and a sentence χ, we may assert that χ is meaningful by virtue of it being either verifiable empirically, or a logical truth or falsehood. If χ fails to satisfy both of these conditions, it is meaningless. For χ, there are four possible cases: (i) χ is not a logical truth or falsehood, but it is verifiable; (ii) χ is a logical truth or falsehood, but is not verifiable; (iii) χ is both verifiable and a logical truth/falsehood; or (iv) χ is neither verifiable, nor is it a logical truth or falsehood. In the first two cases, χ is a meaningful sentence. In the fourth case, χ is meaningless. The third case is perhaps more subtle, and whether or not there is such a sentence that satisfies both conditions is open to discussion, however if there is such a sentence, it must be meaningful, since it satisfies either one of the preceding two conditions. In any of these exhaustive cases, χ is either meaningful or meaningless, and so meaningfulness can be said to be bivalent, and from this it follows that a sentence being both meaningful and meaningless would lead to contradiction.[7]

Let us now consider the following sentence:

ψ: 'This sentence is meaningless if and only if it is self-referential'

The sentence ψ may remind the reader of the infamous paradoxes of self-reference. In fact, the argument that follows is very much in the same spirit. Can we allow ψ to be a member of the set of all well-formed sentences? In

---

[7] If χ is both meaningful and meaningless, then it is both unverifiable and logically contingent; while also being either verifiable, a logical truth/falsehood, or both, contradicting the supposition of bivalence.

order to answer yes to this question, we must first ask whether self-referential sentences are well-formed grammatically. There is, I submit, no good reason to suppose otherwise. The trouble with self-referential sentences (in extreme cases) is that they may not have a clear truth-value, but the lack of a truth-value does not preclude the assignment of a meaningfulness-value. A: 'This sentence contains eight letters' is false, but it is not meaningless, since its falsity can be verified. B: 'The number of letters in this sentence is divisible by six' is both meaningful and true. Other self-referential sentences may well be devoid of meaning; C: 'This sentence sweats melancholy unconstitutionally' is meaningless, as is D: 'This sentence is false', the infamous Liar sentence.[8] That ψ is self-referential, therefore, is not relevant to the question of whether it belongs to the set *S*, and since ψ is well-formed I see no reason to deny that it should belong. It might be noted, however, that a sentence such as ψ is different from A, B, C, and D, in that it refers to its own meaningfulness, rather than its own falsity, or any of its other properties. As such, one could raise the objection that a sentence referring to its own meaningfulness cannot be assigned a meaningfulness value.[9] This would be a mistake; the sentence 'This sentence is meaningless' is meaningful, and simply false.[10]

---

[8] That the Liar sentence cannot be assigned a single truth-value is not universally accepted. Dialetheism, for instance, maintains that there are true contradictions (see Priest, *In Contradiction*) however this does not affect the possibility of assigning it a single meaningfulness-value.

[9] Such an objection would recall Tarski's approach to the Liar Paradox (see Tarski, 'Der Wahrheitsbegriff in den formalisierten Sprachen' 261–405)

[10] There is no contradiction or paradox here: the sentence 'This sentence is meaningless' is meaningful because it is *false*. The sentence (given our assumptions) must either be meaningful or meaningless. Suppose on the one hand that it is meaningless, in which case it is true, and therefore meaningful, contradicting the supposition that it is meaningless. It must therefore be meaningful, and since it is not verifiable, it must either be a logical truth or falsity. Again, if it is a logical truth, it is true that it is meaningless, and so it is meaningless: contradiction. The only remaining option is that it is a logical falsehood, in which case it is false that it is meaningless, and so it is meaningful (without contradiction).

The fact that ψ refers to its own meaningfulness should therefore not be taken as a presupposition of its lack of a meaningfulness value by virtue of it being structured similarly to the Liar sentence.

With all the pieces in place, I assume the following premises:[11]

1. There exists a countably infinite set of well-formed sentences S.

2. The property of meaningfulness is bivalent and has an excluded middle.

3. For any well-formed sentence χ in S, χ is meaningful if and only if it is empirically verifiable, or is a logical truth or falsehood (i.e., the VCM).

4. ψ: 'This sentence is meaningless if and only if it is self-referential', is a member of S.

From the two starting assumptions (1), (2), it follows that, given two sentences in S, α and β, there exists a relation R over S such that $\alpha R \beta$ if and only if α has the same meaningfulness-value as β (i.e., if α and β are both meaningful or are both meaningless). Furthermore, it follows from the assumptions that R is an equivalence relation, and by the Equivalence-Partition Correspondence Theorem,[12] that the set S is thus partitioned into two disjoint, non-empty, and exhaustive subsets of S; M and N, as it were. Given 1, 2, and 3; we may now add:

---

[11] Strictly speaking I am assuming the axioms of Zermelo-Fraenkel set theory, which have been omitted from this exposition for the sake of brevity.

[12] A relation R is an equivalence relation over a set S if and only if it is reflexive (for any member x of S, x is R-related to itself), symmetric (for any two members x and y of S, if x is R-related to y, then y is R-related to x), and transitive (for any three members x, y, and z of S, if x is R-related to y, and y is R-related to z, then x is R-related to z). The Equivalence-Partition Correspondence Theorem states that any equivalence relation over a set partitions that set into equivalence classes. For example, the relation 'has the same eye-colour as' is an equivalence relation over a set P of people, and partitions P into equivalence classes, such as the set of people with green eyes, the set of people with brown eyes, and so on.

5. For any well-formed sentence χ in S, χ is a member of exactly one of either M or N.

Where M is the set of meaningful sentences, and N is the set of meaningless sentences. From 4 and 5 follows 6:

6. ψ is a member of exactly one of either M or N.

Is ψ a member of M or is it a member of N? It cannot be a member of both, it must be a member of one, and since S is partitioned into M and N, ψ is a member of one if and only if it is not a member of the other. We consider two cases:

In the first case, ψ is in M. From 5, it follows that ψ is a meaningful sentence, and from 3, that it is either verifiable empirically or a logical truth or falsehood. Now, ψ cannot be verified empirically, so it is either a logical truth or a logical falsehood. Is the sentence a logical falsehood? No, since it does not contradict itself: meaninglessness does not necessitate self-reference, or vice-versa; as we have seen, the two are independent of one another. The only remaining possibility is that it is a logical truth, or, equivalently, that its negation, $\sim\psi$: 'Either, this sentence is meaningless and not self-referential, or, this sentence is meaningful and self-referential,' contradicts itself. Again we consider two cases and find that the two disjuncts are contradictory: the first since it denies its own self-reference, and the second because it denies its own meaninglessness.[13] The sentence ψ, then, is *true*. As such, since it is undeniably self-referential and asserts its own meaninglessness, it must be meaningless, and thus belongs to the set N.

In the second case, let us suppose ψ is a member of N, and so is a meaningless sentence. However, since the sentence is self-referential, and correctly refers to itself as meaningless, it is *true* that it is meaningless if and only if it is self-referential, and so the sentence is in fact meaningful, and is therefore a

---

[13] 'This sentence is meaningful and self-referential' is meaningless (according to the VCM), since it is not verifiable empirically, nor is it a logical truth or falsehood, contradicting the reference it makes to its own meaningfulness.

member of the set *M* of meaningful sentences. We then derive the following contradiction:

7. ψ is a member of *M* if and only if it is not a member of *M*.

III. Contradiction

What are we to make of this? Let us follow standard protocol and consider the premises one by one in order to identify the source of the contradiction. While I have already argued for premises 1, 2, and 4, there are a few further points pertaining to them that are worth mentioning.

Premise 2, I propose, is indisputable. In addition to the arguments for the bivalence and excluded middle of meaningfulness put forward in the preceding section, it seems clear that the project of distinguishing the meaningful from the meaningless is not of any use if it is possible for a sentence be both. What is more, the very notion of meaningfulness ought to be (arguably, at least as it used in a general sense for the purposes of this type of distinction) more clearly demarcated than even that of truth. In other words, if it is at all possible for sentences to be neither true nor false (or both at once), then at the very least we should hope that such sentences are classifiable as meaningless.[14] Furthermore, permitting a 'gap' in between the possible two meaningfulness values is counterproductive to the project of separation of meaningful propositions from nonsensical ones: what status would we to grant to those sentences that are neither meaningless nor meaningful? What would it mean for a sentence to be meaningful? Giving up this principle would mean giving up any hope of coherently separating fact from nonsense, subsequently banishing these questions to the realm of the unanswerable.

Fortunately, we need not worry at present about giving a precise definition for meaningfulness. The concern of this paper is only the inadequacy of the VCM of providing such a definition. With this in mind, let us return to our sentence ψ. As we have seen, 'This sentence is meaningless if and only if it is

---

[14] Of course, much more can be said about this. The interested reader should consult Beall and Glanzberg, 'Liar Paradox'.

self-referential' is, according to the VCM, both meaningful and meaningless, a blatant contradiction in the face of the principle assumptions. Since premise 2 will not be abandoned, we turn to premise 1, that there exists a set of all well-formed sentences.

Unlike in the case of Russell's Paradox, this contradiction is not due to the positing of a set that is 'too large' to exist consistently. The set S of well-formed sentences, I maintain, exists as the closure of the set of words (in this case, English words) under the generator of well-formed sentences.[15] As such, its existence should not pose a problem even for constructivist philosophers of mathematics who might object on the grounds that a mathematical object can only be proven to exist if it can be constructed. Moreover, the existence of S is difficult to deny given that its elements can be enumerated: begin with the complete list of words, add the first well-formed sentence after one minute, the second after half a minute, the third after a quarter of a minute... and so on. Neither the question of the existence of S nor of what its constituent members are is in any way problematic. A refusal to accept this is a refusal to accept the principles of set theory.

With this said, I anticipate a response that questions our fourth premise. Perhaps the fact that ψ is a member of S is the problem. But if we are to deny ψ's membership, we must do so without appealing to the fact that it references its own meaninglessness as a justification, since, as we have seen, 'This sentence is meaningless' (hereafter: 'φ') is not problematic in this respect. S, we recall, is the set of all well-formed sentences. If ψ is not a member of S, then the grammatical conditions that must be satisfied in order for a sentence to be well-formed include some condition that ψ does not satisfy, but φ does. While it is the burden of the verificationist to specify what this condition is, I will consider what I see as the most plausible candi-

---

[15] Under this definition, single-word strings such as 'The' count as well-formed sentences. Perhaps some readers may find it more palatable to begin with the set $V \subset W$ of verbs, and follow the same procedure. In any case, single-word strings such as 'The' are clearly meaningless, so whether we begin with W or V will not make much difference to our results

date. It could be argued that ψ and φ are different because of the presence and content of the biconditional that occurs in the former and does not in the latter. Of course, the biconditional cannot, in itself, be the problem, unless we are to relegate any sentence involving a biconditional to the same class as 'apple apple apple jazz', so the trouble must be with its content, and in particular the fact that it ascribes the property of being meaningless-if-and-only-if-self-referential. Can we then suppose that, among the other necessary conditions for well-forming, a sentence cannot contain a biconditional of this type? No: consider the sentence, ω: 'φ is meaningless if and only if it is self-referential'. The sentences ω and ψ are structurally alike in the relevant way, even though the referent of 'this sentence' is different in each of them, since they both ascribe the same property to some self-referential sentence (that it is meaningless if and only if it is self-referential). The sentence ω, however, is false, and therefore meaningful, since φ, as we have seen, is meaningful despite its self-reference. If we propose that a sentence must also satisfy the above condition in order to qualify as well-formed, then we must admit that there are sentences such as ω that are meaningful, even though they are not well-formed. Now, those who wish to hold on to the VCM can argue that if ω is not well-formed under this definition, and the VCM is not meant to apply only to well-formed sentences, then ψ is no longer a problem, since it can be denied to be a member of *S*, and therefore a member of neither *M* nor *N*. However, in order to pursue this argument further, one must accept that there are meaningful, non well-formed sentences, sentences of the very type that the VCM was created to banish as nonsense. This is both implausible and counterproductive. Moreover, this ad hoc manoeuvre is not enough to rule out the possibility that there are other members of *S* that, unlike ψ, do not contain such a biconditional, but are equally impossible to place in the partition.

What we have learned, then, is that the verificationist conception of meaningfulness cannot cope with certain cases of self-reference, without discarding too many sentences as nonsense. For those of us who are not committed to the VCM and therefore have no need to add a dubious condition that must be

satisfied by well-formed sentences, these considerations are enough to close the questions of premises 1, 2, and 4. We conclude that the VCM must be abandoned.

Conclusion

Whether or not this all means that the project of demarcating fact from nonsense is futile in a more general sense remains an open question. The aim of this paper was to show that given the verifiability criterion of meaning, the set of all well-formed sentences cannot be partitioned into the cell containing all and only the meaningful sentences, and the cell containing all and only the meaningless ones, because there exists at least one sentence of the type, 'This sentence is meaningless if and only if it is self-referential', that cannot be included in one cell without being included in the other, or cannot be included in either. Since the contradiction follows from the four principle assumptions, and only the VCM is expendable, I submit the above argument in support of the claim that the verifiability criterion of meaning is inadequate as an arbiter of meaningfulness.

References

[1] Ayer, A. J. 'The Elimination of Metaphysics', in *Language, Truth and Logic*, London: Penguin Books, 1936.

[2] Beall, J. C. and Glanzberg, Michael. 'Liar Paradox', in Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), http://plato.stanford.edu/archives/fall2014/entries/liar-paradox/.

[3] Carnap, R. 'The Elimination of Metaphysics Through Logical Analysis of Language', trans. Pap, A. in Ayer, A. J. (ed.), *Logical Positivism*, New York: The Free Press, 1932: 60-81.

[4] Creath, R. 'Logical Empiricism', in Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). http://plato.stanford.edu/archives/spr2014/entries/logical-empiricism/.

[5] Priest, G. *In Contradiction* (second ed.), Oxford: Oxford University Press, 2006.

[6] Roberts, B. *Nonsense Generator*, available at http://personal.lse.ac.uk/ROBERT49/teaching/ph201/lectures/lecture2.html.

[7] Tarski, A. 'Der Wahrheitsbegriff in den formalisierten Sprachen'. Trans. Woodger, J. H., 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics* (second ed.), Indianapolis: Hackett, 1983.

Luc Batty is an MSc student in the Philosophy of Science at the London School of Economics and Political Science (2014-2015). He holds a BA in Philosophy from the University of Warwick (2010-2013). His main interests are the Philosophy of Logic, Mathematics, and Language. You can contact him at [luc.batty@gmail.com]