Is a Verification Machine Really a Problem for the Verifiability Principle? A Vindication of Lycan's Scepticism

Gavin Lee

In *The Concept of Meaninglessness,* Edward Erwin raises a novel objection against the verificationist program. William G. Lycan revisits Erwin's objection in *Philosophy of Language*. While he is critical of Erwin's claim of a decisive victory, he is unable to account for where Erwin's objection goes wrong. In this paper, I will critically examine Lycan's version of Erwin's objection, and using Lycan's framework, establish two ways in which the objection fails. I will start with a discussion of the verifiability principle, direct confirmation, and indirect confirmation. I will then present Lycan's version of the objection and demonstrate two ways in which it fails. Lastly, I will analyse how Lycan's presentation of the objection differs from Erwin's and explain why the differences are important.

The target of Erwin's objection is the verificationist program. The verificationist theory of meaning hinges on the verifiability principle, which states that a non-analytic statement is meaningful if and only if, in principle, it can be shown to be true or false through experience. A statement's meaning is its method of verification; we come to know the meaning of a statement when we know which conditions make it true and which conditions make it false.[1] For a verificationist, if we cannot conceive of an empirical method of verifying the truth or falsity of a statement, then that statement is meaningless.[2] Erwin distinguishes between the need for direct verification and indirect verification. Direct verification would be confirmation provided by immediate experience. Indirect verification is less stringent, allowing the use of induction and

[1] Moritz Schlick, *Positivism and Realism*, 86.
[2] *Ibid.,* 87.

probabilistic confirmation.[3] As Erwin points out, the verificationist theory is not tenable if it is taken as strictly demanding direct confirmation.[4] Therefore, indirect confirmation must be deemed as sufficient to provide meaning.[5] With this in mind, Erwin's objection takes aim at indirect verifiability.

Lycan's interpretation of the verifiability principle differs from Erwin's interpretation. In addition to the basic framework Erwin established, Lycan extends the principle by adding that a statement is meaningful "if and only if its being true would make some difference to the course of our future experience."[6] This requirement comes from Moritz Schlick, who argued that a statement is meaningless if its truth or falsity does not affect our experience.[7] For example, imagine I posit the existence of *entity-Z.* If somebody were to ask me, "In what empirically observable way does *entity*-Z's existence or non-existence make a difference in our world?" and my response was, "None whatsoever," then my statement about *entity*-Z is meaningless. With direct confirmation, it appears that this requirement is not an additional requirement at all; it is a restatement of the verifiability principle's primary claim, that a statement must be confirmable through experience. If we are able to verify a statement through direct observation, then what we are verifying simply *is* the difference in our world that the statement is making.[8]

When considering indirect confirmation, however, the situation is less clear. In the case of directly unobservable or undetectable entities, such as those posited by physics, we cannot always observe the difference they make in the world. When one of our physical theories posits a

[3] By probabilistic confirmation, I mean *degree* of confirmation, or probability of certainty.

[4] If indirect confirmation is not permitted by the verificationist principle, then the majority of our scientific statements would be evaluated as meaningless. Since any knowledge that we obtained using induction would be meaningless, we would only be left with the formal sciences.

[5] See Edward Erwin, *The Concept of Meaninglessness,* 34–5. Erwin provides examples of how direct confirmation is too demanding. Verificationist thinkers like Rudolf Carnap and A. J. Ayer readily ceded this point.

[6] Lycan, 98.

[7] Shlick, 88–9.

[8] This can be represented as a bi-conditional, where if we have verifiability, we have a verifiable difference in the world, and vice versa.

hypothetical entity, for which it cannot provide any account of *what* it does, simply that it is *predicted* or *required* by the theory, in what way can we even indirectly satisfy the requirement of it affecting our sensible experience? It may, in principle, be possible, but its 'in principle' seems to differ from the 'in principle' type of possibility that is raised by a question like, "Is there a mountain on Mars that is taller than Mt. Everest?" In principle, if the hypothetical entity did not exist, would that affect our future experience in any way? It could be argued that its non-existence would be counted as disconfirming evidence against our belief in the physical theory that predicted its existence.

Lycan and Schlick disagree about whether this type of effect on our experience satisfies the requirement. Does a change in our beliefs regarding a theory count as a change in our future experience? Under Schlick's account, the answer would be no, since he explicitly states that it must make a verifiable difference in the *sensible* world.[9] Schlick, then, is referring to direct confirmation in his statement.[10] A change in our understanding of a physical theory would not have an impact on what we observe, and that is what is required for Schlick. The change would only occur to our internal beliefs.[11]

Lycan broadens Schlick's conception to admit indirect confirmation when he says that a statement merely has to "make some kind of difference to thought and to action."[12] This widening allows for changes in internal beliefs to satisfy the requirement. For Lycan, the answer would presumably be yes, then, since there would be a verifiable difference in how we *think* about the affected physical theory. For Schlick, the requirements of verifiability and difference in future experience always

[9] *Ibid.*
[10] There is one sense in which an indirect confirmation could depend on a sensibly different world. A sensibly different world could affect a direct confirmation that informed a particular indirect confirmation. However, this indirect confirmation's reliance on the sensibly different world is derivative and would not be sufficient for Schlick.
[11] An objection could feasibly be raised here on the account of theory-ladenness. While it is clear that the theories we believe can affect our observations, they only affect them insofar as how we consider them. They do not affect what we actually observe in sensation, which is what Schlick cares about.
[12] Lycan, 99.

run together, but for Lycan, they can come apart.[13] Since Lycan's objection aims at indirect confirmation, we can safely set aside Schlick's conception.[14]

In his response to Erwin, Lycan presents his version of the objection through a thought experiment. Imagine we stumble upon a mysterious machine. Whenever we code a punch card with a declarative statement and submit it to the machine, it does some processing and then lights up as either TRUE or FALSE. We independently investigate the declarative statements that we submit to the machine, and we find that the machine's answers are always correct. Consider, then, that we coded an arbitrary statement *S* onto a punch card, and submitted it to the machine. Even if *S* is gibberish, if the machine returned TRUE, then we would have an empirical experience that would act as indirect confirmation for statement *S.* Since *S* could be any statement, we can confirm any statement, and the verifiability principle is trivial.[15]

With the context sufficiently sketched out, I will now move on to evaluating Lycan's version of the objection. According to Lycan, the only extant criticism of Erwin's objection comes from Brian R. Clack, who is suspicious of its "science-fiction character."[16] I hope to establish a more substantive attack on the soundness of Lycan's version. I will present two avenues of attack: An argument regarding the machine's logical consistency, and an argument about whether its assertions actually qualify as meaningful under the verifiability principle.

The first way that Lycan could attack his version of the objection is to focus on the fact that the machine is a *marvellous predictor* and is always right. In order to have the machine's assertions trusted as

[13] This would be possible if we had a statement that we could indirectly confirm or disconfirm without its truth or falsity making an appreciable difference to our future experience. In this case the bi-conditional would fail, as verifiability would not entail a difference.

[14] As the machine's judgments are examples of indirect confirmation (they are trusted because of an inductive inference), it follows that the objection is attacking indirect confirmation.

[15] William G. Lycan, *Philosophy of Language,* 104. If all statements can be confirmed, then all statements are meaningful. If all statements are meaningful, and the verifiability principle is supposed to demarcate meaningful statements from meaningless ones, then it ceases to do any work.

[16] Brian R. Clack, *Religious Belief and the Disregard of Reality*, 270.

confirming evidence, the people in the thought experiment would need to first establish with indirect confirmation that it is, indeed, always right.[17] Something like the following would suffice:

1. Provide the machine with declarative statements, which they believe to be true, and have it respond TRUE.
2. Provide the machine with declarative statements, which they believe to be false, and have it respond FALSE.
3. Provide the machine with declarative statements, of whose truth or falsity they are uncertain, and have it respond TRUE or FALSE, and then after further investigation, confirm its findings.

The people in the thought experiment may also provide the machine with statements they believe to be true or false, have it respond to the contrary, and upon further investigation, they would confirm its findings. The machine does all of this perfectly, with no mistakes. This brings us to the example Lycan uses in the thought experiment, *ST4*:

4. Provide the machine with a statement of gibberish, which they believe to be meaningless, and have it respond TRUE.

They might then ask: why is the machine evaluating this statement as TRUE? The only reason the thought experiment provides is that the machine is *always right.* We can use this fact to deduce the following:

I. If the machine is always right, it must evaluate all types of statements accurately, whether they take the form of 1, 2, 3 or 4.
II. Since we know through *ST1-3* that the machine's evaluations accurately reflect our empirical investigations, we can conclude that its evaluations cannot be arbitrary, and that they must reflect the world.[18]
III. Therefore, the machine evaluates statements like *ST4* the same way it does for statements like *ST1-3,* by returning TRUE or FALSE.
IV. Therefore, when the machine returns TRUE or FALSE to

[17] This is similar to a famous mathematician having to prove several important theorems before we would regard her authorship in itself as confirming evidence.
[18] In *ST1-3* we see that the results of our empirical investigations in the world perfectly match the assertions made by the machine. If our assertions track the world, and the machine's track ours, then it follows via transitivity that the machine's assertions are tracking the world.

> statements like *ST4*, those gibberish statements must actually reflect the world in the same way that statements like *ST1-3* do.[19]

With this in mind, let us imagine that we continue to feed the machine statements like *ST4*. We continuously feed the machine with an indefinite succession of unique statements of gibberish. Each time the machine processes a new statement, if it returns TRUE or FALSE, then the statement must have been about something in the world, even if we cannot understand how.[20] If what we submit ever happens to be incomprehensible or does not pertain to the world, then the machine would not be able to respond with TRUE or FALSE. This makes intuitive sense, for how could the machine consistently evaluate statements like *ST4* as TRUE or FALSE if they were not connected to the world at all? In this case, the machine would respond with something like NULL. For the objection to succeed, it requires every arbitrary statement to be meaningful, which leads to the trivialization of the verificationist principle. Since the statements evaluated by the machine as NULL would not be meaningful, if the NULL response ever occurs, then the objection to the verificationist principle fails.

The only way in which the objection can be saved from this attack, then, is if every conceivable statement is both comprehensible and about something in the world.[21] Even if we grant that our knowledge is limited, and that some statements could be comprehensible and be about something in the world in a way we do not currently understand, it seems inconceivable that every possible statement we do not understand is like this. This intuition can be confirmed by the following. The set of 'conceivable statements,' *CON*, that could be used in statements of the

---

[19] Restating this negatively may make this clearer: if statements like *ST4* were not about the world, that is, if they were unlike in kind to statements like *ST1-3*, then the machine could not return TRUE or FALSE in response to them, as it does for *ST1-3*, as neither evaluation would be accurate. And the machine *must* be accurate.

[20] This is not a problem for the verifiability principle, as something's being verifiable *or* falsifiable is sufficient for meaning.

[21] If every statement can be meaningful, then for meaningfulness to be non-trivial (which the verificationists need to be the case) every statement must be about something in the world.

kind *ST4* is really the set of *all* possible statements.[22] However, each statement in the set of all possible statements is not a coherent statement about the world. Therefore, the set of 'coherent statements that can be made about the world,' *COH,* is a proper subset of *CON*.[23] Therefore, there must be statements in *CON* that are not in *COH,* and when one of these statements is submitted to the machine in the form of *ST4,* the machine would not return TRUE or FALSE, and the objection fails.

There are a few counterpoints to address here. First, someone could claim that the machine is a perfect predictor for statements like *ST1-3*, but not for statements like *ST4*. This is inconsistent with the foundational premise of the thought experiment, as it is stipulated that the machine is always correct. Second, and more interestingly, someone could claim that up until t_1, the point in which we first submit a statement like *ST4* to the machine, the machine is always right, and that after that point in time it becomes inconsistent. The people in the thought experiment would be able to establish the consistency of the machine pre-t_1 while avoiding the possibly of potential NULL responses post-t_1.[24] The problem here is that they would still continue to provide statements like *ST1-3* to the machine and they would still follow up on them from time to time. They would discover that the machine was returning incorrect results to statements like *ST1-3*, and then would come to no longer view its answers as providing confirmation. Once this occurs, the objection fails. This relates back to Erwin's point about renowned mathematicians.[25] If a famous mathematician published a long string of flawed work, we would no longer view the fact that it was her making an assertion as evidence for that assertion being true.

The second way Lycan could attack the objection stems from his own widening of the verifiability principle. Let us grant that it is possible

[22] Since the verificationists are not metaphysical realists about statements, statements are the kind of thing that can only be created by a mind. In order to create a statement, a mind would first need to conceive of it.

[23] B is a proper subset of A if and only if B is a subset of A and there exists at least one element of A which is not in B. In this case, *COH* □ *CON.*

[24] In this case, the machine could respond TRUE or FALSE to statements that would normally prompt a NULL response. Since the machine does not have to be always correct after t_1, wrong answers are tolerated.

[25] Erwin, 36.

for the machine to always be right and return TRUE or FALSE for any arbitrary *ST4*-like statement. Let us call such an arbitrary statement *S.* These TRUE or FALSE evaluations allow for any *S* to have an empirical verification condition, that is, the people in the thought experiment will see the machine respond TRUE or FALSE to their submitted *S*, and that acts as indirect confirmation for *S*. This satisfies the first statement of the verifiability principle, which simply states that the statement must be empirically verifiable.

For Lycan's second requirement, however, things are less clear. Does the truth or falsity of any *S* change the future experience of any of the people in the thought experiment? Does it make some kind of difference to their thoughts and actions? It does not appear that it would. The truth or falsity of any *S* does not influence anyone's beliefs about any theory, the world, or even the beliefs they hold about the machine itself. They cannot independently investigate any arbitrary *S* like they can for statements like *ST1-3.* Therefore, the recursive mechanism that enables the independent corroboration of statements like *ST1-3* to affect the people's beliefs about the machine's reliability is not available. The TRUE or FALSE evaluation of any *S* does not tell the people anything about the world which could impact their future experience. Lycan describes the difference requirement as a question: What will happen depending on whether the statement is true or false?[26] In this case, posed to any arbitrary *S,* the answer would be nothing.

One could challenge that an arbitrary statement *S*'s truth or falsity does make a difference, and that the difference simply is the machine responding with TRUE or FALSE.[27] This raises a question of priority. When Lycan says a statement's truth or falsity must make some difference to the course of future experience, I am reading 'future' to mean *after* a statement's truth or falsity is determined. This must be the correct reading, because to assert that the difference in *future* experience simply *is* the response, we would have to presuppose that its truth or falsity is already known to recipients before they receive it from the machine. This is not the case. Once they receive the TRUE or FALSE response, their future

[26] Lycan, 104.
[27] To be clear, the difference simply *is* receiving the response from the machine that *S* is TRUE or FALSE.

experience begins, and it continues without difference.[28] Even if we did grant that the response of TRUE or FALSE itself is a difference in future experience, then it would render Lycan's second requirement vacuous – there could be no true or false statement that did not satisfy it, since its very evaluation would qualify.

Someone could further object, "Well, surely the recipients will think about S's truth or falsity in some way after receiving the response from the machine, even ever so briefly, to discard the answer as useless, and that very thought satisfies Lycan's requirement." This falls short because Lycan stipulates that the truth or falsity must make a difference both for thought and for action, and just discarding S makes no difference for a person's future action.[29] This points to a bigger problem. This challenge takes Lycan to mean literally any thought when he says 'thought'. While Lycan does not make it explicitly clear what exactly he means by thought, a reasonable account should presume a more nuanced conception than the loosest one possible.

It is interesting that Erwin's version of the objection differs from Lycan's version. Erwin's version of the objection is presented in the form of a thought experiment as well. He asks us to imagine that there exists a computer that knows all the information there is to know. The nature of the computer is such that it only makes true assertions. Suppose we start to receive assertions made by the machine, and every time we independently attempt to confirm them, we find them to be true, and we can repeat the process indefinitely. After enough time has passed, if the machine were to make any assertion *A,* and we did not know whether *A* was true, false, or meaningless, the mere fact that the machine was asserting *A* would be sufficient to indirectly confirm *A*. As *A* could be any statement, it follows that for any conceivable statement made by the

[28] There is one sense in which this is not true. In the future, a statement of the kind, "At Time *X*, the machine read TRUE about statement *S*" will become true. However, as I explain, this type of difference is insufficient as it would render Lycan's second requirement vacuous.

[29] There is room for here for an objection. Instead of merely mentally discarding the result, the person could instead crumple up and throw away the printed judgment while exclaiming "This is useless!" Similar to the discussion that follows regarding the looseness of the term 'thought,' I think it is justified to take Lycan to be meaning some more robust sense of 'action' than what is exemplified by this example.

machine, we would have a method of verifying its truth. Therefore, any conceivable statement made by the machine would be meaningful. If *any* conceivable assertion can be meaningful, then the verifiability principle is trivial.[30] Erwin argues that if his objection is sound, then it proves fatal to the verifiability principle.

In isolation, the thought experiments differ in at least one important way. In Erwin's version, we do not actively interact with the machine. That is, we only receive assertions from the machine, and we cannot have it judge any assertion that we provide to it. Lycan's presentation of the thought experiment allows us to be active participants, with us submitting assertions for the machine to consider. My first attack on Lycan's version of the objection heavily relies on *ST1-3* and *ST4* in order to deduce I-IV. Since we cannot actively submit statements to the machine in Erwin's version, this same approach may not work.[31] For my second attack, Erwin's statement of the verifiability principle does not include anything about a requirement for a difference to be made in future experience. While further investigation would be required in order to determine whether Erwin's version is susceptible to my attacks, it seems likely that it may not be.

If my two criticisms of the objection are successful, and I believe that I have demonstrated that they are, then Lycan's scepticism is well-placed. A careful analysis of the machine's nature shows that the claimed results of the thought experiment would not occur. Further, even granting that the claimed results would occur, they would not satisfy Lycan's requirements for any arbitrary *S* being a meaningful statement. The result is that Lycan's presentation of the objection does not threaten the verificationist principle, which renders verificationism safe from this particular thought experiment. However, Lycan's scepticism is only well-

[30] *Ibid.,* 36–8. Erwin points to real-world examples in mathematics and physics in order to demonstrate how we use this type of induction. Mathematical and physical assertions made by prize-winning mathematicians and physicists are lent some measure of confirmation simply in virtue of the past success their proposers have had in their respective fields.

[31] The machine may never output statements like *ST4*, or it might only do so very sporadically. Without being able to systematically study the machine by submitting statements to it, we may not be justified in treating its random outputs as indirect confirmation.

placed as applied to his own version of the objection. Since it is not clear whether my attacks would be successful against Erwin's version of the objection, the verificationist principle may still require additional defense.

Works Cited

Brian R. Clack, "Religious Belief and the Disregard of Reality" in Joseph Carlisle, James Carter, and Daniel Whistler, eds., *Moral Powers, Fragile Beliefs* (New York: Continuum International Publishing Group, 2011).

Edward Erwin, *The Concept of Meaninglessness* (Baltimore: The John Hopkins Press, 1970).

Moritz Schlick, "Positivism and Realism" in A. J. Ayer, ed., *Logical Positivism* (New York: The Free Press, 1959).

William G. Lycan, "Verificationism" *in Philosophy of Language: a Contemporary Introduction* (New York: Routledge, 2000).