JEFFREY A. BARRETT

ON THE NATURE OF EXPERIENCE IN THE BARE THEORY

ABSTRACT. Quantum mechanics without the collapse postulate, the bare theory, was proposed by Albert (1992) as a way of understanding Everett's relative-state formulation of quantum mechanics. The basic idea is to try to account for an observer's beliefs by appealing to a type of illusion predicted by the bare theory. This paper responds to some recent objections to the bare theory by providing a more detailed description of the sense in which it can and the sense in which it cannot account for our experience.

The bare theory, quantum mechanics without the collapse postulate but with the standard eigenvalue-eigenstate link, has several suggestive properties.[1] These properties tell us what an observer would report concerning her experience in various situations if the theory were true. Here I will discuss just one of these properties in some detail.

Suppose that M is a perfect x-spin measuring device in the following sense: it is constructed so that the position of its pointer becomes perfectly correlated with the x-spin of a spin-1/2 system S without disturbing it. That is, M has the following two dispositions: (1) if S is initially in an x-spin up eigenstate, then M will report that the result of its measurement is x-spin up and will leave S in the x-spin up state

(1)

and (2) if S is initially in an x-spin down eigenstate, then M will report that the result of its measurement is x-spin down and will leave S in the x-spin down state

(2 $|r\rangle_M|\downarrow\rangle_S \rightarrow \downarrow\rangle_M|\downarrow\rangle_S$

According to the bare theory, the time-evolution of every physical system is always correctly described by the time-dependent Schrödinger equation, which is linear. It follows from the linearity of the dynamics and from how M has been constructed (from the two dispositions described above) that if M begins in a ready-to-make-a-measurement state and S begins in a superposition of x-spin up and x-spin down

(3) $r\rangle_M(\alpha \quad _S + \beta|$

Synthese **113:** 347–355, 1998.

then the state of $M + S$ after M's x-spin measurement will be

(4) $\alpha|\uparrow\rangle_M|\uparrow\rangle_S + \beta|\downarrow\rangle_M|\downarrow\rangle_S$

(since the result of applying the linear dynamics to a linear superposition is the linear superposition of applying the dynamics to each term). If α and β are nonzero, then this is not an eigenstate of M reporting any determinate x-spin result; rather, it is a state where M is entangled with S and in a superposition of reporting mutually contradictory results. While M would not be in an eigenstate of reporting any particular determinate x-spin result, one can argue that M would nonetheless report that it recorded *some* determinate x-spin result, either x-spin up or x-spin down.

Suppose that M is constructed in such a way that it has the disposition to answer the question "Did you get some determinate result to your x-spin measurement, either x-spin up or x-spin down?" with "Yes" if it recorded x-spin up (if $M + S$ ended up in the state $|\uparrow\rangle_M|\uparrow\rangle_S$) and with "Yes" if it recorded x-spin down (if $M + S$ ended up in the state $|\downarrow\rangle_M|\downarrow\rangle_S$). Then if M in fact recorded a superposition of the two possible x-spin results (if $M + S$ were in the state described by (4) above), then it follows from the linearity of the dynamics that M would nonetheless answer "Yes" to the question; that is, M would falsely report that it got a determinate x-spin result, either x-spin up or x-spin down.

One reaction to this conclusion has been to argue that the Hermitian operator that is taken here to correspond to asking the question "Did you get some determinate result to your x-spin measurement?" is simply an identity operator and that such a simple operator cannot possibly represent such an interesting question (Weinstein (1996) provides an example of this line of argument). This seems, however, to miss the point. While it is true that the question asked of M here might be represented by the identity operator, the fact remains that if M were constructed in the way described above, then it would be disposed to report that it got a determinate x-spin result when it was actually in an entangled superposition of recording mutually incompatible x-spin results. How one represents the observable corresponding to asking M the whether it got a determinate result is irrelevant. That M would report that it got a determinate x-spin result follows directly from the dispositions that M would have on the bare theory if it was wired to report that it recorded a determinate result in those situations where it did in fact record a determinate result, in those situations where it was in an eigenstate of recording a determinate x-spin result.

Given the way that it is wired, one would expect that M would reliably answer the question "Did you get some determinate result to your x-spin measurement?", but in the world described by the bare theory it does not (it gives the right answer only when it is in an eigenstate of recording one or the other of the two possible x-spin results). The natural question to ask then is whether there is any way to wire M so that it would always provide reliable answers concerning its records of past measurements. It turns out that if the bare theory is true, then the answer is *No*.

Suppose that we wanted to be sure that measuring device N has the disposition to answer the question "Did you get a determinate result to your x-spin measurement?" with "No" if it ends up in the state

$$(5) \qquad \psi_1 = \frac{}{\sqrt{2}} \ \uparrow\rangle_N|\uparrow\rangle_S + ||\downarrow\rangle_N|\downarrow\rangle_S)$$

and "No" if it ends up in the state

$$(6) \qquad \psi_2 = \frac{}{\sqrt{2}} \qquad \uparrow\rangle_S \quad \theta||\downarrow\rangle_N|\downarrow\rangle_S$$

If its measurement record involves more than a few particles, then this would be a very difficult device to build since it would have to perform something akin to an interference experiment on the physical system that records the initial measurement result in order to answer this question reliably in these states. But suppose that we succeed in wiring N accordingly. While N would reliably report whether it recorded a determinate x-spin result in states ψ_1 or ψ_2, N would mistakenly report that it failed to record a determinate result when it was in fact in an eigenstate of recording a determinate result, say x-spin up. f asked whether it recorded a determinate result, N would answer "No' in the state ψ_1 and "No" in state ψ_2, so by the linearity of the dynamics, it would also (though this time mistakenly) answer "No" in the state $|\uparrow\rangle_N|\uparrow\rangle_S = 1/\sqrt{2}(\psi_1 + \psi_2)$, where it in fact did determinately record x-spin up.

The bare theory places a very strong constraint on the reliability of measurement reports: If a measuring device must answer a question the same way in two orthogonal states but differently in a superposition of the two orthogonal states in order to answer the question reliably, then no measuring device can answer the question reliably in general since, by the linearity of the dynamics, it will always answer the question the same way in the superposition as it does in the orthogonal states.

Since one cannot design a universally reliable measuring device, Bub, Clifton, and Mouton (1996) (henceforth BCM) have argued that the bare theory, like many hidden variable theories, requires one to choose a preferred basis, where the preferred basis in the bare theory is the one that tells us when a good measuring device will be able to reliably answer a certain type of question. Since some hidden variable theories are arguably among our best formulations of quantum mechanics, this is not a very serious objection to the bare theory. Indeed, rather than see this as a serious argument against the bare theory, BCM prefer to see it as an argument that undermines one of the most compelling arguments *for* the bare theory: its simplicity. I think that they are at least to some extent right, but there also seems to me to be a significant difference between the role played by the preferred basis in a hidden variable theory and the role played by the preferred basis in the bare theory. In a hidden variable theory the preferred basis determines which out of all possible physical quantities always has a determinate value. In the bare theory, however, the preferred basis is simply stipulated as a part of what it means to be a good measuring device or observer in the theory: more specifically, in discussions of the bare theory it is usually assumed that, whatever else it might do, a good measuring device or observer would be able to reliably answered the question "Did you get a determinate result to your x-spin measurement?" *when it had in fact recorded a determinate result.* It is certainly true that a measuring device or observer might be wired differently. There is something akin to a choice of a preferred basis here, but it is a choice about how to model observers, not about what observable as a fundamental fact about the ontology of the world always has a determinate value, and the former seems a more modest role for a preferred basis.

Further, while BCM are right to insist that the mathematical formalism of quantum mechanics fails to tell us how to model observers and that it is conceivable that an observer might be wired differently than what is usually assumed in discussions of the bare theory (wired like N above, for example), it does not follow that the choice of how to model *human* observers is arbitrary. We tend to believe that if human observers are wired to answer any question reliably, then they are wired to answer the question "Did you get a determinate result to your measurement?" reliably *when they determinately did or determinately did not record a result.* And we tend to believe that humans are not wired to perform the type of measurement that would reliably determine whether they are in superpositions of having recorded mutually incompatible results (Indeed, if the records involve more than a few correlated particles, as brain records certainly do, then we believe that it would be almost impossible to build a device that could perform this measurement reliably).

The upshot is that while the usual way of trying to account for our experience in the bare theory does require one to assume that observers are wired in a special (though plausible) way, a hidden variable theory typically requires one to make the stronger assumption that a particular observable is ontologically privileged, that it picks out the one and only physical property that is always determinate *and* such a theory also requires one to assume that observers are wired in a special way so that every measurement is ultimately a measurement of the always-determinate observable.[2] So while BCM's preferred basis argument does show that the bare theory's account of our experience must depend on a contingent fact about human nature, it might still be argued that this commitment to a preferred basis is less troubling than the commitment typically required by a hidden variable theory. In any case, we are left with the fact that if the bare theory is to have any chance of accounting for our experience, then we must make some assumption about how we are wired. We will assume that human observers are wired like M above (not like N).

BCM have a second more serious objection to the bare theory: even granting the usual way that human observers are modeled on the theory, the bare theory still cannot account for our actual experience and is thus not a serious candidate for a satisfactory formulation of quantum mechanics. While I agree that there is a sense in which the bare theory cannot account for our experience, there is also an interesting sense in which it *can*. Whether one finds this second type of account compelling depends on how seriously one is willing to take what the bare theory tells us concerning the basic nature of experience.

Suppose that an observer M measures the x-spin of a system in an eigenstate of z-spin and thus ends up in a superposition of recording x-spin up and x-spin down. It follows immediately from the standard eigenvalue-eigenstate link that in this state M does not believe x-spin up, does not believe x-spin down, does not believe both, and does not believe neither. Thus, the bare theory is logically incompatible with the claim that either M would believe x-spin up or M would believe x-spin down after the measurement. If *this* is what M experiences, then the bare theory cannot account for M's experience. But the bare theory denies that this is in fact what M experiences.

The bare theory tells us that although M did not in fact record a determinate x-spin result, she would nonetheless believe she did. If M believed her own report, the report that she would with certainty be disposed to make, then it would *seem* to her that the final state had somehow collapsed to $|\uparrow\rangle_M|\uparrow\rangle_S$ *or* $||\downarrow\rangle_M|\downarrow\rangle_S$ when in fact the dynamical evolution of the composite system had been perfectly linear and there had been no collapse. "You are looking at the pointer right now." "Yes." "Do you clearly see it either pointing to x-spin up or to x-spin down?" "Yes." "You are as certain as you can be on the basis of direct empirical evidence that the pointer is pointing at a particular, specific, and fully determinate result?" "Yes."

Again, this is how the conversation *must* go if M has the disposition to reliably answer such questions when she is in an eigenstate of recording x-spin up and when she is in an eigenstate of recording x-spin down. Hence M would believe that she had recorded a perfectly determinate x-spin result when there was no determinate x-spin result that M believed that she recorded.

It is important to be clear about this. The claim that a proponent of the bare theory would make is not that M would believe that she recorded x-spin up or that she would believe that she recorded x-spin down, but rather that she would believe that she recorded x-spin up or x-spin down. To distinguish it from an ordinary phenomenal experience, one might call the event leading to this disjunctive belief a *disjunctive experience*. M's disjunctive experience here would be phenomenally indistinguishable from either getting x-spin up or getting x-spin down (because that is precisely what M would be disposed to report), but it would be wrong to claim that it would be phenomenally indistinguishable from getting x-spin up and it would be wrong to claim that it would be phenomenally indistinguishable from getting x-spin down since the observer would not be in an eigenstate of making either of these judgements (since M would not determinately report either of these). And, for the same reason, it would be wrong to claim that the observer *could* distinguish the disjunctive experience from getting x-spin up or that the observer could distinguish the disjunctive experience from getting x-spin down. Again, the right thing to say is that the observer would be unable to distinguish the disjunctive result from x-spin up *or* x-spin down. If what one takes as standing in need of an explanation is the belief that one's experiences are typically perfectly ordinary and determinate, then the bare theory provides it. (The observer would say that her experience was perfectly ordinary and determinate in the eigenstate cases and would consequently have the sure-fire disposition to say the same thing in the superposition).

Perhaps the following thought experiment will help to clarify the nature of disjunctive experience. Suppose that an observer measures the x-spin of three object systems: the first is in an x-spin up eigenstate, the second is in an x-spin down eigenstate, and the third is in a superposition of x-spin up and x-spin down. The observer will believe that she has a determinate phenomenal result in all three cases. Moreover, she will believe that the result of her last measurement was phenomenally indistinguishable from the result of exactly one of her first two measurements. But the observer's disjunctive experience will be neither distinguishable nor indistinguishable from the experience of getting x-spin up (the observer will fail to have a determinate belief concerning whether her first and third results agree) nor will it be distinguishable nor indistinguishable from the experience of getting x-spin down (she will not have a determinate belief concerning whether her second and third results agree); rather, the disjunctive experience is indistinguishable from either the experience of getting x-spin up or the experience of getting x-spin down (the observer will determinately believe that the result of the third measurement is indistinguishable from exactly one of the first two results). It also follows from the linearity of the dynamics that if an observer can correctly identify those experiences that are specific and ordinary when she has them, then she will believe and report that her disjunctive experiences are perfectly specific and ordinary.

The upshot is that if the bare theory were true, then an observer would be under the illusion that her empirical experience was perfectly determinate, specific, and ordinary, when it would typically be devoid of any ordinary specific content. If the bare theory were true, then first-person authority concerning whether particular experiences and beliefs had ordinary specific content would be routinely violated in a striking way – an observer would typically believe that she had an ordinary determinate experience when there would be no such experience that she believed that she had.

This is odd enough to be philosophically interesting, but not so odd as to be logically incoherent. Let $B(x)$ represent that M determinately believes x, and let *up* and *down* represent "The measurement result was x-spin up" and "The measurement result was x-spin down", respectively. After M measures a system in a superposition of x-spin eigenstates, B(*up* or *down*), not B(*up*), and not B(*down*) would all be true, but these are mutually consistent. What must be sacrificed to avoid an inconsistency here is that if B(*up* or *down*) then either B(*up*) or B(*down*). But this is something that we have never believed anyway. I believe, for example, that Tibet either has or does not have a population greater than ten million, but since I have not performed the relevant observations, I do not believe that it does and I do not believe that it does not. What makes M's situation odd here is that we believe that *M did perform the relevant observations that would provide the basis for either one belief or the other.* What is violated by the bare theory's account of M's experience then is not a basic logical principle but rather a basic intuition about what type of observation would lead a good observer from B(*up* or *down*) to either B(*up*) or B(*down*).

When someone objects that we do in fact know that our experience has specific content *and that we know what that content is*, it is difficult not to agree. But the bare theory predicts that if we were wired like M, then this is just what we would (mistakenly) believe about our experience. "Yes", I complain, "but my experience does in fact have specific content". "Right now I *see*, before my own eyes, the pointer pointing at a particular, fully determinate result. And *it is x-spin up*!" A proponent of the bare theory must respond, "Yes, you believe that the pointer is pointing in a determinate direction just as it seems to me that I heard you make a determinate report just now telling me what direction it is pointing, but the pointer nonetheless fails to point in a determinate direction and I did not hear a determinate report. Regardless of how strongly you might insist that your experience has specific content, the bare theory can explain why you are so insistent, yet wrong".

There are clearly serious problems with the bare theory,[3] but whether BCM are right in claiming that it fails to account for our experience depends on how much scope one is willing to give the theory in characterizing the fundamental nature of experience. One might have little patience for the bare theory if one believes that one already knows the true nature of experience and belief (if one believes, for example, that experience always has specific content that is directly accessible via introspection). But when BCM insist that we do in fact sometimes see *up* when the bare theory predicts that we are actually in a superposition of seeing *up* and *down* the reliably of this judgment requires the reliability of introspection, which is exactly what the bare theory denies.

NOTES

[1] Some of these properties were first suggested by Everett (1957). There have been several subsequent attempts to clarify the properties and to determine their significance: see Hartle (1968), DeWitt (1971), Everett (1973), Graham (1973), Albert and Loewer (1988), Albert (1992), and Barrett (1995) for examples. The bare theory itself was proposed as a reading of Everett's relative-state formulation of quantum mechanics by Albert (1992).

[2] See Barrett (1996).

[3] For further discussions of some of its problems see Albert (1992) and Barrett (1994).

REFERENCES

Albert, D. Z.: 1992, *Quantum Mechanics and Experience*, Harvard University Press, Cambridge, Mass.

Albert, D. Z. and B. Loewer: 1988, 'Interpreting the Many Worlds Interpretation', *Synthese* **77**, 195–213.

Barrett, J.: 1996, 'Empirical Adequacy and the Availability of Reliable Records in Quantum Mechanics', *Philosophy of Science* **63**, 49–64.

Barrett, J.: 1994, 'The Suggestive Properties of Quantum Mechanics without the Collapse Postulate', *Erkenntnis* **41**, 233–252.

Bub, J., R. Clifton, and B. Monton: forthcoming, 'The Bare Theory has no Clothes', *Minnesota Studies in the Philosophy of Science*.

DeWitt, B. S.: 1971, 'The Many-Universes Interpretation of Quantum Mechanics', in *Foundations of Quantum Mechanics*, Academic Press, New York; reprinted in DeWitt and Graham (1973).

DeWitt, B. S. and N. Graham (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.).

Everett, H., III: 1973, 'Theory of the Universal Wave Function', in DeWitt and Graham (eds.). ,

Graham, N.: 1973, 'The Measurement of Relative Frequency', in DeWitt and Graham (eds.), 1973.

Hartle, J. B.: 1968, 'Quantum Mechanics of Individual Systems', in *American Journal of Physics* **36**(8), 704–12.

von Neumann, J.: 1932, *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin; reprinted and translated as *Mathematical Foundations of Quantum Mechanics*, translated by R. Beyer, Princeton University Press, Princeton, 1955.

Weinstein, S.: 1996, 'Undermined', *Synthese* **106**(2), 241–251.