Many Minds are No Worse than One

David Papineau

1 Introduction

The Everett-style interpretation of quantum mechanics developed by Michael Lockwood (the many minds view henceforth) is certainly very weird. Nevertheless, it may be true. Its strength is that it promises to save the appearances without wave function collapses or non-local interactions. Its weakness is that it seems to deny some obvious common-sense truths.

It is worth noting, however, that the common-sense assumptions denied by the many minds view involve either *consciousness* or *objective probability*. And the striking thing about these two notions is that neither is well integrated into the rest of our world view. How do conscious facts relate to non-conscious facts? And how do probabilistic facts relate to non-probabilistic facts? These are two of the most baffling questions in philosophy, and nobody has any good answers. In practice we use certain operational links to tie conscious and probabilistic facts to non-conscious and non-probabilistic facts respectively. Yet we lack any cogent philosophical justification of these links.

In this note I want to suggest that our understanding of consciousness and objective probability would be no less satisfactory on the many minds view than on a more conventional view of reality. Indeed the situation would be quite comparable. The many minds view would preserve the 'operational links' which connect conscious and probabilistic facts to the rest of the world, and would therefore leave us with the problem of explaining these links. Moreover, the many minds view would not offer any obvious solution to this problem. But then, as I said, neither does our conventional view of the world.

True, the many minds view would force us to change *some* familiar common-sense assumptions about consciousness and probability (indeed rather more, I shall suggest, than Michael Lockwood recognizes). But this is arguably a cost worth paying. For it seems to me that the threatened assumptions are backed by nothing except familiarity. As we shall see, they do not matter to our operational use of the notions of consciousness or probability. And they have no theoretical backing, since we have no good theories of consciousness or probability to start with.

So I want to suggest that the common-sense assumptions rejected by the many minds view are unmotivated, free-floating 'danglers', which are therefore up for grabs. Some people will no doubt take the weirdness of denying these assumptions as itself an argument against doing so. But mere unfamiliarity seems a poor argument against the only view that promises to explain the appearances without positing ad hoc or physically impossible mechanisms.

2 Consciousness

According to the many minds view as developed by Lockwood, if you observe a cat in a superposition of 'live' and 'dead', say, then your brain will itself become a superposition of 'registering live' and 'registering dead'; there is no physical point at which the wave function 'collapses'. Moreover, Lockwood's version of the many minds view holds that conscious mental events supervene on physical events. So at the conscious level too you will register 'live' and also register 'dead'.

On the face of it, this seems to contradict our experience. Surely we either see a live cat or a dead cat, but not both. However, we need here to consider what it *would be like* to have a superposed brain. A first thought might be that it would be like seeing a superposed photo of a live cat and a dead cat. And of course this is not what we experience. However, the many minds view rejects this account of what it would be like to have a superposed brain, and says that it would rather be like being two different people who know nothing of each other, one of whom sees a live cat and the other a dead cat. If the many minds view is right about this, then conscious experience would be just as we find it, and the many minds view would save the appearances.

Why should we accept the many minds account of what it would be like to have a superposed brain? Rather than address this question immediately, let us first consider the conventional account of the relation between brains and conscious experience. The conventional account supposes that when a (non-superposed) brain registers a live cat, there is a conscious experience as of seeing a live cat, and when it registers a dead cat, there is a conscious experience as of seeing a dead cat. More generally, it supposes that whenever certain kinds of physical systems are in certain states, then there are corresponding conscious experiences.

This is the 'operational link' between non-conscious and conscious facts which I gestured at above. Perhaps this operational link is less than fully precise, but for present purposes it will suffice for us to assume that conventional thought holds that conscious experience is characteristic of spatio-temporally continuous physical systems with enough complexity to form lasting records of past events and to generate behaviour on the basis of those records.

Why, on the conventional view, is it consciously *like* something to be such a physical system? For example, why do you have a conscious experience as of a live cat, when your visual cortex is in such-and-such a state? I take it that nobody has a good answer to this kind of question.[1] This is why I said earlier that we lack any further theoretical justification for the operational link which connects conscious to non-conscious states.

I now want to observe that the many minds view will preserve this operational link between conscious and non-conscious states. Along with conventional thought, it will hold that there are conscious experiences in any spatio-temporally continuous physical system with sufficient complexity to form records and use them to guide behaviour. The only difference is that on the many minds view there are a lot more such systems than conventional thought supposes. What is more, these systems will 'branch' over time, characteristically yielding a multiplicity of later systems where there was just one before.

This way of seeing things hinges on the possibility of viewing the quantum mechanical evolution of any measuring system as a sum of branching histories, the branchings occurring whenever the measuring system interacts with other systems which are in superposition of eigenvectors of appropriate observables. Provided we are dealing with suitable ('decoherent') observables, we can view the measuring system after any such interaction as a sum of (effectively) non-interfering elements, in each of which the measuring system is correlated with an eigenvalue of observable being measured, with these records then remaining relatively stable over time (and therefore meriting the term 'measurement'). Over a sequence of such interactions, the measuring system will therefore evolve into a tree in which each path traced back (from the tip of a branch to the original trunk) will display a history of measured values of the relevant observables.

Let us now consider a human being as such an evolving measuring system. If we then 'look back' along any of the branching histories involved in its evolution, we will find a spatio-temporally continuous physical system with all the properties of a conventionally viewed human body. So, if we apply the conventional 'operational link' between physical and conscious states, we should conclude that this physical system

[1] In my view, there is no good answer because it is a bad question. (Papineau [1993], ch. 4.6.) However, I am in a minority among contemporary philosophers of mind, most of whom think the question is good, but very hard. We are all agreed, at least, that no good answer is currently available.

enjoys just the mental life that conventional thought takes human beings to enjoy. If we then apply the same principle to all the other histories in the relevant branching tree, we get the requisite many minds account of what it would be like for your brain to become a superposition of different measured values of some macroscopic observable, like 'live cat' and 'dead cat', say—namely, that it would be like becoming two disconnected conscious awarenesses, which share memories, one of whom sees a live cat and the other a dead cat.

The point I want to stress here is that this desired conclusion simply falls out of the conventional operational link between complex physical systems and their conscious experiences, once this assumption is conjoined with the many minds view of physical reality. As to the question of *why* there should be such conscious experiences in such physical systems, the many minds theory can simply confess that it has no more of an answer than anybody else.

Note that this way of motivating the many minds account of consciousness yields a natural explanation of why our conscious awareness seems to pick out a 'preferred basis'. As Lockwood stresses, from a purely physical point of view there is nothing special about the basis corresponding to the observables manifested in our conscious experience. However, if we describe the quantum mechanical evolution of a complex system in terms of some different, non-decoherent basis, then it will not be possible to represent this evolution as a sum of 'decoherent histories', since states of the measuring system which are correlated with eigenvalues of the measured system will not generally be stable over time, because of quantum interference effects. Without such stability, there will be no records of past results, no behaviour guided by such records, and so, given the conventional 'operational link' between physical and conscious states, no conscious awareness of such results.

While I take it that the above remarks are largely in the spirit of Lockwood's many minds theory, there is one respect in which I think they diverge. Lockwood draws an analogy between the distribution of a conscious 'Mind' over the different elements of its superposed brain, and the distribution of a (conventionally viewed) mind over different points in time (p. 179). As he points out, we conventionally think of our conscious selves as wholly present at each of the different temporal points in our life histories. Similarly, he suggests, we should think of our Minds as wholly present at each of the superpositional points in our superposed brain. But this seems wrong to me. It is surely part of the identity condition for a conscious self that it display some kind of continuity (either causal or qualitative) in memories. But this continuity will be destroyed by the kind of split in consciousness that the many minds view takes to be occasioned by a superposed brain. Once you have observed the cat, the 'live cat' branch no longer shares memories with the 'dead cat' branch. So it seems that we should deem there to be two selves after the observation, not one Mind that is somehow present in both branches.

True, if there was one self before the observation, and two after, then there is no straightforward sense in which the earlier self can be identical with any later selves. But here the many minds view can simply adopt Derek Parfit's notion of personal survival as an alternative to that of personal identity (see Parfit [1984]). Of course Parfit's account of personal survival was not designed to cater for the strange hypothesis that our conscious selves multiply, like amoebae, every time we observe anything chancy. Still, Parfit's account shows us how to talk coherently about this hypothesis, and I don't see what else argues against it, except its unfamiliarity.

3 Probability

From the point of view of the many minds theory, physical reality contains nothing but the deterministic evolution of the universal wave function. In the special case of observables that are part of the decoherence basis, this function will associate numbers with eigenvalues in a way that (to a very high degree of approximation) satisfies the probability calculus.

Even so, many commentators insist that the many minds view is not entitled to view these numbers as objective probabilities (I shall take 'objective' as read from here on). After all, they point out, many ways of assigning numbers to events satisfy the probability calculus, not all of which therefore represent the genuine probabilities of those events. Why suppose the quantum mechanical coefficients are genuine probabilities? Moreover, they observe, on the many minds view, *all* the different eigenvalues (such as 'dead' and 'alive') associated with non-zero coefficients will actually occur (albeit in different observer-relative branches). So it is hard to see how these coefficients could possibly specify the probabilities of the relevant values occurring, rather than not—the probability of the cat being alive, rather than dead, say.

Despite these considerations, I think that the many minds view should simply take it as given that the relevant coefficients are objective probabilities. Sceptics will continue to insist that this assumption stands in need of further justification. But before they insist on this, they would do well to consider what kind of justification we have for treating certain numbers as probabilities on the conventional view of things.

Let us begin by asking what it is to treat certain numbers as probabilities. There are two operational links between probabilities and non-probabilistic facts that matter here. (1) *The Inferential Link.* We use frequencies to estimate probabilities. If we observe a frequency of p for some type of result R in a finite sequence of trials of type T, then this is evidence that the probability of R in T is close to p. (2) *The Decision–Theoretical Link*. We base rational choices on our knowledge of objective probabilities. In any chancy situation, a rational agent will consider the difference that alternative actions would make to the objective probabilities of desired results, and then opt for that action which maximizes objective expected utility.

Perhaps surprisingly, conventional thought provides no agreed further justification for either of these links. Let us consider them in turn. First, the 'Inferential Link'. *Why* are frequencies evidence for probabilities? The law of large numbers tells us (roughly) that, if we have a long sequence of independent trials on which R has probability p, it is very *probable* that the frequency of R will be close to p. But the underlined mention of probability in this law means that it yields no obvious rationale for the 'inverse' inference that, if the observed frequency is q, then the probability of R is (probably?) close to q. Nor is there any agreed theoretical justification of such inverse inferences. True, there are various alternative attempts to systematize such inferences (Fisherian, Neyman–Pearsonian, Bayesian), but none of these is generally agreed to show us what justifies the inverse inferential move from frequencies to probabilities.

Nor does conventional thought provide any good justification for the 'Decision–Theoretic Link'. Note in this connection that what agents want from their choices are desired *results*, rather than results which are objectively *probable* (a choice that makes the results objectively probable, but unluckily doesn't produce them, doesn't give you what you *want*). This means that there is room to ask: *why* are rational agents well advised to choose actions that make their desired results objectively probable? However, there is no good answer to this question (after all, you can't assume you *will* get what you want if you choose the probabilistically indicated action). Indeed many philosophers in this area now simply take it to be a primitive fact that you ought to weight future possibilities according to known objective probabilities in making rational decisions. In a sense, the 'Decision–Theoretic Link' is even worse off than the 'Inferential Link'. It is not just that philosophers can't agree on the right justification; many have concluded that there simply isn't one.

I said above that the many minds view should simply take it as given that the relevant quantum mechanical coefficients are objective probabilities. We now see that taking a number to be a probability involves treating it in accord with the two 'operational links' just outlined. So my suggestion is that the many minds view should simply stipulate that the quantum mechanical coefficients (1) have their values evidenced by frequencies, and (2) provide a decision–theoretic basis for rational decisions. As to a justification for these stipulations, the many minds theory can simply retort that it provides as good a justification as conventional thought does for treating its probabilities similarly—namely, no good justification at all (c.f. Papineau [1995]).

It is true that the many minds view requires us to think about probabilities in a way we are quite unused to. Normally we think that just one of a set of chancy outcomes will occur, with the probabilities therefore indicating the outcomes' differing prospects of becoming actual. On the many minds view, by contrast, all the outcomes will definitely occur, on some branch of reality, and the probabilities therefore need to be read as attaching weights to these different branches. But it seems to me that this contrast is a 'dangler', which makes no difference to the rest of our thinking about probability. It does not disrupt either of the 'operational links' connecting probabilistic to non-probabilistic facts. And it does not contradict the theories which underlie these links, since we have no such theories.

It might not be obvious that the two operational links involving probability are consistent with all chancy outcomes occurring. Let us consider the two links in turn. The Inferential Link may seem to be threatened by the fact that, in a repeated sequence of trials, all frequencies will be observed (on different branches), not just those that are close to the probability. How then can we infer the probability from the observed frequency? However, note that, even on conventional thinking, it is *possible*, though improbable, that the frequency will diverge from the probability. So conventional thinking already qualifies its advice about inferring the probability from the observed frequency, by admitting that this inference will go astray if you have observed an improbable sample. The many minds view can simply follow suit. That is, it can follow conventional thought and advise that, if you want to know the probability, you should note the frequency and infer that the probability is close—and then hope that you are not the unlucky victim of an improbable sample. As to the justification for this attitude, the many minds theorist is no worse off than conventional thought. For on neither metaphysical view do existing statistical theories offer any cogent rationale for this inverse inference.

The occurrence of all chancy outcomes may also seem to threaten the Decision–Theoretic Link. If every action is sure to be followed by all its possible results, then what does it matter what we do? Any action with the same set of possible results would seem to have the same outcome. But note that conventional thinking could be similarly challenged: since all possible results *may* follow any action, aren't all actions with the same possible results equally good? Of course conventional thinking responds by urging that we should weigh the worth of actions, not just by which results they make *possible*, but by the *probability* of their producing those results.

However, as we saw, conventional thinking offers no further justification for this Decision–Theoretic Link. So again the many minds view can simply follow suit. It can stipulate, without further justification, that (even though all possible results will be actual) actions ought still to be chosen according to the *probability* of desired results.

Let me conclude by contrasting Lockwood's comments about probabilities with the above remarks. Lockwood too is concerned to explain why the relevant quantum mechanical numbers should be viewed as probabilities. His first move is to postulate an infinity of minds within any sentient being, plus a 'natural measure' over subsets of this infinity; he thereby aims to justify ascribing probabilities to different possible states of mind, via the thought that each token experience is randomly sampled from the relevant infinity (pp. 172–3). I don't see that any of this helps (and so see no need for the postulated infinities). For even if we go along as far as the existence of a 'natural measure', we still need to explain why this measure should be considered a *probability* measure. (Talk of 'random sampling' doesn't help, since this presupposes the connection with probability we are trying to explain).

Lockwood recognizes this difficulty on p. 182, at which point he appeals to Albert's and Loewer's version of the many minds view. His thought is that, since he has the same measure on sets of minds as Albert and Loewer do, and since Albert and Loewer are clearly justified in viewing these numbers as probabilities, he must be justified too.

This seems to me a bad strategy for Lockwood. Albert's and Loewer's version of the many minds view is motivated precisely by their conviction that chancy outcomes can't possibly have (non-unitary) probabilities if they are fated to occur in some branch of reality, as Lockwood's view requires. So Albert and Loewer construct a dualistic metaphysics of stochastically evolving minds, to give us chancy outcomes that might or might not occur (will *this* mind see the cat alive, or dead?).

Because of this, Albert and Loewer will deny that Lockwood is entitled to treat his measure as a probability measure just because it matches their measure. From their point of view, he has thrown away just the aspect of their theory that makes the measure a probability measure.

So the more basic issue is whether outcomes which are fated to occur can have non-unitary probabilities. Because of this, I think Lockwood gains nothing by trying to piggy-back on Albert and Loewer. He would do far better to argue directly, as I have done, that the fact that all outcomes with non-zero measure will occur (on some branch of reality) is no reason to deny that this measure is a probability distribution.

I know it flies in the face of common sense to hold that all chancy outcomes occur. Still, I have tried to show that this supposedly obvious truth floats free of anything else we do or think about probability. So, as before, it seems that nothing else argues against the many minds view except unfamiliarity.

Department of Philosophy
King's College London
Strand
London WC2R 2LS
UK

References

Papineau, David [1993]: *Philosophical Naturalism*, Oxford, Basil Blackwell.

Papineau, David [1995]: 'Probabilities and the Many Minds Interpretation of Quantum Mechanics', *Analysis*, 55, 4, pp. 239–46.

Parfit, Derek [1984]: *Reasons and Persons*, Oxford, Oxford University Press.

Comment on Lockwood

Simon Saunders

Michael Lockwood has, with his usual elegance and fluency, laid out a careful and accessible overview of Everett's ideas and some of the ways in which they have been developed, not least by himself. Although there are important areas of disagreement between us, there are a number of essential things on which we agree. The criticisms that follow, then, may give a misleading impression of the debt that I for one owe to Lockwood's writings.

The areas of disagreement are important, however; one concerns the overall aspect of the account, focusing as it does on mentality and the nature of consciousness. The other concerns probability, and in particular the hypothesis, originally formulated by Deutsch, that for each component of the state there exists a *continuous infinity* of physically identical worlds, or, as understood by Lockwood, of identical minds. The same hypothesis, arrived at by slightly different reasoning, is made by Albert and Loewer. In my view Everett's views can be consistently developed without either of these features; mentality *per se* has no more fundamental a role here than in classic physics, and the infinite multiplicities can be dispensed with.

In what follows I will focus on the former claim. I have argued for the latter elsewhere:[1] here I will only say that once Albert's and Loewer's

[1] See my [1995a]. For the parallels with tense and 'passage' in the context of space-time theory, see Saunders [1995b], [1996].