Review of: "Free will and the paradox of predictability"

Ronald Meester[1]

1 Free University Amsterdam

Potential competing interests: No potential competing interests to declare.

The concept of the paradox of predictability (PoP) is as follows. Suppose a demon predicts the future of the universe, and somewhere in the universe there is a device, referred to as a “frustrator” or “counterpredictive device, which is such that it acts counter to any prediction of its behavior that is revealed to it. Suppose then that the demon is somehow forced to reveal his prediction to the device; the device then acts counter to the demon’s prediction, and hence the prediction is proved wrong. But how could this be if the demon knew precisely the initial state of the universe and the laws, and these determine the future states of the universe, including those of the frustrator?

Alexander Syrakos (hereafter “the author”) has written an interesting and readable account on this paradox, and he convincingly shows that there is not a paradox, really. Instead of trying to summarize his arguments, I instead will explain why I find it surprising that the paradox seems to attract so much attention of philosophers to begin with.

The whole problem seems to arise because of some form of self-reference. However, such self-reference routinely results into paradoxes. Popular forms of such self-references include the barber of the village who shaves everybody who does not shave himself, or the sentence “This sentence is not true”. Not everything that is expressible in works makes sense logically, of course. Deeper forms of self-reference arise when ons speaks about the “set whose members consist of all sets”. Such a set-theoretical notion jeopardizes the whole enterprise of set-theory based system of axioms of mathematics, and the only way out seems to be to explicitly forbid such self-references. In addition, as a mathematician and researcher in probability theory, I have come across many examples of stochastic processes that were easy to describe in words, but principally impossible to formally construct in a mathematical way. The author also gives an example in the form of a computer program.

So, what to think of a prediction, say expressed by me, which refers to myself? The author gives the following example:

“So, suppose I had the appropriate equipment (a molecular scanner and a powerful computer equipped with molecular dynamics simulation software) and I used it to predict, via such a simulation of my whole body, that after two minutes I will get up from my chair, go to the fridge, open it and get something to eat. Having acquired this knowledge, I then deliberately decide to instead stay seated at my desk for a whole hour and watch YouTube videos on my laptop, despite my hunger. On first glance, there doesn’t seem to be anything inconsistent with this scenario.”

I am not sure that I would call this a problem of inconsistency, but of course the “deliberate” decision to stay seated at the desk, should have been part of the earlier simulation. So if there is an inconsistency, then it must be the fact that one can overrule an earlier prediction/simulation, thereby making that very simulation completely useless. Indeed, the very

overruling should be part of the simulation. Similar remarks apply to other examples in the paper. For instance, the example with the light bulb is not convincing, since if a demon has full knowledge, it also has knowledge about the wiring of the device.

It is clear to me that the author is fully aware of such observations, and I agree with him that the notion of free will is not really at stake in this discussion.

He analyses the PoP in various contexts, distinguishing between the demon being internal or external. To be honest, I think it is clear that the two situations (internal versus external) are nor really different, and if I understand the author, then this indeed is the conclusion he draws.

So, I think that the PoP is easily seen to be a more or less “ordinary” paradox, very similar to the many paradoxes that we routinely have around self-references. Spending so many words on it seems perhaps a bit overkill to me, although it might certainly be found useful for many readers through the explicit and very clear exposition. In addition, the paragraph in which it is explained that self-prediction of any future state of itself by a computer, automatically implies that the full infinite future is revealed, is an interesting example of another, related, problem around self-prediction. If the author makes anything clear, it is that self-reference is always problematic, and that any conclusion around the concept of free will based on it, must be mistaken and fallacious.

Apart from this, the manuscript contains a number of important observations about the nature and philosophy of simulations and mathematical models. I understand that this is not the main focus of the paper, but as it happens, this is the part that I found perhaps more important than refuting an alleged paradox. Simulations and predictions require a mind to interpret them, since the actual physics inside the processor of the computer has in itself absolutely nothing to do with the actual physical system it tries to simulate or predict. The link between simulation and reality is entirely mental. Clearly this has vast philosophical consequences for the interpretation of models and their simulations.