Studies in History and Philosophy of Modern Physics 57 (2017) 98–103

Contents lists available at ScienceDirect

Studies in History and Philosophy of Modern Physics

journal homepage: www.elsevier.com/locate/shpsb

On the notion of free will in the Free Will Theorem ☆

Klaas Landsman

Institute for Mathematics, Astrophysics, and Particle Physics, Faculty of Science, Radboud University, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 17 October 2015
Accepted 2 November 2016
Available online 24 November 2016

Keywords:
Free Will Theorem
Local miracle compatibilism

ABSTRACT

The (Strong) Free Will Theorem (FWT) of Conway and Kochen (2009) on the one hand follows from uncontroversial parts of modern physics and elementary mathematical and logical reasoning, but on the other hand seems predicated on an undefined notion of free will (allowing physicists to "freely choose" the settings of their experiments). This makes the theorem philosophically vulnerable, especially if it is construed as a proof of indeterminism or even of libertarian free will (as Conway & Kochen suggest).

However, Cator and Landsman (*Foundations of Physics* 44, 781–791, 2014) previously gave a reformulation of the FWT that does not presuppose indeterminism, but rather assumes a mathematically specific form of such "free choices" even in a deterministic world (based on a non-probabilistic independence assumption). In the present paper, which is a philosophical sequel to the one just mentioned, I argue that the concept of free will used in the latter version of the FWT is essentially the one proposed by Lewis (1981), also known as 'local miracle compatibilism' (of which I give a mathematical interpretation that might be of some independent interest also beyond its application to the FWT). As such, the (reformulated) FWT in my view challenges compatibilist free will à la Lewis (albeit in a contrived way via bipartite EPR-type experiments), falling short of supporting libertarian free will.

1. The Free Will Theorem

The Free Will Theorem (FWT) of Conway and Kochen (2006, 2009) shows that some small and uncontroversial corner of quantum mechanics (i.e., the response of massive particles with spin one to measurements of their spin) combined with a rather weak locality condition suggested by Einstein's theory of special relativity (which effectively forbids superluminal signaling), is incompatible with the conjunction of determinism and the ability of experimentalists to "freely choose" the directions along which they measure spin. The FWT was published in two versions, of which the second, called the *Strong* Free Will Theorem by Conway and Kochen, has superseded the first (which may therefore be discarded). Conway and Kochen (2009, p. 226) paraphrase their theorem in the following way:

> 'if indeed we humans have free will, then elementary particles already have their own small share of this valuable commodity. More precisely, if the experimenter can freely choose the directions in which to orient his apparatus in a certain measurement, then the particle's response (to be pedantic—the universe's response near the particle) is not determined by the entire previous history of the universe. (…) our theorem asserts that if experimenters have a certain freedom, then particles have exactly the same kind of freedom. Indeed, it is natural to suppose that this latter freedom is the ultimate explanation of our own. (…) Granted our three axioms [i.e., the physical ones and freedom of choice], the Free Will Theorem shows that nature itself is nondeterministic.'

It is clear from Conway's recent biography (Roberts, 2015) that the authors saw their FWT as a major contribution to science (perhaps even to philosophy), and indeed it has generated considerable publicity. Part of this interest has been rather critical (cf. Bassi & Ghirardi, 2007; Cator & Landsman, 2014; Goldstein, Tausk, Tumulka, & Zanghi, 2010; Hemmick & Shakur, 2012; Hermens, 2014, 2016; 't Hooft, 2007; Wüthrich, 2011), mainly on the following grounds:

1. Lack of novelty compared with the famous paper by Bell (1964), whose assumptions and conclusions are at least quite similar to

☆Dedicated to Henk Barendregt, on the occasion of his official retirement (October 1, 2015).

E-mail address: landsman@math.ru.nl

http://dx.doi.org/10.1016/j.shpsb.2016.11.001
1355-2198/

those of the FWT, although the underlying proofs are mathematically distinct from those in the FWT.
2. Lack of novelty even within its own terms: almost identical results, even based on very similar mathematical reasoning, had previously been published by Heywood and Redhead (1983), Stairs (1983), Brown and Svetlichny (1990), and Clifton (1993).
3. Circularity, in that indeterminism is presupposed (namely in the assumption that 'experimenters have a certain freedom') instead of derived.

I only discuss these criticisms here in so far as they justify my own contribution; my take is that all of the above criticism is deserved, but that nonetheless the FWT is an interesting result, which triggers further discussion (of which the present paper is an instance).

1. The difference between earlier literature (of which, incidentally, Conway and Kochen only cite Heywood and Redhead) and the FWT is almost exclusively one of emphasis, namely on free will. Given this emphasis, it is striking that one looks in vain for serious philosophical analysis in Conway and Kochen (2006, 2009). All one finds is:

> 'The tension between human free will and physical determinism has a long history. Long ago, Lucretius made his otherwise deterministic particles swerve unpredictably to allow for free will. It was largely the great success of deterministic classical physics that led to the adoption of determinism by so many philosophers and scientists, particularly those in fields remote from current physics. (This remark also applies to "compatibilism", a now unnecessary attempt to allow for human free will in a deterministic world.)' (Conway & Kochen, 2009, p. 230).

Also elsewhere, one finds little respect for the philosophical debate on free will, e.g.,

> 'Compatibilism is an old viewpoint from previous centuries when philosophers were talking about free will. The were accustomed to physical theory being deterministic. And then there's the question: How can we have free will in this deterministic universe? Well, they sat and thought for ages and ages and ages and read books on philosophy and God knows what and they came up with compatibilism, which was a tremendous wrenching effect to reconcile 2 things which seemed incompatible. And they said they were compatible after all. (…) But in my view is it's really nonsense. (Conway, quoted in Roberts, 2015, p. 361).

Thus the main goal of the present paper is to relate the FWT (whatever its novelty compared to its predecessors) to the philosophy of free will. However, the (negative) relationship we are going to establish will be with compatibilist free will à la Lewis, as opposed to the (positive) relationship with libertarian free will envisaged (but not analyzed in any detail) by Conway and Kochen; the floor remains open for the latter.
2. Regarding the earlier work of Bell, Conway and Kochen (2006) acknowledged that:

> 'Our result is by no means the first in this direction. It makes use of the notorious quantum mechanical entanglement brought to light by Einstein, Podolsky, and Rosen, which has also been used in various forms by J.S. Bell, Kochen and Specker, and others to produce no-go theorems that dispose of the most plausible hidden variable theories. Our theorem seems to be the strongest and most precise result of this type.'

The precise relationship between the FWT and Bell's Theorem was analyzed in detail in Cator and Landsman (2014), with the following conclusion:

(a) Both Bell (1964) and the original version of the FWT in Conway and Kochen (2009) use an informal way of talking about free settings, granting which both establish a contradiction between determinism, locality (in the sense of Bell, which in the presence of determinism reduces to what is called parameter independence), and quantum mechanics. The difference lies in three facts:
   i. Bell relies on probability theory (whereas the FWT does not).
   ii. The (optical) corner of quantum mechanics used in Bell's Theorem may be replaced by the corresponding experimental results, whereas the FWT uses uncontroversial yet untested predictions about massive spin-1 particles.
   iii. The FWT must assume perfect (EPR) correlations, which are difficult to realize and hence are avoided by later versions of Bell's Theorem (i.e. through the well-known CHSH inequalities rather than the original Bell inequalities).

(b) The same three differences persist also in the new versions of both Bell's Theorem and the FWT proposed by Cator and Landsman (2014), in which the experimentalists' "freedom" of choosing settings is defined rigorously (in a probabilistic and a deterministic framework, respectively).

3. Conway and Kochen (2006) themselves already anticipated the criticism of circularity on the very first page of their first paper:

> "'I saw you put the fish in!" said a simpleton to an angler who had used a minnow to catch a bass. Our reply to an analogous objection would be that we use only a minuscule amount of human free will to deduce free will not only of the particles inside ourselves, but all over the universe.'

This did not stop Wüthrich (2011) from concluding that:

> 'Their case against determinism thus has all the virtues of theft over honest toil. It is truly indeterminism in, indeterminism out.'

Both are right: the FWT is far from circular, but its conclusion would be much more transparent if Wüthrich's charge could be dispelled. This is exactly what has been achieved in Cator and Landsman (2014), at least mathematically: we show that rather than "*indeterminism in*, *indeterminism out*", the thrust of the FWT is really: "*determinism in*, *constraints on determinism out*".

What is missing, then, from both the original papers by Conway and Kochen (2006, 2009) and the reformulation of the FWT by Cator and the author, is a serious analysis of the (philosophical) kind of free will assumed in the theorem, and thence of the implications of the theorem for that specific kind. The present paper attempts to fill this gap. In fact, it bridges a canyon, in relating the philosophical prose typical of at least the Lewisian corner of the free will literature (which I briefly review in Section 2) to elementary mathematics of the kind relevant to the FWT. This is done in Section 3, upon which the actual application to the FWT in Section 4 is straightforward. Finally, Section 5 contains my conclusions.[1]

2. Compatibilist free will à la David Lewis

The first question is which philosophical notion of free will is

[1] Since I base my analysis on our own revised FWT, any conclusions from this analysis about the original version can only be indirect, but in my opinion the potential link between indeterminism in the quantum world and free will in humans is so feeble that even if we grant that the original FWT is non-circular (in that it proves such indeterminism, as claimed by Conway and Kochen), its implications for free will are at best speculative.