Review for *European Journal of Philosophy*

The Quantum Mechanics of Minds and Worlds

Jeffrey A. Barrett, *The Quantum Mechanics of Minds and Worlds*. Oxford: Oxford University Press, 1999. Pp. xv + 267. price 30.00 UK Pounds. ISBN: 0 19 823838 X.

Review by: Jeremy Butterfield, All Souls College, Oxford OX1 4AL, England

This book is about the central interpretative problem of quantum theory, the measurement problem, and the strategy for solving it that was suggested by Everett (1957). Barrett aims to explain these issues, with a minimum of technical apparatus; emphasizing instead the philosophical, especially metaphysical, aspects. In particular, he aims to survey in an even-handed way the various versions of the Everettian strategy. Even without technicalities, that is ambition enough for one book. For not only are the versions connected with central metaphysical topics such as the nature of transtemporal identity and the relation between mind and matter; they also involve some dizzying speculations about such matters. Thus some versions envisage an ontology of many 'worlds'; i.e. they claim that the universe (philosophers' 'actual world') contains a plethora of Everettian 'worlds', where each such 'world' is something like a familiar macroscopic realm—but the worlds differ among themselves about e.g. the location of macroscopic objects. Other versions envisage an ontology of many 'minds'; i.e. they claim that to each sentient brain (a human's, a cat's) there corresponds a plethora of minds (or if you prefer, mental states), their experiences differing about such matters as the location of macroscopic objects. Hence Barrett's alluring title.

Overall, Barrett succeeds admirably in his aims. To be sure, there are some lacunae and rough edges; but by and large, his arguments are cogent, clear and often convincing. The book is the more valuable for two other reasons. First, there is (so far as I know) no other monograph undertaking the same kind of survey. (Although the last 10 years has seen a growth in the number and quality of philosophical analyses of Everettian ideas, these have been in articles; by such authors as Albert, Donald, Lockwood, Loewer, Saunders—and Barrett himself.) Second, I think it is important that philosophers assess dizzying speculations such as those I have mentioned. Agreed, many philosophers will think such speculations, when they first hear them, to be so incredible as to be not worth pursuing. But their being dizzying, or even incredible, is no *reason* for dismissing them. Agreed, there are such reasons: one might be some sort of instrumentalist about physical theory, and so unwilling to take metaphysical suggestions from even so outstandingly successful a theory as quantum theory; or one might have positive reasons for some other strategy than Everett's for interpreting quantum theory. But setting aside such reasons, philosophers should judge such speculations on their merits: after all, they could be a refreshing stimulus to metaphysical disputes. Accordingly, in this review, I will sketch the contents of the book, and end by mentioning some criticisms.

Barrett begins with two Chapters about the measurement problem and standard quantum theory's response to it. In short, the measurement problem is the threat that according to quantum theory itself, the lack of values for physical quantities such as position, momentum and energy, which is characteristic of quantum theory's description of microscopic systems such as electrons, should also infect the macroscopic realm. The threat is clearest if one considers a measurement situation: quantum theory apparently predicts that measuring say the momentum of an electron, when it is in a state that is not definite for momentum (a 'superposition of momentum eigenstates') should lead to the pointer of the apparatus having no definite position—it should be in a superposition of position eigenstates.

As Barrett discusses in Chapter 3, this problem has been known since the discovery of quantum theory in the 1920s. The founding fathers—Bohr, Einstein, Heisenberg and Schrödinger—were aware of it, and of a cluster of other interpretative problems, such as non-locality, which are bound up with it. They struggled mightily with these problems (and with each other!), but without agreeing on any solution. However, by about 1935 a consensus formed around a minimal interpretation of quantum theory, that postulated that the quantum state of both the microscopic system and the apparatus changed discontinuously after the measurement interaction, so that the apparatus' pointer got a definite position. (This postulate was called the 'projection postulate'; the change of state was called, more colloquially, the 'collapse of the wave-packet'.) Barrett, very reasonably, dubs this the 'standard formulation' of quantum theory—and points out how unsatisfactory it is, since 'measurement' is vague, and there are difficulties of principle in reconciling quantum theory's usual law of how states change over time, the Schrödinger equation, with the projection postulate. (Technically, the difficulties centre around replacing an improper mixture with a proper one.)

Nevertheless, for some twenty-five years (from about 1935 to 1960) almost all quantum physicists ignored the measurement problem (and quantum theory's other interpretative problems). There were at least two main reasons for this. The stunning theoretical development and empirical success of quantum theory from the 1920s onwards, in ever wider domains of applications, made the interpretative problems seem unimportant, or at least not 'ripe for solution'. There was also the immense authority of such figures as Bohr and Heisenberg, who defended the standard formulation, or something close to it, with philosophical arguments—albeit misty ones! (Barrett's historical discussion of the establishment and hegemony of the 'standard formulation' is judicious, but very short; for an excellent recent monograph, cf. Beller (2000).)

So it was into this forbidding climate that Everett in 1957 launched his suggestion that one could solve the measurement problem without any recourse to the projection postulate. Barrett's Chapter 3 gives a detailed exposition of Everett's writings. In brief: Everett claims that (i) the universe has a quantum state, $\Psi$ say, that always evolves according to the Schrödinger equation; he admits that (ii) the measurement problem suggests $\Psi$ will be a superposition corresponding to many different definite macroscopic realms ('macrorealms'); but he then argues that (iii) one can recover the subjective appearance of a definite macrorealm by postulating that all the various definite macrorealms are actual—'we just happen to be in one rather than the others'.

Stated so briefly, the Everettian strategy stands in dire need of being clarified; and Barrett's succeeding Chapters, 4 to 9, undertake to do so. (Indeed, the details of Everettian interpretations turn out to differ so much that the common slogan 'the various definite macrorealms are actual' is really ambiguous.) Broadly speaking, Barrett sets up two 'probes' (Chapters 4 and 5). Then he discusses three types of interpretation (Chapters 6, 7 and 8). Finally, in Chapter 9 he gives a short summary of some common problems shared by most of the interpretations discussed in Chapters 6 to 8; and he tentatively endorses one of Chapter 8's interpretations (a version of the 'many minds' interpretation, mentioned in my first paragraph).

Barrett's first probe, in Chapter 4, is the so-called 'bare theory' ; (so named by David Albert; cf. his (1992)). Roughly speaking, this is the standard formulation of quantum theory, *minus* the projection postulate. A bit more precisely: of claims (i) to (iii) above, it asserts (i) and (ii), but not (iii)—and it explains away the subjective appearance of a definite macrorealm as an illusion. The main tactic in this explanation is to postulate a toy-model of how quantum states of the brain (or perhaps of the brain and parts of its environment) underly beliefs in e.g. the positions of pointers. Quantum theory then allows one to argue that: if (a) a person reliably comes to believe that the pointer reads '1 unit', and so has a definite value for position, after the apparatus measures an electron initially in a '1 unit of momentum' eigenstate, and (b) similarly for 2 units; then the person will also come to believe, after the apparatus measures an electron initially in a superposition of 1-unit and 2-unit momentum eigenstates, that the pointer has a definite value—even though it doesn't! In short, this is what philosophers call an 'error theory', with a vengeance—it takes almost all our everyday beliefs about the macrorealm to be utterly false.

Overall, Barrett rejects the bare theory, not least because if it were true, we would not have an empirical justification for holding it—since the apparent evidence for quantum theory, e.g. statistics drawn from experimental results, would be illusory (p. 116). But for someone expounding and assessing Everettian interpretations, the bare theory is useful in various ways. In general, it enables one to discuss (i) and (ii) separately from (iii). And more specifically: the bare theory has illuminating formal properties, for example about the relation between probabilities and frequencies; and the toy-model of quantum states underlying beliefs is useful in connection with the 'many minds' interpretations of Chapter 8.

Barrett's second probe, in Chapter 5, is a comparison of the Everettian strategy with pilot-wave interpretations of the type developed by deBroglie and Bohm. So far as I know, the first person to make this comparison was John Bell (famous for Bell's theorem which concerns non-locality); and here Barrett's discussion—like the whole literature about the interpretation of quantum theory!—is much indebted to Bell's writings. Bell admired the pilot-wave interpretations, and urged that they were superior to Everettian interpretations in three respects. (1): They specify precisely which physical quantity is preferred in the sense of having definite values in addition to those assigned by standard quantum theory. (The specifications vary somewhat; but in short, for non-relativistic quantum theory they choose the position of every point-particle.) On the other hand, Everettians have traditionally left it vague how to define 'macrorealm' or even 'pointer-position'. (2): They specify precisely how the postulated extra values evolve over time (in what is called the 'guidance equation'); whereas again, Everettians have traditionally left it vague how macrorealms evolve over time. (3): They take only one position to be actually possessed (by each point-particle); whereas Everettians claim that somehow *each* of the extra values is possessed by some sort of 'copy' of the system in question (claim (iii) above)—which seemed to Bell at best otiose, and at worst an ontological extravagance (cf. Bell 1987, pp. 97, 133).)

Barrett does not share Bell's preference for pilot-wave interpretations. But in his ensuing Chapters, he agrees that the Everettian strategy must address the three topics Bell raises—defining the preferred quantity, specifying the evolution over time, and justifying the 'plurality' of values. By and large, he agrees with Bell that Everettians have yet to fully address the first and second; but he thinks that the 'plurality' of values is justified by arguments about the interpretation of probability.

In Chapters 6, 7 and 8, Barrett discusses *seriatim* three types of Everettian interpretation, under the names 'many worlds', 'many minds' and 'many histories'. I shall only discuss the first two, since the distinctive aspects of the third concern a technical proposal (the 'consistent histories formalism') about evolution over time.

'Many worlds' interpretations are closest to the Everettian strategy as I have so far described it: the 'worlds' are the macrorealms, each definite as regards such matters as the position of macroscopic objects. But the 'worlds' terminology is due, not to Everett, but to DeWitt and Graham; who revived Everett's ideas in the 1970s. Barrett expounds their articles, and uses his previous Chapters' probes to press the advocate of many worlds about such questions as how exactly one should define 'world', and how exactly they evolve over time. In particular, he distinguishes two proposals about evolution over time. (i) The 'splitting worlds theory' (p. 149f.) holds that when the number of 'world components' in the universal quantum state increases, say because a measurement occurs so that an apparatus' pointer goes from being in a 'ready' state to being in a superposition of two possible positions, the universe literally splits to give two 'daughter-worlds'; (cf. cases of fission in discussions of personal identity). (ii) The 'many-threads theory' (p. 179f.) holds that there is a fixed total population of worlds that each persist over time, different sub-populations inhabiting the various 'world components'; when the number of 'world components' increases, the total population is simply re-partitioned; (cf. cases of divergence of possible worlds in discussions of indeterminism). Furthermore, the total population is taken to be enormously large—most authors assume it has at least the cardinality of the continuum. This postulated population is often taken to support a broadly frequentist interpretation of quantum theory's probabilities, compatibly with the over-arching deterministic evolution of the universal quantum state.

The main difference between 'many minds' and 'many worlds' interpretations lies in the definition of the preferred quantity. The 'many minds' Everettian suggests that to solve the measurement problem, we do not need to secure what I have called 'a definite macrorealm': we only need to secure the *appearance* of such. A bit more precisely: the idea is that the preferred quantity is whatever physical quantity, defined on brains (or brains and parts of their environments), has definite-valued states (eigenstates) that underpin such appearances, i.e. underpin our mental states of belief in, or sensory experience of, the familiar macroscopic realm. This idea returns Barrett to Chapter 4's bare theory, and its toy-model of how quantum states of the brain might underly mental states. Indeed, in some ways the 'many minds' Chapter (Chapter 8) is the climax of the book. For most of Barrett's framework for stating and assessing Everettian interpretations now comes into play. In particular, Barrett's discussion invokes each of: the toy-model, Bell's comparison of Everett with pilot-wave interpretations and the questions which Bell's comparison raises, and Chapter 6's distinction between (i) 'splitting' and (ii) 'many threads'.

Finally, some criticisms. I said that there were some lacunae and rough edges. The lacunae I have in mind concern the literature. By my lights, some significant authors are given too short a discussion. For example, Saunders has developed a distinctive version of the 'splitting' version of 'many worlds', in papers in *Synthese* from 1995 to 1998; but Barrett cites only the first paper, as part of his discussion of 'many minds'. Besides, that discussion has other significant lacunae. (a): Donald's papers are well-known for going significantly beyond Chapter 4 's toy-model of how quantum states of the brain might underly mental states. Barrett cites this work briefly (in three footnotes) but does not discuss it: he sticks to the toy-model. (b): More important, Barrett does not cite Squires (1990) whose version of 'many minds' is close to the 'single mind' version (pp. 186-192) that Barrett tentatively endorses in his final conclusion (p. 248). (c:) Finally, Zeh seems to have been the first to think of a 'many minds' Everettian interpretation: a slightly revised version of his original paper, not cited by Barrett, has recently been published (Zeh (2000)) . (Incidentally, most of the authors Barrett cites are American; is it a coincidence that the authors I have just complained to be neglected by him are all European?)

As to rough edges, the second half of the book seems to me in various ways less polished. In particular, several passages, even sections, seem to me to be in the wrong place. Two examples are worth describing since they mean that readers might misunderstand central points which underly much of the discussion. They concern (i) the distinction between a proper and an improper mixture, and (ii) the process of decoherence. Barrett postpones both these topics till Chapter 8's discussion of 'many histories'. But that seems to me too late. For (as I said in paragraph 4) a proper understanding of the measurement problem involves (i); and (although I have not discussed it) decoherence is very relevant to all versions of the Everettian strategy. A bit more precisely: all parties agree that for 'many worlds' and 'many minds', decoherence bears upon all three of the topics raised by Chapter 5's probe: defining the preferred quantity, specifying the evolution over time, and justifying the 'plurality' of values. (These are not the only examples. Another is the discussion of dynamics on p. 202f., in the 'many minds' Chapter 7; surely this would be better placed in Chapter 6, or even 5?)

So to sum up: criticisms apart, this is a good book. Future work on the philosophical aspects of Everettian interpretations starts here.

References:

Albert D., 1992. *Quantum Mechanics and Experience*. Cambridge MA: Harvard University Press.
Bell J., 1987. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.
Beller M., 2000. *Quantum Dialogue*. Chicago: University of Chicago Press.
Everett H., 1957. 'Relative State' Formulation of Quantum Mechanics, *Reviews of Modern Physics* **29**, pp. 454-462.
Squires E., 1990. *Conscious Mind in the Physical World*, Bristol: Adam Hilger.
Zeh H-D., 2000. The Problem of Conscious Observation in Quantum Mechanical Description, *Foundations of Physics Letters* **13** pp. 221-233; quant-ph/9908084.