

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

Jan-Boje Frauen¹

¹ Xiamen University

Potential competing interests: No potential competing interests to declare.

This is a very fine article, which I believe is publishable in its present form. I will go through the text and comment on its separate parts, occasionally highlighting points that remained a bit unclear to me. Revisions should be considered optional, rather than mandatory, with one exception: I would like the author to take a close look at my discussion of 5.1. To my mind, the author's line of argument is correct here but what the author derives from it is not.

ABSTRACT

The abstract clearly states the background, agenda and thesis of the article. Clearly nothing to improve here.

INTRODUCTION

I liked the Spinoza quote that leads into the topic. The alleged "paradox" is introduced and exemplified in a very clear and succinct fashion, not hiding the fact that this is a purely philosophical argument due to our current (and possibly future) technological and thus epistemic limitations.

I find it a bit peculiar that implications from quantum mechanics are brushed aside quickly in the introduction by writing that one can attempt to defy probabilities in the same way as one can attempt to frustrate certain outcome predictions: while in the second case there would actually be a proof against determinism, the first case can merely serve as an indicator that determinism may be wrong, to my mind. However, this is beside the point the article is trying to make and the article is already on the long side, so I understand why the author does not go into great detail.

2.

The second chapter successfully lays out the argument that *understanding* computation is always a mental factum of the mind, which has no existence in the physical world. The author therefore calls it "metaphysical," which I would agree with. After all, epiphenomenalists never tire of highlighting that consciousness is an insubstantial surplus to the physical world and that our actions would be performed by "zombies" or "computers" in the same way. It should be noted here perhaps that Roger Penrose's "orchestrated objective reduction" theory of consciousness makes precisely this point, also quoting (Gödel's second incompleteness theorem and) the Turin halting problem. There is, of course, a huge debate about this claim, but I do understand that there are always length issues to be taken into consideration when one writes an article.

3.

The author highlights our practical limitations due to ignorance of initial states etc. and cites meteorological as an

example. Looking out of the window and seeing that it is raining again, I certainly understand the author's point (my weather forecast confidentially stated that it would be sunny all day). However, the author again highlights that this is beside the point in a thought experiment that makes the assumption that in theory perfect predictability would be possible because of the fully determined nature of the physical realm. Perhaps the author dwells a bit too much on the point only to dismiss it here.

The author's summary of determinism, to my mind, is well-written and well-advised to keep the article open to an interdisciplinary audience.

I wonder if the author would want to illustrate his definition of "indeterminism" by an example from classical philosophy: Kantian freedom happens when the moral law overrules physical determinism without the agent being able to choose between the two. In performing the "free" act, an agent is ironically more strongly determined than in performing a "determined" act. Freedom is thus the independence from the causal law of the physical realm, but does not necessarily entail freedom of choice (as the author also highlights by pointing to quantum indeterminism).

4.

Being a very visual thinker, I really appreciated the graphics. However, I wonder if the author might not have meant to write that in case (a) domino 5 will remain standing and fall in case (b) in the captions? I get that the point is that you change the setting so that the prediction becomes false, of course, but I found the captions a bit confusing put this way.

I completely understand what I take to be the author's main points here: the argument is tautological in that the demon's free will is supposed in the thought experiment's initial conditions as an "external interference" with the closed determined system. Being circular, it therefore does not prove anything really.

Secondly, since what is asked of the demon as an "external interference" is impossible, there is no free act. It is like asking someone to prove her freedom by drawing a round square. If the proof is logically (as opposed to merely practically) impossible, there is no freedom either. And no paradox.

5.

Now the author sets out to show that it is not possible to reload the paradox by making the "demon" part of the closed, determined system.

5.1

If the demon is supposed to be endowed with free will to begin with, the author claims, it will have the choice between two equally unsatisfactory outcomes (meaning its behavior will become unpredictable). I wonder if the author should not highlight here that the assumption of free will within the closed system leads to a *reductio ad absurdum* which in fact disproves this assumption instead: that the demon cannot logically do what it is determined to do does not falsify determinism in a closed system. If the contradiction arises from tautologically assuming its free will to begin with, it falsifies free will in a closed determined system instead (as one would indeed expect)!

5.2

I wonder if the author does not assign a bit too much space to possible ways to computational solutions of the problem here only to conclude that, as expected, there are none.

6.

The author first introduces the thought here that self-prediction is impossible as every self-prediction changes the conditions upon which the prediction is based. So one has to predict again and runs into a regress. The point that the observer (“representation”) is not included in what is observed, making prediction a subjective enterprise that is nowhere to be found in a physical system, is repeated. Once again, there is a fierce debate about what consequences this actually has (the famous Turing test, for instance, claims that it has none). However, I understand the author’s desire to bring a long paper to a close without unnecessary distractions.

The point that there is a limit to what a computer can process and that this limit makes prediction impossible when the computer itself needs to be included in the simulation is well taken. This really is a logical argument, which makes it practically impossible to realize the thought experiment in the first place.

7.

I am not too sure if the author employs a correct definition of “compatibilism” here. To my mind, compatibilism means that I am free to do what I want, even though what I want may be determined. Do compatibilists actually believe that determination can emerge freedom in its traditional interpretation as independence from causal determination, as the author seems to suggest (paragraph 3)?

I very much enjoyed the discussion of causes vs. reasons (which is similar to my own research about the difference between the “why?” of theoretical reason that asks for causes and is hence reducible to “how?” and the “why?” of practical reason that is concerned with “wherefore?” a course of action is selected and hence irreducible to theoretical knowledge of causal connections[1]) and the discussion of subjective self-prediction and how, as an emergent phenomenon, it might be different to a computational machine predicting its future actions.

Overall, I think this is a very well-executed article that is both interesting and readable. I wish the author best of luck with getting it published!

[1] Frauen, JB. Survival, freedom, urge and the absolute: on an antinomy in the subject. *Int J Philos Relig* 91, 63–85 (2022). <https://doi.org/10.1007/s11153-021-09812-z>.

