Consider the following argument:

(P1) If I know I'm reading, then I know I'm not a brain in a vat.

(P2) I don't know I'm not a brain in a vat.

(C) Therefore, I don't know I'm reading.

Do you know whether you are currently reading?

In this essay, I will argue that I do not know whether I am currently reading. Specifically, I will refute the argument that Nozick would have made against (C) by showing that Nozick has made a mistake with the counterfactual case. I will argue that the possibility that we are a brain in a vat (BIV) makes it impossible to evaluate the truth value of counterfactuals. First, I will explain Nozick's definition of knowledge, which I will accept for the sake of argument. Then, I will explain why Nozick would, from his theory of knowledge, accept (P2) but not accept (C) (even though he would accept (P1) as well). Finally, I will explain the mistake that Nozick made with the 'possible worlds' account of subjunctives and why the BIV scenario makes it impossible for us to evaluate the truth value of counterfactuals.

Nozick's theory of knowledge

Nozick's (slightly simplified) theory of knowledge is the following.
S knows p if:

(1) p is true
(2) S believes that p
(3) If p was not true, S would not believe that p [not-p → not-(S believes that p)]
(4) If p is true, S believes that p [p → S believes that p]

Here, I will focus mainly on condition (3), the counterfactual, as it is the most important for Nozick's argument against (C). Nozick argues that this condition is necessary because of, among others, the following scenario. If we suppose that Henry is riding through special countryside where all the barns are made out of papier-mâché, which are indistinguishable from real barns. Suppose now that Henry randomly looks out of his window and sees a barn, which happens to be the only actually real barn there, and comes to believe that it is a barn. If knowledge is just a justified true belief, we should recognize Henry's belief as knowledge. However, to call this knowledge would be ridiculous: if Henry had looked out of his window at any other time, he would have seen a fake barn and he would have still come to believe that he saw a barn. This is why condition (3) is necessary. Henry only knows that he sees a barn if, when he would see a fake barn, he would not believe that it is a real barn. In this scenario, this is not the case and, therefore, we should not count Henry's belief as knowledge.[1]

[1] (Nozick, 1981, pp. 172–175)

Met opmerkingen [OB1]: Good essay, except for the fact that my objection to Nozick is a bit stupid. Nozick rejects KK, so it does not matter that we don't know that.

Met opmerkingen [MOU2]: Good focus on one theory!

Met opmerkingen [MOU3]: He does not really accept P1

Met opmerkingen [MOU4]: 4 is also a counterfactual

Met opmerkingen [MOU5]: No if

Met opmerkingen [MOU6]: Good tot use this

Met opmerkingen [MOU7]: Condition 3 delivers the intuitive result... and talk about possible world. It is not the requirement that he has a fake-barn detector, there should not be a close possible world in which there is a fake barn but he thinks there is.

Examining the premises

I will, for the sake of argument, accept (P1). This, I think, is reasonable given what the argument is trying to achieve. It tries to prove that I cannot be certain that I am reading because I cannot be certain that I am not a brain in a vat (BIV). Therefore, it implicitly defines 'reading' as an activity which requires an external object (such as a book or an article) which is not simulated and requires me to read *that* external object. Consequently, the implicit definition of 'reading' requires me not to be a BIV. As a result, (P1) should be accepted.

Met opmerkingen [MOU8]: Idem

Nozick would accept (P2). This is because of condition (3). For me to know that I am not a BIV, I need to know what I would know if I was a BIV. If I am a BIV (the counterfactual), I would not know that I am. Everything could be simulated in such a way that it would be exactly the same as the external world. Therefore, condition (3) is not satisfied. I do not know whether I am not a BIV because I would not be certain that I was a BIV if I was a BIV. Therefore, Nozick's theory of knowledge would accept (P2).

Met opmerkingen [MOU9]: Wrong sentence.

Nozick's approach to counterfactuals

At first sight, it would seem that it is a simple modus ponens: (C) follows logically from (P1) and (P2). However, Nozick would argue that I do know whether I am reading. He argues that condition (3) for knowledge would in this case be satisfied. If I am not reading, he argues, I would know that I am not reading. If I was, for instance, watching a video instead of reading, I would clearly know that I was not reading.

Met opmerkingen [MOU10]: You also need to show that it satisfies condition 4

Then what about the possibility that I am a BIV? It seems like Nozick does not seem to account for that possibility, only for more practical reasons why I would know that I am not reading (e.g. watching a video). This seems wrong, if I was a BIV and was, therefore, not reading, I still would not know whether I was reading or not because I could be in a simulation which made me experience the act of reading. This is because he uses a 'possible worlds' account of subjunctives.[2] To illustrate this view, consider the following statement: 'if kangaroos had no tails, they would topple over'. This statement does not consider *all* the possible worlds in which kangaroos have no tails (surely, there must exist one in which the kangaroos would not topple over). It only considers the possible world which is closest to ours as much as kangaroos having no tails permits it to.[3] Therefore, one should not evaluate the truth value of counterfactuals by considering all possible worlds, but by only considering the possible worlds which are the *closest* to our world.

Met opmerkingen [MOU11]: The closest world, or the set of worlds which are closest. It is ambiguous.

[2] (Nozick, 1981, pp. 173–174)
[3] (Lewis, 1973, p. 1)

Hence, Nozick would argue that in order to evaluate the statement 'if I was not reading, I would know I was not reading' we would only need to examine the worlds which are the most similar to our worlds. An example of one of those worlds could then be one in which I was watching a video. Thus, Nozick would conclude, if I was not reading, I would know that I was not reading so (C) is not true.

The problem with Nozick's application of the 'possible worlds' account of counterfactuals

I argue that there is a problem with the way Nozick applies the 'possible worlds' idea. It is impossible to know which world is the *closest* to our world. This is because we do not know whether or not we are a BIV (Nozick agrees with this by accepting (P2)). For instance, if we assume that we are a BIV, then the closest world in which we are not reading is still a world in which we are a BIV. And as we are a BIV, we do not know whether we are actually reading or not (it might be all simulated). Only if we assume that we are not a BIV, then the closest world in which we are not reading might be one in which we are, for example, watching a movie. Therefore, in arguing that the closest world is one in which we might be watching a movie, Nozick actually assumes that we are not a BIV. Since Nozick himself agrees that we do not know whether we are a BIV, he is making an implicit logical error here.

Since we do not know whether we are a BIV or not, we also do not know what the closest world to our world is. And since we do not know what the closest world is in which we are not reading, we do not know whether we are reading or not. Therefore, (C) is, in fact, true.

The scepticism about being a BIV or not has a further consequence: we can no longer evaluate any counterfactuals through the 'possible world' method. Take for instance the kangaroo example. If we assume that we are a BIV, it is possible that in the closest possible world, real, external kangaroos will actually not topple over, only the ones that are being simulated would topple over (the simulation does not have to be identical to the external world). Thus, as we do not know whether we are a BIV or not, we do not know which possible world is the closest to ours. As a result, a sceptic who accepts (P2) should also accept the logical conclusion which follows from it:

(C2) I do not know which world is the closest possible world.

Therefore, as I do not know which world is the closest possible world, it is not possible for me to evaluate a counterfactual (at least through the 'possible worlds' method).

In conclusion, I have showed that a sceptic (someone who accepts (P2)) cannot accept any knowledge (with knowledge defined in the way that Nozick defines it), including the knowledge that I am reading. This is because Nozick's definition of knowledge requires

Met opmerkingen [MOU12]: There is another good challenge here: it is very hard to say which worlds are closest to us.

Met opmerkingen [MOU13]: An externalist response: Nozick could object and say that he only proves that we know something, not that we know that we know something. Therefore, the burden of proof is on me to argue that we really do not know whether we have hands, because (as Moore argues) we can just simply argue that we do know that we have hands, just that we cannot prove it. However, I could argue that we should suspend judgement to whether we are reading or not.

one to be able to evaluate the counterfactual. And it is impossible to evaluate the counterfactual through the ‘possible worlds’ method.

Olivier Berg

Bibliography

Lewis, D. (1973) *Counterfactuals*. Cambridge: Harvard University Press.

Nozick, R. (1981) ‘Knowledge and Skepticism’, in *Philosophical Explanations*. Oxford: Oxford UP, pp. 167–185.