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

Julio Gea-Banacloche[1]

1 University of Arkansas - Fayetteville

Potential competing interests: No potential competing interests to declare.

It is hard to believe that “professional” philosophers are still seriously considering the “paradox of predictability” as relevant in any way to the question of free will, but I know for a fact that at least some of them do, since I remember running into it, years ago, in the book by Ismael that is this paper’s Ref. 13. The rather clear and thorough analysis presented here may therefore be helpful to some people who might otherwise be tempted to take the PoP seriously as well.

As other reviewers have pointed out, the paper is very long, and this may discourage some potential readers. If the author wants to trim it down, I would suggest removing most of the detailed examples of how the paradox would work with a (deterministic) counterpredictive device, as the basic nature of the difficulty can be easily grasped from a single instance.

I basically agree with all the points made by the author, and differ only on matters of emphasis. As an example, his presentation of the PoP with counterpredictive device appears (at least, at first) to emphasize the physical impossibility of the task, when the real difficulty is purely logical: the “task” assigned to the machine is simply an embodiment of the liar’s paradox (“everything I say, including this, is false”). The author eventually does, correctly, point this out, placing the difficulty squarely in the domain of *meaning*: as with other logical paradoxes, the problem is that it is sometimes impossible to assign meaning (in this case, what would be meant by a “true prediction”) to some formally correct, typically self-referential statements or instructions. This is good, but it might have been better to start there, or arrive there with fewer distractions along the way.

The author also correctly notices (although I don’t think he mentions this explicitly until the conclusions section) that the problem is “subtly different” when the possible “frustrator” is actually a human being, as opposed to a pre-programmed device. I, however, would not have called the difference subtle: I think the problem is, in fact, *completely* different, because now it is not a logical paradox, but a genuine research question, since we cannot *in principle* be sure of whether the prediction will fail or not. If all our decisions are determined by low-level (e.g., molecular) processes, assumed deterministic and computable, it is conceivable that we would have no actual choice but to do what the predictor tells us, even if it feels “against our will.” Such an outcome is certainly counterintuitive, but not *logically* impossible. Hence, the way out of the difficulty this time is not by proving a logical contradiction, but a true *physical* impossibility (inasmuch as the storing and processing of information is a physical process), namely, the impossibility of self-prediction. That this is a nontrivial result should be apparent from the fact that it took Popper a couple of rather dense papers to argue the point back in 1950. (I notice that one of the reviewers mentions it as if it were a well-known information-theoretic result, although unfortunately they do not provide a reference.) The author does present the argument clearly enough, in modern

computer simulation terms, but I would definitely have liked to see the difference between the two versions of the "paradox," and their resolutions, emphasized much more strongly.

Finally, the author leaves the issue of human free will vs. determinism open (although he states his own opinion on the matter), which I guess is fair, since the paper is strictly about the PoP and the correct conclusion (nontrivial, to the extent it depends on the impossibility of self-prediction) is that the matter is undecidable on these terms. I do, however, sympathize with the reviewer who would like to see at least a mention of the fact that the physical world is, as best as we can ascertain, very much non-deterministic. (Disclosure: I am a physicist who works with quantum mechanics constantly…)