Chapter 10
The Many-Worlds Theory

The last version of quantum mechanics that we will explore in detail was developed by Hugh Everett III while he was a graduate student under John Wheeler in the 1950s. Everett's basic idea is at once beautifully elegant and uncomfortably radical. Max Jammer rightly described it as "one of the most daring and most ambitious theories ever constructed in the history of science" [1].

Some idea about the nature of Everett's proposal can be gleaned by the different titles used for various draft versions of his PhD thesis: "Quantum Mechanics by the Method of the Universal Wave Function", "Wave Mechanics Without Probability", and "On the Foundations of Quantum Mechanics [2]." Everett's thesis advisor, John Wheeler, was a strong proponent of Bohr's Copenhagen interpretation and was thus sensitive not only about the radical nature of Everett's proposal, but also about Everett's sharp criticisms of the Copenhagen philosophy. Wheeler thus demanded that Everett produce a significantly toned-down presentation of his ideas; this was ultimately published in 1957 with the somewhat cryptic title "[The] 'Relative State' Formulation of Quantum Mechanics [3]." The somewhat muted nature of the presentation in this published version probably contributed to Everett's ideas not being widely understood or appreciated for several subsequent decades, and his near-complete departure from the world of theoretical physics. But Everett did inspire a few early followers such as Bryce deWitt who, along with his own graduate student Neill Graham, published the original, full-length version of Wheeler's thesis, as well as some other commentary, as "The Many-Worlds Interpretation of Quantum Mechanics [4]". This title is probably the most accurately descriptive of Everett's ideas, and is the one by which the theory has largely come to be described today.

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture
Notes in Physics, DOI 10.1007/978-3-319-65867-4_10

10.1 The Basic Idea

As with the Spontaneous Collapse theory of the last chapter, Everett's theory is principally motivated by the Measurement Problem that we studied in Chap. 3. In Everett's description, the usual quantum formalism contains two incompatible rules, "Process 1" and "Process 2", for how the states of quantum systems evolve. Process 1 is the discontinuous and random change that is postulated to occur when an observer or measuring instrument from outside the quantum system interacts with it in an appropriate way, whereas Process 2 is the continuous and deterministic state evolution described by the Schrödinger equation. Everett bemoans the fact that, since measuring instruments (and ultimately observers) are just physical systems like any other, it is simply not clear when the two very different Processes are supposed to apply. Discussing an isolated system that includes an observer or measuring instrument, Everett writes:

> Can the change with time of the state of the *total* system be described by Process 2? If so, then it would appear that no discontinuous probabilistic process like Process 1 can take place. If not, we are forced to admit that systems which contain observers are not subject to the same kind of quantum-mechanical description as we admit for all other physical systems. [And note that when we speak of an "observer", we really mean things like] photoelectric cells, photographic plates, and similar devices where a mechanistic attitude can hardly be contested [3].

Moreover, if one wants to apply quantum mechanics to the universe as a whole (which is natural in cosmology and in particular in the quest to unify quantum theory and gravitation) the idea of an "outside observer" becomes obviously incoherent:

> No way is evident to apply the conventional formulation of quantum mechanics to a system that is not subject to *external* observation. The whole interpretive scheme of that formalism rests upon the notion of external observation. The probabilities of the various possible outcomes of the observation are prescribed exclusively by Process 1. Without that part of the formalism there is no means whatever to ascribe a physical interpretation to the conventional machinery. But Process 1 is out of the question for systems not subject to external observation [3].

Everett's central idea, therefore, is to simply abandon Process 1:

> This paper proposes to regard pure wave mechanics (Process 2 only) as a complete theory. It postulates that a wave function that obeys a linear wave equation everywhere and at all times supplies a complete mathematical model for every isolated physical system without exception. It further postulates that every system that is subject to external observation can be regarded as part of a larger isolated system. The wave function is taken as the basic physical entity.... [3]

Let us think through what that means, in the context of our standard example: the measurement of the energy of a particle-in-a-box (whose spatial coordinate we call x) using an energy-measuring-device (whose pointer has center of mass coordinate y). We would describe the measuring device as faithfully and accurately measuring the energy of the particle if, when the particle is initially in an energy eigenstate $\psi_k(x)$ (with eigenvalue E_k), the interaction causes the apparatus pointer to move by a distance proportional to E_k. That is, we assume that Process 2 – Schrödinger's equation – evolves the joint quantum state of the particle and pointer as follows:

$$\psi_k(x)\phi_0(y) \to \psi_k(x)\phi_0(y - \lambda E_k t) \tag{10.1}$$

where t is the duration of the interaction. It then immediately follows from the linearity of Schrödinger's equation that, if the particle-in-a-box is initially in a superposition of several different energy eigenstates, the system will evolve into an entangled superposition as follows:

$$\left[\sum_i c_i\, \psi_i(x)\right] \phi_0(y) \to \sum_i c_i\, \psi_i(x)\, \phi_0(y - \lambda E_i t). \tag{10.2}$$

All three versions of quantum theory that we have studied so far have regarded this last formula as problematic, and have thus proposed some way of resolving the problem. The orthodox/Copenhagen view, for example, would say that it was inappropriate to try to describe the measurement interaction in terms of a quantum mechanical wave function; measuring devices are classical objects, so we should have treated the interaction instead using Process 1, according to which the post-interaction state involves a collapsed wave function for the particle-in-a-box and a pointer with a definite, classical position. The pilot-wave theory accepts that there is a wave function associated with the particle-pointer system, and that the wave function indeed evolves into a state like that in Eq. (10.2), but insists that the real physical state of the pointer is not to be found in this wave function but instead in the actual positions of the associated particles, which will be unambiguous and unproblematic. Finally, the spontaneous collapse theory insists that the wave function for a system including a macroscopic pointer will simply not obey Schrödinger's equation, and so the troubling macroscopic superposition state, Eq. (10.2), simply will not arise (or at least, will not arise for long enough to notice!).

In contrast to all three of these views, Everett wants to say, about Eq. (10.2), that it is fine; there is no problem. To understand this, though, it will help to briefly review what the problem with Eq. (10.2) was supposed to be. In short, the problem was that Eq. (10.2) involves a superposition of different positions for the (macroscopic, directly observable) pointer. It's frankly not even exactly clear what this means, or what it would look like, but apparently it is some kind of state in which the pointer somehow has several different positions at the same time. It seems it should appear in some sense "blurred" among the several different positions. But, of course, nobody has ever seen a pointer in a state like that. Real pointers always point this way or that. And so Eq. (10.2) simply cannot provide a faithful, direct, literal, complete description of the physical state of the pointer. Or at least that is what we have been taking for granted until now.

Everett, though, invites us to consider in more detail what – according to quantum mechanics – the pointer in a state like Eq. (10.2) would look like. To analyze this, we should consider the different possible states that an observer might get into upon interacting with a pointer. Suppose, to begin with, that the observer – whose (many!) degrees of freedom we call z – begins in a "ready" state, $\chi_0(z)$. Suppose he interacts with a pointer with a reasonably well-defined position $y = y_k$. Then the observer should get into a state $\chi_k(z)$ in which he has seen (and, for example, remembers seeing, and will, if asked, report having seen) that the pointer has position y_k. That is, during the time of interaction between the pointer and the observer, Schrödinger's equation should evolve the quantum state as follows:

$$\phi_0(y - y_k)\, \chi_0(z) \to \phi_0(y - y_k)\, \chi_k(z). \tag{10.3}$$

But then, just as before, it immediately follows from the linearity of Schrödinger's equation that if the observer observes the position of the pointer in the particle-pointer system, described by Eq. (10.2), the quantum state of the particle-pointer-observer system will evolve as follows:

$$\left[\sum_i c_i\, \psi_i(x)\, \phi_0(y - y_i)\right] \chi_0(z) \to \sum_i c_i\, \psi_i(x)\, \phi_0(y - y_i)\, \chi_i(z). \tag{10.4}$$

So, in the same way that the pointer failed to pick out some one particular outcome from the set of superposed "possibilities", but instead got tangled up in the superposition, so now the observer (of the pointer) does not end up in a state that corresponds to seeing some one particular location for the pointer. Instead he, too, joins the entangled superposition. It is, of course, unclear exactly what to make of this. But notice right away one thing that this definitely does *not* say: it does not say (or at any rate, does not seem to say) that the observer will be in a state in which he definitely experiences (and remembers experiencing and will report, if asked, having experienced) the pointer as "looking blurry" or "being smeared out among several different positions". Everett explained this as follows in his thesis:

> Why doesn't our observer see a smeared out needle? The answer is quite simple. He behaves just like the apparatus did. When he looks at the needle (interacts) he himself becomes smeared out, but at the same time correlated to the apparatus, and hence to the system.... [T]he observer himself has split into a number of observers, each of which sees a definite result of the measurement.... As an analogy one can imagine an intelligent amoeba with a good memory. As time progresses the amoeba is constantly splitting [2].

Whatever else one wants to say, there is a suggestion here that our assumption that there was some kind of fatal problem with Eq. (10.2) – that the particle-pointer system just obviously wouldn't look right if this were the correct and complete state description – was perhaps too hasty.

It will be helpful here to follow Everett in introducing the concept of a "relative state". As he points out, in a system described, for example, by Eq. (10.2), neither of the components – the particle or the pointer – can be attributed a definite state of their own. That's essentially what "entanglement" means. But Everett points out that we can define a "relative state" for each component, relative, in particular, to the other component's being in a particular state. For example, the state of the pointer relative to the particle-in-a-box being in state $\psi_k(x)$ is defined to be

$$\phi^{\text{rel. to } \psi_k(x)}(y) \sim \int \psi_k^*(x)\Psi(x, y)dx \tag{10.5}$$

where $\Psi(x, y)$ is just the joint particle-pointer state given in Eq. (10.2). (The "$\sim$" is because the right hand side here is not properly normalized.) Plugging in, we find

$$\begin{aligned}\phi^{\text{rel. to } \psi_k(x)}(y) &\sim \int \psi_k^*(x) \sum_i c_i \psi_i(x)\phi_0(y - \lambda E_i t) \\ &= c_k \phi_0(y - \lambda E_k t)\end{aligned} \tag{10.6}$$

since the different $\psi_i(x)$'s are orthonormal: $\int \psi_k^*(x)\psi_i(x)dx = \delta_{i,k}$. So the (properly normalized) relative state is just

$$\phi^{\text{rel. to } \psi_k(x)}(y) = \phi_0(y - \lambda E_k t). \tag{10.7}$$

In words: relative to the PIB being in a particular energy eigenstate, the pointer ends up in a perfectly definite and appropriate state, namely, one in which it indicates the energy E_k of the PIB.

The converse also holds: relative to the pointer indicating outcome E_k, the PIB is in the state $\psi_k(x)$. And we can generalize this concept to bring in the observer as well: when the overall particle-pointer-observer wave function is given by the right hand side of Eq. (10.4), relative to the PIB being in the state $\psi_k(x)$, not only does the pointer indicate that its energy is E_k, but the observer *sees* (and remembers seeing and will report having seen) the pointer indicating that its energy is E_k.

This idea of "relative state" provides a way of capturing the fact that, although it remains puzzling, a state like the right hand side of Eq. (10.4) is not just utter chaos. It is not just some kind of incomprehensible blur in which everything is happening in a completely mixed-up way. Rather, there are definite *correlations* built into the state: it is an orderly mixture, in some sense, of several individually perfectly reasonable situations, in which the PIB has a definite energy, and the pointer indicates correctly what its energy is, and the observer sees the pointer indicating its energy and correctly and validly infers what its energy is.

It is clear that subsequent interactions will work exactly the same way, and just bring more and more of the world into the mixture. For example, the air molecules in the vicinity of the pointer get jostled around in slightly different ways depending on how fast the pointer moves during its journey from its "ready" position to its final position, and where exactly that final position is. Of course, there are several distinct final positions in the mix, and so, just like the observer, the air molecules join the entangled superposition. They cannot be said to have any particular definite state of their own, but relative to the pointer being in a particular final position, their configuration is clear and definite and sensible. Similarly, if the observer's mom calls him on the phone to ask how his energy measurement turned out, she will now also join the entangled superposition, as will the ink molecules in the physics journal where he publishes his results. There will not be any one particular fact of the matter about what his mom hears or what is printed in the journal; but "relative to" the observer seeing the pointer at $y = \lambda E_k t$ (and relative to the pointer being at $y = \lambda E_k t$ and relative to the energy of the PIB being E_k) mom will hear, and the journal will report in print, that the energy measurement came out E_k. And so on.

It is as if the big (and ever-expanding) entangled superposition, which we previously took as just somehow obviously wrong, is actually describing all of these perfectly coherent stories playing out in parallel. Except that, for Everett, it is not merely "as if" this were the case. Everett's idea is precisely that this is literally the case. By eliminating "Process 1" and letting the universe be described by a single wave function, evolving always exclusively according to the linear Schrödinger equation (Process 2), we arrive at the following picture: whenever we would have said (according to one of the previously considered formulations of QM) that there were several distinct possibilities, only one of which is in fact realized, instead in Everett's theory *all* of the possibilities are realized; the world splits into several branches, each of which realizes one of the possibilities. Further interactions then produce further branchings in each original branch, and so on. The overall pattern of iterative branching is indicated schematically in Fig. 10.1.

A few words of clarification are in order. First, although Everett's theory is often called the "many worlds" theory and the different branches are sometimes referred to as (for example) "parallel universes", these turns of phrase can also cause confusion

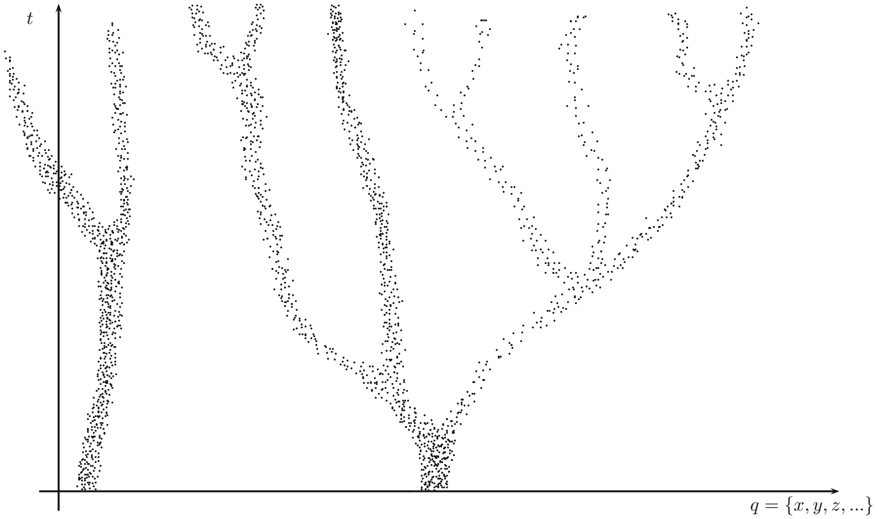

Fig. 10.1 Schematic depiction of the wave function of the universe, evolving in time, with an iterative branching structure

and suggest misleading pictures. Really, according to Everett's theory, there is only one universe, only one world. It's not, for example, that every time a quantum event (triggering a branching) occurs, a whole new copy of the physical universe is created, ex nihilo, "next to" the old one, such that, over time, more and more and more universes are all existing, in some sense separately. (People who misunderstand Everett's idea in this way often complain, for example, that the multiplication of worlds flagrantly violates the idea of mass or energy conservation.)

Instead, it is supposed to be the case that the matter in the one, only-existing universe just has these different patterns going on in it, all, so to speak, on top of one another. Perhaps a good analogy here would be to light waves: if you're driving your car during the daytime and turn on the headlights, the region in front of the lights has (let's say) some light waves, propagating east, emitted by the headlights – and also some light waves, headed down, emitted by the sun. These two things are happening in the same place and are associated with the same one underlying field. They are distinct structural patterns in that field. But the dynamics of the field is such that the two patterns do not affect each other. The light waves from the sun just do their thing, passing downward, in the same way they would if the light from the car headlights were not there, and vice versa. The non-interaction of these two light waves explains why it is appropriate to think of what's going on in terms of these two overlapping but distinct patterns.

One should remember, though, that unlike electromagnetic waves which propagate in 3-dimensional physical space, the quantum mechanical wave function exists in a very high-dimensional space. So one should for example recognize that the horizontal, "spatial" axis in Fig. 10.1 is a very schematic simplified way of representing what is in fact a space of enormous dimension. This is also relevant to understanding why the different branches, once formed, do not interact. In principle, packets can interact, by *interfering* with each other, if they overlap. But here "overlap" means "overlap in configuration space" – because that is where the wave function lives. If one is thinking about a single particle moving in one dimension, it may seem very probable that, for example, if the wave packet splits into two "branches", one of which moves off to the left and the other to the right, it might occur (for example if one of the packets reflects off something and moves back the other way) that the two packets might come again to overlap, producing some interference effects. *In princple* this can happen, but due to a phenomenon called "decoherence" this basically never happens in practice once the difference between two branches becomes macroscopic (which by the way is when you'd first be justified in thinking of them as distinct branches).

You can think of it this way: configuration space is *really high-dimensional*. So there's just a lot of room there. If a branching event occurs, like when our energy measuring device interacts with the PIB, it's not just – as our schematic treatment in terms of the center of mass coordinate y might suggest – that the two wave function packets separate by a small macroscopic distance d (say, a centimeter). In fact, there are some enormous number – 10^{23} or something – of particles in the pointer. So the two wave packets in configuration space are not just separated by distance d. Rather, they are separated by distance *d in each of some 10^{23} distinct coordinates*. So, by a high-dimensional analog of the Pythagorean theorem, the packets are actually separated by something like a centimeter times $\sqrt{10^{23}}$, i.e., about *two million miles*. The packets are, for all practical purposes, permanently and irreparably separated by a vast distance in configuration space, never to interact again. (And note that the separation and its irreparability only continue to increase as the pointer interacts with air molecules in its vicinity, which then in turn interact with further degrees of freedom that they are in contact with, and so on.)

So that is a nice way to think about the process, decoherence, that makes these individual branches in the wave function very stable, separate, non-interacting. What might seem like a very small difference between two branches actually (when we remember the enormous and ever-increasing number of particles that are involved) implies that the branches are extremely well-separated in the vast open wilderness of configuration space and will hence never see each other again.

Everett summarizes the overall idea as follows:

> We thus arrive at the following picture: Throughout all of a sequence of observation processes, there is only one physical system representing the observer, yet there is no single unique *state* of the observer (which follows from the representations of interacting systems). Nevertheless, there is a representation in terms of a *superposition*, each element of which contains a definite observer state and a corresponding system state. Thus with each succeeding observation (or interaction), the observer state 'branches' into a number of different states. Each branch represents a different outcome of the measurement and the *corresponding* eigenstate for the object-system state. All branches exist simultaneously in the superposition after any given sequence of observations.
>
> [In a footnote he adds:] From the viewpoint of the theory *all* elements of a superposition (all 'branches') are 'actual', none any more 'real' than the rest. It is unnecessary to suppose that all but one are somehow destroyed, since all the separate elements of a superposition individually obey the wave equation with complete indifference to the presence or absence ('actuality' or not) of any other elements. This total lack of effect of one branch on another also implies that no observer will ever be aware of any 'splitting' process [3].

10.2 Probability

In the last section, we started to come to grips with Everett's central idea of simply omitting, from the axioms of quantum theory, the measurement postulates (such as the Born rule) which seem difficult to reconcile, at the fundamental level, with Schrödinger's equation. However, in the conventional interpretation, these measurement postulates provide practically the entire *testable* content of the theory – they tell us, in particular, about the probabilities for various possible measurement outcomes. And it is precisely the fact that these probabilities match up with the empirically observed outcome frequencies, that we believe in the quantum formalism in the first place. So if Everett's "many worlds" theory is to be worth taking seriously at all, it will need to be able to account for these conventional probabilistic claims.

From his very first presentation of the many worlds idea, Everett recognized the importance of being able to somehow derive the Born rule (in the context of his new theory which adamantly does not just posit it as an axiom). Indeed, Everett claimed to provide such a derivation/explanation already in 1957:

> The new theory is not based on any radical departure from the conventional one. The special postulates in the old theory which deal with observation are omitted in the new theory. The altered theory thereby acquires a new character. It has to be analyzed in and for itself before any identification becomes possible between the quantities of the theory and the properties of the world of experience. The identification, when made, leads back to the omitted postulates of the conventional theory that deal with observation, but in a manner which clarifies their role and logical position [3].

However, Everett's claim has been met with skepticism and in general this issue has remained a highly controversial one ever since Everett's original proposal.

Let's try to understand what's at issue here, starting with a simple example. Suppose an experimenter prepares a spin 1/2 particle in the "spin-up along the x-direction" state,

$$\psi_{+x} = \frac{1}{\sqrt{2}} \left(\psi_{+z} + \psi_{-z} \right), \tag{10.8}$$

and then performs a measurement of the z-component of the particle's spin. According to conventional quantum mechanics, we'd say that there is a 50% probability that the measurement comes out spin-up (let's call that "heads" for simplicity here) and a 50% probability that it comes out spin-down ("tails"). But of course, in Everett's view, that's not right. Instead, according to Everett, both things happen: the act of measuring the z-spin (i.e., setting up a coupling between the z-component of the particle's spin and some eventually-macroscopic degrees of freedom that include those belonging to the observer himself) triggers a branching of the universal wave function, and each outcome occurs in one of the two branches. As it is sometimes put, the observer has two "descendants" – one who sees the experiment come out "H" and one who sees it come out "T".

Now suppose the experimenter does this N times – that is, suppose he prepares a *bunch* of spin 1/2 particles in the state ψ_{+x} and then measures their z-spins, one at a time. The branching structure that will be produced is illustrated, for the case $N = 4$, in Fig. 10.2. At the end, there are $2^4 = 16$ different branches, and the experimenter observed a different sequence in each one: $HHHH$ for the branch on the left, $HHHT$ for the next one over, and so on, all the way over to $TTTT$ on the right. All 16 of these branches appear, in the expression for the total wave function, with the same amplitude, so aside from each involving a distinct sequence of outcomes, they all seem to be on an equal footing.

However, for purely combinatoric reasons, certain statistical patterns of outcomes occur in more branches. For example, there is only the one branch in which the observer saw 4 Hs, and similarly there is just the one branch in which the observer saw 4 Ts. But there are *four* branches in which the observer saw 3 Hs and 1 T. (These four branches have the following sequences: $HHHT$, $HHTH$, $HTHH$, and $THHH$.) Similarly, there are four branches in which the observer saw 1 H and 3 Ts. And finally there are *six* branches in which the observer sees 2 Hs and 2 Ts. One begins to see the overall pattern: although every possible sequence of outcomes

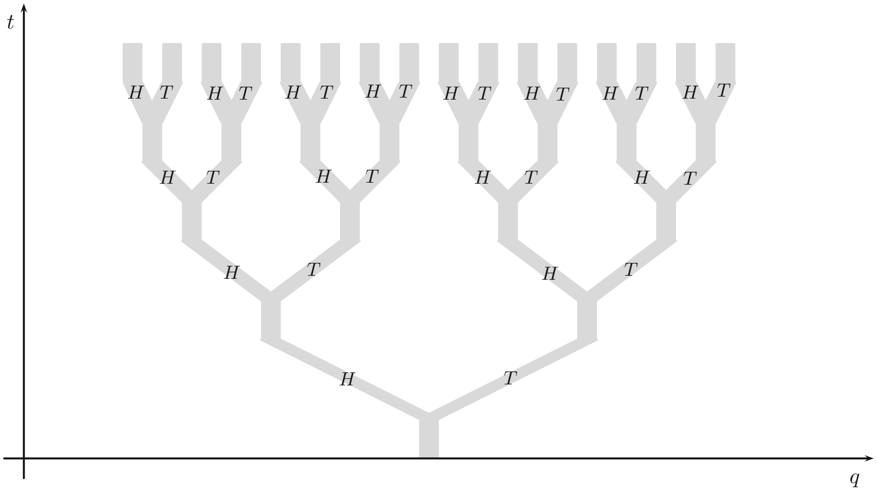

Fig. 10.2 The branching structure created by an experiment in which a "quantum coin" is flipped 4 times

occurs in precisely one world, *most* of the worlds will exhibit statistics that are close to those associated with the Born rule (here, equal numbers of Hs and Ts).

In the general case of N binary quantum measurements (which we'll continue to think of as coin flips for simplicity), the number $g_N(n)$ of worlds in which exactly n Hs are observed will be "N choose n":

$$g_N(n) = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}. \tag{10.9}$$

And so (there being 2^N worlds at the end of the sequence of experiments) the *fraction* $f_N(n)$ of worlds in which exactly n Hs are observed will be

$$f_N(n) = \binom{N}{n}\left(\frac{1}{2}\right)^N = \frac{N!}{n!\,(N-n)!}\left(\frac{1}{2}\right)^N. \tag{10.10}$$

For large N, this function of n is well-approximated by a normalized Gaussian whose center point is at $n = N/2$ and whose half-width σ is $\sqrt{N}/2$. That is,

$$f_N(n) \sim e^{-(n-N/2)^2/(N/2)}. \tag{10.11}$$

See Fig. 10.3 for a sketch.

So again, although each possible sequence is realized in exactly one branch, and no sequence is in any way preferred over any other, it is the case that, at the end of the experiment, the overwhelming majority of observers – that is, the overwhelming majority of descendants of the original observer – will observe a sequence in which

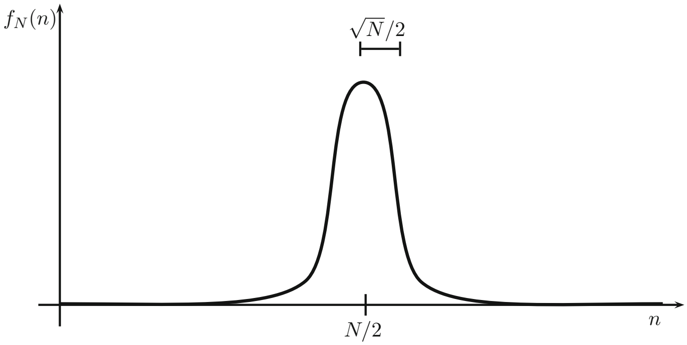

Fig. 10.3 In an experiment involving N binary measurements (in which, according to conventional quantum theory, the probability for each possible outcome is 50%), the fraction $f_N(n)$ of Everettian worlds in which n Hs are observed will be sharply peaked around $n = N/2$. That is, the overwhelming majority of observers in the different worlds (i.e., the overwhelming majority of descendants of the original experimenter) will see approximately Born rule statistics

there are roughly equal numbers of Hs and Ts. That is, we can reproduce the Born rule by looking at the statistical patterns that are *typical* for universes, i.e., present in *most* of the universes. Of course, there will some universes in which very non-Born-rule statistics (e.g., a string of N consecutive Hs!) will be observed. But so long as such rogue universes represent a vanishingly small fraction of the total number of universes it perhaps seems somewhat reasonable to ignore them and claim that, according to Everett's theory, observers should typically expect to see Born-rule statistics.

But there is a serious problem with this line of thinking: it only works for the special case that, in the conventional way of describing the situation, the outcome probabilities are 50/50. Or, to put the same point in Everettian terms, it only works for the special case in which the two branches created by each individual measurement event appear (in the overall expression for the wave function) with equal amplitudes. To see this, let's consider the more general case in which, say, the initial preparation of each spin 1/2 particle has it being spin-up along a direction $\hat{n}$ such that

$$\psi_{+n} = \sqrt{p}\psi_{+z} + \sqrt{q}\psi_{-z} \tag{10.12}$$

where $p + q = 1$, i.e., p is what would ordinarily be called the probability of H (i.e., the particle coming out spin-up along z), but which is in the context of Everett's theory instead called the *branch weight* of the "spin-up along z" branch that the measurement creates.

It is easy to see that, for the $N = 4$ case, the "tree of outcome sequences" is exactly the same as what was already displayed in Fig. 10.2. The only difference is that now the branch weights are not all equal. For example, the $HHHH$ branch has a branch weight p^4; the four branches with three Hs and one T each have branch weight p^3q; the six branches with two Hs and two Ts have branch weight p^2q^2; and so on. In general, the weight of a branch with n Hs and $(N - n)$ Ts will be

$$w_N(n) = p^n q^{(N-n)} = p^n (1-p)^{(N-n)}. \tag{10.13}$$

Now, the Born rule tells us that (in conventional terms) the probability of a H for each flip is p. So in a sequence of N flips, the expected number of Hs will be Np. For example, if $N = 100$ and $p = 90\%$ we should expect to see about 90 Hs. But if we just naively count worlds the way we did before, it remains true that the overwhelming number of worlds have approximately 50 Hs and approximately 50 Ts.

Therefore, in order to continue accounting for the usual Born rule statistics in the Everettian model, it is necessary to *weight* the worlds differently – and in particular to weight each branch by, what else, its branch weight – when we compute the world-fraction which displays a certain characteristic. We thus define the weighted world fraction as follows:

$$f_N^w(n) = g_N(n) w_N(n). \tag{10.14}$$

(Note that what we called $f_N(n)$ before is just this same formula but for the special case $p = q = 1/2$ in which the weight function $w_N(n)$ is equal to $1/2^N$ independent of n.) One can show that this weighted world fraction function is, for large N, sharply peaked around $n = Np$. (See the Projects.) That is, when we include the non-equal weightings, we can still say that the overwhelming majority of worlds (in the weighted-by-their-branch-weights sense) will exhibit approximately Born rule statistics. The idea here is visualized in Fig. 10.4.

That sounds good, but also raises a number of questions. For example: what, exactly, are these "branch weights" that we've been talking about? Well, they are nothing but the (absolute) squares of the amplitudes of the different branches, i.e., the different terms in the universal wave function. If, that is, after some sequence of measurements, the universal wave function has the structure

$$\Psi = \sum_i c_i \psi_i \tag{10.15}$$

(where the index i is labeling particular sequences of measurement outcomes, or whatever) the branch weight associated with the ith branch is just

$$w_i = |c_i|^2. \tag{10.16}$$

That is, the formula for the branch weights – the equation telling us how much to "care" about each individual branch in the tree – is really just the Born rule. So the overall argument has a strong air of circularity about it: if you weight the branches using the Born rule, then (the Born-rule-weighted-sense-of) "most" of the branches will display Born rule statistics. It seems that we get the Born rule out (as a description of the statistics that will be observed in typical branches) only because we put the

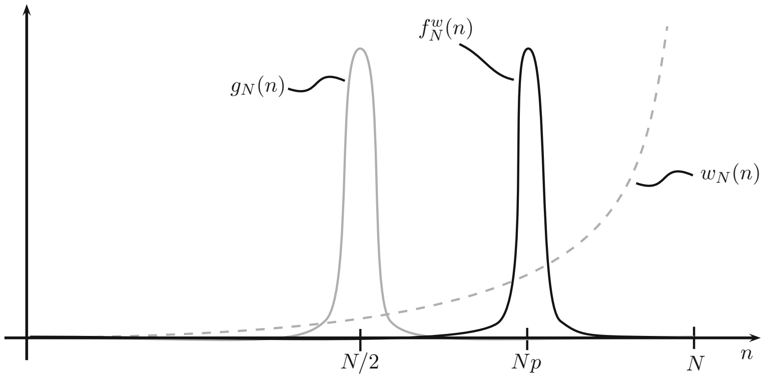

Fig. 10.4 Graphical illustration of that fact that, although the (raw, unweighted) number of worlds in which n "Heads" results appear in a sequence of N quantum coin flips is strongly peaked around $N/2$, we can nevertheless say that (in a weighted sense) "most" worlds will display Born rule statistics, i.e., $n \approx Np$. Here the raw world-counting function $g_N(n)$ is shown as the solid gray curve; the weighting function $w_N(n)$, according to which worlds with larger amplitude or "branch weight" are more strongly emphasized in the accounting, is shown as the dashed gray curve; and the weighted world fraction $f_N^w(n) = g_N(n)w_N(n)$ is shown as the solid black curve. The case $p = 3/4$ is shown

Born rule in (as a measure of how much each branch should count in our assessment of what is typical).

There is a long history of proponents of the many worlds interpretation trying to give further arguments to prove that Eq. (10.16) is somehow the only mathematically reasonable way to weight the different branches. If this could be convincingly established, it would significantly reduce the feeling of circularity. You are invited, in the Projects, to analyze and assess the argument that Everett presented already in 1957.

But there are some deeper concerns as well. For example, the very idea that we should use this un-equal weighting seems somewhat in conflict with Everett's overall idea. Recall, for example, Everett's statement that "none [of the branches is] any more 'real' than the rest". But what is this non-trivial weighting function, other than some kind of measure of how real, exactly, each (supposedly equally real) branch is? Suppose there are just two branches, one – say, in which a red light flashes – with a branch weight of $1/100$, and the other – in which, say, a blue light flashes instead – with a branch weight of $99/100$. Everett would have us say that the vast majority of worlds – namely, 99% of them – include a flashing blue light. This may make some kind of sense from the point of view of an external God-like observer, who can somehow "see" that the blue-light world is *brighter* (more intense? heavier?? more real???) than the red-light world. But in some sense Everett's whole program is to abandon, as non-existent and meaningless, the idea of such external God-like observers, and instead to exclusively consider what the world is like "from the inside", i.e., according to observers who are part of the world and governed by its fundamental laws.

There has indeed been a trend in the recent literature on this issue, away from treating the "branch weights" as somehow objective facts that we (the theorists analyzing the merits of Everett's theory) must acknowledge, and toward treating them instead as measures of how much individual observers within an Everettian world should care about their various descendants. As Simon Saunders has summarized this point,

> In recent years, with the development of decision-theory methods for quantifying subjective probability in quantum mechanics, the link between probability in the subjective sense and an objective counterpart has been greatly clarified. Specifically, it can be shown that agents who are rational, in order to achieve their ends, have no option but to use the modulus squared branch amplitudes in weighting their utilities. In this sense the Born rule has been *derived* [5].

This claim, though, remains controversial, with questions proliferating about: the necessity of defining (or, indeed, expecting) the "rationality" of agents in the way required in the derivation; the appropriateness and relevance of focusing on how agents who believe in Everett's theory should behave as opposed to explaining why we should interpret our empirical observations in Everettian terms; and even whether the concept of "probability" can possibly mean *anything* in a picture where, with certainty, everything that can happen *will* happen.

We will not be able to resolve these issues here. What should be clear, though, is that the unusual, many-worlds character of Everett's proposal forces us to reimagine certain concepts – like "probability" – that play an important role in quantum theory. As we will see, this uncomfortable "stretching" of concepts previously thought to be well-understood is a theme that will re-appear as we continue our exploration of Everett's proposal.

10.3 Ontology

So far we have followed Everett (and his followers) in essentially taking for granted that each branch of the universal wave function can be understood as describing a sensible physical world, with stars and planets and trees and cats and measuring equipment and observers with brains (whose physical states give rise to appropriate conscious experiences) and so on. That is, we have just been assuming that (at least at an appropriately macroscopic coarse-grained level) each branch of the wave function corresponds to a physical world basically of the sort we take ourselves to experience.

But we should remember, from Chap. 5, that the wave function is a funny and abstract kind of mathematical object. There is no obvious and straightforward sense in which the wave function can be understood to directly describe physical goings-on in ordinary three-dimensional space, because the wave function is something like a field on an abstract, high-dimensional configuration space. So we should not simply take for granted that the wave function (or any individual branch of the wave function) describes a three-dimensional physical world of the sort we are accustomed to imagining exists. Instead, we should ask: if it does, how, exactly, does the description work?

One possibility is Schrödinger's original idea that the wave function can be used to compute a mass density field (on physical, 3D space). Recall that, in this scheme, the mass density of the ith particle would be given by

$$\rho_i(x,t) = m_i \int |\Psi(x_1, x_2, \ldots, x_N, t)|^2 \, \delta(x - x_i) \, dx_1 dx_2 \cdots dx_N \tag{10.17}$$

and the total mass density would then be

$$\rho(x,t) = \sum_i \rho_i(x,t). \tag{10.18}$$

In the context of the GRW theory we discussed in Chap. 9, in which only one branch of the universal wave function survives the spontaneous collapses, we were able to recognize this mass density field as corresponding to a world that "looks right" at the macroscopic scale. But in the context of Everett's proposal – in which all branches of the universal wave function survive – the mass density field becomes a big incoherent mess. In an illuminating discussion of this idea [6] an analogy has been given to an old TV set which is badly tuned and is therefore receiving and displaying the programs from several different channels all at once. Indeed, this was the primary reason that Schrödinger himself abandoned this idea as a possible way of understanding the ontology associated with the quantum wave function.

But is the mess really so incoherent? Just like, in the TV set analogy, the different programs (being displayed on top of one another) do not *interact* with each other, so the contributions to the mass density field from different branches of the wave function remain dynamically independent. That is, just like two characters from one of the TV programs will interact with each other (but neither can in any sense interact with the characters from one of the other simultaneously-displayed programs), so with the different contributions to the mass density field associated with different branches of the wave function. We should recognize, that is, that the total mass density field, Eq. (10.18), only looks like an incoherent mess to some God-like external observer (and only then, perhaps implausibly, if She is unable to disentangle the overlapping programs). In keeping with the Everettian philosophy, though, we recognize this as irrelevant. To an observer living in that universe, himself made out of some portion of $\rho(x,t)$ which only interacts with other portions of $\rho(x,t)$ arising from the same branch of the universal wave function, the world looks entirely coherent. Such an observer would, in effect, be happily oblivious to the fact that there were countless alternative programs playing out literally right on top of him, but the (limited part of the) world he actually experiences would indeed "look right" – i.e., have the same kind of overall macroscopic coherence we are familiar with from our actual experiences.

Let us explain, more formally, how and why the different contributions to $\rho(x,t)$ arising from different branches of the wave function can be thought of as non-interacting and causally independent. Suppose the wave function can be written as a linear combination of macroscopically-distinct packets

$$\Psi(x_1, x_2, \ldots, x_N, t) = \sum_\alpha c_\alpha \Psi_\alpha(x_1, x_2, \ldots, x_N, t) \tag{10.19}$$

where, as discussed in Sect. 10.1, the individual packets are well-separated in configuration space so that

$$\int \Psi^*_\beta(x_1, x_2, \ldots, x_N, t)\Psi_\alpha(x_1, x_2, \ldots, x_N, t)\, dx_1,\; dx_2 \cdots dx_N = 0 \tag{10.20}$$

if $\alpha \neq \beta$. (The requirement that the different terms be *macroscopically* distinct effectively ensures that two terms which are orthogonal in this sense at one time will remain orthogonal in the future.) It then follows from Eq. (10.17) that the mass density associated with the ith particle can be written as

$$\rho_i(x, t) = \sum_\alpha |c_\alpha|^2 \rho_i^\alpha(x, t) \tag{10.21}$$

where

$$\rho_i^\alpha(x, t) = m_i \int |\Psi_\alpha(x_1, x_2, \ldots, x_N, t)|^2\, \delta(x - x_i) dx_1 dx_2 \cdots dx_N \tag{10.22}$$

is the mass density of particle i arising specifically from the α branch of the wave function.

The total mass density can then similarly be written as

$$\rho(x, t) = \sum_\alpha \rho_\alpha(x, t) \tag{10.23}$$

where

$$\rho_\alpha(x, t) = \sum_i \rho_i^\alpha(x, t) \tag{10.24}$$

is the total mass density associated with the α branch of the wave function.

The important point here is captured by Eq. (10.23), which says that the total mass density can be broken apart into distinct pieces that (like the programs from different TV channels that are being displayed simultaneously) each play out independently of the others. In a sense, there is nothing new here compared to the way we were thinking about Everett's many-worlds proposal previously. The point is just that the mass density ontology provides a definite, viable way of extracting, from the evolving universal wave function Ψ, a coherent (many worlds!) story about physical goings-on in ordinary three-dimensional space, i.e., this is a way to give a precise meaning to the way we were already talking about Everett's idea in earlier sections.

A few contemporary proponents of the Everettian picture (for example, Lev Vaidman) seem to basically understand the theory in this way. But for the most part, Everett's contemporary followers resist the idea that some special, explicit postulate about the ontology of the theory is required.

The reason for this resistance is the idea that one of the main virtues of Everett's approach is its elegance, its parsimony: there is just the wave function, obeying Schrödinger's equation, full stop. This is supposed to be in contrast, for example, to the pilot-wave picture, which followers of Everett would regard (because it posits not only the wave function obeying Schrödinger's equation, but in addition particles moving in accordance with some further dynamical law) as ontologically cluttered and cumbersome. That is, "wave function monism" – the idea that the wave function is all there is – plays a very important role, for Everettians, in explaining and justifying their preference for the Everettian theory.

One can indeed appreciate how an explicit endorsement of something like Schrödinger's mass density ontology – Eqs. (10.17) and (10.18) – would feel dangerously and suspiciously similar to the pilot-wave theory's explicit postulation of additional ontology. But, of course, the problem is that it is very difficult to understand what to make of Everett's theory if one just says "the wave function is everything" and leaves it at that. Independent of whatever worries one might have about the many worlds idea, such a position would mean that the theory suffers rather acutely from the ontology problem we discussed in Chap. 5.

So the problem faced by proponents of Everett's theory is to, on the one hand, avoid the ontology problem by finding some way of extracting, from the theory, an explanation for our experience of material objects moving and interacting in three-dimensional space, while at the same time avoiding the need to postulate additional things, distinct from and additional to the wave function itself. One approach to this problem has been to argue that familiar macroscopic structures in 3D can be understood to emerge from the structure in the wave function, in the same way that complicated macroscopic objects like, say, tigers can be understood as complex macro-patterns of more basic ontological posits. As David Wallace elaborates,

> It is simply untrue that any entity not directly represented in the basic axioms of our theory is an illusion. Rather, science is replete with perfectly respectable entities which are nowhere to be found in the underlying microphysics.... Tigers [for example] are (I take it!) unquestionably real, objective physical objects, but the Standard Model [of particle physics] contains quarks, electrons and the like, but no tigers. Instead, tigers should be understood as patterns, or structures, *within* the states of that microphysical theory.... The moral of the story is: there are structural facts about many microphysical systems which, although perfectly real and objective (try telling a deer that a nearby tiger is not objectively real) simply cannot be seen if we persist in describing those systems in purely microphysical language. Talk of zoology is of course grounded in cell biology, and cell biology in molecular physics, but the entities of zoology cannot be discarded in favour of the austere ontology of molecular physics alone. Rather, those entities are structures instantiated within the molecular physics, and the task of almost all science is to study structures of this kind [7].

The idea is that, in something like that same way, the ordinary world of macroscopic objects (including tables and chairs and planets and trees and cats and human observers) is already there, instantiated within the complicated ripplings in the structure of the universal wave function. In particular:

> Structurally speaking, the dynamical behaviour of each wavepacket [i.e., each decoherent branch of the wave function] is the same as the behaviour of a macroscopic classical system. And if there are multiple wavepackets, the system is dynamically isomorphic to a collection of independent classical systems [7].

That, I think, is exactly correct, but seems also to miss the point of the ontology problem.

It is true that a relatively narrow and well-isolated wave packet propagating through $3N$-dimensional configuration space is equivalent to an (approximate) *trajectory* through $3N$-dimensional configuration space and hence, in turn, isomorphic to (i.e., mathematically interchangeable with) a description of N particle trajectories in 3-dimensional space, i.e., a classical system. But surely this mathematical isomorphism does not imply that the real physical existence of a propagating wave packet in $3N$-dimensional space somehow brings about the additional real physical existence of N particles moving and interacting in 3D. A single billiard ball, bouncing around on a square two-dimensional billiards table, for example, is mathematically isomorphic to two beads (one small enough to pass through the hole of the other so they don't interact with each other) bouncing back and forth from the ends of a wire. Does each really-existing billiard ball on a table thus somehow call into existence a pair of beads on a wire somewhere? Nobody believes this, yet it seems like a perfectly fair analogy to what would be required for a wavepacket in configuration space to genuinely give rise to a classical system of particles in three-dimensional, physical space.

The sticking point is really the trans-dimensional character of the required sort of emergence. If, for example, the fundamental quantum mechanical description were in terms of N single-particle wave functions propagating in 3D space, there would be *no difficulty at all* in understanding how a rough macroscopic description in terms of atoms, molecules, and ultimately tigers, could be appropriate and entirely consistent with that fundamental ontology. There is no problem, that is, in understanding how something like a tiger can be understood as emerging from a fundamental ontology involving waves. The problem is in understanding specifically how something like a tiger (which is a certain complex pattern of microscopic goings-on *in three-dimensional space*) could be understood as emerging from a fundamental ontology involving only waves that live in an entirely different, much higher-dimensional space.

Perhaps some ultimately-satisfying account of the needed sort of trans-dimensional emergence could be given. Or perhaps this is the wrong way to think about it. Wallace, for example, seems to have suggested that the very appearance that we live in a three-dimensional world could itself be emergent:

> Note firstly that the very assumption that a certain entity [namely, a certain branch of the universal wave function] which is structurally like our world is not *our world* is manifestly question-begging. How do we know that space is three-dimensional? We look around us. How do we know that we are seeing something fundamental rather than emergent? We

> don't; all of our observations ... are structural observations, and only the sort of a prioristic knowledge now fundamentally discredited in philosophy could tell us more [7].

He means here that the idea that we live in a three-dimensional world should not be taken as some kind of a priori dogma which has to appear, in stone, at the most basic level. This, too, could be emergent from some very different more elementary processes, as they appear to rough creatures like us. But (although Wallace would certainly deny that this is an appropriate way to express it) there is a suggestion here that the three-dimensionality of the world is then something like an illusion. And if we can be deceived, through our ordinary direct perceptual experience of the world, about something so basic as that, one worries that it might become difficult to hold off concerns about what else we might have been deluded about, and so why we should believe the quantum formalism in the first place.

Again, the goal here is not to resolve these issues, but to help make you aware of their existence. Suffice it, then, to note that (just as with "probability"), controversial questions about ontology persist for Everett's theory. Can the theory explain the existence (or, at least, the appearance) of familiar three-dimensional material worlds of the sort we ordinarily take ourselves to inhabit and of which, according to the theory, there are actually many? And in particular, can it account for such worlds exclusively on the basis of the universal wave function? The needed structures are unquestionably present there – as shown by the possibility of understanding the ontology in terms of Schrödinger's mass density field. But does the singling-out of, for example, that particular bit of structure – as the thing we should look at to understand what the theory says about goings-on in three-dimensional space – constitute the postulation of additional ontology, beyond the wave function, as is done unapologetically in the pilot-wave theory? If so, it is hard to understand why one would not then just prefer to adopt the pilot-wave theory and skip the difficulties (pertaining, for example, to "probability") that arise due to the many-worlds character of Everett's theory. But if not, why only that particular structure? And are certain things we took as basic facts about our world (like its three-dimensionality) then rendered merely illusory, and, if so, is that even a problem?

These are some of the questions that would, I think, need to be addressed before an Everettian theory could be considered to be as ontologically satisfactory as the pilot-wave theory, or GRWm, or GRWf.

10.4 Locality

The question of whether Everett's theory respects relativistic local causality is yet another subtle and controversial one. Among the theory's supporters, it is widely believed that the theory is – uniquely among available options – locally causal. And so this claim, that Everett's theory is somehow uniquely compatible with relativity, is a big part of the reason why the theory's supporters support it.

The claim is generally based on two different lines of reasoning. The first is that, in ordinary quantum mechanics, the non-locality originated from the wave-function collapse postulate. So, by retaining only (an appropriately relativistic generalization of) the Schrödinger equation – i.e., by simply eliminating the collapse postulate from the dynamics – Everett's theory supposedly retains the local part, and abandons the nonlocal part, of ordinary QM, and is therefore itself perfectly local.

The second line of reasoning addresses the question of how Everett's theory supposedly eludes Bell's proof (discussed in Chap. 8) that *any* empirically viable theory must include nonlocal dynamics. The claim here is that Bell's arguments involve a previously-unacknowledged assumption which does not apply to Everett's theory – namely, Bell assumes that the spin measurements (made by Alice and Bob at opposite ends of the experimental setup) *have definite outcomes*. That is, Bell assumes that, for each individual spin measurement, there is a particular unambiguous result – the particle is either found spin up along the axis in question, or spin down. Bell's inequality is then a constraint on the statistical correlations that are possible, if locality is respected, within this set of unambiguous particular measurement outcomes. But, the Everettians point out, it is simply not true in Everett's theory that each individual spin measurement has a single particular outcome; instead, there is a branching point in the structure of the world and both "possible" outcomes are realized, one in each branch.

It is certainly true that Bell wasn't anticipating that kind of possibility and that his theorem does indeed tacitly assume that experiments have particular, definite, single realized outcomes. And there is some value in pointing out precisely how the many-worlds theory manages to elude Bell's general argument. But in a way it is unnecessary to ask, of any extant candidate theory, whether and how Bell's theorem applies to it. (The use of Bell's theorem is that it allows us to diagnose, as either non-local or in conflict with the experimental facts, all of the *non-extant* theories – the theories that nobody has managed or bothered to think of yet.) We can instead assess the theory's status vis-a-vis locality directly, by just seeing whether or not the theory respects our explicitly formulated notion of locality from Chap. 1.

When we attempt to do this for Everett's theory, however, we immediately realize that the difficulties we reviewed in the last two sections – pertaining to the ontology of the theory and the role and meaning of probability within it – preclude anything like a straightforward diagnosis. If, for example, we understand the theory as positing the existence of nothing but the wave function – thought of as a kind of field in a 3N-dimensional abstract space, and with the appearance of three-dimensionality being some kind of emergent delusion within our conscious experiences – then it will be completely impossible to say anything meaningful about whether the theory does or does not respect locality. Locality, remember, is the idea that causal influences between physically real objects in ordinary 3-dimensional space never propagate faster than the speed of light. If, according to a theory, there *are* no physically real objects in ordinary 3-dimensional space, then concepts like "local" and "non-local" are simply, radically, fatally, inapplicable. The theory, so understood, would be "not even non-local" in precisely the sense introduced back in Chap. 5.

Of course, this is just an extreme example, intended to make a pedagogical point. Probably no actual proponent of Everett's theory would endorse the perspective described in the previous paragraph. Still, it is a crucial and under-appreciated point that a theory has to clearly articulate an ontology of physical objects in three-dimensional space (and, if that ontology is not openly posited like the particles of the pilot-wave theory, must explain clearly how the ontology of physical objects in three-dimensional space relates to and emerges from whatever *is* openly posited) before the theory's status vis-a-vis local causality can be meaningfully assessed.

Ambiguities surrounding the concept of "probability" also prevent a straightforward application, to Everett's theory, of Bell's formulation of locality. Recall that, in Bell's formulation, "locality" was the requirement that the probability assigned to each event in space-time, conditioned on a complete description of events in a slice across the past light cone, should be independent of events with suitable space-like separation. In earlier chapters, we have always been concerned in particular with events that correspond to definite, observable occurrences such as a certain experiment having a particular outcome. Such events can of course still be said to occur in Everett's theory, but (what would previously have been described as) the different possibilities are not related to one another in the familiar way in Everett's theory, and this undermines and obscures the applicability of certain probabilistic ideas.

For example, it would normally be assumed that, since a given spin measurement on a spin-1/2 particle has two possible outcomes, two probabilities like $P(\text{up}|\lambda)$ and $P(\text{down}|\lambda)$ – where "up" and "down" mean, respectively, that the "spin-up" and "spin-down" outcomes are manifested in the macroscopic ontology in the appropriate space-time region – should sum to 100%:

$$P(\text{up}|\lambda) + P(\text{down}|\lambda) = 1. \tag{10.25}$$

But, in the context of Everett's theory, each of these probabilities is (for generic λ) already 100%: *both* outcomes will, with certainty, be instantiated, right on top of one another, in the appropriate space-time region. Thus, for Everett:

$$P(\text{up}|\lambda) + P(\text{down}|\lambda) = 2. \tag{10.26}$$

This illustrates the sense in which certain basic assumptions about how probabilities work – having to do with the mutual exclusivity of the possibilities to which we conventionally assign probabilities – take a very different form in the context of Everett's theory and thus prevent certain probability assignments from working in the familiar ways.

In any case, you can perhaps begin to see why the question of whether the many worlds theory is local, is a subtle and controversial one. Not only is the many-worlds character of the theory *weird* in a profound and radical way (so that normal everyday assumptions, like that experiments always have definite specific outcomes, as well as more technical assumptions like that the probabilities associated with what we normally think of as distinct possible outcomes should sum to one, fail to apply), but it also remains very obscure how/whether the theory's postulates relate to and/or account for and/or give rise to physical objects and process in ordinary 3-dimensional physical space.

Still, let us attempt to set these abstract worries aside, and get a concrete feeling for how the many worlds theory talks about one of the example situations we've used to discuss the non-locality of other theories.

To make things as definite as possible, we'll consider a version of the theory in which a mass density field $\rho(\vec{x}, t)$ is explicitly postulated as the way to understand what 3-space ontology the wave function is describing. And let's analyze the Einstein's Boxes scenario from the point of view of this version of the theory. Suppose, then, that there is a single particle, split between two "half boxes" located at widely separate locations. Suppose further that Alice is stationed near the half-box on the left and decides to implement, at $t = 0$, a measurement to see whether or not the particle is contained in her half box; the measurement outcome is registered on a pointer which swings to the right by some distance d if the particle is detected, and stays stationary if the particle is not detected. Meanwhile, Bob is stationed near the half-box on the right and also decides to implement, at $t = 0$, a similar measurement. Thus, prior to $t = 0$, the wave function for the particle-and-two-pointers system can be written

$$\Psi(x, y, z) = \frac{1}{\sqrt{2}} \left[\psi_L(x) + \psi_R(x)\right] \phi_0(y)\chi_0(z) \tag{10.27}$$

where $\psi_{L/R}(x)$ are wave functions with support exclusively in the left/right half-boxes, and $\phi_0(y)$ and $\chi_0(z)$ are narrow wave packets centered at $y = 0$ and $z = 0$, the undeflected "ready" positions of the two pointers. The mass density ρ associated with this wave function will have contributions at the undeflected positions of the two pointers and will also involve half of the split particle's worth of mass density in each of the half boxes.

After $t = 0$, when both measurements have gone to completion, the wave function will evolve into

$$\Psi(x, y, z) = \frac{1}{\sqrt{2}} \left[\psi_L(x)\, \phi_0(y - d)\, \chi_0(z) + \psi_R(x)\, \phi_0(y)\, \chi_0(z - d)\right] \tag{10.28}$$

which is a superposition of two terms: one in which the particle is on the left, the pointer on the left has deflected (indicating that the particle was detected there), and the pointer on the right remains undeflected (indicating that the particle was not detected there) – and another in which the particle is on the right, the pointer on the left has not deflected (indicating that the particle was not detected there) and the pointer on the right has deflected (indicating that the particle was detected there). These two terms are well-separated in configuration space (especially when one remembers that our schematic degrees of freedom y and z are really proxies for some huge macroscopic number of individual particle positions) and so the two terms can be understood as describing distinct, no-longer-interacting worlds.

The mass density is then simply the sum of the individually reasonable mass densities associated with each individual world. That is, $\rho = \rho_1 + \rho_2$ where ρ_1 has the mass density associated with the particle being contained exclusively in the left box, Alice's pointer having swung to the right indicating that the particle is there on the left, and Bob's pointer remaining in its undeflected position... and where ρ_2 instead has the mass density associated with the particle being contained exclusively in the *right* box, Alice's pointer remaining in its undeflected position, and Bob's pointer having swung to the right indicating that the particle is there on the right. See Fig. 10.5 for a sketch of how the mass density evolves during the process.

As shown in the Figure, from this "God's eye" perspective, nothing particularly dramatic happens here and there isn't much of a suggestion of nonlocality. The mass density associated with the particle-in-the-two-half-boxes is initially split between the two half-boxes and the pointers are both sitting in their ready positions. Then, as the interactions proceed around $t = 0$, the mass density associated with the two pointers splits in half so that both pointers now have "split" positions in the same way that the particle did initially. It perhaps seems plausible to say that Alice's pointer splits into these two different positions in response to the (purely local) fact that the particle is only half-contained in her half-box, and similarly for Bob's pointer. And so it may seem plausible to say that (weird though the many-worlds character here may be, with each pointer pointing to two different positions!) there is not really any suggestion of nonlocality here.

But this appearance is somewhat misleading since there are *relational facts* about the various pieces of mass density which are not captured in Fig. 10.5. In particular,

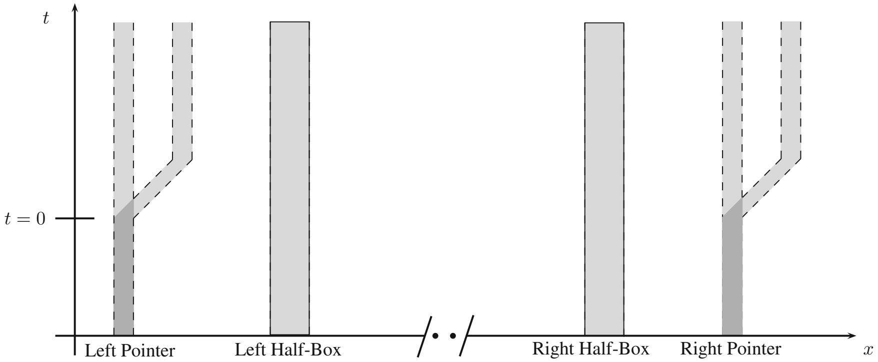

Fig. 10.5 Alice and Bob perform simultaneous measurements to detect the presence of a particle which is initially "split" between their two locations. Alice's and Bob's pointers are initially in their undeflected, "ready" positions, and the mass density associated with the particle is split between the two half-boxes. After $t = 0$, when both Alice and Bob each initiate an interaction which causes their pointer to deflect if the particle is present in their half-box, the two pointers "split", with half of each pointer's mass density remaining in the undeflected position (indicating the non-detection of the particle) and the other half of each pointer's mass density deflecting to the right (indicating the successful detection of the particle). Overall, it perhaps appears that there is no hint of nonlocality here: Alice's choice to initiate a measurement procedure causes her pointer (and shortly thereafter, her self!) to split, and similarly for Bob and his pointer, and the contents of the two half-boxes never change and hence appear unaffected by any of the measurements, distant or otherwise

there are facts, implied by Eq. (10.28), about which pieces of mass density are in the same world – the same branch – as each other. Remember here that the decomposition of the total mass density field ρ into the sum, $\rho_1 + \rho_2$, is robustly implied by the fact that the wave function is itself the sum of two disjoint terms, i.e., the two terms which are extremely well-separated in configuration space. So the very fundamental concept of Everett's theory not just allows, but requires, us to consider the separate branch-identities of these terms. See Fig. 10.6 for an attempt to visualize the same process again, but now including these relational facts about which pieces of mass density are "in the same universe as" which other pieces.

The point of this further elaboration is to stress the following: it is *not* the case that Alice's measurement merely causes a "local splitting" of her pointer and her self, with Bob's measurement also causing a second, *independent*, "local splitting" of his pointer and his self. If the two splittings were independent in this sense, you would expect that, if for example Alice and Bob get together later in the day to compare notes on the outcomes of their experiments, the interaction between the two Alices and the two Bobs would generate *four* branches: (i) one in which "yes-Alice" (i.e., the Alice who detected that the particle was present at her location) meets "yes-Bob"; (ii) one in which "yes-Alice" meets "no-Bob"; (iii) one in which "no-Alice" meets

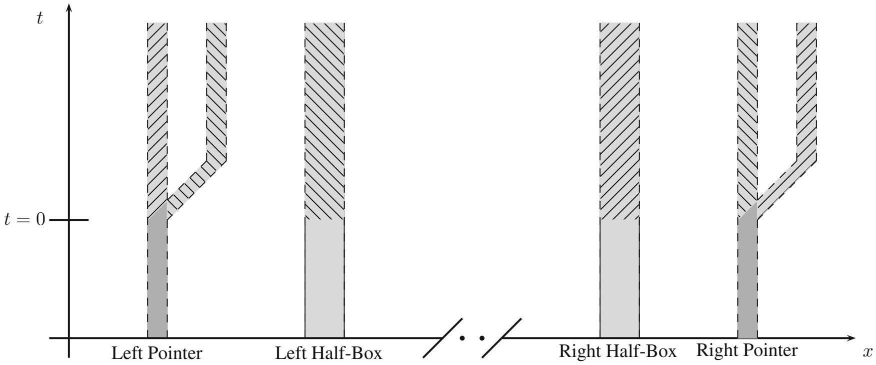

Fig. 10.6 Same as Fig. 10.5 but with the various contributions to the mass density ρ now marked to indicate the splitting into two distinct worlds or branches: The down-to-the-right striping indicates the world in which the particle is found on the left, by Alice, while the up-to-the-right striping indicates the world in which the particle is found on the right, by Bob. (Note that, prior to the measurements, the mass density associated with the particle is not identified with one or the other of these worlds, even though there would be a rather obvious way of doing this and it wouldn't be terribly misleading to do it. The reason is that, as long as the splitting of the wave function into two terms remains based on a purely microscopic difference, we don't really have separate worlds at all, in Everett's sense. (This is immaterial as long as Alice and Bob are inevitably going to carry out their measurements as we have been describing. But in principle, prior to their doing this, they could decide instead to bring the two half-boxes back together in the middle and perform some kind of interference experiment instead, and we would expect that they would indeed be able to see interference. This remains a possibility precisely because no actual branching in Everett's sense occurs until a more macroscopic number of degrees of freedom is involved in the decomposition of the wave function into disjoint terms

"yes-Bob"; and (iv) one in which "no-Alice" meets "no-Bob". But this is not right. There will not be a branch in which "yes-Alice" meets "yes-Bob" and there will not be a branch in which "no-Alice" meets "no-Bob". Only branches (ii) and (iii) will actually exist later, if Alice and Bob get together to chat, because *already*, after they have each completed their measurements but not yet gotten together to chat, there are just two distinct worlds.

This is inherent in the mathematical structure of the wave function even though it does not appear in the mass density field ρ. In general, it should be clear that, in using Ψ to compute ρ, we lose (by integrating) a lot of information. (Remember the examples, from Chap. 5, of very different wave functions which all produce the same mass density fields.) So while ρ is supposed, on this understanding of Everett's theory, to tell us what is going on in 3D physical space, there is in some sense much more that is true about goings on in physical space than is contained in the mass density field. In particular, as we are seeing here, there are relational facts – about which contributions to the mass field are in the same world as which other contributions – that really exist and have dynamical implications for how events will play out in the future as different sub-systems continue to interact with one another.

For our example here, the situation seems to be as follows. Alice's measurement of the contents of her half-box induce a splitting; she has two descendants, one of whom ("yes-Alice") sees the particle in her half-box and the other of whom ("no-Alice") fails to see the particle. Simultaneously, but at a distant location, Bob's measurement induces another splitting, and he too has two descendants, one of whom ("yes-Bob") sees and one of whom ("no-Bob") fails to see the particle in his half-box. But the two splittings are correlated despite their spatial separation: "yes-Alice" and "no-Bob" are, so to speak, born into the same post-measurement world, and "no-Alice" and "yes-Bob" are also born into the same post-measurement world. These spatially-separated birthings are correlated in a way that seems impossible to understand in any purely local way.

That said, I do not think it is possible to really argue cleanly, the way we have done for both the pilot-wave and spontaneous collapse theories, that there is an unambiguous violation of Bell's formulation of local causality. This is partly because that formulation is built around the concept of probability (it demands, remember, that a certain probability not change when distant events are specified) and the question about how to understand the usual quantum mechanical probabilities in the context of Everett's many worlds theory remains rather murky. It is also in part a result of the murkiness of the ontology posited by the theory. Thinking in terms of the mass density ontology at least gives us something reasonably clear that we can draw pictures of and think about (even though it is perhaps somewhat contrary to the Everettian spirit of insisting that the wave function *alone*, evolving in accordance with Schrödinger's equation *always*, is sufficient). But things still remain murky, with the intuitive non-locality somehow being associated with the relational facts which are not captured by the mass field.

And so the conclusion of this discussion will, unfortunately, but like our discussions of Probability and Ontology, be somewhat anti-climactic. It simply is not clear, in the context of Everett's version of quantum theory, how we should understand and formulate the concept of “local causality”, whether we should think of the theory as local or non-local, or, indeed, whether we should care about whether the theory is in some sense local or not. (Note that the closely related – but, as we saw in Chap. 9, not identical – question of the theory’s compatibility with fundamental relativity also remains, I think, an open question.)

By way of bringing our discussion of the many-worlds theory to a close, I think it is helpful to acknowledge the truly shocking nature of the idea, in Bryce de Witt’s description, that the

> ...universe is constantly splitting into a stupendous number of branches, all resulting from the measurementlike interactions between its myriad components. Moreover, every quantum transition taking place on every star, in every galaxy, in every remote corner of the universe is splitting our local world on earth into myriads of copies of itself [8].

As de Witt continues:

> I still recall vividly the shock I experienced on first encountering this multiworld concept. The idea of 10^{100+} slightly imperfect copies of oneself all constantly splitting into further copies, which ultimately become unrecognizable, is not easy to reconcile with common sense. Here is schizophrenia with a vengeance [8].

I think it should be admitted, however, that although our intuitions recoil at this suggestion, the picture is compelling and elegant as an approach to addressing the measurement problem of ordinary quantum mechanics, and should be regarded (despite its initially shocking character!) as seemingly compatible with experience.

On the other hand, I think it must also be admitted that the theory does not yet provide sufficient clarity regarding the several issues we focused on in this chapter – probability, ontology, and locality – and that it therefore remains impossible to assess in anything like a final or conclusive way. It remains, to a greater extent than the pilot-wave theory or the spontaneous collapse theories, a work-in-progress.

Projects:

10.1 According to Everett’s theory, if your friend measures the z-component of the spin of a spin-1/2 particle that is initially in the state ψ_{+x}, he gets into an entangled superposition (with the spin-1/2 particle and the measuring equipment) in which he experiences, in some sense, *both* outcomes: spin-up and spin-down. So, how will your friend respond if you ask him which outcome he experienced? Explain.

10.2 True or false: according to Everett’s theory, matter is made of particles. Explain.

10.3 Suppose a measurement of the energy of a particle-in-a-box produces the joint PIB-pointer state

$$\psi(x, y) = \frac{1}{\sqrt{2}} [\psi_1(x)\phi_1(y) + \psi_2(x)\phi_2(y)] \tag{10.29}$$

where $\psi_n(x)$ is the nth energy eigenstate for the PIB and $\phi_n(y)$ is a pointer state that is sharply peaked around $y = Y_n$, the position that indicates the nth outcome for the energy measurement. What is the state of the pointer *relative to* the PIB having state $\psi_1(x)$? What is the state of the PIB *relative to* the pointer having state $\phi_2(y)$? What is the state of the pointer *relative to* the PIB having state $\frac{1}{\sqrt{2}}\left[\psi_1(x) + \psi_2(x)\right]$?

10.4 It was claimed in Sect. 2 that, if the worlds are appropriately weighted in the counting, we can still say that the overwhelming number of worlds display outcome statistics that are compatible with the Born rule. Show in particular that, if Eqs. (10.9) and (10.13) are plugged into Eq. (10.14), the resulting function $f_N^w(n)$ does indeed peak at $n = Np$. (Hint: set d/dn of $\ln[f]$ to zero, and use the Stirling approximation, $\ln(m!) \approx m \ln(m)$.)

10.5 Consider the case of $N = 4$ "quantum coin flips", as discussed in Sect. 10.2, but with the branch weight for the H outcomes being $p = 3/4$. There is one world in which the sequence $HHHH$ is observed; its weighted world-fraction is therefore $f^w(4) = 1 \times (3/4)^4 \approx 0.316$. Calculate in a similar way $f^w(n)$ for $n = 0, 1, 2,$ and 3. Which value of n produces the largest weighted world-count? Is this what you would expect?

10.6 Suppose an observer measures some quantity (on a system that is not initially in an eigenstate for that quantity), and then subsequently re-measures the same quantity again, but using a different measuring apparatus. (Thus, there will be three degrees of freedom involved – say, "x" for the system whose property is being measured, "y" for the position of the pointer of the first measuring apparatus, and "z" for the position of the pointer of the second measuring apparatus.) Will the results of the two measurements agree in each branch of the wave function, or will there be some branches (i.e., some worlds) in which the measurements disagree? Explain how this relates to the collapse postulate of ordinary QM.

10.7 Suppose a spin-1/2 particle is in the state ψ_{+z}. First its z-spin is measured, then its x-spin is measured, and then its z-spin is measured again. Will the results of the two z-spin measurements agree in each branch of the wave function, or will there be some branches (i.e., some worlds) in which the measurements disagree? Explain how this relates to the collapse postulate of ordinary QM.

10.8 Imagine that you live in Everett's universe and are about to perform a biased quantum coin flip with $p = 3/4$. Explain what is problematic with each of the following statements you might consider making: (i) "The probability that I will see H is $3/4$." (ii) "Of all my descendants, the probability that the one who is really *me* sees H is $3/4$." (iii) "There will be descendants in branches with all possible outcomes, but the probability that I will end up *experiencing* a branch with outcome H is $3/4$." Can you construct a similar statement, assigning a $3/4$ probability to *something*, that would actually make sense in the Everettian point of view?

10.9 In Everett's 1957 paper, Ref. [3], he gives the following derivation of the branch weighting rule. The goal is to find a function w which assigns weights to the different terms in $\psi = \sum_i c_i \psi_i$. If the individual factors ψ_i are properly normalized, then the weight assigned to a given term can only depend on the complex number c_i. But a pure phase can be absorbed into the states ψ_i, so that the weight function should only depend on the modulus of the expansion coefficients: $w\,(c_i) = w\,(|c_i|)$. Now, with further time-evolution, the nth branch will split into additional sub-branches: $c_n\psi_n = \sum_j a_j\phi_j$. Assuming again that all the states (ψ_n and the ϕ_j) are properly normalized, this implies $|c_n|^2 = \sum_j |a_j|^2$. Now, If we require that this further splitting preserves the total weight of the involved branches, we have

$$w(c_n) = \sum_j w(a_j). \tag{10.30}$$

Everett calls this the "additivity requirement". Using the above, it implies

$$w\left(\sqrt{\sum_j |a_j|^2}\right) = \sum_j w\left(|a_j|\right) \tag{10.31}$$

which implies that $w(x) = cx^2$ for some constant c that will be 1 if the total weight is normalized to 1. To see this, Everett suggests defining a new function $g(x) = w(\sqrt{x})$, in terms of which the previous equation reads:

$$g\left(\sum_j |a_j|^2\right) = \sum_j g\left(|a_j|^2\right). \tag{10.32}$$

One can see that this requires $g(x) = cx$. Fill in the gaps in the mathematics and reasoning here to make the argument fully clear. Then assess it. What, exactly, does it prove in the context of Everett's theory? Does the argument completely remove the circularity alluded to in the text?

10.10 In our preliminary discussion of the Einstein's Boxes scenario, depicted in Fig. 10.5, we said that "it perhaps appears that there is no hint of nonlocality here". Make this a little more formal by applying our modification of Bell's locality condition from Chap. 1, Eq. 1.28. Let χ_1 denote, say, the presence of a nonzero mass density, for the Left Pointer, at its undeflected position (after the measurement has gone to completion). And let χ_2 and χ_2' represent, respectively, nonzero mass densities, for the Right Pointer, at its undeflected and deflected positions. ($\mathcal{C}_\Sigma$ here just includes everything that was true prior to $t = 0$.) Show that the condition is formally respected.

10.11 The attitude of Everettians toward the issue of quantum non-locality is pretty-well captured by Everett's comments from 1957: "Consider the case where the states of two object systems are correlated, but where the two systems do not interact. Let one observer perform a specified observation on the first system, then let another observer perform an observation on the second system, and finally let the first observer repeat his observation. Then it is found that the first observer always [i.e., in each branch] gets the same result both times, and the observation by the second observer has no effect whatsoever on the outcome of the first's observations. Fictitious paradoxes like that of Einstein, Podolsky, and Rosen which are concerned with such correlated, noninteracting systems are easily investigated and clarified in the present scheme." [3] Why do you think Everett calls the EPR paradox "fictitious"? Explain how you understand Everett to be thinking about this kind of situation. Do you agree that there is just clearly nothing non-local going on here, according to the Everett theory, such that it makes sense to call EPR's suggestion of non-locality "fictitious"? Explain.

10.12 Everettians, starting with David Deutsch, have accused the pilot-wave theory of being a "parallel universe theor[y] in a state of chronic denial [9]." The basis for this accusation is the fact that the pilot-wave theory also has the wave function of the universe obeying Schrödinger's equation all the time. So the many-worlds structure that Everettians find in that wave function is, they argue, just as present in that wave function in the pilot-wave picture, as it is in the wave-function monist Everettian picture. What do you make of this accusation? How do you think a proponent of the pilot-wave theory would or should respond?

10.13 Tim Maudlin pointed out in Ref. [10] that GRWm (but, interestingly, not GRWf) also has a kind of many-worlds character: since the GRW localizations involve multiplication by a Gaussian function which is small (but never quite zero) far from the Gaussian's center, the mass density field associated with the "un-selected" branches of the wave function is, while very small compared to the "selected" branch, not zero. What do you think? Is GRWm really a single-universe theory (because those other, "un-selected" worlds are so dim that it is reasonable to ignore them), or is it really a many-worlds theory in denial (because, dim or not, and anyway the dimness isn't visible from the inside, those "un-selected" worlds have all the right structure to count as real worlds)?

10.14 Proponents and critics of Everett's theory both sometimes appeal to Occam's razor in support of their position. The proponents say that, because the theory dispenses with the measurement axioms of ordinary QM (and because it doesn't replace those with anything like additional dynamical laws for "hidden variables"), Everettism is by far the simplest, most parsimonious version of quantum theory. On the other hand critics say that Everett's worldview, with the huge number of "parallel universes" that are totally unobservable to us, is ridiculously extravagant. Explain precisely how each side interprets and applies Occam's razor, i.e., explain what leads the two sides to these two opposite conclusions even though they are allegedly appealing to the same criterion. What do you think? Is Everett's theory clean and elegant, or ugly and complicated?

10.15 David Deutsch has argued that evidence for an Everettian multiplicity of universes is ubiquitous:

> The point that theorists tend to miss is that the multiplicity of reality is not only, or even primarily, a consequence of quantum *theory*. It is quite simply an observed fact. Any interference experiment (such as the two-slit experiment), when performed with individual particles one at a time, has no known interpretation in which the particle we see is the only physical entity passing through the apparatus. We know that the invisible entities passing through obey the same phenomenological equations of motion ... as the single particle we do see. And we know from [EPR] type experiments, such as that of Aspect, that these not-directly-perceptible particles are arranged in extended 'layers' each of which behaves internally like an approximately classical universe. Admittedly all these observations detect other universes only indirectly. But then we can detect pterodactyls and quarks only indirectly too. The evidence that other universes exist is at least as strong as the evidence for pterodactyls or quarks [9].

What do you think of this argument? Is there really no single-universe theory that can explain the results of the double-slit experiment?

10.16 In the last section of Chap. 9, we saw that certain regions of the $\{\lambda, \sigma\}$ parameter space for spontaneous collapse theories were empirically refuted, and certain other regions were considered "Perceptually/Philosophically Unsatisfactory". The many-worlds theory can be thought of as a spontaneous collapse theory, but with collapse rate $\lambda = 0$. So is the many-worlds theory "Perceptually/Philosophically Unsatisfactory"? Explain the assumption that is made in diagnosing small-λ versions of spontaneous collapse theory as unsatisfactory, and how Everett would challenge this assumption.

References

1. M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974)
2. B. Peter, Everett and wheeler, the untold story, in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)
3. H. Everett, [The] 'Relative State' formulation of quantum mechanics. Rev. Mod. Phys. **29**(3), 454–462 (1957)
4. B.S. de Witt, N. Graham, *The Many-Worlds Interpretation of Quantum Mechanics* (Princeton University Press, Princeton, 1973)
5. S. Saunders, Many worlds? An introduction in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)
6. S. Goldstein, V. Allori, R. Tumulka, N. Zanghi, Many-worlds and Schrödinger's first quantum theory. Br. J. Philos. Sci. **62**(1), 1–27 (2011), arXiv:0903.2211
7. D. Wallace, Decoherence and ontology in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)
8. B. de Witt, Quantum mechanics and reality. Phys. Today **23**(9), 30–35 (1970)
9. D. Deutsch, Comment on lockwood. Br. J. Philos. Sci. **47**, 222–8 (1996)
10. T. Maudlin, Can the world be only wavefunction? in *Many Worlds? Everett, Quantum Theory, and Reality*, ed. by S. Saunders, J. Barrett, A. Kent, D. Wallace (Oxford University Press, Oxford, 2010)