



'Many Minds' Interpretations of Quantum Mechanics: Replies to Replies

Michael Lockwood

The British Journal for the Philosophy of Science, Vol. 47, No. 3 (Sep., 1996), 445-461.

Stable URL:

<http://links.jstor.org/sici?sici=0007-0882%28199609%2947%3A3%3C445%3A%27MIOQM%3E2.0.CO%3B2-9>

The British Journal for the Philosophy of Science is currently published by Oxford University Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/oup.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact jstor-info@umich.edu.

DISCUSSION

‘Many Minds’ Interpretations of Quantum Mechanics: Replies to Replies

Michael Lockwood

Let me begin by saying how grateful I am to the commentators on my article,¹ who have given me much food for thought, and especially to David Papineau, who made this whole exercise possible. It has forced me to confront important issues which, as I now realise, I have skated over rather too lightly in previous publications.

Loewer

The bulk of Loewer’s article² consists of an admirably lucid and succinct account of the differences between our respective views, with which I essentially concur. Much of his discussion is concerned with probability; and it is that which provides the basis for his main objection which appears in his final paragraph.

Loewer’s argument is best understood in the context of a specific example. Suppose that I find myself with a recollection of having carried out 100 successive measurements of spin, in the x direction, on electrons prepared in the state z spin-up, and having got the results spin-up and spin-down in a ratio of, say, 53 : 47. Were I to favour an interpretation of quantum mechanics according to which quantum measurement involves some sort of stochastic evolution—whether of the observed system, as in dynamical collapse theories, or of my mind, as in Albert’s and Loewer’s view—then I could regard my recollection as confirming quantum mechanics, on this interpretation. The same would be true, were I to favour a hidden variable theory, according to which the square moduli of the coefficients, when the prior state is expanded in the basis of the measured observable, are associated with randomly distributed variables whose actual values measurements reveal. For in either case these square moduli would correspond to probabilities as ordinarily understood. And

¹ Lockwood [1996].

² Loewer [1996].

quantum mechanics, under both types of interpretation, predicts a probability of 0.807 that the difference between the number of measurements yielding spin-up and the number yielding spin-down, here, will be fourteen or less. My present recollection of having got a difference of six is thus well in line with this prediction.

I accept none of these interpretations, however. And Loewer maintains that, given the interpretation I *do* favour, I am not entitled to regard my recollections as confirming quantum mechanics, so interpreted. The reason, according to him, is that the measures on sets of instantaneous minds, which I associate with the square moduli, are simply *not* probabilities. Hence, my current recollection of having got the results spin-up and spin-down in a ratio of 53 : 47 has no tendency whatsoever to confirm the prediction, by my theory, of a measure of 0.807 for the set of instantaneous minds which have recollections of having found a difference of fourteen or less between the number of measurements yielding spin-up and the number yielding spin-down. The upshot, Loewer argues, is that although my interpretation *might* be true I can have no good reason for believing it.

Loewer's argument would be valid, it seems to me, if the measure, here, simply had *nothing to do* with probability, as ordinarily conceived. But that is far from being the case. 'How are we to understand this measure?' Loewer asks (p. 230), 'What does the measure measure?' My answer is that the measures postulated in my theory measure something physical: specifically, those features of the Mind in virtue of which our experience of quantum measurement has the *phenomenologically* probabilistic character that it does. In short, they are what explain the *appearance* of stochastic evolution. Loewer is quite right to insist that they are not probability measures either in the sense of probability that is associated with a stochastic law, or in the *epistemic* sense, associated with limited ignorance. And he is correct, also, in pointing out that they do not satisfy his conditions (1)–(3). Nevertheless, the characterization offered in my article seems to me to link these measures to our ordinary concept of probability in a perfectly intelligible fashion, such that they could reasonably be regarded as probability measures. (Why not think of probability, in the context of quantum measurement, as whatever in Nature is responsible for the corresponding probabilistic *appearances*, without prejudice to the question of what this might ultimately turn out to be?) Nothing of substance, however, hinges on what we *call* these measures. And once this is appreciated, a curious hiatus is revealed in Loewer's argument. For what Loewer needs to establish, but, strangely, does not even try to establish, is that failure, on the part of my measures, to satisfy his conditions (1)–(3) is sufficient to block the statistical reasoning whereby I take my recollections to confirm the theory.

Not only does Loewer in fact make no attempt to show this; were he to do so, he would surely fail. Loewer's claim is that the *experiential* equivalence between our respective interpretations does not entail *evidential* equivalence. But how can that be? For, given the experiential equivalence, any set of recollections which Albert and Loewer would be entitled to regard as confirming their interpretation would *a fortiori* be just the kind of thing I would expect on my interpretation. My interpretation, in effect, predicts that experience will be such as to confirm Albert and Loewer's interpretation; and it is precisely *because* it predicts this that I am *pragmatically* justified in applying the standard statistical reasoning. Indeed, I can go further. Loewer concedes (p. 229) that if my interpretation were 'empirically adequate', then I would be 'right to prefer' it to the Albert and Loewer view, 'since its metaphysical commitments are far less problematic'. Quite so. Given, as I have been arguing, that my interpretation *is*, after all, empirically adequate, and hence evidentially equivalent to Albert's and Loewer's, it follows that it is *their* interpretation, not mine, which has the feature that although it might be true, one could have no good reason for believing it. For any experiences which confirmed it would, by the same token, confirm mine, which, as Loewer acknowledges, is then to be preferred on philosophical grounds.

Saunders

At the beginning of his stimulating and wide-ranging essay,³ Saunders identifies two areas of disagreement between us. One, which he has dealt with elsewhere, has to do with probability. The other, on which he mainly concentrates here, concerns my overall approach, focusing, as it does, on the relationship between consciousness and the quantum state description. Now I agree that, in this regard, our respective approaches differ markedly in methodology and emphasis. In my play, consciousness is a leading character from the outset. In Saunders' play, by contrast, consciousness appears, if at all, only in an optional epilogue. What is less clear to me, however, is how far these differences should be taken to imply disagreement on matters of philosophical substance.

Anyone who wishes to give a realist interpretation of quantum mechanics, while repudiating dynamical collapse and hidden variables, is faced with the problem of reconciling ordinary appearances with quantum reality, thus construed. Two strategies may then suggest themselves. One is to look for something 'out there' in the external world, as described by quantum mechanics, which for the purpose of explaining the appearances,

³ Saunders [1996].

sufficiently resembles physical reality as common sense conceives it, to serve as a surrogate for that common-sense reality. And a promising candidate, as Saunders points out, is the notion of a *consistent history* within a *quasi-classical domain*, such as emerges from recent work on decoherence. The idea then (of which Saunders has become an eloquent and persuasive advocate) is that it is in just such a quasi-classical domain that we effectively ‘live and move and have our being’. And although, from a God’s eye view, the corresponding consistent histories are only subplots within the overall story associated with the evolution of the universal wave-function, we are nevertheless perfectly entitled to regard as real any given such history in which we find ourselves caught up. Indeed, Saunders thinks that when, in common parlance, we talk about what ‘really happens’ or is ‘actually the case’, it is to specific such histories (or states *qua* components of such histories) that we are implicitly referring. Saunders calls these *relative realities*, where ‘relative’, here, is to be understood in a sense which is analogous to Everett’s use of the term when he speaks of *relative states*. All ordinary uses of the terms ‘real’ or ‘actual’ correspond, he thinks, to relative realities in his sense. And what they are relativized to are states of ourselves considered as ‘concrete physical forms’, capable of high-level information processing, storage, and exchange, and of forming ‘epistemic communities’ with shared views of the world. Nowhere is there any need to appeal to consciousness. And (to a fair approximation) a preferred basis will arise naturally, within the quasi-classical domain, as the basis which, under coarse-graining, diagonalizes the decohering variables.

On the alternative strategy, adopted in my own writings, one starts by assuming that there is some subsystem of the brain, which I call the Mind, on whose states consciousness directly supervenes. Then one explains the appearances by way of a hypothesis about *what it is like to be* in those *mixed* states of the Mind, which quantum mechanics tells us would be generated by the interactions with surrounding systems that underlie sensory perception. This hypothesis involves the postulation of a basis, for the Mind, which, though not necessarily *physically* preferred, is nevertheless preferred from the standpoint of consciousness, in that it defines both the way we appear to ourselves, and the way we view our surroundings.

As I see it, these two strategies should be regarded as complementary. Saunders, in effect, starts with the world and works inwards towards the observer. I, by contrast, start with the observer, and work outwards towards the world, by way of the quantum correlations which I assume that the mechanisms of perception establish between basis states of consciousness and states of external objects. Convergence between these two

approaches would be achieved at a stroke, if it could be established, for example, that the consciousness basis of the Mind just is its decoherence basis. But I regard it as a virtue of my account that it does not require me to hitch my wagon to any specific scientific programme, and hence allows me to remain uncommitted on the question of what the consciousness basis amounts to, in purely physical terms. I prefer to start with the mind, not because I have idealist leanings, in contrast to Saunders' realist ones, but because I take it that no interpretation of quantum mechanics can really claim to have accounted for the appearances, unless it is, at least tacitly, making some assumptions about the way in which consciousness maps on to, or fits into, the physical world as quantum mechanics describes it. I wish to make these assumptions explicit at the outset, and put them to work.

Not only do I, in fact, take a robustly realist view of quantum mechanics. As a realist, I have serious reservations about Saunders' *relativized* conception of reality, which strikes me as a misguided attempt to reinterpret common sense, in such a way that Everett's momentous insight no longer contradicts it. Saunders and I both appeal, in our writings, to the analogy between the impact of the Everett interpretation of quantum mechanics on our concept of reality, and that of the 'block universe' interpretation of relativity on our common-sense concept of time. But I see both an analogy and a disanalogy here. The analogy is this. Just as it is a mistake, from the standpoint of the block universe conception, to regard an event one thinks of as happening now as possessing some intrinsic attribute of 'presentness', which is denied to events occurring at other times, so it is a mistake, from the standpoint of the Everett interpretation, to regard that outcome of a quantum measurement which one thinks of as actual as possessing some intrinsic attribute of 'actuality', which is denied to the other possible outcomes. Where, then, is the disanalogy? Well, it is clearly a mistaken account of the block universe conception to say that it represents *all* events as being equally *present*: it emphatically does not say that all these events are happening *now*. But by contrast, I judge it to be a *correct* account of the Everett interpretation, to say that it represents *all* outcomes of a quantum measurement as being equally real. With his talk of relativized reality, Saunders seems to me to obscure the true philosophical message of Everett's work, which is that there is a hitherto unsuspected *perspectival* dimension to language, thought, and perception.

In recognition of this, one ought, ideally, to coin a new word: one which stands to distinct terms of the state vector as 'now' stands to distinct times. But *faute de mieux*, I can see the pragmatic value of letting such words as 'real' and 'actual' serve this role. This would mean using them, in ordinary contexts, as *indexicals*: in a way, that is, which makes their reference a function of the points of view from which statements containing them are

made. Indeed, that, and the more familiar forms of indexicality, such as are involved, for example, in tenses and the use of the first person, all have a natural interpretation in terms of what I call the *experiential manifold*. ‘I’ is indexed to a specific such manifold, and hence conscious subject; ‘now’ is indexed to positions on the vertical dimension of this manifold; ‘real’ and ‘actual’, used as just suggested, are indexed to positions on the horizontal (or superpositional) dimension; and finally, ‘here’ is a derived term, definable as where *I really* am (or my body really is) *now*. That way of using ‘real’ (and cognate expressions) would be in outward conformity both with what Saunders recommends, and with what people ordinarily say. But in *outward* conformity only. For I take both common sense and Saunders to be in error here. Common sense, as I see it, is mistaken in taking outcomes to be unique; and Saunders is mistaken in taking reality—the real McCoy—to be merely ‘relative to ourselves’. Indeed, were we to use ‘real’ in the indexical way just proposed, we should then need another, nonindexical term to do duty for what we currently understand by ‘real’. (As regards the written word, ‘Real’ could perhaps, by analogy with ‘Mind’, be made to occupy this particular semantic niche.)

To end, however, on a note of agreement, I entirely share Saunders’ scepticism about the possibility of developing relativistically covariant dynamical collapse or hidden-variable interpretations of quantum mechanics, or of providing a satisfactory interpretation of relativistic quantum field theory, in anything other than the Everett framework. In fact, as regards hidden variable theories, it now appears that, thanks to Hardy [1992], we have a proof that no such theory which reproduced the predictions of conventional quantum mechanics could possibly be Lorentz invariant. This supports what I have thought for as long as I have been attracted to the Everett interpretation: that from beyond the grave Einstein is effectively telling us that Everett’s is the only viable approach.

Papineau

I don’t know how reassured I should be by Papineau’s claim⁴ that my views on consciousness and probability are in no *worse* shape than their more conventional rivals, given that he thinks that these, in their turn, are in *terrible* shape! Nevertheless, I found Papineau’s summary of the current state of play in the foundations of probability extremely illuminating—and in certain respects, very helpful to my cause. For one thing, his own characterization of probability, by contrast with Loewer’s, allows the things *I* call probabilities genuinely to count as such. Whilst they fail to

⁴ Papineau [1996].

satisfy Loewer's principles (1)–(3), I am safe in claiming, I think, that my probability measures will support Papineau's Inferential Link and Decision-Theoretic Link, if anything will.

I believe, in fact, that I can give an *argument* as to why these measures do indeed support the Decision-Theoretic Link. Papineau asks: '*why* are rational agents well-advised to choose actions that make their desired results objectively probable?' Given my concept of the *successors* of a given instantaneous mind, an answer is immediately forthcoming, in respect of quantum-mechanical probabilities: 'Because choosing those actions which maximise the *expected* return, means maximising the total *actual* return, as integrated over the *successors* of whatever instantaneous mind is making the decision.' On the assumption that it is rational for an instantaneous mind to care equally about all its successors (to the extent, at least, that they are equally proximate in time), this is analogous to preferring the longer to the shorter of two alternative pleasures, when they are of equal intensity. Indeed, in my terms, both forms of reasoning involve integrating over the relevant part of the *experiential manifold*. Imagine that I am wired up to a pleasure machine, capable of administering some pleasurable sensation, of a fixed intensity, for different durations, and that these different durations are associated, respectively, with distinct outcomes of a quantum measurement. Suppose that, in advance of being connected up to the machine, I *qua* instantaneous mind, wish to calculate *how much* pleasure, overall, lies ahead of me: the aggregate pleasure, in other words, which the machine will confer on the successors of *this* instantaneous mind. In terms of my measures, the correct way to calculate this is to take, for each possible measurement outcome, the product of the measure corresponding to its square modulus, and the associated duration, and sum the resulting values. What this calculation yields is the total *area* of that region within my experiential manifold, in which the sensation in question occurs.

Note that the above explanation of the Decision-Theoretic Link would be unavailable to me, were I to follow Papineau's recommendation that I posit just a single successor for each distinct quantum outcome. Papineau, no doubt, will point out that I can get the same *results* by replacing measures on the experiential manifold by weighted sums over branches. But I question whether it makes any sense to construe the weights, here, as expressing different probabilities, given that the branches are in one-to-one correspondence with outcomes all of which will actually occur. Indeed, I would stand by what I said in my article: it seems to me that if the underlying reality conformed to this model, the phenomenological probabilities associated with different outcomes of a quantum measurement would all be *equal*. In his own reply to my article, Butterfield (if I may

respond to this point here) objects that I am illicitly appealing to the discredited *principle of indifference*. But I take it that the reason *why* this principle is to be rejected in most contexts where we are in the business of trying to estimate probabilities, is that for the purposes of deploying the principle, there is no preferred way of partitioning the space of possible outcomes. That is a legitimate objection, certainly, *where it applies*. But it surely has no purchase in the present situation, where it is stipulated that the history of a mind, beyond a certain point, has just n discrete continuations, all of which are actual. For there is no freedom, here, to partition this n -fold continuation in any way other than that stipulated.

Regardless, however, of whether I were to stick with my original proposal, or were to adopt, instead, the 'stripped down' version of the theory which Papineau prefers, Papineau's remarks about the connection with Parfit's work on personal identity⁵ would still apply. For my *succession* relation, as it links instantaneous minds (or maximal experiences), creates a structure of branching *biographies*, similar to those which arise in the sort of split-brain and teleportation cases which Parfit discusses, and to which he applies his concept of *survival*. Like Parfit's concept, my succession relation resembles personal identity, except that it allows a mind to have simultaneous multiple successors, or 'survivors', which are not identical with each other.

Finally, it's a misunderstanding of my account of probability in quantum mechanics to see it as trying to 'piggy-back' on Albert's and Loewer's. Albert's and Loewer's theory is useful to me only as a device for fending off the objection that I cannot intelligibly associate, with continuously infinite sets of simultaneous maximal experiences, measures which will be experienced as probabilities. The point is simply this. It cannot be a *logical* prerequisite of such continuous sets of experiences being associated with the appropriate probabilistic phenomenology, that they be generated by any *specific* causal mechanism. But on the other hand, were these sets to be generated by an irreducibly stochastic process of the kind that Albert and Loewer envisage, they clearly *would* be associated with the appropriate phenomenology. So the intelligibility of my original claim is vindicated.

Deutsch

With Deutsch's spirited and incisive defence of what is often referred to as the 'many worlds' or 'many universes' interpretation of quantum mechanics,⁶ I have few, if any, disagreements of substance. Deutsch chides me, however, for my reluctance to adopt this way of talking.

⁵ See Parfit ([1984], Part III). ⁶ Deutsch [1996].

Let's be clear at the outset. The conception which we both favour is the one which results from taking quantum mechanics (without the projection postulate) *completely at face value*—with no extraneous elements added, and with its counterintuitive features squarely faced up to, rather than being excised, ignored, or enveloped in a fog of evasive double talk. It is not, therefore, out of any desire to cushion the shocking impact of Everett's message that I prefer (in many contexts) to avoid 'worlds' talk. Rather, it is because it offends, somewhat, against my philosopher's predilection for using language in as precise a way as possible, and because experience has taught me that, especially with non-physicists, it tends to create unnecessary barriers to understanding. Indeed, these two things are connected. If I want to explain to someone the rationale of my views, I usually start with the electron double-slit experiment. Most of the people I talk to can accept that, in this experiment, the electron goes through both slits. Using this to introduce the notion of a *superposition*, I put it to them that little things, like electrons, and big things, like electric kettles or electricians, should really be expected to obey the same physical laws, and that the predictive success of quantum mechanics in the micro-world creates a presumption, therefore, that big things, including ourselves, ought also to go into superpositions. And I point out that, since we wouldn't be aware of so doing if superpositions of brain states generated distinct parallel streams of consciousness, our actual experience doesn't contradict this suggestion. By this means, I can often get people to see that it is, in fact, *more economical* to suppose this than to suppose that some *deus ex machina* comes into play, as we move from the micro- to the macro-world, to make big things behave roughly as classical physics says they do. I then point out, finally, that this reasoning can be applied to the universe as a whole; and that what we think of as *the* world actually corresponds merely to one term of the universal superposition. That, I say, is what people *mean* when they talk about 'many worlds'. But I then add that there's really no *more* reason for talking about a multiplicity of worlds or universes, here, than there is for talking about a multiplicity of electrons, in the double-slit experiment. In each case, it would be just a picturesque way of talking about a superposition state of a single thing. I hope that Deutsch will think that this is an honest way of presenting my case; I can assure him that it is the most effective strategy I know. Indeed, the hardest part of my task, more often than not, is that of overcoming the prior resistance created by the phrase 'many worlds', with its apparent associations of ontological profligacy gone mad.

Having said all that, I frequently *do* speak in terms of 'worlds' or 'universes' when talking to people, such as Deutsch himself, for whom the explanations, qualifications, and caveats which Deutsch sets out in his

article can be taken as read. So I certainly wouldn't regard myself as observing any 'taboo', in this respect. Indeed, I freely acknowledge that, for the reasons Deutsch gives, this is often a very convenient way of talking. Moreover, as Deutsch himself has elsewhere pointed out, Ockhamist objections to many worlds are, in any case, largely misguided. For it is not so much to the number of postulated *entities* as to the number of mutually independent *assumptions* (and their individual complexity) that the principle of parsimony, in scientific and philosophical theorizing, properly applies.

One final remark—a point impressed upon me by Euan Squires. It seems to me that Deutsch is a little unfair to the proponents of rival views, even if technically correct, in denying that they so much as *count* as rival *interpretations* of quantum mechanics. Doubtless, this is good 'fighting talk'. But whether they qualify as rival interpretations of the same theory, or rather as alternative theories, is clearly not the real issue. The question is: what, if anything, do they have going for them? However we classify them (as I'm sure Deutsch would agree) they deserve to be judged on their merits.

Brown

Brown's aim, in his magisterial survey of the issues,⁷ is not so much to present criticisms of my views, as to set them within a wider context. The relative brevity of my response hence reflects my almost complete agreement with his very perceptive observations. I would, in particular, strongly endorse his remark (p. 190) that 'What is, or should be, important... in Everett's scheme are the instantaneous experiences and memories in each of Alice's minds, and not a *specific* simultaneous relative state of the world outside her brain.' Indeed, the concept of a relative state does not even figure in my own key interpretative hypotheses (I)–(III). The main work, here, is really being done by certain assumptions concerning *psychophysical parallelism*. In my account, these are set out explicitly. But as Brown rightly observes, some such assumptions must be understood as being at least *implicit* in Everett's approach, on pain of his having failed (as Bell⁸ alleged) to *interpret* the formalism of quantum mechanics at all.

The closest Brown comes to a criticism of my views is in his suggestion that developments in quantum gravity and quantum cosmology might require me to adopt a more *radical* metaphysical perspective than that defended in my article: a perspective, perhaps, such as that advocated by Barbour.⁹ Be that as it may. I do not see such developments as invalidating

⁷ Brown [1996].

⁸ Bell ([1981], fn. 9).

⁹ See references in Brown [1996].

my position *qua interpretation of elementary quantum mechanics*. If it is to be taken seriously, any theory of quantum gravity must be capable, in some appropriate limit, of yielding *effective* subsystem and Hilbert space structure, and an *effective* Schrödinger equation. This limiting regime defines a corresponding *level of explanation*; and that, accordingly, is the intended level of *application* of my own proposed interpretation of quantum mechanics.

Nothing I say, therefore, is intended to rule out the possibility of a *deeper* level of explanation at which, for example, the distinction made in my article between the vertical and horizontal dimensions of what I call the *experiential manifold* may cease to be fundamental. So striking, indeed, is the analogy between a multiplicity of instantaneous actualities, associated with different times, and a multiplicity of instantaneous actualities, associated with different terms of a quantum superposition (a dimension of variation which Brown dubs *Everett time*), that it is tempting to wonder whether this may not be more than just an analogy. Maybe the former is somehow a *special case* of the latter.¹⁰ Perhaps, as Brown puts it (p. 197), 'Everett time is the only fundamental time there is'. Such a view follows very naturally from developments in *canonical* quantum gravity, where time, as an external parameter, drops out of the picture altogether, and is replaced by an *internal* time, defined by reference to some subsystem of the universe which is designated as a clock. The temporal ordering of states in general—and of states of Mind and their associated maximal experiences in particular—then becomes a matter of their being correlated with certain clock states or (where the clock is effectively classical) or their being functions of the relevant degree of freedom.

It is, however, a further step to follow Barbour in denying the existence of any objective time order in Nature beyond that which can be *read into* a set of instantaneous configurations on the basis of their contents alone (rather as one might reconstruct a film out of a jumble of individual frames). Some may find this position philosophically appealing, entailing, as it does, a conception of time very similar to that which Leibniz adopted in respect of space. But Barbour has yet to persuade me that there is anything in current physics which makes it remotely compulsory.

Butterfield

Such is the range and depth of the issues which Butterfield raises in his commentary,¹¹ that it is difficult, here, to do full justice to them. Perhaps I

¹⁰ See Lockwood ([1989], pp. 188–90).

¹¹ Butterfield [1996].

should start, in relatively uncontroversial vein, by amending Butterfield's brief history of many minds interpretations. H. Dieter Zeh¹² has clear priority over all the authors he mentions, defending, as he does, an Everett-inspired 'multi-consciousnesses interpretation' in a 1981 article (around the same time, incidentally, that I began to develop my own version of the theory and to present it in public lectures). Indeed, the idea is implicit in other articles by Zeh dating back to 1970. (I'm indebted to Euan Squires for bringing Zeh's articles to my attention.)

Butterfield begins with a clearly formulated summary of my position which is broadly on target, but nevertheless prompts me to correct a couple of apparent misconceptions. First, although I mentioned it in passing (while trying to clarify what many worlds theorists should be understood to mean by 'world'), the concept of a *simultaneous* consciousness basis, available though it is,¹³ plays no role in my theory. That is to say, I make no use of the concept of a basis comprising N-fold tensor products of respective elements of the consciousness bases of N distinct sentient beings. My own starting point is, in a sense, a Cartesian one: I'm endeavouring to make sense of *these* experiences of *mine*, on the assumption that I inhabit a quantum-mechanical universe. In that sense it really is *my* consciousness basis that I'm talking about, in the first instance. But I assume, of course, that every sentient being has its own individual consciousness basis, to which exactly the same principles apply. Thus, when I talk, impersonally, of *the* consciousness basis, I'm not referring to the simultaneous consciousness basis; rather, this is a way of expressing generality—just as an entomologist might speak of the life cycle of the tsetse fly.

From what I've just said, it will be clear that I need not, *qua* philosopher of physics, agonize over the question which beings, precisely, should be credited with consciousness bases: it suffices, for present purposes, that *I* should be and so should my readers. Nevertheless, Butterfield gives a misleading impression of my views when he says of me (p. 203): 'he does not need to give some precise sense to such terms as "sentient being" and "conscious state"... Rather he believes, in liberal vein, that there are various ways to make such terms precise (e.g. including or excluding cats among sentient beings!); and that for any such precise sense, the Everett construction of relative states, branches, and quantities having values relative to branches, is legitimate.' This statement of my beliefs is wide of the mark: I am not the liberal that Butterfield takes me for. Without, of course, being able to *define* them in terms of something more basic—for, conceptually speaking, there *is* nothing more basic—I would

¹² See references in Lockwood [1996].

¹³ Assuming, as Don Page pointed out to me, that the corresponding observables all commute.

nevertheless claim already to know precisely what I *mean* by 'consciousness' and 'sentience', expressions which I use as synonyms. And given what I mean by them, I regard it as a straightforward matter of fact whether cats or (a more seriously contentious issue) prawns are sentient. Thus, it's not, as I see it, a question of how we define our terms, but whether there's a 'what it's like to be' a cat or a prawn—whether, for example, prawns *feel* anything, if they're tossed alive on to a hot griddle.¹⁴ Lacking, as we do, a proper scientific handle on consciousness, we are doubtless unable at present actually to *determine* the answer to the latter question. But that does not prevent its *having* a determinate answer.

Butterfield is right to say that my interpretation of quantum mechanics makes few concessions to our common-sense conception of the world. I might naively suppose that, relative to *this* maximal experience, at least, I'm entitled to regard the centre of mass of a rock on the dark side of the moon as being in a *near*-eigenstate of position—or equivalently, in an eigenstate of some suitably defined coarse-grained position observable. But as Butterfield points out, the state of the rock, in this Everett branch, is likely not to be a pure state at all, but an improper mixture. If so, it cannot, *a fortiori*, be in an eigenstate, or near-eigenstate, of *any* observable! It may be true, by courtesy of decoherence, that this improper mixture bears a certain formal resemblance to a proper mixture of eigenstates of coarse-grained position—inasmuch as the off-diagonal elements of the corresponding reduced density matrix are very close to zero. But that doesn't make it even *approximately* true that this mixed state admits of an ignorance interpretation. On the contrary, the suggestion that, relative to this maximal experience, the rock's centre of mass *does* have a well-defined position, albeit one of which the mixed state gives only probabilistic information, is likely, on my view, to be straightforwardly false. In short, there's little of the common-sense conception of reality to be salvaged here.

But why should this worry me? Butterfield's tone—I have in mind especially his quip (pp. 203) 'believe it if you can'—suggests that he may think it a mark against my theory if it turns out that, in spite of decoherence, I'm stuck with a macroscopically very *indeterminate* world, even within a given Everett branch. I, however, am entirely sanguine about this consequence; I am content here (as with consciousness) to let the cards fall as they may. Harvey Brown put the point very nicely (in a recent conversation) when he said: 'Saving the appearances is the name of the game.' Any theory which has the consequence that, were it to be true, our

¹⁴ In the 1970s, a woman working in a fish and chip shop was accused of doing this, in a prosecution for cruelty to animals brought by the RSPCA. The RSPCA lost the case.

common-sense conception of reality would itself *appear* to be true, is *a priori* as good a theory as common sense itself—and must be judged, along with common sense, by the usual criteria by which we adjudicate between rival theories in science. To accord common sense some unique authority, in this of all areas, seems to me sheer dogmatism.

As Butterfield says, any version of the Everett interpretation, including mine, is faced with the choice of either embracing what he calls ‘nihilism’, or else giving transtemporal identity conditions for Everett branches. The latter, as he rightly remarks, is a task which most advocates of the Everett approach (including myself in the past) have shirked. It is, however, a matter to which (together with David Deutsch) I have recently given a great deal of thought. In the context of my own approach, one starts with time-slices of the experiential manifold. These time-slices can be partitioned into what I call *segments*, which correspond to distinct occurrences of basis states of consciousness, within a canonically represented mixed state of the Mind. We are then faced with a twofold task. First, there is a problem of *synchronic* identity which needs to be addressed; for there can be distinct segments on a given time-slice (corresponding to occurrences of the same basis state on different component projectors of the canonically represented mixed state) which intuitively belong to the same Everett branch, and hence ought to be pooled.¹⁵ Assuming that we have come up with an acceptable criterion for when pooling of segments is appropriate, we can then seek to define an *antecedent–descendant* relation between segments on different time-slices: this relation, appropriately spelled out, will yield transtemporal identity conditions for egocentrically defined Everett branches.

What, then, of the *transition probabilities*, which Butterfield quite reasonably demands? Well, transition probabilities, as I see it, should be defined, in the first instance, between earlier and later *sets* of identical maximal experiences, lying within segments all of which belong to the same set of pooled segments. Such probabilities will be automatically forthcoming once we know (a) the synchronic probabilities of the segments, at the times in question, (b) which segments are to be pooled, and (c), for every pair of temporally separated such segments, whether or not the later segment is a descendant of the earlier one. In the absence of any relevant *merging* of Everett branches (the implications of which, for the concept of a transition probability, I cannot, unfortunately, explore here), the transition probability, associated with the sets of experiences, S_1 , at t_1 , and S_2 , at t_2 , can be calculated as follows. First, we take the measure, m_2 , which results from summing the measures associated with all the segments, at t_2 ,

¹⁵ See Lockwood [1996], p. 23, fn. 10.

which are descendants of members of the set of pooled segments to which the elements of S_1 belong. Next, we take the measure, m_1 , given by summing the measures associated with the members of the set of pooled segments, at t_2 , to which the elements of S_2 belong. The required transition probability is then m_1/m_2 . By extension, we can apply this account to a specific maximal experience, e , which lies within a segment belonging to S_1 . Intuitively, if e is *this* maximal experience, and S_2 is a set of identical maximal experiences five minutes hence, then m_1/m_2 gives the phenomenological probability that, in five minutes time, I shall be having a maximal experience which belongs to the set S_2 .

I cannot, at present, give rigorous and fully general criteria for (b) or (c); so to that extent, Butterfield is correct in saying that there is a lacuna in my interpretation. But I have every confidence that satisfactory criteria can be found. And I'm at a loss to understand why Butterfield takes what I say in my article about the antecedent–descendant relation between segments to imply transition probabilities which are in conflict with the prescribed synchronic ones. I regard it as a condition of adequacy of any proposed criteria that they allow one to regard the 'flow' of probability, within the experiential manifold, as analogous to that of an incompressible fluid, made up of separate streams which can divide and (in principle, at least) merge. I thus take as axiomatic the *conservation of probability*, on which Butterfield rightly insists. With 'maximal experiences' substituted for Butterfield's 'conditions', and maximal experiences themselves regarded as 'evolving' from their antecedents and into their descendants, my model satisfies Butterfield's requirement (p. 215) that 'the measure of a set of final maximal experiences is to be equal to the measure of those initial maximal experiences that evolve into it'. It does so, at least, provided that, in order to allow for the branching of 'streams', we stipulate that the 'set of final maximal experiences' is to be understood as including *all* the maximal experiences, at the time in question, which have evolved from that set of initial maximal experiences. With a parallel proviso on 'initial maximal experiences', to allow for 'streams' to merge, we also have the time inverse of this: the measure of a set of initial maximal experiences is to be equal to the measure of those final maximal experiences which have evolved from it.

Let me end by repeating here what I acknowledged in my article: that it is obviously an oversimplification to associate maximal experiences with instantaneous Schrödinger states, just as it is an oversimplification to describe the world in terms of elementary quantum mechanics. Many minds theorists should certainly be looking, therefore, for ways of formulating their ideas within the framework of relativistic quantum field theory. It is thus entirely appropriate for Butterfield to draw our attention

to the work of Matthew Donald¹⁶ who brings algebraic relativistic quantum field theory to bear on a 'light bulb' model of observers, according to which the contents of consciousness supervene on certain equivalence classes of sequences of sets of localised neuronal switching events. But unlike Butterfield, I do not see Donald's approach and my own as in any sense competitors. For I accept, as I say, that a relativistic generalization of my views is ultimately required, and I agree also that the development, in this connection, of speculative models of the physical substrate of consciousness is to be encouraged. If I remain, at present, merely an interested and sympathetic spectator of such projects, it is because I feel that I could more confidently advance into this more difficult terrain, first having mapped out the nonrelativistic hinterland.

Acknowledgements

I should like to thank David Deutsch, David Papineau, and Euan Squires, for some very helpful suggestions, and for their own illuminating responses to the commentaries on my article.

*Oxford University
Department for Continuing Education
Rewley House
1 Wellington Square
Oxford OX1 2JA
UK*

¹⁶ See references in Lockwood [1996].

References

- Bell, J. S. [1981]: 'Quantum Mechanics for Cosmologists', in C. J. Isham, R. Penrose, and D. Sciama (eds), *Quantum Gravity 2: A Second Oxford Symposium*, Oxford, Clarendon Press. Reprinted in J. S. Bell, *Speakable and Unsayable in Quantum Mechanics*, Cambridge, Cambridge University Press [1987], pp. 117–38.
- Brown, H. R. [1996]: 'Mindful of Quantum Possibilities', *British Journal for the Philosophy of Science*, **47**, pp. 189–200.
- Butterfield, J. [1996]: 'Whither the Minds?' *British Journal for the Philosophy of Science*, **47**, pp. 200–21.
- Deutsch, D. [1996]: 'Comment on Lockwood', *British Journal for the Philosophy of Science*, **47**, pp. 222–8.
- Hardy, L. [1992]: 'Quantum Mechanics, Local Realistic Theories, and Lorentz Invariant Realistic Theories', *Physical Review Letters*, **68**, pp. 2981–4.
- Lockwood [1989]: *Mind Brain and the Quantum: The Compound 'I'*, Oxford, Blackwell.

- Lockwood [1996]: '“Many Minds” Interpretations of Quantum Mechanics', *British Journal for the Philosophy of Science*, **47**, pp. 159–88.
- Loewer, B. [1996]: 'Comment on Lockwood', *British Journal for the Philosophy of Science*, **47**, pp. 229–39.
- Papineau, D. [1996]: 'Many Minds are No Worse than One', *British Journal for the Philosophy of Science*, **47**, pp. 233–41.
- Parfit, D. [1984]: *Reasons and Persons*, Oxford, Oxford University Press.
- Saunders, S. [1996]: 'Comment on Lockwood', *British Journal for the Philosophy of Science*, **47**, pp. 241–8.