Undermind

Steve Weinstein*

Abstract

David Albert and Barry Loewer have proposed a new interpretation of quantum mechanics which they call the Many Minds interpretation, according to which there are infinitely many minds associated with a given (physical) state of a brain. This interpretation is related to the family of many worlds interpretations insofar as it assumes strictly unitary (Schrödinger) time-evolution of quantum-mechanical systems (no "reduction of the wave-packet"). The Many Minds interpretation itself is principally motivated by an argument which purports to show that the assumption of unitary evolution, along with some common sense assumptions about mental states (specifically, beliefs) leads to a certain non-physicalism, in which there is a many-to-one correspondence between minds and brains. In this paper, I critically examine this motivating argument, and show that it depends on a mistaken assumption regarding the correspondence between projection operators and "yes/no" questions.

1 Introduction

The Many Minds interpretation of quantum mechanics first saw the light of day in a 1988 article by David Albert and Barry Loewer entitled *Interpreting the Many Worlds Interpretation* (Albert and Loewer 1988). Albert discusses it somewhat less formally in his recent book, *Quantum Mechanics and Experience* (Albert 1992). A distinctively new and radical interpretation, it appears to be motivated by a straightforward argument, given most clearly by Albert and Loewer in the 1988 paper. In this paper I will be concerned with examining this argument in detail.

Albert and Loewer begin their article with a discussion of many worlds interpretations, which they see as attempting to reconcile the idea that physical states always time-evolve according to

^{*}Thanks to David Albert, Arthur Fine, Martin Jones, and the two anonymous referees for their comments on an earlier version of this paper, and thanks to Jeremy Butterfield, Rob Clifton, Meir Hemmo and David Malament for comments on a revised version. Special thanks to Arthur for several helpful discussions.

¹Various authors, arguably beginning with Everett (1957), have suggested interpretations which are akin to Many Minds in that they restrict themselves to strictly unitary time-evolution while eschewing talk of "many worlds." Zeh (1970) is an early example; more recently Lockwood (1989), Squires (1990) and Donald (1990, 1992, 1995) have put forth such interpretations.

the Schrödinger equation (no "reduction of the wave packet") with the fact that we don't seem to observe superpositions—the world around us seems macroscopically well-defined. Following a discussion of various problems with many worlds interpretations, they formulate an argument designed to motivate their Many Minds interpretation. The Many Minds interpretation is supposed to do the same work as the many worlds interpretations: it is supposed to explain, for example, how Schrödinger's cat can appear to be dead, or appear to be alive, while really being in a quantum-mechanical superposition of dead and alive. Many Minds has it that minds supervene on certain physical states of the brain, and that different minds can be associated with the same brain (since brains can evolve into superpositions of these states). Furthermore, the Many Minds view is that there are, in general, infinitely many minds associated with a brain, though the number of minds associated with a given element of the superposition of brain states (which constitutes the brain) is proportional to the square of the amplitude associated with the element.

The Many Minds interpretation no doubt seems fairly radical and implausible as I have described it, and even its authors consider their proposal to involve an "extravagant dualism." However, Albert and Loewer feel their argument provides compelling reasons for adopting the Many Minds view. Rather than evaluating the interpretation itself, I would like to look at this argument in some detail.

2 The Argument Examined

The strategy of Albert and Loewer's argument is to derive a contradiction from three suppositions regarding beliefs and their physical realizations. Albert and Loewer purport to show that there is at least one belief, one mental state, which could not be a physical state, and they draw the conclusion that minds could not simply be brains.

Here are the three suppositions:

- (i) A is an observer who can perfectly measure x-spin. (That is: subsequent to an x-spin measurement by A, \uparrow_x iff A believes that \uparrow_x and \downarrow_x iff A believes that \downarrow_x .)
- (ii) A can correctly report some of her mental states. Specifically, when A sincerely reports that she has a definite belief about the value of x-spin (i.e., she reports: "It is either the case that I believe that spin is up, or it is the case that I believe that spin is down, but it is not the case that my beliefs about the value of x-spin are in any way uncertain, ill-defined, or superposed"), then A does believe that x-spin has a definite value.
- (iii) The state wherein A believes \uparrow_x and the state wherein A believes \downarrow_x are identical

with certain physical states of A's brain. We will call these states B_{\uparrow_x} and B_{\downarrow_x} .

Supposition (i) initially seems uncontentious, while supposition (iii) is the supposition of some sort of materialism, whereby there is some sort of one-to-one correlation (perhaps type-type identity) between belief-states and brain-states. The precise nature of this correlation is not specified, but it need not trouble us here.

Supposition (ii) is somewhat obscure. The idea seems to be that there are at least some mental states that A can correctly report. Presumably there are other states she cannot correctly or reliably report: for instance, unconscious mental states. The intended extension of the class of reportable mental states is not apparent. Whatever this class is supposed to include, it at least includes the state of having a definite belief about the value of x-spin. One caveat, however. What A actually reports, according to (ii), is that she either believes spin is up, or believes it is down, but that her belief is not "uncertain, ill-defined, or superposed." This would appear to require that she had some way of determining whether she were in a superposition of belief states. We will shortly return to this point.

Now on to the argument. Albert and Loewer write,

[L]et A measure the x-spin of an electron which is initially in the state $c_1 \uparrow_x + c_2 \downarrow_x$. Subsequent to that measurement, if quantum theory and supposition (i) are correct, the state of A+electron is

$$c_1 B_{\uparrow_x} \otimes \uparrow_x + c_2 B_{\downarrow_x} \otimes \downarrow_x . \tag{8}$$

(Albert and Loewer refer to the state above as "state (8).") This state is arrived at by considering the interaction of A, our measurer and (following supposition (i)) belief-former, with an electron which is in the given superposition of spin states. Given supposition (iii), that beliefs $B_{\uparrow x}$ and $B_{\downarrow x}$ are physical states, and using the standard von Neumann formalism for a measurement interaction, the system A+electron will indeed wind up in this state. This all seems fine.

Here is step (1) of the argument (the numbering is mine):

(1) It follows from supposition (ii) that A will answer the question, "Do you believe that the xspin of the electron has a definite value?" affirmatively whenever either of the states $B_{\uparrow_x} \otimes \uparrow_x$ or $B_{\downarrow_x} \otimes \downarrow_x$ obtains.

This is a salvageable misstep. From (ii) alone, nothing follows, because:

²In the original version of (i), the last three A's are M's. I have taken the liberty of correcting this. Also, in (iii), I have changed $B \uparrow$ and $B \downarrow$ to $B_{\uparrow x}$ and $B_{\downarrow x}$, respectively, for consistency with Albert and Loewer's later notation. All quotations in this paper are from Albert & Loewer 1988, 203-205.

- (a) (ii) is of the form "If A reports 'p', then A believes that p," whereas what is required is the converse, "If A believes that p, then A will report 'p'."
- (b) (ii) says nothing about the correlation between physical and mental states.
- (c) As noted above, (ii) seems to require that A know that her beliefs are superposed. We have been given no reason to believe that she has access to this fact.

That A will give an affirmative answer actually follows from the *converse* of a modified version of (ii), in conjunction with a modified version of (iii). Step (1) can be justified if (ii) is rewritten as a biconditional, deleting the reference to A's knowing that her beliefs are not superposed, and if this rewritten (ii) is augmented with a revised supposition (iii). Here is a new version of (ii) which incorporates the changes:

(ii') A sincerely reports, "It is either the case that I believe that spin is up, or it is the case that I believe that spin is down" if and only if A believes that x-spin has a definite value.

This new supposition will lead to the conclusion that A will answer in the affirmative the question, "Do you believe that the x-spin of the electron has a definite value?", but it (the supposition) will do so only if the states $B_{\uparrow_x} \otimes \uparrow_x$ and $B_{\downarrow_x} \otimes \downarrow_x$ represent states in which "A believes that x-spin has a definite value." In the form in which it is stated, supposition (iii) does not imply that this is the case, because it correlates, e.g., the belief that x-spin is up with the state B_{\uparrow_x} , not with the tensor-product state $B_{\uparrow_x} \otimes \uparrow_x$. The latter state is not a state of a brain, but a state of a brain-and-electron. We can revise (iii) to incorporate such states. The new supposition is,

(iii') States of the form $B_{\uparrow x} \otimes \psi$ (or simply $B_{\uparrow x}$) or $B_{\downarrow x} \otimes \phi$ (or simply $B_{\downarrow x}$) (for any states ψ and ϕ) are states where, respectively, A believes \uparrow_x and A believes \downarrow_x , and these are the only states in which A believes \uparrow_x and A believes \downarrow_x , respectively.

Given (ii') and (iii'), we are now in a position to accede to step (1) of the argument.

The motivation for Albert and Loewer's particular phrasing of step (1) ("A will answer the question... affirmatively") is that in quantum mechanics, experiments are often said to ask "yes/no" questions of a system (e.g., "Is the electron in region R?"), and these questions are represented by special linear operators: namely, projection operators, whose 1 or 0 eigenvalues correspond to the "yes" or "no" answers, respectively (following Mackey 1963). In this light, step (1) seems to be suggesting that we treat the question, "Do you believe that the x-spin of the electron has a definite value?" as being representable by some such projection operator. Let us call this operator Q, where Q is an operator of the form $P \otimes I$ on the brain \otimes electron tensor-product space $H_1 \otimes H_2$

, i.e., it is a projection operator on H_1 , the space in which the states of the brain (e.g., B_{\uparrow_x} and B_{\downarrow_x}) are represented, and it is the identity operator on H_2 , where the spin states of the electron are represented.³ From (ii') and (iii'), it follows that the question will yield a "yes" answer with certainty when A is in a state of the form $B_{\uparrow_x} \otimes \psi$ or $B_{\downarrow_x} \otimes \phi$. Provided one makes the further assumption that asking A the question does not affect her beliefs, this implies that states of the form $B_{\uparrow_x} \otimes \psi$ or $B_{\downarrow_x} \otimes \phi$ are eigenstates of Q with eigenvalue 1 ("yes"). Formally, all this amounts to saying that the action of Q is such that:

$$Q(B_{\uparrow_x} \otimes \psi) = yes(B_{\uparrow_x} \otimes \psi)$$

and

$$Q(B_{\mid_x} \otimes \phi) = yes(B_{\mid_x} \otimes \phi).$$

This more precise and explicit formulation of step (1) is useful in understanding the balance of the argument.

The next step draws on the notion, central to the Many Minds interpretation, that the timeevolution of a quantum-mechanical system is given exclusively by the Schrödinger equation:

(2) And now it follows from the linearity of the quantum mechanical equations of motion, that A will answer this question affirmatively whenever any linear superposition of $B_{\uparrow x} \otimes \uparrow_x$ and $B_{\downarrow x} \otimes \downarrow_x$ obtains.⁴

This is to say that

$$Q(c_1B_{\uparrow_x} + c_2B_{\downarrow_x}) = yes(c_1B_{\uparrow_x} + c_2B_{\downarrow_x}).$$

This follows trivially from the conditions on Q above, in particular from the assumption that Q is linear.

Now step (3):

(3) So, in particular, it follows from (i) and (ii) and quantum mechanics that when state (8) obtains, A will affirm that she believes that the x-spin of the electron has a definite value, either up or down.

Actually, neither supposition (ii) nor (ii') is of any particular relevance to this step. Suppositions (ii) and (ii') correlate belief reports (A's affirmation, in this case) with beliefs, and step (3) says

 $^{^{3}}$ In the service of his relative-state/many-worlds interpretation, Deutsch (1985, 32-37) offers an argument quite similar to that of Albert and Loewer, constructing an operator P that is essentially the operator I am calling Q. (This argument is also to be found in Deutsch (1986), though with a misprinted version of P.)

⁴Typographical error in original article corrected.

nothing about A's beliefs. In order to conclude that A will affirm that spin has a definite value, all one needs is the assumption that Q is a linear operator, and so has the action described in step (2).

Finally, step (4):

(4) But if supposition (iii) is true then when (8) obtains it is not the case that A believes that x-spin is up, and it is not the case that A believes that x-spin is down since A's brain is in a superposition of states.

Albert and Loewer's point is that if beliefs are physical (supposition (iii) or (iii')) then A is not in a definite belief state—rather, she is in a superposition of belief states. The "but" which begins the sentence suggests that Albert and Loewer think that this implies a contradiction with step (3). Indeed, their final claim is:

So we have derived a contradiction from (i), (ii), (iii) and quantum theory.

What exactly is the contradiction? It is this. By (iii'), A does not believe that x-spin has a definite value—she is in a superposition of definite belief states, and this superposition is not one of the states which corresponds to a definite belief, according to (iii'). On the other hand, by the linearity of Q, A will affirm that x-spin has a definite value, and by (ii'), this affirmation implies that she does believe that it has a definite value. One line of reasoning concludes that A doesn't believe x-spin has a definite value, and another line concludes that she does. Thus the contradiction.

3 The Argument Dissolved

The key assumption in Albert & Loewer's argument is the assumption that the question "Do you have a definite belief about the value of x-spin?" is represented by some projection operator (which I refer to as "Q" in this paper). The basis for this particular assumption seems to be a general assumption that any "yes/no" question is representable by a projection operator. Now, the constraints on Q are that it act as a projection operator on H_1 , and that among its eigenvectors having eigenvalue 1 are $B_{\uparrow_x} \otimes \psi$ and $B_{\downarrow_x} \otimes \phi$, since these are states which will definitely produce a "yes" answer. This implies that any superposition of the two states, i.e., any vector of the form $c_1(B_{\uparrow_x} \otimes \psi) + c_2(B_{\downarrow_x} \otimes \phi)$, is also an eigenvector of Q, and will also elicit a "yes" answer, just as Albert and Loewer's argument requires.

In order to see what is wrong with the assumption that any "yes/no" question can be represented by a projection operator, consider the classic two-slit experiment.⁵ An electron propagating from

⁵Those readers unfamiliar with this classic thought-experiment will find an excellent presentation in Feynman, et.al. (1965). The experiment has recently been elegantly realized with electrons by Tonomura, et.al. (1989).

the two-slit diffraction grating to the screen may be prepared in state $|1\rangle$ (corresponding to slit 1 open and slit 2 closed), state $|2\rangle$ (vice-versa), or a superposition of the two (both slits open). One can define an operator R on the space spanned by $|1\rangle$ and $|2\rangle$ by the following prescription:

$$R|1\rangle = "yes" |1\rangle$$

 $R|2\rangle = "yes" |2\rangle$.

If one further specifies that this is a linear operator, then

$$R(c_1 | 1\rangle + c_2 | 2\rangle) = "yes"(c_1 | 1\rangle + c_2 | 2\rangle).$$

One might think that R asks the question, "Did you go through one slit, but not the other?", since one gets a "yes" answer when only one of the two slits is open. But as one can see, the catch is that one does not get a "yes" answer only when one of the slits is open. In fact, the interference pattern obtained when both slits are open precludes the possibility that the electron went through one slit or the other.⁶ So R does not in fact ask "Did you go through one slit, but not the other?" In fact, it doesn't really ask any interesting question at all. The "yes" answer conveys no information. But this is not surprising, since R just acts as the identity operator for all vectors in the space spanned by $|1\rangle$ and $|2\rangle$.⁷

The analogy with Albert & Loewer's operator Q is, I hope, transparent. Albert & Loewer are no more entitled to claim that Q asks the question "Do you believe x-spin has a definite value?" than one would be entitled to claim that R asks the question, "Did you go through one slit, but not the other?" The argument therefore fails, because the fact that state (8) is an eigenstate of Q does not mean that A has answered the question "Do you believe x-spin has a definite value?" in the affirmative, since Q doesn't represent that question.

The misstep in Albert & Loewer's argument is their assumption that all "yes/no" questions are represented by projection operators. What one is entitled to assume is the converse: all projection operators represent "yes/no" questions. Their assumption is simply false, as the two-slit example illustrates.

⁶In this discussion I am bracketing the de Broglie-Bohm, realist interpretation of quantum mechanics, in which the particle *does* in fact go through one slit or the other. See Cushing (1994) for an up-to-date presentation of this intriguing outlook.

⁷In the same way, Q simply acts as the identity operator on any vector in the subspace spanned by $B_{\uparrow_x} \otimes \psi$ and $B_{\downarrow_x} \otimes \phi$, i.e., on any superposition of the form $c_1(B_{\uparrow_x} \otimes \psi) + c_2(B_{\downarrow_x} \otimes \phi)$.

4 The Argument Resurrected?

I have claimed that one in effect cannot write down an operator which represents the question Albert & Loewer want to ask. But here is a way out one might try. One can certainly ask the question in the sense that one can utter the words. The physical interaction between this utterance and a system in some superposition of $B_{\uparrow x} \otimes \uparrow_x$ and $B_{\downarrow x} \otimes \downarrow_x$ should, in principle, be representable by some Hamiltonian operator H (though this operator will of course be unimaginably complex in its full glory). Furthermore, it does not seem to be making too strong a claim to say that a system in, for instance, $B_{\uparrow x} \otimes \uparrow_x$, will answer the question with a "yes" and remain in that state, the state (in this instance) of believing x-spin to be up. Why not make the following claims, then:

$$(B_{\uparrow_x} \otimes \uparrow_x \otimes blank) \xrightarrow{H} (B_{\uparrow_x} \otimes \uparrow_x \otimes yes)$$
 (a)

$$(B_{\downarrow_x} \otimes \downarrow_x \otimes blank) \xrightarrow{H} (B_{\downarrow_x} \otimes \downarrow_x \otimes yes),$$
 (b)

where "blank" and "yes" represent, say, the states of a sheet of paper designed to register the answer to the question? Such a formulation implies that

$$[c_1(B_{\uparrow_x} \otimes \uparrow_x) + c_2(B_{\downarrow_x} \otimes \downarrow_x)] \otimes blank \xrightarrow{H} [c_1(B_{\uparrow_x} \otimes \uparrow_x) + c_2(B_{\downarrow_x} \otimes \downarrow_x)] \otimes yes .$$
 (c)

This in turn suggests that A will answer "yes" even when she is in a superposition of belief states. There are two problems with such a reformulation of Albert & Loewer's argument. The first is that it assumes there is a *unique* (up to phase) "yes" state into which both the "believe-up" system and the "believe-down" system evolve. Without this assumption, (a) and (b) should be rewritten as

$$(B_{\uparrow_x} \otimes \uparrow_x \otimes blank) \xrightarrow{H} (B_{\uparrow_x} \otimes \uparrow_x \otimes yes_{\uparrow})$$
 (a')

$$(B_{\downarrow_x} \otimes \downarrow_x \otimes blank) \xrightarrow{H} (B_{\downarrow_x} \otimes \downarrow_x \otimes yes_{\downarrow}).$$
 (b')

This implies that the Hamiltonian (representing the question) will evolve a system in a superposition of "believe-up" and "believe-down" states into a system containing a superposition of yes_{\uparrow} and yes_{\downarrow} . However, this is no good for Albert & Loewer's argument, because the two states can destructively interfere. A simple illustration of this sort of interference is provided by the two–slit experiment, with the states $|1\rangle$ and $|2\rangle$ analogous to yes_{\uparrow} and yes_{\downarrow} . Suppose we are firing a beam of particles through the slits, and suppose that the analogue of a "yes" answer is the detection of one or more of the particles at some point near the center of the screen. If one and only one slit is open, then

there is a non-zero probability for finding a particle in the neighborhood of that point (no matter which point we have selected, and no matter which slit is open). However, if the particles are in a superposition of the two states, and the point we have selected is an interference minimum, then there will be a severely diminished (almost zero) probability of detecting a particle in the neighborhood of that point.⁸ Returning to the Albert & Loewer example, the interference of the two "yes" states gives rise to the possibility of a non-"yes" answer.

The only way for Albert & Loewer's argument to work is if yes_{\uparrow} and yes_{\downarrow} are identical up to a phase factor. But there is absolutely no reason to think that yes_{\uparrow} and yes_{\downarrow} are related in this way. Even if one were to suppose that the "yes" responses generated by observers who "believe-up" and observes who "believe-down" are classically identical, this would be insufficient, since there is no unique map from classical states to quantum states. I conclude that this approach to resurrecting the argument is a dead-end.

5 Many Minds

The argument offered by Albert & Loewer is intended to show how the demand for strictly unitary, Schrödinger time-evolution forces one to something like the Many Minds interpretation. The argument itself fails, relying as it does on a mistaken assumption regarding the relationship between "yes/no" questions and projection operators, but this does not mean that the interpretation itself is untenable—it only means that it must find motivation by another path. In the sections of their article I have ignored, Albert & Loewer point out several important problems with other interpretations. Clearly something else is needed, and perhaps that something else is the Many Minds interpretation. It might be interesting to see how Many Minds could be enriched, and perhaps more soundly motivated, by some of the recent work in decoherent histories, work which attempts to provide a mathematical framework for quantum mechanics with strictly unitary evolution.¹¹

⁸The probability is *almost* zero because one can only assign probabilities to finite regions of the screen, and the wave-function is only zero at one point (the interference minimum). Those who are bothered by this should note that one can construct a similar example with a Mach-Zehnder interferometer which avoids this subtlety.

⁹By "classically identical," I simply mean that the classical descriptions of the two "yes" responses are identical. For example, if a "yes" response is physically instantiated by writing "yes" on a piece of paper, a classically identical "yes" response would be an identical configuration of ink on an identical piece of paper. This seems like an absurdly stringent requirement, but it is necessary, though still not sufficient, for this argument to work.

¹⁰To see this, consider the problem of writing down "the" quantum state corresponding to a particle in onedimension at rest at the origin. One might approximate this by some gaussian $\exp(-x^2/\Delta)$, since this is peaked around the origin, and has an average momentum of zero associated with it. But why a gaussian? And if a gaussian, how does one pick the width Δ ? And what is to prevent one from multiplying by some phase factor $\exp(icx)$ (where c is a constant)? Such a phase factor will not affect the probability of finding the particle in a given region (though it will affect the probability distributions of some other observables).

¹¹See Halliwell (1995) for an excellent survey of work on decoherent histories. Page (1995) is an attempt to construct a Many Minds-like interpretation which utilizes this approach.

References

- [1] Albert, D. and Loewer, B.: 1988, 'Interpreting the Many Worlds Interpretation,' Synthese 77, 195–213.
- [2] Albert, D.: 1992, Quantum Mechanics and Experience, Harvard University Press, Cambridge.
- [3] Cushing, J.: 1994, Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony, University of Chicago, Chicago.
- [4] Deutsch, D: 1985, 'Quantum Theory as a Universal Physical Theory,' Int. J. Theor. Phys. 24, 1-41.
- [5] Deutsch, D: 1986, 'Everett's Interpretation and Experiment,' in *Quantum Concepts in Space and Time*, ed. R. Penrose and C.J. Isham, Clarendon, Oxford.
- [6] Donald, M.J.: 1990, 'Quantum Theory and the Brain,' Proc. R. Soc. London A427, 43-93.
- [7] Donald, M.J.: 1992, 'A Priori Probability and Localized Observers,' Found. Phys. 22, 1111-1172.
- [8] Donald, M.J.: 1995, 'A Mathematical Characterization of the Physical Structure of Observers,' Found. Phys. 25, 529-571.
- [9] Everett, H., III: 1957, 'Relative State Formulation of Quantum Mechanics,' Rev. Mod. Phys. 29, 454-462.
- [10] Feynman, R.P., Leighton, R.B., and Sands, M.: 1965, *The Feynman Lectures on Physics*, Volume III, Addison-Wesley, Menlo Park.
- [11] Halliwell, J.J.: 1995, 'A Review of the Decoherent Histories Approach to Quantum Mechanics,' in *Fundamental Problems in Quantum Theory*, ed. D. Greenberger, Annals of the New York Academy of Sciences, New York.
- [12] Lockwood, M.: 1989, Mind, Brain and the Quantum, Blackwell, Oxford.
- [13] Mackey, G. W.: 1963, Mathematical Foundations of Quantum Mechanics, Benjamin, New York.

- [14] Page, D.N: 1995, 'Sensible Quantum Mechanics: Are Only Perceptions Probabilistic?,' Univ. of Alberta report Alberta-Thy-05-95, LANL preprint quant-ph/9506010.
- [15] Squires, E.J.: 1990, Conscious Mind in the Physical World, Adam Hilger, Bristol.
- [16] Tonomura, A., Endo, J., Matsuda, T., Kawasaki, T., and Ezawa, H.: 1989, 'Demonstration of single-electron buildup of an interference pattern,' Am. J. Phys. 57, 117-120.
- [17] Zeh, H.D.: 1970, 'On the Interpretation of Measurement in Quantum Theory,' Found. Phys. 1, 69-76.