Interview with David Wallace

In M. Schlosshauer (ed.), *Elegance and Enigma: the Quantum Interviews*

July 2011

Q1. What first stimulated your interest in the foundations of quantum mechanics?

I can't really separate that from learning quantum mechanics itself. When I originally came across it as a first year undergraduate, it just felt like a confusing mess — I think the way most students (in the UK, at least) start off with quantum theory is a bunch of unrelated stuff about wave equations, uncertainty principles, wave-particle duality and the like. at that point, the whole subject looked so messy and confused, I struggled to get very interested in it.

But then a bit later I learned it properly and saw how elegant it really was from a mathematical and conceptual point of view — and that just made the quantum measuremetn problem stand out. If the theory was that elegant, it had to be possible to make sense of it somehow. And to a substantial extent, I found I wasn't really able to work on other problems within quantum mechanics as long as I didn't understand how the measurement problem was to be resolved.

Fifteen years later, I now feel pretty confident that the Everett interpretation satisfactorily resolves the measurement problem — and ironically, that's somewhat reduced my interest in the foundations of quantum mechanics. But there are plenty of foundational, conceptual and philosophical questions left in quantum theory, most of which are easier rather than harder to make progress on from the point of view of having a definite interpretation.

Q2. What are the most pressing problems in the foundations of quantum mechanics today?

I think anyone's answer to this is going to depend above all on what they think of the quantum measurement problem. After all, the measurement problem threatens to make quantum mechanics incoherent as a scientific theory — to reduce it, at best, to a collection of algorithms to predict measurement results. So the only reason anyone could have not to put the measurement problem right at the top of the list would be if they think it's solvable within ordinary quantum mechanics. (Someone who thinks it's solvable in some modified version of quantum mechanics — in a dynamical-collapse or hidden-variable theory, say — ought to think that the most pressing problem is generalising that modified version, to account for all of quantum phenomena, including the phenomena of relativistic field theory.)

As it happens, though, I *do* think the measurement problem is solvable within ordinary quantum mechanics: I think the Everett ("many worlds") interpretation solves it in a fully satisfactory way, and while I think there are some philosophical puzzles thrown up by that solution — mostly concerned with probability and with emergence — that would benefit from more thought, I wouldn't call them *pressing.* Not from the point of view of physics, at any rate.

So from my point of view, the "most pressing problems" aren't going to be ultra-broad problems like "what does quantum mechanics as a whole mean" — they're going to be a bit more detailed, a bit more concerned with particular puzzling features of the conceptual and mathematical structure of quantum mechanics. (The advantage of the Everett interpretation — the main *scientific* benefit it's brought, I'd say — is that it allows us to ask those questions without getting tangled up in worries about whether there are hidden variables or dynamical collapses or whatever not included in our equations, and without all sorts of doubletalk about "experimental contexts" and "the role of observers" and "subjective quantum states" and so on.)

All that said, here's the problem that leaps out for me. Just how are we to understand the apparently greater efficiency of quantum computers over classical ones? When I started as a physics grad student in the late 1990s, we had two really great quantum algorithms — Shor's algorithm, which factorises large numbers, and Grover's algorithm, which finds the biggest number in a list — and both of them were dramatically more efficient than the best-known classical algorithms. Shor's algorithm in particular had had a huge impact, because the problem of factorising large numbers *both* is one of the standard examples of a difficult computational problem, *and* is crucial in decoding a lot of codes that were and are thought to be basically un-decodable by classical computers. So everyone who was working in quantum information — including me at the time — was very excited by this, and pretty much all of us thought that Shor's and Grover's algorithms were going to be the tip of the iceberg, that there were going to be dozens or hundreds of these amazing quantum algorithms. But actually, ten years and more later, and those algorithms are still pretty much all we've got. Even if you could solve the technical problems involved in making a quantum computer that would fit on your desktop, at the moment there's not much you could do with it that you can't do with your existing classical desktop.

Now that's embarassing for people writing grant applications. But it's also bizarre from a foundational point of view. It's one thing to discover that quantum mechanics has a completely different computer-complexity theory from classical mechanics. It's quite another to discover that it's almost identical but *not quite.* My hunch is that we're missing something pretty profound here.

The second problem I'd identify is a bit easier to attack, and indeed we've got quite a long way with it already, but there's further to go. It's fairly clear now that the really big mysteries in quantum theory come not so much from superposition as from entanglement (after all, classical electromagnetism admits superpositions) — but getting a detailed quantitative grasp of what's going on in multipartite entanglement is really hard. We've got a variety of tools, and a variety of results, but it feels as if we still haven't found the right way of thinking about it, or maybe the right mathematical framework to use, such that it all becomes less opaque and less mysterious. (I think the very graphical "language" that Bob Coecke and his co-workers are developing is really promising here, but it's early days.)

I'll mention one more thing which might not normally be classified as "quantum foundations" — and which I guess isn't exactly "pressing", because we've been stuck with it for decades. The last two or three decades have made it really clear that quantum mechanics is way, way different from classical mechanics, and that it's possible to understand why the world looks classical without having to keep classical concepts as basic. (I'm thinking, in particular, of the role of decoherence theory, and the way we've basically managed to wean ourselves of the correspondence principle.) But the way we construct quantum theories, particularly in quantum field theory, is still almost invariably to start with a classical theory and then "quantise" it. That really, really shouldn't be necessary, but it seems to be. We need to find some way of thinking about quantum fields that doesn't require this link to classical fields.

Q3. What interpretive program can make the best sense of quantum mechanics, and why?

What interpretation? Everett's "many-worlds" interpretation.

Why? Here's the short version. Normally, we don't get worried about "interpreting" physical theories — we don't really need to interpret general relativity, or classical electromagnetism. We just take the theory as representing (part of) the structure of the world, so that states of the theory correspond to states of the world according to the theory. In quantum mechanics, things like Schrödinger's cat made us think that couldn't be the case — what kind of a state of the world is it in which a cat is a superposition of alive and dead? So we thought we had to give up on the usual story and find some alternative, clever way to think of the theory, or else change it into one which didn't have the same problem. What Everett did was tell us what kind of state of the world it is: it's a state of the world in which there are two cats, and one's alive and the other's dead. (Or really, there are two lots of cats, and one lot are alive and the other lot are dead.) And given the way quantum entanglement works, that pretty quickly means there are two (lots of) copies of the solar system, one with a live cat and one dead cat. Two worlds, in other words — at least locally.

What's the advantage of the Everett interpretation in particular? Here's one way to put it. In trying to interpret quantum mechanics, you've got two yes/no choices to make. Choice one: are you going to change the physics? Are you going to stick with the Schrödinger equation and the quantum state, or are you going to add dynamical collapse processes or hidden variables or backwards-in-time interactions or something? Choice two: are you going to change the philosophy? Are you going to stick with the straightforward way of reading a scientific theory as just telling us what the world is like, or are you going to start saying “a scientific theory is just a predictive algorithm for experiments” or “observers can’t just be modelled as physical systems” or “ordinary logic is wrong” or something? If you answer “no” to both questions, you’re stuck with the Everett interpretation, because the Everett interpretation is just the “take quantum mechanics completely literally” interpretation.

Now, for each choice, there are a lot of people who have said “yes” to each choice. And both those answers have led to interesting insights. But basically, the task they’re setting themselves is pretty challenging. Answering “yes” to the first question basically commits you to redoing the last seventy-five years of developments in quantum theory — not just coming up with alternatives to nonrelativistic quantum mechanics, but coming up with alternatives to QED, to the Standard Model, to neutrino-mass developments of it ... Answering “yes” to the second question basically commits you to overturning a really pretty solid consensus in philosophy of science that scientific theories really do have to be understood as making claims about what the world is like, and aren’t just shorthands for claims about how experimental devices work. (And it’s a consensus that I think pretty much all scientists share when they’re not actively philosophising. Are there really astrophysicists who think that the reason for talking about stars is to model patterns of detections on photoplates, not *vice versa*?)

But here’s the crucial point. You can make those yes/no choices for any scientific theory you care to name. You can do it for *palaeontology* if you want to! Spend time coming up with alternative theories for fossil formation, or decide that dinosaurs are just theoretical constructs used in theorising about fossils. It’s a free country. But the only motivation for answering “yes” to either question in the particular case of quantum mechanics and not in general is that you think there’s some special problem with the no/no answer in quantum mechanics: that is, you think the “take quantum mechanics completely literally” interpretation — the Everett interpretation — *doesn’t make sense.*

Okay, so does it make sense? Well, the main worries people have raised are: what justifies the “many-worlds” description? And what about probability? I don’t have space here to do more than comment briefly, but I think these are both resolvable. The “many-worlds” language follows from decoherence theory — from the various processes that dynamically suppress quantum interference — once we realise that they’re not supposed to be part of the fundamental ontology of the theory, but just something approximate, something emergent. And the probability issue turns out to be a strength, not a weakness, because (i) once you start thinking hard about probability *in Everettian quantum mechanics*, you realise probability is philosophically really mysterious *in general*, but (ii) it turns out that there are ways of understanding probability in quantum mechanics that don’t work in classical mechanics. (I’m thinking of the so-called decision-theoretic approach to quantum probability that goes back to David Deutsch, though really it’s more about symmetries of the quantum state, and the decision-theoretic gloss is just there to operationalise the concept of probability and dodge some philosophical worries.) Actually, probability is just one of several places where what looks like an Everett-specific philosophical problem turns out to be an old problem in a new guise.

Philosophers who come across the Everett interpretation tend to get worried about whether it makes sense, but physicists are more likely to ask, "what's the point of it", or "can it be tested". The answer to both questions is basically that there isn't any point using the Everett interpretation instead of ordinary quantum mechanics, and that you can't test the Everett interpretation against ordinary quantum mechanics, but that's because really, the Everett interpretation just *is* ordinary quantum mechanics — maybe not the quantum mechanics of the undergrad textbooks, which explicitly invoke wavefunction collapse, but the quantum mechanics we mostly use in practice, where we model measurement processes physically and apply non-unitary evolution only because we're tracing out — i.e., deciding to neglect — some environmental degrees of freedom. From that point of view, the point of the Everett interpretation is to allow us to do quantum theory without either taking measurement as some kind of primitive or having to change the formalism. And the right way to test the Everett interpretation is to test the universality of the superposition principle and the unitary dynamics.

Q4. What are quantum states?

At least if you mean pure states, they're states of the world. It's a bit misleading to take that as saying that they're real physical things, though. After all, in classical mechanics, the classical state — the phase-space point, that is — is a state of the world, but that doesn't mean that the world is a point in a really-high-dimensional space. What we mean by saying, of the state in a physical theory, that it's a state of the world, is that it represents, not facts about our knowledge of the world, but facts about the world itself.

Now, we might get worried about just what those facts are. In classical mechanics it's not so hard to answer: they're facts about where the particles are in space and how fast they're moving, or else in field theory they're facts about what the field strengths are in various spatial locations. In quantum mechanics, too, we can talk about the quantum state of a given spacetime region (of course, it's normally a mixed state). I'm not sure how much point there is trying to get an intuitive grip on what the state of that region really represents, beyond "certain features of the structure of that region".

Of course, in saying this I'm rejecting the alternative view, that the state is somehow a codification of our ignorance, somewhat like the statistical-mechanical state. But I don't really think this is viable. In (classical) statistical mechanics, it's pretty easy to see what we're ignorant *of*: we're ignorant of what the real classical microstate is. But we know that making a strategy like that work in quantum theory is going to be incredibly difficult, because of the Bell-Kochen-Specker theorem. The alternative that's most frequently discussed is that the quantum state represents our ignorance of the possible results of measurements, but that forces us to take measurement as some primitive thing that can't be analysed. I don't know how that can be squared with the fact that experimental physicists blatantly do analyse measurement processes all the time. (I actually think this is one place where the very abstract flavour of quantum information can get in the way; I say more about that in my answer to question 9.)

(What about mixed states? By and large I think they represent states of the world too, but not necessarily states of this particular branch of the world (in Everettian — many-worlds — terms). Say we prepare an EPR pair and throw one element of the pair away: the only quantum state available to represent the other element is a mixed state, and I don't see anything particularly wrong with saying that that really is the state of the qubit. Then if we let the qubit get decohered — say, if we measure it but don't look at the result — then it gets entangled with the macroscopic degrees of freedom of its vicinity, but *it* is still in a mixed state. Of course, relative to us, it's in some unknown pure state, so in that more limited sense the mixed state represents ignorance.)

Q5. Does quantum mechanics imply irreducible randomness in nature?

Okay, so that can't be answered without saying something about the measurement problem. Hidden variable theories typically (not always) reject irreducible randomness; so do interpretations that take the wavefunction as just a measure of our ignorance. Dynamical collapse theories typically build in randomness explicitly.

My own view is that the only interpretative strategy that currently makes sense of quantum mechanics is the Everett (many-worlds) interpretation, for reasons I spell out in my answer to question 2. And probability is really interesting from a many-worlds perspective, because there's clearly a sense in which nature is not random at all: the Everett interpretation says that the Schrödinger equation always holds, and the Schrödinger equation is deterministic. And yet there's clearly a sense in which the world at least *looks* random: when we do an experiment, we can't predict the outcome. And in fact the Everett interpretation guarantees that we can't predict the outcome, because it tells us that different outcomes happen in different branches.

Now, there's a line of argument that says that this just points to something incomprehensible, something unacceptable, about the Everett interpretation — that it tells us that probability in Everettian quantum mechanics doesn't make sense. And that's a serious line of argument and deserves a serious response, which I'm not going to give here in detail, but the short answer (here I'm repeating part of my answer to question 2) is that (i) once you start thinking hard about probability *in Everettian quantum mechanics*, you realise probability is philosophically really mysterious *in general*, but (ii) it turns out that there are ways of understanding probability in quantum mechanics that don't work in classical mechanics.

So if that's right, whether there's irreducible randomness in nature according to quantum mechanics depends on your vantage point. From the third-person vantage point — put metaphorically, from God's perspective — there's no randomness in nature, everything just plays out according to the Schrödinger equation. But whether or not there's a God, *we* can't achieve that perspective. From our point of view, the randomness is irreducible.

Q6. Quantum probabilities: subjective or objective?

Objective, definitely. I'm much more confident of that than I am of any particular interpretation.

I think this is another place where the abstractness of quantum information can be a bit misleading. It can seem kind of tempting to suppose that when we talk about the probability of a qubit being measured to be in a certain state, we're just talking about our subjective assessment. But quantum probability doesn't just apply to qubits, it applies to the *half-life of uranium-235*, and I really can't make sense of the idea that the decay rate of uranium isn't some fact about the world. When I say "you can make nuclear weapons out of plutonium because it has a really high probability to undergo fission in such-and-such situations, so we shouldn't let terrorists get hold of it", am I *really* not saying anything objective about plutonium? This is one of the places where I find the situation in the field kind of confusing, because some really smart people who I respect a lot seem happy with saying this, and I can't understand that. But then, I think people often say that about supporters of the many-worlds theory...

I suppose I should point out that there are ways and ways for probability to be objective. According to Everettian quantum mechanics, it's identified with mod-squared-amplitude of the branches, so it's objective, but can't be defined except in situations where decoherence gives us a branching structure. According to (most) hidden-variable theories, it's derived from the probability distribution over the hidden variables (and so *that* has to be objective). According to dynamical collapse theories, it's written into the equations.

Q7. The quantum measurement problem: serious roadblock or dissolvable pseudo-issue?

In a sense I don't think those are incompatible possibilities. Working out that something is a dissolvable pseudo-issue can be really hard work — just look at the difficulties that Einstein had thinking about general covariance, or that people thinking about black holes had thinking about the co-ordinate singularity on the event horizon. Something can be a serious roadblock *until* the conceptual insight that lets us dissolve it.

I actually think that's basically what the situation is in quantum mechanics, in that the measurement problem arises because it looks like you can't take the wavefunction literally as a description of reality without getting a flat contradiction with observations, because of Schrödinger cats and the like: the theory predicts that we ought to see the world in a superposition of macroscopic states, and it seems that we don't. And the relevant conceptual insight was Everett's: what would it really look like if the world *was* in a superposition of macroscopic states. Once you start thinking that way, you start to see that it's not *obvious* that we don't see the world looking like that, because of course if we looked at a cat in a superposition, we'd end up entangled with it and becoming part of the superposition. That doesn't dissolve the *problem*, but you might say that it dissolves the *paradox*, it changes it from "the world can't possibly be like quantum mechanics says it is, what do we do??!!!" to, "okay, what *exactly* does quantum mechanics say the world is like, and is it like that". And that gets us into decoherence theory and the like.

Maybe the thing I should say is that it's not an *easily* dissolvable pseudo-issue! The measurement problem maybe ought to be called the "macro-reality problem" — how can quantum mechanics be reconciled with observed macroscopic reality? It's not at all *obvious* that it can. I think if you think hard about it, along Everettian lines, and play around with decoherence theory and the quantum theory of big open systems, you can basically establish that it can. But that took a lot of hard work by a lot of people. And if it turns out all to fall apart for some reason, then I'd go right back to thinking of the measurement problem as a roadblock.

Q8. What do the experimentally observed violations of Bell's inequalities tell us about nature?

Well, for them to tell us anything about nature, we have to accept that it's legit to ask questions about nature (and not just about our experimental apparatus) in the first place. I think it's *obviously* legit, that's what science is for. But I guess a really hardline operationalist about quantum mechanics wouldn't care one way or another about the Bell inequalities.

Having got that out of the way, what the violations of Bell's inequalities *seem* to tell us is that the dynamics of the micro-world allows interactions that are faster than light (or slower than light but backwards in time, I guess, if that really means anything). If the only interactions in the world are subluminal, Bell's inequalities would be satisfied; they're not, so systems can interact superluminally. End of story. Sometimes people talk about Bell inequalities as if what they rule out is just local hidden variable theories — maybe even just deterministic local hidden variable theories — but I think Bell's later work makes it clear just how general they are.[1]

But I said that's what they *seem* to tell us. I don't think they *actually* tell us that, because there's a tacit premise in Bell's argument: that the results of measurements actually have definite outcomes. That looks pretty innocuous, because if measurement outcomes are macroscopic results, *of course* they're definite. But of course, that's exactly what the Everett interpretation of quantum mechanics denies. Or more accurately, measurement outcomes are relative to a branch. And branching (because it's just a dynamical process, the process of decoherence) is a local effect, and spreads out at lightspeed (actually, it spreads out at the speed of

---

[1] I'm thinking of his paper "Bertlmann's Socks and the Nature of Reality", in particular.

the fastest interaction that entangles regions with their neighbours, but in practice that's always lightspeed). So if I'm at one end of a Bell-type experiment, and you're at the other, I won't be able to ascribe any definite measurement outcome to your measurement until the branching caused by that measurement has reached me, and that happens at lightspeed.

So a better way of putting it is: *if the Everett interpretation is wrong*, violation of Bell's inequalities tells us that there are faster-than-light interactions. And this isn't particularly controversial among people that try to build realist (usually dynamical-collapse or hidden-variable) alternatives to quantum theory. What's slightly more controversial is whether that faster-than-light interaction requires a violation of Lorentz covariance. At first sight it looks like it has to — if we have superluminal interactions and we have Lorentz covariance, it looks as if we can construct closed causal loops — but actually it's a bit subtler, and people — notably Wayne Myrvold and Roderich Tumulka — have played around with so-called "hyperplane-dependent collapse theories" that try to get a relativistically covariant version of wavefunction collapse that's compatible with Bell's result. (Tumulka actually has a concrete version, albeit for non-interacting particles). Of course, if you buy the Everett interpretation then it's of rather theoretical interest if this works. But I'm told not everyone does buy the Everett interpretation.

Q9. What contributions to the foundations of quantum mechanics have or may come from quantum information theory? What notion of 'information' could serve as a rigorous basis for progress in foundations?

Quantum information theory brought something completely new to foundations of physics, in that it was the first time people had combined foundationally careful attention to the specifically quantum mechanical aspects of quantum mechanics with detailed, quantitative exploration of the theory's implications in particular situations.

People had done one or the other before. In particle physics, say, people were absolutely asking foundational questions, but they were mostly using quantum mechanics as a calculational tool — come up with a classical field theory, plug it into the machinery of Feynman diagrams and renormalisation group flows, and see what comes out. And what came out was wonderful, of course, but the quantum mechanics was largely functioning as a black box. Conversely, people in foundations of physics and philosophy of physics were asking foundational questions about quantum mechanics itself, but they were either not doing mathematics at all, or they were proving rather general theorems. They weren't playing with toy models, they weren't calculating much, they weren't exploring quantitatively just what the theory was capable of in various specific situations.

Then quantum information came along, and suddenly we discovered a huge range of things that *could* have been discovered in the 1940s, but weren't — teleportation, dense coding, the no-cloning theorem, entanglement swapping, Shor's algorithm, etc, etc. And those things haven't just been *practically* relevant — they've really deepened our understanding of what quantum mechanics is as a theory. And that's ongoing, and I'm sure other people answering this question are much better placed than I to go into details.

So, quantum information theory is an amazing tool to explore quantum mechanics. But there's a more ambitious project, which is to say that quantum information theory *is* quantum mechanics — or rather, that quantum mechanics just is a theory about information. Slogans like "physics is information" start getting mentioned at this stage.

I'm much more skeptical about this project. Partly that comes from worrying about whether it even makes sense — we don't think that the world could coherently be made of *opinion* or *belief* or *rumour*, and I'm not at all sure *information* is any better as a building block. (That's not to say that it's a scientifically useless concept — no more is belief a scientifically useless concept — it's just not obviously the sort of concept that can do as a fundamental-level description of reality.)

But more seriously, I don't really see how physics=information squares with what we use quantum mechanics for 99% of the time, which is to calculate physical properties of rather specific systems — crystals, metals, plasmas, atomic excitations, mass spectra of hadrons, etc, etc. Quantum information hides that away from us, because we study it in an incredibly abstract way which hides the ultimately physical, dynamical nature of whatever the Hamiltonian is of any given system.

That's exactly what quantum information should do, of course. The brilliant thing about it is that precisely because it *does* abstract away all those aspects of the system, it lets us see general features we'd never have spotted if we'd kept all the messy details in play. But it may be a mistake to treat that as an insight into the nature of reality itself, rather than into the nature of information flow in that reality.

Q10. How can the foundations of quantum mechanics benefit from approaches that reconstruct quantum mechanics from fundamental principles? Can reconstruction reduce the need for interpretation?

I should say first that I'm not up to speed with recent work on reconstruction. It's a field where there's been lots of very exciting progress in recent years, and I'm not well positioned to comment on the details.

But in general, I think reconstructions can tell us something interesting about the structure of quantum mechanics, but maybe not as much as their proponents sometimes hope. They certainly do a lot to help us understand the logical structure of quantum mechanics, and what happens to that structure if, say, we use reals or quaternions instead of complex numbers, or swap tensor products for Cartesian products, or whatever. And it's very often the case that something that's fairly opaque from one perspective on quantum mechanics is much more transparent from another perspective. The equivalence principle in general relativity is like that — once you understand that principle, various results that would have been calculationally horrific become really obvious.

But beyond that, I'm not sure how much we gain by rederiving the theory from "fundamental" principles, or even what it means for those principles to be "fundamental" in the first place. Take the analogy with special relativity, which often gets used in these discussions. Yes, we can understand why the Poincaré symmetry group applies by deriving it from the relativity principle and the light postulate. But we can equally well understand the relativity principle as a consequence of the dynamical fact that the symmetries of fundamental physics include the Poincaré group. Which route is more fundamental? I'm not sure that's a very fruitful question.

(It's tempting to say that the fundamental principles are in some sense 'natural' or 'intuitively reasonable'. But our intuitions about what's reasonable and natural don't have such a great track record at predicting how fundamental physics turns out.)

I'd also say that I don't see how reconstruction could reduce the need for interpretation. Ultimately, however we reconstruct quantum mechanics, we're either going to end up saying (i) that the mathematical structure thus reconstructed represents physical reality faithfully (in which case we end up with the Everett interpretation or something like it), or (ii) that it represents physical reality incompletely or inaccurately (in which case we need to fix it, which leads us to hidden-variable or dynamical-collapse theories), or (iii) that it's not in the business of representing physical reality at all (which leads us to operationalist or neo-Copenhagen or physics-is-information approaches). I say a bit more about this in my answer to question 3.

Q11. If you could choose one experiment, regardless of its current technical feasibility, to help answer a foundational question, which one would it be?

I'd do the two-slit experiment, but using gravity waves.

Of course, that's taking the "regardless of its current technical feasibility" clause *really* seriously! We haven't yet succeeded in detecting classical gravity waves, still less seen if they're quantized. But it would be one of those experiments which would be astounding whatever happened. If we see what effective field theory predicts — quantization of detection events, interference continuing even when the wave amplitude is so low that gravitons only pass through every few seconds — that would be an incredible triumph for quantum theory and quantum field theory. And of course, if we *didn't* see that, it would be unambiguous evidence that not only general relativity, but quantum mechanics too, stands in need of modification.

Q12. If you have a preferred interpretation of quantum mechanics, what would it take to make you switch sides?

I do have a preferred interpretation of quantum mechanics: the Everett interpretation. I take Everett's basic insight to be that we don't have to treat quantum mechanics differently from any other physical theory: we can just regard the theory's mathematical models (in quantum mechanics, unitarily-evolving quantum states) as representing physical states of affairs, just as in classical mechanics, or classical field theory, or general relativity. (From that point of view, the "measurement problem" arose because we erroneously thought quantum mechanics couldn't be understood that way.)

So to "change sides", I'd have to be convinced either that (a) something was wrong with the theory itself, or (b) that for some reason Everett's insight is wrong, and that after all we can't just take quantum mechanics as a straightforward physical theory like classical mechanics.

I'm prettty clear what it would take to persuade me of (a): empirical evidence in contradiction with quantum mechanics. If we find a violation of the superposition principle, in particular, that would be pretty good reason to reject the Everett interpretation of quantum mechanics — but we'd be rejecting quantum mechanics (and the Everett interpretation along with it), rather than rejecting an interpretation but keeping the theory.

I'm less sure what it would take to persuade me of (b). At one point I'd have said, "strong philosophical reasons to think that probability doesn't make sense in the Everett interpretation". But I've become more and more convinced that probability doesn't make any more sense in non-Everettian contexts (and indeed, that probably it makes less sense). And certainly, *mathematically* probability works fine in Everettian quantum mechanics, at least where decoherence is applicable.

Overall, I think in most cases I'd be more willing to revise a philosophical principle that was in conflict with the Everett interpretation than revise my interpretation of quantum mechanics. I guess I'm just not that confident that we'd have reasons to believe some given philosophical principle that were so persuasive that they'd require us to modify quantum mechanics or to interpret it completely differently from the way in which we normally interpret scientific theories. Perhaps that's just my lack of imagination, though.

Q13. How do personal beliefs and values influence one's choice of interpretation?

I'd like to say that they don't, but what I really mean is that they *shouldn't.* This is an objective question, even a scientific one: what does our best theory of the microscopic tell us about the physical world? That might not be a question directly answerable by experiment — though experiment bears on it quite a lot — but it shouldn't be a matter of taste. My beliefs and values shouldn't influence my take on the quantum measurement problem any more than they should influence my take on global warming or gamma ray bursts.

Dynamical-collapse theories are a really strong example here, of course, because they really are testable. But even in the case of the pilot wave theory, which makes the same predictions as quantum mechanics in normal circumstances, it's still a different theory with a very different formalism, and adopting it would have pretty major consequences for how we go beyond the Standard Model, or how we quantize gravity. And even something like a "pure" interpretation, like the approaches based on information, are mainly explored by people who think this is really telling us something important about the objective structure of the world, maybe even something with experimental consequences sooner or later. In every case, the choice of "interpretation" — "theory" would be a better word — is actually influencing the science that people are doing.

I haven't mentioned the Everett interpretation — which is the interpretation I think is correct — because oddly enough I think it's the conservative option, the one that doesn't really require any (well, much!) change in how we do or think about quantum physics. That's because the Everett interpretation, at least as I see it, basically just tells us to take quantum mechanics literally, and reassures us that there's no immediate paradox in doing so, that macroscopic superpositions aren't in contradiction with our observations. And in day-to-day physics, we basically do take quantum mechanics literally — we regard the quantum state of a system as something about that system, something we can prepare and modify and interact with. We use "collapse of the wave-function" as a shorthand, but when pushed we quickly retreat to saying that decoherence makes the superposition unobservable, not to regarding the collapse as some objective non-unitarity.

Now, it's true that we don't always *believe* what we think our theories *say.* So even if someone acknowledges that quantum mechanics *says* that after a measurement the world is still in a superposition but the superposition is unobservable, they might not believe that that's *true*, that the theory really can be trusted when it says that. And maybe that is a matter of our "beliefs and values", maybe there isn't any knock-down argument to convince somebody who uses quantum mechanics as a predictive tool that he ought to believe what the theory says about the physical world. But that isn't anything specific about quantum mechanics — any scientist is at liberty to carry on using his theory but not really believe its claims, if that's what he wants to do.

(It seems to me a pretty strange thing to want to do, but maybe that's *my* beliefs and values talking!)

Q14. What is the role of philosophy in advancing our understanding of the foundations of quantum mechanics?

I guess I ought to be in the ideal place to answer this question, since my original training was in physics and I moved into philosophy after my doctorate. But actually I don't have a systematic answer to give — ultimately, you make progress with problems by applying whatever the needed techniques and tools are, and whether those tools, or the departmental affiliation of the tool-user, count as "physics" or as "philosophy" isn't that important.

That said, what a philosophy training tends to give you is not so much a body of relevant knowledge, so much as a certain way of analysing a problem. Philosophy teaches you to be very careful, very attentive to whether the steps of your argument really do follow from one another, very concerned about what the conceptual assumptions are in your reasoning, very worried about whether ideas you're using have been properly defined.

Now, quite often the style of reasoning in physics — even theoretical physics — is a lot more free-wheeling than that. There's generally an impatience to get to the point at which concrete calculations can be done, and a willingness to cut corners — mathematically and conceptually — in doing so. As a rule, the proof of the pudding is in the eating — if you've managed to calculate something accurately, you must have been doing things right.

That might sound as if I'm building up to criticise the physics way of doing things, but in general I'm not — in fact, I think philosophers sometimes both underestimate its power, and confuse lack of mathematical rigor with lack of conceptual clarity (for example, a lot of people in philosophy of physics seriously underestimate how much conceptual progress quantum field theory has made with issues of renormalisation, just because that progress doesn't lend itself to rigorous axiomatisation). However, when you're trying to understand the foundational structure of a subject, and not just do calculations with it, the philosophy style of reasoning can be a useful complement to the physics style.

(I'm generalising, of course. Plenty of physicists — Einstein, most famously — can reason in both styles, according to what's needed at the time. But I think it's often the case that some formal philosophy training can help develop a more conceptually careful style.)

That's not to say that there aren't places in foundations of physics where philosophical knowledge, not just philosophical technique, can come in handy. In particular, physicists — or some physicists, at any rate — can end up saying very silly things about some philosophical issues. Free will, and the problem of consciousness, are pretty key examples: these are topics that have been thought about for a very long time, and while there's not a consensus on the *right* way to think about either, there's a lot that's been learned about superficially plausible but actually plain *wrong* ways of thinking about them. So sometimes, unfortunately, you get situations where someone claims that quantum theory has profound implications for (say) freedom of will, where actually they're just working with a philosophically naive and uninteresting notion of free will. You get the same problem in discussions of operationalism in quantum mechanics (to a lesser extent, though: physicists tend to be better informed on these closer-to-home topics).

As always when work gets interdisciplinary, the solution is to find a cooperative colleague in the other discipline, and talk to them. That happens less in physics-philosophy interdisciplinary work than in, say, the physics-biology case. I think that's partly because philosophy has a bit of a bad reputation among scientists and partly because — I'm sorry to say — that bad reputation is often deserved: too many philosophers end up saying really silly things about science in general and physics in particular because they haven't done their homework and haven't consulted a colleague. But that's not true for everyone in philosophy, any more than it's true that everyone in physics is ignorant of relevant philosophical ideas.

Q15. What new input and perspectives for the foundations of quantum

mechanics may come from the interplay between quantum theory and gravity/relativity, and from the search for a unified theory?

If I knew that, I think I'd probably be most of the way to having that unified theory myself!

Seriously, I think it's very interesting how *little* modification either string theory (as the current leading candidate for a quantum theory of gravity) or loop quantum gravity (as the current runner-up) make to the basic conceptual structure of quantum theory. In both cases, we basically hold on to unitary dynamics, transition amplitudes, Hilbert spaces, and the like.

I don't think we should be terribly surprised by that. Most of the great advances in theoretical physics come from a kind of radical conservatism: we try to push the basic principles of our extant theories as far as we can and see where that leads us. String theory and loop quantum gravity adopt that kind of conservatism towards quantum mechanics. (One too-glib way of characterising the difference between them is that string theory also adopts it towards particle physics, and loop quantum gravity also adopts it towards general relativity).

Whether it's sensible to bet on that strategy is going to depend pretty strongly on your take on the measurement problem. It's hard to make any sense of quantum gravity unless you understand it in an observer-independent way — that is, in Everett's way. If you think that notions of observation and measurement play an essential role in quantum theory — or if you think that quantum theory doesn't really make sense as a theory, and needs to be supplemented with hidden variables or modified to introduce a collapse of the wave-function — you should probably be sceptical about mainstream quantum gravity research. (Roger Penrose is probably the most famous example of someone who accepts this way of thinking: he sees dynamical collapse as something which we should expect to be caused by trying to create superpositions of spacetime geometries.)

On the other hand, if you think Everett's approach to quantum mechanics is basically satisfactory — which I do — then we don't have any reason to expect the foundations of quantum mechanics to be particularly illuminated by the search for quantum gravity. And, if string theory or loop quantum gravity turns out to be basically correct, the general structure of quantum theory won't really be modified at all by the incorporation of gravity. (The specific quantum mechanics in question, of course, will be modified a lot.)

Does that mean I'd bet on those programs succeeding? Not especially. Making progress so far ahead of the experimental data is bound to be chancy at best. But at any rate, I don't think we have much positive reason to reject their shared assumption that quantum theory continues to be applicable even in the general-relativistic regime — nor, if that shared assumption fails, much of a clue as to what will take its place.

Incidentally, this is a way in which the Everett interpretation is almost disappointing, at least compared to strategies like dynamical collapse that change the quantum formalism. If we really did expect some failure of quantum theory in the vicinity of the measurement process, that would be an amazing experimental regime to probe — hard, but way easier than quantum-gravity experiments — and might give us the experimental clues we need to make progress on quantum gravity. But the universe isn't designed for our convenience, and the fact that it would be useful for dynamical collapse to occur doesn't give us any reason to think it does occur.

Q16. Where would you put your money when it comes to predicting the next major development in the foundations of quantum mechanics?

I wouldn't — not much of it, anyway. Major changes in a field are by their nature pretty much impossible to predict in advance, and I'm not close enough to the detailed work in quantum information and computation to predict the relevant next steps there.

That said, I might wager a small sum on our making some fairly substantial breakthrough before *too* long in how to think about quantum computation and information flow in quantum systems — something that would give us a better handle on why quantum computers seem almost-but-not-quite equivalent to their classical counterparts. That's no more than a hunch, though, and it's largely driven by the exciting progress in recent years on diagrammatic ways to think about quantum mechanics. (Never underestimate the power of a new notation!)

I'd also put quite a bit of money on our *not* finding any experimental failure of unitarity, or any other evidence that quantum theory breaks down (anywhere outside the general-relativistic regime, at any rate). Given the coherence of the Everett interpretation as a solution to the measurement problem, and given the problems with relativity and with field theory involved in changing the quantum formalism, I strongly suspect that unitarity, and the universality of the superposition principle, are here to stay.

Come to think of it, though, it probably makes sense for me to hedge, and put my money on finding violations of unitarity after all. If my preferred approach to quantum mechanics were to be empirically falsified, at least I'd be rich.

Q17. What single question about the foundations of quantum mechanics would you put to an omniscient being?

I'm going to cheat and offer him the choice of two.

First question: "Is quantum theory — not quantum field theory, or any other *particular* quantum theory, but the general dynamical framework of quantum theory — ultimately correct? Or is the quantum framework, like the classical framework, just something to be superseded in due course?"

(A quick comment on this one: it's fashionable to say that scientific theories are always being superseded and replaced, and at some level that's true. But at a deeper level, we've only *ever* had two dynamical frameworks which were well-enough developed to actually do any proper calculations: classical physics, and quantum physics. (And the problem with quantum gravity is that we only know how to write a relativistic theory of gravity in the classical framework). It's a live option — the option string theory bets on — to suppose that the quantum framework really is the ultimate dynamical framework. But either answer would be fascinating.)

Second question: “Do the unobserved branches in macroscopic superpositions represent physically real states of affairs, as real to their inhabitants as our surroundings are to us? If not, why not?”

In a way, the first question would be more sensible, as I’m more confident I know the answer to the second already. On the other hand, the second would settle a lot of arguments! (And it sneakily combines the question “is the Everett interpretation the right way to understand *unitary quantum mechanics*?” with the question “is unitary quantum mechanics *true*?”.)