On What It Takes To Be a World

David Z. Albert and Jeffrey A. Barrett

ABSTRACT. A many-worlds interpretation is of quantum mechanics tells us that the linear equations of motion are the true and complete laws for the time-evolution of every physical system and that the usual quantum-mechanical states provide complete descriptions of all possible physical situations. Such an interpretation, however, denies the standard way of understanding quantum-mechanical states. When the pointer on a measuring device is in a superposition of pointing many different directions, for example, we are to understand this as many pointers, each in a different *world*, each pointing in a different determinate direction. We ask here whether such talk makes any genuinely intelligible sense of the term "world". We conclude that it does not.

This note will consist, more or less, of a single remark about something that follows from supposing that the linear quantum-mechanical equations of motion are the true and complete equations of motion for the entirety of the universe – which is something that has been explicitly supposed by Everett (1957 and 1973) and many who have tried to make sense of Everett's formulation of quantum mechanics. The remark is a pretty trivial one, but we want to draw attention to it here because – notwithstanding its simplicity – it seems to us to raise a rather urgent question about what the proponents of a many-worlds interpretation can possibly mean by the term "worlds". The remark may also be relevant to formulations of quantum mechanics where "histories" or "branches" are meant to do the work of "worlds" – examples may include such formulations as those found in Zeh (1970), Gell-Mann and Hartle (1990), and Omnés (1992).

There is, of course, a whole tradition of arguments in the physical literature (arguments that go by names like "The Story of Schrödinger's Cat", "The Story of Wigner's Friend", "The Story of Einstein's Camera", etc.) to the effect that the linear quantum-mechanical equations of motion cannot possibly be the true and complete equations of motion of everything there is. Those arguments run, more or less, like this. Suppose that there are linear quantum-mechanical equations of motion which *are* both true and complete. And suppose that *M* is a measuring device for measuring, say, the *z*-spins of electrons (we can think of *M* as nothing more than a constructed measuring instrument or as a composite system consisting of such an instrument together with a human observer). And suppose that *M* is working properly, which means nothing more than that

$$(1)\quad |r\rangle_M|z=+1\rangle_e \Rightarrow |\text{indicates that } z=+1\rangle_M|z=+1\rangle_e$$

and

$$(2)\quad |r\rangle_M|z=-1\rangle_e \Rightarrow |\text{indicates that } z=-1\rangle_M|z=-1\rangle_e,$$

where $|r\rangle_M$ represents the so-called "ready" state of *M*, the state in which the instrument and the observer (if there is one) are fully calibrated and plugged in and inclined and generally prepared to do whatever it takes to carry out a measurement of the *z*-spin of an electron, and the symbol $\Rightarrow$ represents the time-evolution of the composite system.

And now here comes the punch line. In the event that the above device is used to measure the *z*-spin not, as above, of a *z*-spin *up* electron or of a *z*-spin *down* electron but of an *x*-spin *up* electron, then the above definition of what it is for such a measuring device to be working properly, together with the assumption that the quantum-mechanical equations of motion are the true and complete equations of motion of the entirety of the universe, together with the fact that those equations are *linear*, will entail that

$$(3)\quad |r\rangle_M|x=+1\rangle_e = \frac{1}{\sqrt{2}}(|r\rangle_M|z=+1\rangle_e + |r\rangle_M|z=-1\rangle_e)$$

$$\Rightarrow \frac{1}{\sqrt{2}}(|\text{indicates that } z=+1\rangle_M|z=+1\rangle_e$$

$$+\ |\text{indicates that } z=-1\rangle_M|z=-1\rangle_e$$

And this amounts to a problem since the standard dogma

Topoi 14: 35–37, 1995.

concerning what it is to be in a quantum-mechanical superposition entails that when states like this obtain there simply fails to be any determinate matter of fact at all about what the measuring device is indicating, and yet what our *experience* tell us about the ends of the above sorts of measuring processes is that they are invariably situations in which there *is* some such matter of fact – since we believe that we get determinate results to our measurements, any satisfactory formulation of quantum mechanics is going to have to account for this one way or another.

The familiar response to this, the response that seems more or less inevitable, has been to conclude either that the linear quantum-mechanical equations of motion are not the true equations of the time-evolution of the quantum state of the universe (that quantum states sometimes undergo nonlinear reductions) or else that quantum states afford less-than-complete descriptions of physical systems (that those state-descriptions need to be supplemented by specifications of the values of "hidden variables").

But there has for some time been another response on the table too, one which affirms that quantum states afford complete descriptions of all possible physical situations and *also* that the linear quantum-mechanical equations of motion are indeed the true and complete laws of the evolution of those states. The idea here is that what needs to be amended is neither the linearity of the equations of motion nor the representation of reality by means of quantum-mechanical wave-functions, but the standard dogma about precisely what sort of reality it is that those wave-functions represent, the dogma, that is, about precisely what it is to be in a quantum-mechanical superposition. The idea, more particularly, is that the above quantum state ought to be read as a description not of *one* physical world but of *two*, in one of which there is an electron whose z-spin is *up* and an M that indicates that spin to be "up", and in the other of which there is an electron whose z-spin is *down* and an M that indicates that spin to be "down".

A number of rather difficult questions have been raised over the past few years about the internal coherence of talking this way, and what we want to do is to raise another particularly simple one: we want to ask whether this talk makes any genuinely intelligible sense of the term "world".

The difficulty arises as follows. Suppose that the measurement-interaction described in (3) occurs in the presence of another measuring device (precisely what *this* device measures will be described shortly) called M', and suppose that M' is initially in its ready state, and suppose that during the course of the interaction between M and e described in (3), M' has no interaction whatever with M or with e or, for that matter, with any other system. In that case, once the interaction is over, the overall quantum state is going to be given by

$$(4)\quad \frac{1}{\sqrt{2}}\,(\,|\text{indicates that } z = +1\rangle_M |z = +1\rangle_e + |\text{indicates that } z = -1\rangle_M |z = -1\rangle_e)|r\rangle_{M'},$$

which is the same as

$$(5)\quad \frac{1}{\sqrt{2}}\,(|\text{indicates that } z = +1\rangle_M |z = +1\rangle_e |r\rangle_{M'} + |\text{indicates that } z = -1\rangle_M |z = -1\rangle_e |r\rangle_{M'})$$

which according to the interpretations of quantum mechanics under discussion here is to be read as depicting two physical worlds, in one of which there is an electron whose z-spin is *up* and an M that indicates that spin is to be "up" *and* an M' in its ready state, and in the other of which there is an electron whose z-spin is *down* and an M that indicates that spin is to be "down" *and*, once again, an M' in its ready state.

Now, suppose that the observable that M' is designed to measure – let's call it O – happens to be a complete observable of the composite system consisting of e and M of which the state given by (5), and the state on the right-hand side of (3), are eigenstates with the associated eigenvalue, say, +1. And suppose that at the end of the little story told above, when the state that obtains is the one in (5), M' now measures O.

Consider what ought to happen. Let's think it through from two different angles. First, focus on the fact that, the first term in (5) is supposed to describe a *world* in which the z-spin of e is *up*. Well, within that world, the world where the actual z-spin of e is *up*, the observable O (since it doesn't commute with the z-spin of e) patently has no determinate value, and so the outcomes of O-measurements will necessarily be matters of chance. But, and this is the crucial point, the hypothesis that the overall quantum state of the entire *collection* of worlds invariably evolves in accordance with the *linear equations of motion* straightforwardly demands otherwise. What *that* hypothesis demands is that the overall state at the conclusion of the O-measurement will *with certainty* be

$$(6)\quad \frac{1}{\sqrt{2}}\,(\,|\text{indicates that } z = +1\rangle_M |z = +1\rangle_e + |\text{indicates that } z = -1\rangle_M |z = -1\rangle_e)\,|\text{indicates that } O = +1\rangle_{M'}.$$

which is the same as

$$(7)\quad \frac{1}{\sqrt{2}}\,(\,|\text{indicates that } z = +1\rangle_M |z = +1\rangle_e\,|\text{indicates that } O = +1\rangle_{M'} + |\text{indicates that } z = -1\rangle_M |z = -1\rangle_e\,|\text{indicates that } O = +1\rangle_{M'})$$

And so the hypothesis that the overall quantum state invariably evolves in accordance with the linear equations of motion apparently demands that when a state like (5) obtains and a measurement of O is carried out by M', then the outcome of that measurement, in the world in which the z-spin of e is up *will* with certainty *be* +1. What those equations demand, to put it a little differently, is that the outcome of that measurement, in the world in which the z-spin of e is *up*, will, emphatically, *not* be a matter of chance.

Let's put it one more way. What the linear equations of motion demand (if we are really taking the quantum-mechanical state to be a complete description of every physical system and if we are really taking the linear dynamics to be a complete and accurate description of the time-evolution of these systems) is that, when states like (5) obtain, certain experiments carried out within the world in which the z-spin of e is *up* will come out as if they had *not* been carried out within a world like that *at all* – that is, they will come out as if they had been carried out in a world very different from that one, a world in which O is, with certainty, +1.

And at this point it becomes exceedingly unclear (or at any rate it becomes exceedingly unclear to *us*) what sort of sense it can make to say that the first term of (5) describes a "world" at all!

References

DeWitt, B. S. and Graham, N. (eds.): 1973, *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton University Press, Princeton.

Everett, H., III.: 1957, '"Relative state" formulation of quantum mechanics', *Reviews of Modern Physics* **29**, 454–62, reprinted in DeWitt and Graham (eds.) (1973).

Everett, H., III.: 1973, 'Theory of the universal wave-function', in DeWitt and Graham (eds.) (1973).

Gell-Mann, M. and Hartle, J. B.: 1990, 'Quantum mechanics in the light of quantum cosmology', in W. H. Zurek (ed.), *Complexity, Entropy, and the Physics of Information*, Proceedings of the Santa Fe Institute Studies in the Sciences of Complexity, Vol. VIII, Addison-Wesley, Redwood City, CA, pp. 425–58.

Omnès, R.: 1992, 'Consistent interpretations of quantum mechanics', *Reviews of Modern Physics* **64**(2), 339–82.

Zeh, H. D.: 1970, 'On the interpretation of measurement in quantum theory', originally in *Foundations of Physics*, 1; reprinted in *Quantum Theory and Measurement*, by Wheeler and Zurek (eds.) (1983), Princeton University Press, Princeton.

Columbia University and University of California, Irvine, USA