Chapter 9
The Spontaneous Collapse Theory

Commenting on the quantum measurement problem as illustrated by Schrödinger's infamous cat, Bell remarked: "Either the wavefunction, as given by the Schrödinger equation, is not everything, or it is not right [1]." We have seen, in Chap. 7, how the measurement problem can be avoided if the wave function is not everything: by *supplementing* the wave function with additional objects (like the always-definite positions of particles in the pilot-wave theory) we can have a theory which actually predicts that definite things should happen, without anything like *ad hoc* and ill-defined exceptions to the usual dynamical rules.

In this chapter, we explore the other possibility mentioned by Bell – that the wavefunction, as given by the Schrödinger equation, isn't right. Recall that the essence of the measurement problem was that, according to Schrödinger's equation, interactions between (for example) a particle and a measuring apparatus do not typically result in the measuring apparatus pointer having a definite post-interaction position. Instead, the measuring apparatus gets infected with whatever quantum superposition was present in the initial state of the particle being measured. In ordinary quantum mechanics, this seemingly problematic result is already avoided in the way suggested by Bell – the "collapse postulate" is precisely a claim that the wave function *as given by Schrödinger's equation* isn't right. In particular, when a *measurement* occurs, the wave function ceases to evolve according to Schrödinger's equation, and (momentarily) does something entirely different instead. This avoids the seemingly problematic idea of superpositions of macroscopically distinct situations, but at a heavy price: it seems unbelievable that there is a fundamental distinction between "measurement" and "non-measurement" processes. Somehow, the true fundamental theory should treat all processes in a consistent, uniform fashion.

The "spontaneous collapse" theory is, at root, an attempt to remove this troubling dualism by positing, for the wave function, a single, universally-applicable dynamical evolution law which will somehow accomplish, in a single stroke, the two jobs done respectively by the Schrödinger equation and the collapse postulate in ordinary QM. The idea, more specifically, is to *modify* Schrödinger's equation with stochastic

T. Norsen, *Foundations of Quantum Mechanics*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-3-319-65867-4_9

non-linear terms which will have the effect of preserving the Schrödinger evolution for microscopic systems (where we know it is correct) but also ensuring that macroscopic things like pointers (and cats!) end up in the sorts of definite, non-superposed states we observe them to always end up in.

9.1 Ghirardi, Rimini, and Weber

The main idea of the spontaneous collapse theory is sometimes traced back to a 1966 paper by David Bohm and Jeffrey Bub, which explores a type of hidden variable theory rather unlike Bohm's 1952 pilot-wave theory. In the pilot-wave theory, of course, the additional variables (namely, the positions of the particles) are controlled by the wave function, which thus plays a somewhat mysterious background role. In the 1966 paper, by contrast, it is the hidden variables – here something like a background field – which influence the evolution of the wave function and give rise to deviations from the normal Schrödinger-equation evolution thereof.

This motivated Philip Pearle and also, somewhat later, Nicolas Gisin – both of whom were very concerned by the measurement problem of ordinary QM – to begin exploring stochastic modifications of the usual Schrödinger equation. Some progress was made toward the goal of reconciling wave function dynamics with the appearance of definite outcomes, but no systematic method of achieving the desired ends was identified, and several difficulties (including for example the apparent inevitability of conflicts with relativity when deviations from the Schrödinger evolution were contemplated) were brought into sharper focus.

A breakthrough appeared in 1986, when three Italian physicists (Ghirardi, Rimini, and Weber – hereafter "GRW") took fuller advantage of the fundamentality of *position*: if you can get the *positions* of macroscopic things right, then you will automatically get other properties right as well, since the outcomes of measurements of other properties (such as energy, momentum, spin, etc.) are always registered in the position of some macroscopic object (like our ubiquitous pointer) [2]. So whereas the earlier proposals had struggled with the problem of deciding which basis to use in narrowing the wave function (does one narrow in momentum space when a momentum measurement is happening?), GRW proposed the simple and elegant idea that wave functions should occasionally (randomly, spontaneously) localize exclusively in position space.[1]

In the theory, it is as if, at randomly selected moments, some outside observer makes a (somewhat rough) position measurement and thus collapses the particle's wave function (but, due to the roughness, to a finite-width Gaussian wave packet

[1]Note that there is an interesting parallel here to the pilot-wave theory, which eludes the "no hidden variables" theorems by letting non-position properties (such as momentum, energy, and spin) be "contextual". This difference, between the way position and other properties are treated by the theory, does not prevent the theory from generating correct empirical predictions for measurements of non-position properties since even the outcomes of momentum/energy/spin measurements are registered, at the end of the day, in the position of some pointer.

rather than a perfectly sharp delta function). But of course the whole point of the theory is to avoid the idea of some mysterious "outside observer" whose interventions imply exceptions to the usual dynamical behavior... hence "as if". According to GRW, the occasional collapses or "localizations" of the wave function should be considered as purely natural – part of the ordinary, universal way that wave functions evolve in time.

In just a moment, we'll talk through the technical details of these spontaneous localizations, starting first, in the present section, with the simple case of a single particle (in 1-D for simplicity). Then in the following sections we will explain how the theory describes multi-particle systems, including those that we would commonly describe as involving "measurements".

But first, let me just acknowledge that the theory, as it will be explained, maybe doesn't seem to do a very good job of truly *unifying* the two different types of wave-function time-evolution posited by ordinary QM. The GRW evolution will amount to: wave functions just evolve according to Schrödinger's equation most of the time, except for these occasional random moments when they instead suffer a spontaneous localization. The supposed unification here perhaps feels a bit like taking these two allegedly incompatible dynamical evolution laws, wrapping a bow around both of them together, and saying "Voila!" Putting this point another way, it may feel like there is somehow not much difference between the GRW theory and standard textbook QM: whereas orthodox QM says "Wave-functions evolve according to Schrödinger's equation, except during measurements, when they instead collapse" GRW says "Wave-functions evolve according to Schrödinger's equation, except at certain random moments, when they instead collapse." Other than a minor change in the words, is there really any difference?

It's a fair question, in response to which two things might be said.

One is that while the GRW process does indeed have a somewhat implausibly dualistic character, this can to at least some degree be eliminated. For example, models have been developed (especially by Pearle) in which, instead of abrupt intermittent wave function collapses, one has gentler localizations that are occuring continuously in time. The net effect (that is, something like the total amount of localization that happens per unit time) is roughly the same, but – because the Schrödinger-type and collapse-type evolutions are simultaneous and omnipresent – the dynamics feels a little more natural, coherent, and plausible. These so-called "continuous spontaneous localization" (CSL) models also solve some other technical issues with the GRW process as we will explain it. So, to some degree, one can take our discussion of the GRW process merely as a kind of pedagogical simplification of an overall concept which can perhaps be implemented somewhat more elegantly.

A second point, though, is that to some extent, no matter how you slice them up, "spontaneous collapse" theories just are somewhat dualistic. They are, after all, literally designed to unify the two dynamical postulates of ordinary QM. The duality, then, is somehow a problem only if one is expecting something that is somehow radically different from ordinary QM. But perhaps we should not expect that, and we should instead view the spontaneous collapse idea simply as an attempt to replace the "loose talk" (about "measurements" and "observers") in ordinary QM, with sharp mathematics. From this point of view, we should not regard the GRW theory so much as an alternative to ordinary, textbook QM, but rather as something like "ordinary QM v2.0". In this vein, Bell said about the GRW theory: "I do think [the spontaneous collapse theories have] a certain kind of goodness... in the sense that they are honest attempts to replace the wooly words by real mathematical equations – equations which you don't have to talk away – equations which you simply calculate with and take the results seriously [3]."

All right. With all of that as preamble, let's finally jump into exploring in mathematical detail how the GRW theory works.

So, as has been said, the theory posits that the wave function of a single particle evolves according to Schrödinger's equation

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial x^2} + V(x,t)\psi(x,t) \tag{9.1}$$

most of the time. But the Schrödinger evolution is interrupted by occasional localizations. The Schrödinger-equation part of the evolution is already well-understood, so we will focus our exposition on the localizations.

First, when do they happen? For a single particle, there is supposed to be a constant probability per unit time, $\frac{dP}{dt} = \lambda$, for a spontaneous localization to occur. This will give rise to a (Poisson-distributed) sequence of times $t_1, t_2, t_3, \ldots$, with an average "waiting time" $\tau = t_{n+1} - t_n$ between the localizations given by $\tau = 1/\lambda$. For reasons that will be discussed as we proceed, GRW suggest that the constant λ should have a value in the neighborhood of

$$\lambda \approx 10^{-16}\, s^{-1} \tag{9.2}$$

so that the average time between localizations is

$$\tau = \frac{1}{\lambda} \approx 10^{16}\ \text{s} = 3 \times 10^8\ \text{years.} \tag{9.3}$$

That's three hundred million years – a very long time! So, for a single particle, the spontaneous localizations are quite rare. It may even appear that localizations occuring at such a slow rate would be totally negligible. But, as we will see in the next section, they will become quite important in the evolution of macroscopic systems containing a large number of particles. Before turning to that, though, let's understand in more precise detail what exactly happens at one of these intermittent localization events.

Consider the Gaussian function

$$g_r(x) = \frac{1}{(2\pi\sigma^2)^{1/4}} e^{-(x-r)^2/4\sigma^2} \tag{9.4}$$

which has a half-width of about σ, is centered at the point $x = r$, and is normalized in the following sense:

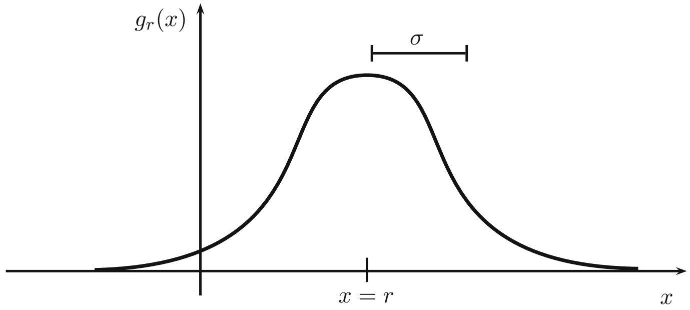

Fig. 9.1 The localization of a wave function in the GRW theory basically consists of its being multiplied by the Gaussian function $g_r(x)$ shown here

$$\int_{-\infty}^{\infty} |g_r(x)|^2 \, dx = 1. \tag{9.5}$$

(Since $g_r(x)$ is real-valued, the absolute value bars here are unnecessary but, of course, totally harmless.) The function $g_r(x)$ is shown in Fig. 9.1. Note that, again for reasons that will emerge as our presentation proceeds, the value of the constant σ is postulated by GRW to have a value in the neighborhood of

$$\sigma \approx 10^{-7}\,\text{m}. \tag{9.6}$$

It is important that this is fairly small on the macroscopic scale, but fairly large compared to, for example, the size of an atom.

The basic idea is then that, during an episode of "spontaneous localization", the wave function gets suddenly multiplied by $g_r(x)$. Suppose one of the localizations happens at time t. Then the wave function $\psi(x, t^+)$ just after time t is given by

$$\psi(x, t^+) \sim g_r(x)\psi(x, t^-) \tag{9.7}$$

where $\psi(x, t^-)$ is what the wave function was right *before* time t. That is the basic idea, but there are a couple of mathematical details to iron out.

First of all, I wrote "$\sim$" rather than "$=$" just above because the product on the right hand side will not generally be a properly normalized wave function. This is easy enough to fix by writing instead

$$\psi(x, t^+) = \frac{g_r(x)\psi(x, t^-)}{N(r)} \tag{9.8}$$

where the re-normalization factor $N(r)$ given by

$$N(r)^2 = \int \left|g_r(x)\psi(x, t^-)\right|^2 \, dx \tag{9.9}$$

ensures that $\psi(x, t^+)$ is properly normalized:

$$\int \left|\psi(x, t^+)\right|^2 dx = \frac{1}{N(r)^2} \int \left|g_r(x)\psi(x, t^-)\right|^2 dx = 1. \tag{9.10}$$

The second mathematical detail addresses the question: what is the value of r, i.e., what point does the wave function get localized *around*? The answer is that r is *random*, with a probability distribution

$$P(r) = N(r)^2 = \int \left|g_r(x)\psi(x, t^-)\right|^2 dx. \tag{9.11}$$

This says, basically, that the wave function is most likely to localize around some point $x = r$ where the wave function modulus is large to begin with. In a little more detail, it says that the probability for localization at the point $x = r$ is proportional to what would usually be regarded as the total probability associated with the new (but not yet normalized) localized state, if the localization did occur at $x = r$. (You are invited to prove that $P(r)$ as defined here really is a valid probability distribution in the Projects.)

Let's work through a couple of simple examples to clarify the idea.

To begin with, suppose the wave function is initially extremely spread out so that it has (say, over some region of width $L \gg \sigma$) a *constant* value:

$$\psi(x, t^-) = \frac{1}{\sqrt{L}}. \tag{9.12}$$

(The actual value here doesn't matter much for our purposes, but we might as well take the wave function to be properly normalized.) Now, at time t, let's say a spontaneous localization happens to occur. It is (approximately) equally likely to occur at any point r where the wave function $\psi(x, t^-)$ has support, since

$$P(r) = N(r)^2 = \int \left|g_r(x)\psi(x, t^-)\right|^2 dx \approx \begin{cases} 1/L & \text{where } \psi(x, t^-) = 1/\sqrt{L} \\ 0 & \text{where } \psi(x, t^-) = 0 \end{cases}. \tag{9.13}$$

The reason for the "$\approx$" is that technically, at the edges of the width-L region where the wave function is initially non-zero, P(r) will be a little smaller than $1/L$, and similarly it'll be a little bigger than zero just outside that region where $\psi(x, t^-) = 0$. That is, $P(r)$ will have a smooth transition at the edges, as shown in the lower graph of Fig. 9.2. But still, leaving aside the edge effects, we can say that the localization is equally likely to occur at any point in the width-L region where the wave function was nonzero.

And of course, after the localization, the multiplication of the initially-constant wave function by the Gaussian $g_r(x)$ produces a Gaussian wave function, centered at the randomly-selected point r. The transition from $\psi(x, t^-)$ to $\psi(x, t^+)$ is sketched in the upper graph of Fig. 9.2.

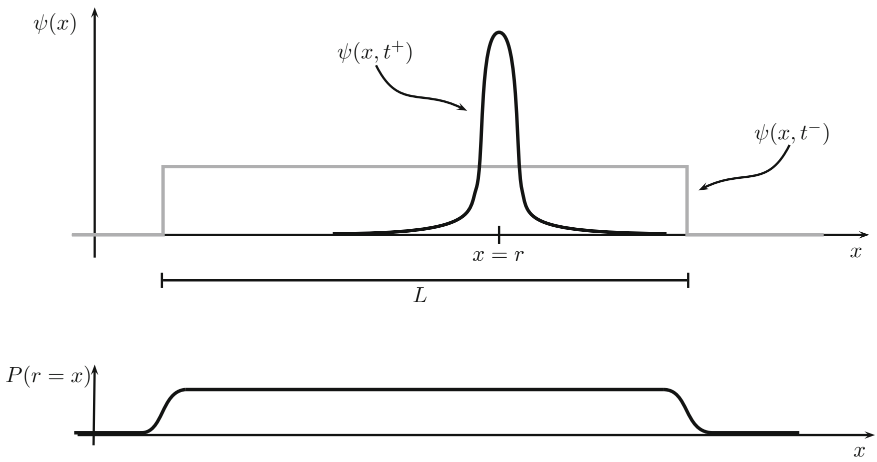

Fig. 9.2 If the wave function $\psi(x, t^-)$ is roughly constant over some region, a spontaneous localization will narrow it down to a Gaussian, of width σ, centered at some point r which is (approximately) equally likely to be anywhere in the region where $\psi(x, t^-)$ was nonzero

OK, so, if the wave function is initially very spread out compared to the length scale σ, a spontaneous localization does exactly what is advertised – it *localizes* the wave function around some new, randomly selected point where the wave function was originally big.

As a second example, let's take the opposite limit, where the wave function is already, initially, very narrowly peaked. It is convenient to take the extreme limiting case of a position eigenstate, i.e., a wave function which is a Dirac delta function:

$$\psi(x, t^-) \sim \delta(x - a). \tag{9.14}$$

Of course, this is not a properly normalized state, and (while not really making any difference at the end of the day), that will be slightly annoying as we try to figure out what our general formulas imply for things like the probability distribution $P(r)$. With apologies to any mathematicians who are reading, we can elude these problems in a simple way by writing

$$\psi(x, t^-) = \frac{1}{\sqrt{\delta(0)}}\,\delta(x - a). \tag{9.15}$$

Now suppose a spontaneous localization happens. This means the wave function will be multiplied by a Gaussian centered at some random point r. What is the probability distribution for this point? Well,

$$P(r) = N(r)^2 = \int \left| \frac{g_r(x)\delta(x - a)}{\sqrt{\delta(0)}} \right|^2 dx = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(r-a)^2/2\sigma^2} = g_a(r)^2. \tag{9.16}$$

That is, $P(r)$ is a Gaussian function, of width σ, centered at the same point $x = a$ where the wave function is initially concentrated. That makes sense.

So some point r (within about σ either way from $x = a$) is randomly chosen. What does the wave function look like after multiplication by $g_r(x)$ and re-normalization? Well,

$$\psi(x, t^+) = \frac{g_r(x)\psi(x, t^-)}{N(r)} = \frac{g_r(x)}{N(r)} \frac{\delta(x-a)}{\sqrt{\delta(0)}} = \frac{g_r(a)}{N(r)} \frac{\delta(x-a)}{\sqrt{\delta(0)}} = \psi(x, t^-) \tag{9.17}$$

since, as we just showed, $N(r) = g_r(a)$. In this case as shown in Fig. 9.3, the spontaneous localization actually doesn't change the wave function at all! (And note in particular that the wave function stays the same no matter which value of r was selected.) This actually makes sense: a δ-function wave function is already as localized as it is possible for a wave function to be, so the spontaneous localization doesn't change it at all.

Another interesting case to consider is an initially Gaussian wave function of width w_0. It turns out that, for $w_0 \gg \sigma$, the width after the localization decreases to about σ, whereas if $w_0 \ll \sigma$, the width is basically unaffected by the "localization". That is, this case smoothly interpolates between the two extremes embodied in our two examples. Rather than pursue that here, though, I'll let you work it out in the Projects.

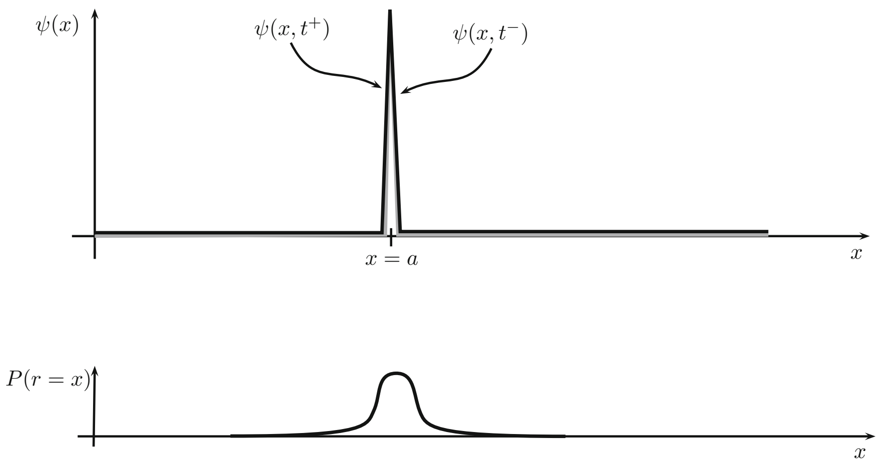

Fig. 9.3 If the wave function $\psi(x, t^-)$ is a δ-function (centered at $x = a$), then the probability distribution $P(r)$ is a width-σ Gaussian centered at $x = a$. But that turns out to be irrelevant because, no matter what value of r is chosen, the wave function is unaffected by the spontaneous localization: $\psi(x, t^-) = \psi(x, t^+)$

For a third and final example here, let's consider an "Einstein's boxes" kind of situation in which a particle is in a superposition of two relatively-sharply-defined positions. Concretely, suppose that

$$\psi(x,t^-) = \frac{1}{\sqrt{2}}\left[\frac{\delta(x+a)}{\sqrt{\delta(0)}} + \frac{\delta(x-a)}{\sqrt{\delta(0)}}\right] \tag{9.18}$$

so that the particle is in a 50/50 superposition of "being at $x = -a$" and "being at $x = +a$". And let us assume that the two possible positions here are very distant, i.e., $a \gg \sigma$.

The spontaneous localization process in this example is illustrated in Fig. 9.4. The probability distribution $P(r = x)$ for where the localization will be centered consists of two symmetric Gaussian functions centered at $x = +a$ and $x = -a$ respectively. Suppose that, by chance, $r \approx +a$. Then we have that

$$N(r)^2 = \frac{1}{2}\left[g_a(-a)^2 + g_a(+a)^2\right] = \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-2a^2/\sigma^2} + 1\right]. \tag{9.19}$$

For $a \gg \sigma$, the first term is extremely small compared to 1 and we may thus take

$$N \approx \frac{1}{\sqrt{2}}\frac{1}{(2\pi\sigma^2)^{1/4}}. \tag{9.20}$$

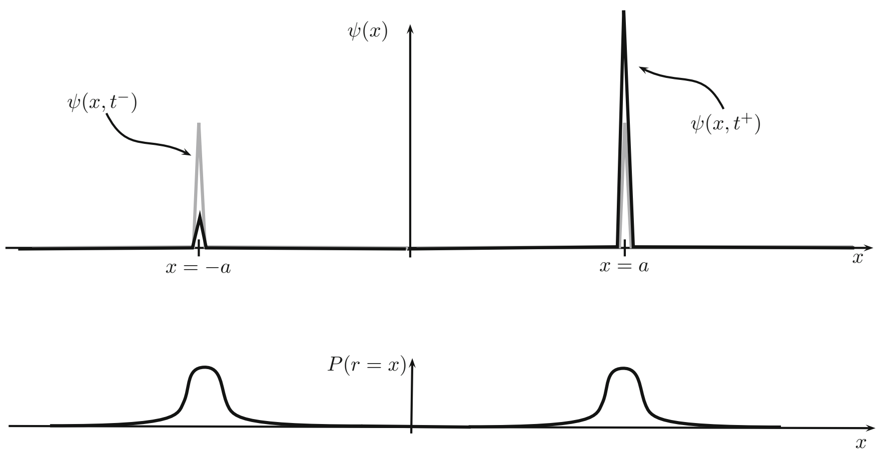

Fig. 9.4 If the wave function $\psi(x, t^-)$ is a superposition of two δ-functions, separated by a distance much larger than σ, the localization promotes one of the δ-functions while greatly suppressing the size of the other. For all practical purposes, the post-localization wave function is just one or the other of the previously-superposed spikes, so the localization has the effect of erasing spatial superpositions over a length scale greater than σ. Note that, as shown in the $P(r = x)$ graph below, the localization is equally likely to promote the $x = a$ or the $x = -a$ term. What is shown above is, obviously, the case in which $r \approx +a$ so that $\psi(x, t^+)$ is basically $\delta(x - a)$ – but with just a tiny bit of $\delta(x + a)$ remaining as well

The post-collapse wave function is then given by

$$\begin{aligned}\psi(x,t^+) &= \frac{g_{+a}(x)\psi(x,t^-)}{N} \\ &= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{\delta(0)}}\frac{1}{N}\left[g_{+a}(x)\delta(x+a)+g_{+a}(x)\delta(x-a)\right] \\ &\approx \frac{\frac{1}{\sqrt{2}}g_a(a)}{N}\frac{\delta(x-a)}{\sqrt{\delta(0)}} \\ &= \frac{\delta(x-a)}{\sqrt{\delta(0)}} \end{aligned} \tag{9.21}$$

where we have again thrown out a term that is small by a factor like e^{-2a^2/σ^2} which is extremely small if $a \gg \sigma$.

So basically, if $r \approx +a$, the localization completely annihilates the delta function spike at $x = -a$ and leaves only a (re-normalized) spike at $x = +a$. (Of course, it was equally probable that instead we would have had $r = -a$ in which case the fates of the two spikes would have been reversed.) A particle which is in a superposition of two distinct locations (separated by a distance greater than σ) will not remain in that superposition forever; instead, according to GRW, the particle will eventually be located *definitely on the left* or *definitely on the right* – and this transition will happen spontaneously, without the need of anything like an external intervention or observation.

It might occur to you to worry that this spontaneous localization could destroy the interference that is observed in, for example, the two-slit experiment: if the two slits are separated by a distance greater than $\sigma \approx 10^{-7}$ m – and in typical demonstrations of interference, they are! – then the wave function of a particle which happens to suffer a spontaneous localization while it is traversing the 2-slit apparatus would *not* form an interference pattern at the screen, but would instead form something like a single-slit diffraction pattern. Does this mean that the GRW theory contradicts the observation of interference? No, for recall that, according to the theory, an individual particle only suffers a spontaneous localization every 300 million years or so. So unless you have a *lot* of time on your hands and send particles through the apparatus *very* slowly, you would never expect to see deviations from the usual quantum mechanical predictions in this kind of situation. Virtually all of the particles sent through would remain uncollapsed during the entire duration of their journey from source to screen.

9.2 Multiple Particle Systems and Measurement

As hinted at before, the incredible slowness/rarity of the GRW localizations might make one think that the localizations can just be completely ignored and will play no role whatever in the theory's predictions. But that is only true as long as we are thinking of individual particles. To understand the role of the localizations in the

GRW theory's solution to the measurement problem, we therefore need to see how the theory describes multi-particle systems.

The generalization of the theory to many-particle systems is pretty straightforward. In a nutshell, the idea is just that each individual particle suffers spontaneous localizations in the same way that we described in the previous section. How things play out then depends importantly on whether or not there is *entanglement*. Let's begin by discussing the simpler case in which there is no entanglement.

Consider, then, a two-particle system which, at the moment t^- just before a spontaneous localization occurs, is in the (non-entangled, i.e., factorizable) quantum state

$$\Psi(x_1, x_2, t^-) = \psi(x_1, t^-)\phi(x_2, t^-). \tag{9.22}$$

In a multi-particle situation like this, it is supposed to be irreducibly random *which* particle suffers the first localization (in addition to being irreducibly random exactly when and where that localization occurs). But, for definiteness, suppose that particle 2 suffers a localization at time t. Then, just as in the previous section, we have

$$\Psi(x_1, x_2, t^+) = \frac{g_r(x_2)\Psi(x_1, x_2, t^-)}{N(r)} \tag{9.23}$$

where, in the obvious generalization of what we saw previously,

$$N(r)^2 = \int \left| g_r(x_2)\Psi(x_1, x_2, t^-) \right|^2 dx_1\, dx_2. \tag{9.24}$$

And note that, also just as before, the probability density for the localization to be centered at the point r is $P(r) = N(r)^2$.

Plugging in the non-entangled two-particle state, Eq. (9.22), we see that

$$\Psi(x_1, x_2, t^+) = \psi(x_1, t^-)\frac{g_r(x_2)\phi(x_2, t^-)}{N(r)}. \tag{9.25}$$

The important point here is that when the overall quantum state involves no entanglement, an individual spontaneous localization only affects the particular particle that is "hit" by it. The overall state remains an unentangled product state, and the factors representing the wave functions of the *other* particles (in our example here, particle 1) are in no way affected by the localization.

But let us now turn to the more interesting case where there *is* entanglement between the two particles. Take, for simplicity, the same sort of example we considered in the previous section, in which a particle is located either at $x = +a$ or at $x = -a$, but suppose now there are *two* particles in a superposition of "both particles are at $x = +a$" and "both particles are at $x = -a$". Suppose in particular that, at a time t^- just before one of the particles happens to suffer a spontaneous localization, the two-particle wave function is

$$\Psi(x_1, x_2, t^-) \sim \frac{1}{\sqrt{2}} \left[\delta(x_1 - a)\delta(x_2 - a) + \delta(x_1 + a)\delta(x_2 + a)\right]. \tag{9.26}$$

Now, what happens to this wave function if one of the particles suffers a spontaneous localization? We cannot, as before, just say that "the wave function of one of the particles gets localized, while that of the other is unaffected"... For an entangled state like this the particles cannot even be said to possess their own individual wave functions! So let's just let the math tell us what happens, supposing, again arbitrarily, that it is particle 2 which nominally suffers the localization:

$$\Psi(x_1, x_2, t^+) = \frac{g_r(x_2)\Psi(x_1, x_2, t^-)}{N(r)} \tag{9.27}$$

where r is random, with probability distribution $N(r)^2$. Here, just as before, the definition of $N(r)$ is

$$N(r)^2 = \int \left|g_r(x_2)\Psi(x_1, x_2, t^-)\right|^2 dx_1\, dx_2. \tag{9.28}$$

This will be large in a small (size-σ) neighborhood around $r = +a$ as well as a small neighborhood around $r = -a$. That is, the spatial probability distribution for the center of the localization will look exactly like it did in the last example of the previous section.

Suppose that, for this particular localization, it happens that $r \approx +a$. Then (leaving out the uninteresting re-normalization factor) the two-particle wave function after the localization will look like

$$\begin{aligned}
\Psi(x_1, x_2, t^+) &\sim g_a(x_2)\Psi(x_1, x_2, t^-) \\
&= g_a(x_2)\frac{1}{\sqrt{2}} \left[\delta(x_1 - a)\delta(x_2 - a) + \delta(x_2 + a)\delta(x_2 + a)\right] \\
&= \frac{1}{\sqrt{2}} \left[\delta(x_1 - a)\delta(x_2 - a)g_a(x_2) + \delta(x_1 + a)\delta(x_2 + a)g_a(x_2)\right] \\
&= \frac{1}{\sqrt{2}} \left[\delta(x_1 - a)\delta(x_2 - a)g_a(a) + \delta(x_1 + a)\delta(x_2 + a)g_a(-a)\right]
\end{aligned} \tag{9.29}$$

Now here is the crucial point. The factor $g_a(a)$ (in the first term in the square brackets) is "big" – it is just the value of g at exactly the place where g peaks. But the factor $g_a(-a)$ (in the second term in the square brackets) is vanishingly small, if the separation ($2a$) between the two places the particles might have been is large

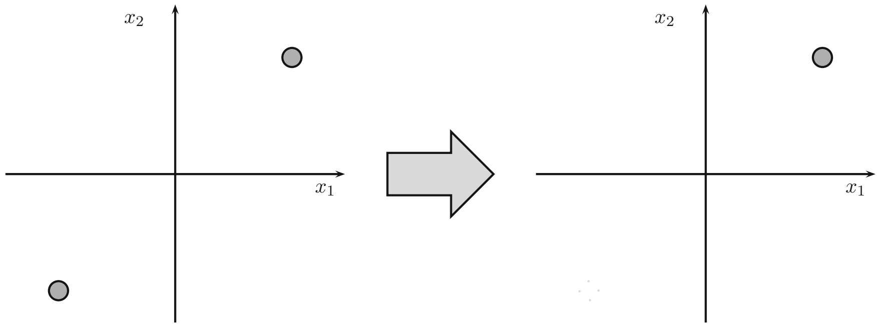

Fig. 9.5 The *left* graph is a configuration space map of the two-particle wave function $\Psi(x_1, x_2)$ for two particles in a superposition of "both particles on the *left*" and "both particles on the *right*". The *right* graph is the same map, after a single spontaneous localization. Despite nominally acting on just one of the two particles, the spontaneous localization gives rise to a wave function which has both particles localized together

compared to the width σ of the localization function g. So, to an excellent approximation, we have that, after the spontaneously localization (which, remember, was nominally associated with just one of the two entangled particles)

$$\Psi(x_1, x_2, t^+) = \delta(x_1 - a)\delta(x_2 - a) \tag{9.30}$$

which is of course a state in which *both* particles are definitely located at $x = a$. This process is illustrated in Fig. 9.5.

That was of course just one of several possibilities. We assumed arbitrarily that particle 2 happens to suffer the first spontaneous localization, and that this localization happens to be centered around $r = +a$. If you think through it, though, it should be clear that the final state would have been exactly the same had it been instead particle 1 that suffered a collapse centered near $r = +a$. So it doesn't actually matter which particle gets "hit" – *either* one getting localized localizes *both* because their positions started out in the special entangled state. And it should also be clear that, if either particle instead suffered a localization centered near $x = -a$, then both particles would have ended up definitely localized at $x = -a$.

Now we are finally in a position to understand how this spontaneous collapse theory solves the measurement problem. Suppose that, instead of just two particles being in an entangled state that binds their positions together (even as they remain in a superposition of, say, "all being on the left" or "all being on the right") it is instead some macroscopically-large number, like $N \approx 10^{23}$ particles whose positions are so bound. That is, suppose the initial state is something like

$$\Psi(x_1, x_2, \ldots, x_N, t^-) \sim \frac{1}{\sqrt{2}} \left[\delta(x_1 - a)\delta(x_2 - a)\cdots\delta(x_N - a) + \delta(x_1 + a)\delta(x_2 + a)\cdots\delta(x_N + a)\right]. \tag{9.31}$$

Then we can see that, as soon as *any one* of the N particles suffers a spontaneous localization, the *entire set* of particles will localize along the following lines:

$$\Psi(x_1, x_2, \ldots, x_N, t^+) \sim \delta(x_1 - r)\delta(x_2 - r) \cdots \delta(x_N - r) \tag{9.32}$$

with $r = +a$ or $r = -a$ with 50/50 probability. Everything is just the same as before, with one important exception. With just one particle, we would typically need to wait around $\tau = 300$ million years for the particle to spontaneously localize. For two particles, we would typically need to wait around $\tau/2 \approx 150$ million years. But for $N \approx 10^{23}$ particles, we would typically need to wait $\tau/N \approx 30$ nanoseconds. That is, because of the enormous number of individual particles comprising anything remotely macroscopic, a macroscopic object (like, say, a pointer or a cat) will suffer a constant barrage of spontaneous localizations (millions or billions or trillions of them per second), which will for all practical purposes prevent it from ever getting into the kind of macroscopic superposition state which so worried Einstein and Schrödinger. As Bell expressed this point: "Quite generally any embarrassing macroscopic ambiguity in the usual theory is only momentary in the GRW theory. The cat is not both dead and alive for more than a split second [1]."

Let us illustrate this one last time with our standard example of a quantum measurement process: a single "particle in a box" which begins in a superposition of several different energy eigenstates, but which is then coupled to an energy measuring device, represented schematically as a pointer whose position moves by an amount proportional to the energy of the particle. As should be familiar from earlier treatments, if the coupling begins at $t = 0$, the Schrödinger equation dictates that the wave function at time t will be given by

$$\Psi(x, y, t) = \sum_i c_i \psi_i(x, t)\phi_0(y - \lambda E_i t) \tag{9.33}$$

where the ψ_i are the energy eigenfunctions (with corresponding energy eigenvalues E_i) for the particle-in-a-box and ϕ_0 is a narrow Gaussian wave packet representing the position of the center of mass of the (roughly 10^{23}) particles composing the pointer.

The particle-in-a-box is just a single particle, so the probability that it will happen to suffer a spontaneous localization during the time of the experiment is negligible. The pointer, on the other hand, being macroscopic, will suffer repeated localizations. If, however, the width of the wave packet ϕ_0 describing its center-of-mass location is small compared to σ, these localizations will have essentially no effect on the overall wave function for early times during which the individual wave packets – corresponding to terms with different values of i in Eq. (9.33) – remain overlapping in configuration space.

However, as soon as the different terms begin to fail to overlap, such that the spacing between them is of order σ, the situation will be just like that discussed earlier in this section: because all 10^{23} particles composing the pointer are bound together (by the usual sorts of intra- and inter-atomic forces) a single spontaneous

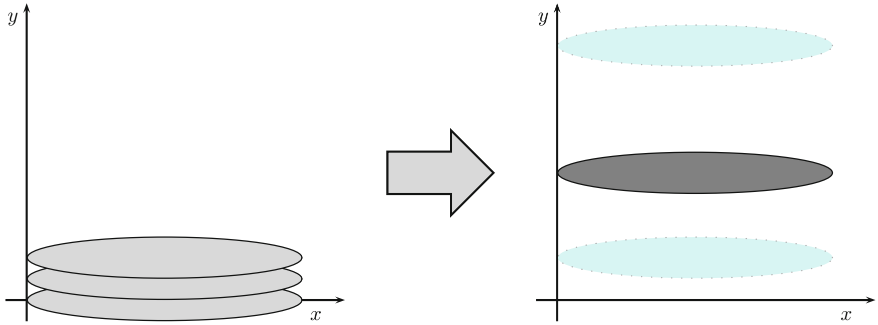

Fig. 9.6 Evolution of the wave-function (in the schematic, two-dimensional configuration space whose axes are the position x of the particle-in-a-box and the position y of the center-of-mass of the pointer) for our toy measurement example. As soon as the individual terms in the superposition (which are separating along the y-direction of configuration space) have separation of order σ, the superposition collapses to just one term, randomly selected from all the possibilities, with probability $|c_i|^2$. Thus, before there would be time for anybody to notice or worry about a troubling macroscopic superposition, the overall wave function describes the pointer as having a well-defined center-of-mass position (here, arbitrarily, $y \approx \lambda E_2 t$) and the particle-in-a-box as having the correct, associated energy E_2

localization of any of the particles will localize all of them, i.e., will localize the entire macroscopic pointer, to just one of the terms. The others will, for all practical purposes, disappear. This is illustrated in Fig. 9.6.

Notice, in particular, that although the spontaneous localizations exclusively localize the particles in *position* space, the particle-in-a-box (whose energy is being measured in this example) ends up in a state of definite energy, and, indeed, the particular state corresponding to the final position of the pointer on the energy-measuring device. Thus, no special/additional/contradictory prescription is required to cause sub-system wave functions to collapse (in the way that ordinary QM says they do) when some arbitrary measurement is performed on them. The measurement *outcome* being displayed in the macroscopic spatial configuration of some aspect of the measuring device (here, the pointer, but one could just as well think of the distribution of ink droplets on a computer printout, or the distribution of photons emitted from a computer screen, for example) is perfectly sufficient in general. This should help clarify the earlier comments about the importance of recognizing the fundamentality of position.

9.3 Ontology, Locality, and Relativity

As we have explained it so far, the GRW theory describes the world in terms of a wave function. Because the collapse/localization mechanism is built into the dynamical evolution law for the wave function, the theory has no need to follow orthodox QM in postulating a separately-existing macroscopic world and associated exceptions to the usual dynamical laws. That is, by providing a uniform description of the world that avoids (noticeable) macroscopic superpositions, the GRW theory avoids the measurement problem that plagues ordinary QM. But what about the other two problems associated with standard QM that we reviewed in earlier chapters?

We begin with the ontology problem. The wave function for an N-particle system (where, for GRW, ultimately N is the total number of particles in the entire universe) is something like a field on $3N$-dimensional configuration space. This does not, in any obvious or straightforward way, attribute definite properties to particular locations in regular, 3-dimensional physical space. Since the ultimate goal must be to provide a coherent description and explanation of the observable 3-dimensional physical world, it is clear that more needs to be said about what, according to the theory, the physical world is made of, and how it relates to the universal wave function whose evolution we have already discussed.

Two possibilities have gained traction in the literature. The first is, interestingly, just the early idea of Schrödinger that we discussed in Chap. 4. Recall that Schrödinger's idea was that the wave function (on configuration space) could be used to define (for example) a *mass density* associated with each individual particle, according to

$$\rho_i(x,t) = m_i \int |\Psi(x_1, x_2, \ldots, x_N, t)|^2 \, \delta(x_i - x) \, dx_1 \, dx_2 \cdots dx_N. \tag{9.34}$$

The total mass density could then be written

$$\rho(x,t) = \sum_i \rho_i(x,t) \tag{9.35}$$

and the original hope was that this field $\rho(x,t)$ would contain, at least at an appropriately coarse-grained level, an image of the familiar macroscopic world of everyday perception, including things like pointers with definite positions and unambiguously alive or dead cats.

Schrödinger himself gave up this interpretation of the wave function (as representing a continuous matter density in physical space) because it simply did not work the way he had hoped for. If the wave function obeys Schrödinger's equation all of the time, then the mass field $\rho(\vec{x},t)$ inherits (or one might say, makes ontologically clear) whatever problematic superpositions arose in the wave function itself. For example, in the Schrödinger's cat kind of situation, the mass field would not contain just a living cat or just a dead cat, but both – superimposed on top of one another, so to speak. If one imagines extrapolating to a description of the entire world, with frequent splittings of the universal wave function into different "branches" each of which corresponds to some more or less definite macroscopic situation, *all* of these different "possibilities" would be superimposed in Ψ and hence $\rho(\vec{x},t)$ would be, for lack of a better term, a complete and utter mess. The ρ generated by the theory, that is, simply wouldn't look anything like what we know the world is supposed to look like. So the theory seems clearly wrong.

But by altering the rules of wave-function evolution, the GRW dynamics avoids precisely this sort of trouble. That is, if the wave function of the universe evolves, not according to Schrödinger's equation, but instead according to the GRW process, the mass density field $\rho(\vec{x}, t)$ associated with the wave function will correspond to just one of the sensible macroscopic possibilities. (Or at least, any appreciable non-sensible macroscopic blurriness will not last for more than a split second.) The three dimensional world, consisting of a mass field $\rho(\vec{x}, t)$ produced by a universal wave function obeying the GRW dynamics, that is, *will look right*. There will be tables and chairs and trees and planets with adequately-sharp shapes, structures, and trajectories; cats will be unambiguously alive or dead; and so on.

So that is one possible way of understanding the ontology of the physical world according to this theory. In the literature, this has come to be called "GRWm", meaning: the universal wave function evolves according to the GRW dynamics, and the ontology is understood as a *m*ass density field.

The other possible way was suggested by Bell:

> There is nothing in this theory but the wavefunction. It is in the wavefunction that we must find an image of the physical world, and in particular of the arrangement of things in ordinary three-dimensional space. But the wavefunction as a whole lives in a much bigger space, of $3N$-dimensions. It makes no sense to ask for the amplitude or phase or whatever of the wavefunction at a point in ordinary space. It has neither amplitude nor phase nor anything else until a multitude of points in ordinary three-dimensional space are specified. However, the GRW jumps (which are part of the wavefunction, not something else) are well localized in ordinary space. Indeed each is centered on a particular spacetime point $[\vec{x}, t]$. So we can propose these events as the basis of the 'local beables' [Bell's term for the physical space ontology] of the theory. These are the mathematical counterparts in the theory to real events at definite places and times in the real world.... A piece of matter then is a galaxy of such events [1].

The idea, then, is that each spontaneous localization, which happens at a particular time and is centered at a particular location in 3D space, produces a kind of "matter point" at that location in space-time. These "matter points" have, in the subsequent literature, come to be called "flashes", and so this version of GRW has come to be called "GRWf".

It is helpful to visualize the two options here, so in Fig. 9.7. I have sketched, on spacetime diagrams, the story of what is going on with the pointer in our toy measurement example, according to GRWm and GRWf.

With two definite proposals for understanding the ontology of the GRW theory, we are in a position to ask: does the theory (in either version) respect the idea of relativistic locality, i.e., no (spooky, faster-than-light) action at a distance? The answer, simply and unambiguously, is: no. GRW (with either of the proposed ontologies) is a non-local theory. This, of course, is not surprising given that we know, from Bell's theorem (Chap. 8), that *any* theory which agrees with the quantum mechanical predictions will have to be non-local. (But see Chap. 10!) Still, it is worthwhile to understand in more detail how the non-locality appears.

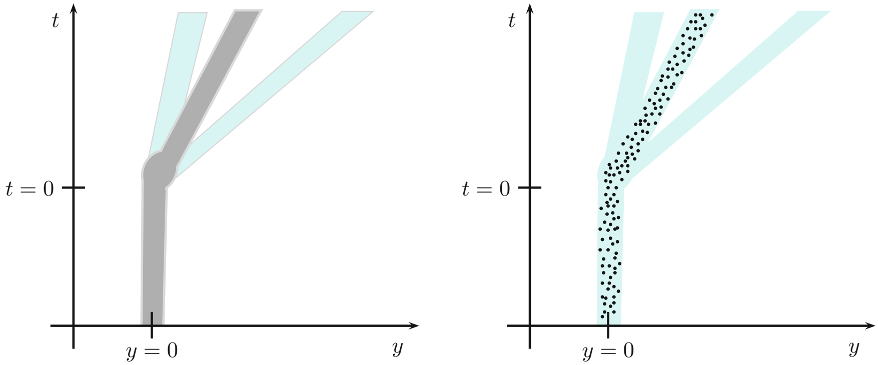

Fig. 9.7 The panel on the *left* shows the mass density field $\rho(y, t)$ associated with all the particles composing the pointer, for the familiar toy measurement example in which the energy of a particle-in-a-box is measured and the outcome, E_2, is indicated by the position of a pointer. Here "the pointer" consists of a lump of nonzero mass density that begins near $y = 0$ and then starts moving to the *right* at a certain well-defined rate just after the measurement interaction begins at $t = 0$. The slight "bulge" around $t = 0$ is meant to suggest that, as the individual terms in the wave function begin to separate, there is a brief period of time in which $\rho(y, t)$ includes several superimposed possibilities. But after a tiny fraction of a second, a spontaneous localization picks just one of the possibilities, the rest disappear, and $\rho(y, t)$ contains just the one realized possibility. The *right* panel shows the same situation, but for the flash ontology. The *black dots* represent the discrete, space-time point flashes and something like the overall motion of the pointer to the *right*, at a basically well-defined rate, can indeed be understood as "a galaxy of such events". It is interesting to contemplate, however, the fact that (in the same way that a real galaxy is mostly empty space), most of the time the pointer is, according to GRWf, literally nothing. That is, for the overwhelming majority of horizontal slices you could draw through the diagram (corresponding to particular moments), the slice would intersect precisely zero of the *dots/flashes*. The physical world, according to GRW, is curiously sparse and pointillistic at the micro-scale... though it coarse-grains to produce a sensible image of the familiar world at the macro-scale

It is easiest to see and understand in the case of GRWm, so let us begin there. Consider a kind of double Einstein's boxes situation, in which two particles are each split between two pairs of half-boxes. In particular, suppose that Alice, a million miles to the left, has a particle which is split between the half box in her left hand (state ψ_L^A) and the half box in her right hand (state ψ_R^A). And suppose that Bob, a million miles to the right, has a second particle which is similarly split between the half box in his left hand (state ψ_L^B) and the half box in his right hand (state ψ_R^B). And suppose that, by some prior careful arrangement, the two particles are in the following entangled state:

$$\psi(x_1, x_2) = \frac{1}{\sqrt{2}}\left[\psi_L^A(x_1)\psi_L^B(x_2) + \psi_R^A(x_1)\psi_R^B(x_2)\right]. \tag{9.36}$$

Suppose also that Bob is prepared with a position measuring device (the center-of-mass position of whose macroscopic pointer we denote y) which can interact with the two half-boxes he's holding and determine whether particle 2 is in the half box

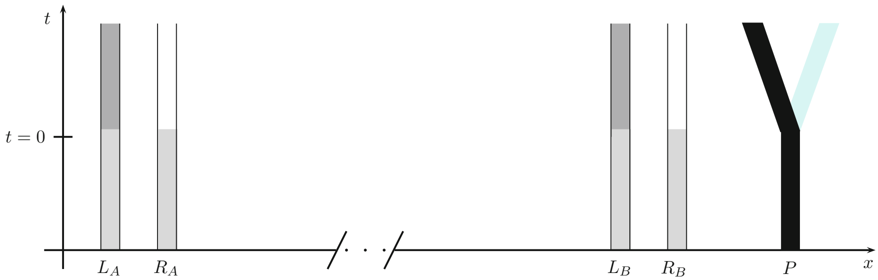

Fig. 9.8 Space-time diagram showing the mass densities of the various objects described in the text: the contents of the half-boxes in Alice's left and right hands (L_A and R_A), the contents of the half boxes in Bob's *left* and *right* hands (L_B and R_B) and the pointer P on Bob's position measuring device. At $t = 0$ Bob initiates the measurement of the position of his particle; very shortly after $t = 0$, a spontaneous localization in one of the (many!) pointer particles collapses the wave function in such a way that subsequently, say: (*i*) the entire mass density associated with the pointer moves unambiguously to the left, (*ii*) the mass density associated with Bob's particle coalesces entirely into L_B (i.e., the density there doubles while the density in R_B suddenly goes to zero), and (*iii*) the mass density associated with Alice's particle (millions of miles away!) also coalesces entirely into L_A. The change in the mass density distribution associated with Alice's particle, as a consequence of Bob's measurement on his particle, is a clear-cut case of non-local action-at-a-distance. Note, though, that as in the analogous case in the pilot-wave theory, even though what's happening in Alice's boxes is affected by Bob's distant actions, Alice has no way to observe this change. She could open her boxes and see where the particle is, and she would of course find it somewhere. But she would have no way to know whether it was her own observation that triggered her particle to randomly coalesce either in her right hand or her left hand, or whether, instead, the particle had already coalesced in one place or the other as a result of Bob's distant actions. So although there is nonlocal action-at-a-distance, according to the theory, the nonlocality cannot be used to transmit messages faster than light, and so avoids the most blatant sort of conflict with relativity theory

in his left hand, or instead the one in his right hand. The measuring device is initially in its ready state, with the pointer at $y = 0$, and we assume the pointer moves to the right/left if particle 2 is found in the right/left-hand box.

Now suppose that at $t = 0$ Bob decides to proceed with the measurement, i.e., to let the measuring device begin interacting with his half-boxes. A space-time diagram showing the mass densities associated with the two particles and the pointer is shown in Fig. 9.8. The important point is as follows. Prior to $t = 0$, the mass density associated with Alice's particle is genuinely split 50/50 between her two half-boxes. And it would (with extremely high probability) have remained so split (for millions of years!) had Bob not initiated the position measurement on his own particle. But when he does initiate this position measurement, it has the effect, by the mechanism we discussed in the previous section, of causing (in some very short period of time) a collapse to one or the other of the definite, initially superposed states. Thus not only Bob's particle, but also Alice's distant one, will switch from being "evenly smeared" between the two half boxes, to being definitely in one or the other of the two half boxes, as a direct result of Bob's decision to initiate his measurement procedure. Bob's decision – a million miles to the right – thus (almost) instantaneously influences the distribution of mass (associated with Alice's particle) even though Alice's particle is a million miles to the left. It is a clear-cut case of nonlocal action-at-a-distance. (You are invited, in the Projects, to render this diagnosis in a more formal way by applying Bell's locality condition or one of our modifications of it.)

The case of GRWf is basically the same, although it is slightly harder to draw a nice picture to capture the nonlocality since, as mentioned, for individual particles, the space-time diagram would be almost completely empty. Still, the same kind of analysis applies. Suppose, for example, that the two particles are prepared, just as before, in the state

$$\psi(x_1, x_2) = \frac{1}{\sqrt{2}}\left[\psi_L^A(x_1)\psi_L^B(x_2) + \psi_R^A(x_1)\psi_R^B(x_2)\right]. \tag{9.37}$$

Now, there is a certain (very small, but nonzero) probability that, according to the theory, there will be a "flash" inside the box in Alice's left hand in, say, the next one minute. However, if Bob measures the position of his particle in the same way we described before, this probability (for a flash to appear in L_A) will either double (if Bob's measurement "finds" his particle on the left) or will go to zero (if Bob instead "finds" his particle on the right). Thus, the probability for a certain event over where Alice is, a million miles to the left, will be different depending on what happens over where Bob is, a million miles to the right, even when we are conditionalizing those probabilities on a complete specification of events (including, here, in particular, the fact that there have been no prior flashes associated with Alice's particle!) in the past light cone of the event in question.

So, with either the "m" or "f" ontology, the GRW theory is nonlocal, just like the pilot-wave theory, and just like we should have expected on the grounds of Bell's theorem.

However, as first pointed out by Bell, there is a sense in which the spontaneous collapse theories seem to be more compatible with relativity – or at least a little more promising in that respect – than the pilot-wave theory [1]. This has to do with the fact that the spontaneous collapse theories are irreducibly stochastic (unlike the pilot-wave theory, which is deterministic). The technical details are somewhat beyond the level of this book, but it should be noted that spontaneous collapse theories with both "mass field" and "flash" ontologies have been constructed which, despite being non-local, appear to be more plausibly consistent with a more serious notion of fundamental Lorentz invariance than appears to be possible with pilot-wave type theories [4, 5]. That is, the spontaneous collapse theories warrant a somewhat hopeful attitude toward the project – which you may not until this very moment even have conceived of as a possibility – of *reconciling* non-locality (which we know, from Bell's theorem, must be present) with some satisfying notion of fundamental relativity (which, needless to say, there is strong reason to demand).[2]

---

[2]The idea that it might be possible, after all, to reconcile relativity with non-locality may perhaps suggest that earlier chapters have over-stated the extent to which Bell's formulation of locality successfully captures the idea of "no faster-than-light causal influences" that we ordinarily take to be an implication of relativity theory. Let me assure you that this is not the case. The quantum

The technical details involved in these issues definitely render them beyond the scope of the present book. But one can nevertheless appreciate that certain seemingly simple questions – for example, "What precisely does it *mean* for a theory to be fundamentally relativistic?" – turn out to be surprisingly difficult to answer in a context in which they are entangled with the possibility of irreducibly stochastic (non-deterministic) laws, unclarity about ontology (how quantum wave functions relate to goings-on in 3+1-dimensional space-time), and other issues we have grappled with in this book. Suffice it to say here that this remains an area of continuing controversy and ongoing research, but that the spontaneous collapse theories have put on the table, for further analysis and contemplation, the previously-unrecognized possibility that Bell's theorem (and the associated experiments) could live in harmony with fundamental relativity.

9.4 Empirical Tests of GRW

So far we have presented the spontaneous collapse theory as a way of reformulating quantum mechanics so that it (i) posits a clear set of unambiguous and universal dynamical rules and (ii) provides a coherent ontology in terms of which directly observable macroscopic features of the real world can be recognized. The goal has been basically to clean up the foundational problems that plague ordinary quantum theory, while maintaining (as closely as possible) quantum theory's seemingly accurate empirical predictions. But, as hinted at the beginning of this chapter, because they predict that spontaneous collapses will occur with specific length- and time-scales, the spontaneous collapse theories do, in principle, make slightly different empirical predictions from ordinary quantum mechanics. And of course this is nice, because it means that spontaneous collapse theories can be tested, experimentally, against other versions of quantum mechanics.

The easiest type of test to understand involves something like two-slit interference. Recall that, for a single particle, spontaneous collapses, which localize the particle's wave function to a distance scale of order σ, happen with frequency λ, i.e., on a timescale $\tau = 1/\lambda$. It should thus be clear that, in an interference experiment with individual particles, spontaneous collapse theories will predict that the interference should start to disappear if the spatial separation between the individual components of the wave function exceeds σ for a time period greater than τ. Thus, the successful observation of interference places experimental constraints on the values of σ and τ, or equivalently, σ and λ.

(Footnote 2 continued)
non-locality really does mean that there are causal linkages between space-like separated events, of a sort normally thought to be prohibited by relativity. The possibility of perhaps reconciling non-locality with relativity does not mean that we previously misunderstood or misformulated locality. Rather, it means that there may have been something rather deep and subtle wrong with the way we were thinking about relativity (or causality or both) that fooled us into thinking that relativity was incompatible with space-like separated events being causally linked.

Of course, for interference experiments involving single particles, the particles are typically in a state of spatial superposition for only some small fraction of a second. So the experimental constraint is something like $\tau \gg 1$ s, i.e., $\lambda \ll 1\ \text{s}^{-1}$. The value of λ proposed by GRW, recall, was $\lambda \approx 10^{-16}\ \text{s}^{-1}$. So the experimental constraint coming from, say, single-neutron interferometry, is almost completely useless: it tells us only that, if the spontaneous collapse theories are right, the frequency of collapses must be much smaller than something that is already 16 orders of magnitude bigger than what we guessed the frequency might be!

However, as we saw previously, the effective collapse rate for an object consisting of N particles is $N\lambda$. So, by performing interference experiments with atoms, molecules, and even larger objects, we can start to get experimental constraints that are at least in the neighborhood of the hypothesized values of the collapse parameters. For example, in 1999, a group led by Markus Arndt and Anton Zeilinger in Vienna demonstrated interference using "buckyballs", which are C_{60} molecules [6]. Sixty carbon atoms, each containing 12 nucleons (6 protons and 6 neutrons) and 6 electrons, is roughly a thousand elementary particles. So the spontaneous collapse rate for buckyballs should be about a thousand times faster than the fundamental (per particle) collapse rate, and so the experimental constraint on the GRW parameters is about three orders of magnitude closer to relevancy.

In subsequent years, interference with even bigger molecules has been demonstrated, and there are plans for pushing this particular envelope even further [7]. In addition, experimental limits on the spontaneous collapse parameters can also be extracted from other kinds of observations. For example, spontaneous localizations add high-momentum Fourier components to wave functions and thereby add energy to systems that would not otherwise be present. Such additions of energy might be observed as anomalous warming of otherwise-thermally-isolated systems, or perhaps anomalous emission of high-energy particles such as X-rays. Some of these processes can be explored in more detail in the Projects.

In a very nice recent paper, Tumulka and Feldmann have organized the various known experimental constraints on the spontaneous collapse parameters, and produced a "parameter diagram" showing ranges of values that are excluded by the different types of observations [8]. We reproduce one of their diagrams here as Fig. 9.9. As might have been anticipated from our previous discussion, the most stringent constraints actually do not come from interference experiments, but arise instead from observations of systems (like the intergalactic medium!) with considerably more particles. Note also that these kinds of observational constraints tend to exclude the "upper left" portion of the parameter space, i.e., large values of λ and small values of σ.

A rather different kind of constraint tends to exclude the opposite region of the parameter space, i.e., very small values of λ and very large values of σ. The idea here goes back to one of the original motivating goals of the spontaneous collapse theories, which is to avoid the embarrassing sort of macroscopic superposition that is illustrated by Schrödinger's cat. More specifically, the idea is that we *know*, just from ordinary direct perceptual experience of the physical world around us, that macroscopic things do not appear "blurry". So their positions must be sharply defined at length scales

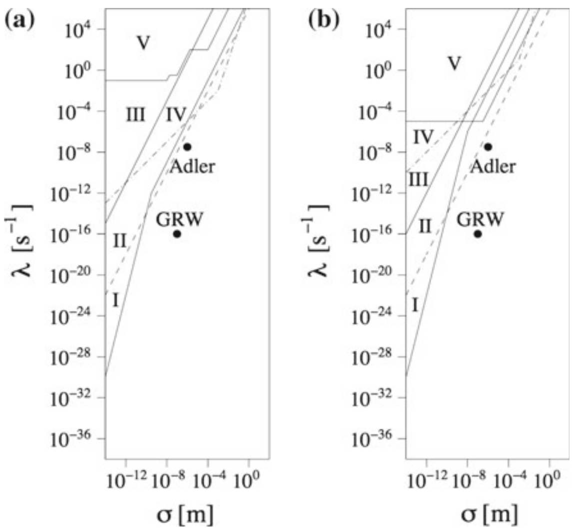

Fig. 9.9 A map of the parameter space, for both (**a**) the GRW theory we've discussed in detail as well as (**b**) the related "continuous spontaneous localization" [CSL] theory that was alluded to earlier, showing the values of σ (the spatial width of the collapse function $g_r(x)$) and λ (the collapse frequency) that are excluded by various sorts of observations, from Ref. [8]. The five numbered categories of experimental/observational constraints are "I = spontaneous x-ray emission, II = spontaneous warming of the intergalactic medium (dashed line), III = spontaneous warming of air, IV = decay of supercurrents (dashed-and-dotted line), V = diffraction experiments [8]." Note that, on each panel, the two dots represent the parameter values suggested originally by GRW and another slightly different suggestion by Stephen Adler. Figure © IOP Publishing. Reproduced with permission. All rights reserved. https://doi.org/10.1088/1751-8113/45/6/065304

where any blurriness would be perceptually evident – say, something of order a millimeter. Or at least, visible macroscopic things should not remain blurry at a distance scale much larger than a millimeter, for a time long enough for us to notice the blurriness! One can see in principle here how small values of λ and large values of σ can be excluded as "perceptually unsatisfactory". See Fig. 9.10 for Tumulka and Feldmann's nice diagram showing both the "Empirically Refuted Region" and (what they call the "Philosophically Unsatisfactory Region" but I would prefer to call) the "Perceptually Unsatisfactory Region" of parameter space for both GRW and CSL.

We close this section and this Chapter with one final Figure from the paper by Tumulka and Feldmann. In Fig. 9.11 we reproduce their diagram showing the progression of experimental constraints, coming from interference experiments, over time. The visual implication is that we are perhaps only two or three decades away from the ability to experimentally probe the parameter values originally suggested by GRW. Thus, the "open window" – between the ERR and PUR in Fig. 9.10 – may close in the near future, and we will know, once and for all, whether or not the spontaneous collapse models are right.

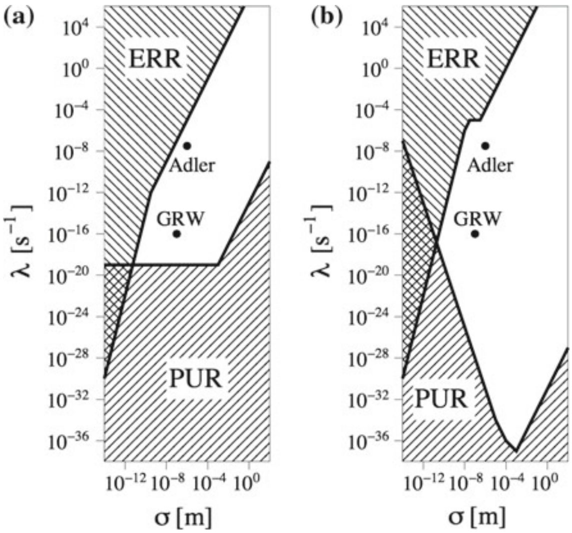

Fig. 9.10 Map of parameter space, again for both GRW and CSL theories, showing now both the "Empirically Refuted Region" (ERR) and the "Perceptually Unsatisfactory Region" (PUR) as discussed in the text. From Ref. [8]. Figure https://doi.org/10.1088/1751-8113/45/6/065304

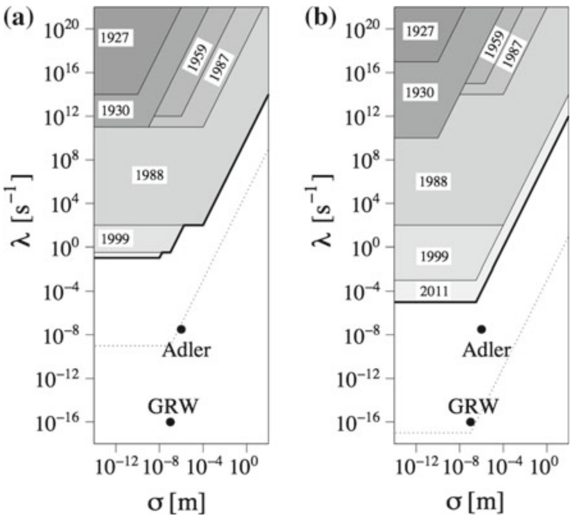

Fig. 9.11 The "Empirically Refuted Region" (ERR) of the GRW and CSL parameter spaces has steadily advanced, in recent decades, leaving an ever-narrowing window of parameter values which are compatible both with experimental and perceptual evidence. This suggests that, within perhaps a couple of decades, we will either have direct experimental evidence in support of the spontaneous collapse models, or the models will have been ruled out as either empirically or perceptually unacceptable. From Ref. [8]. Figure https://doi.org/10.1088/1751-8113/45/6/065304

Projects

9.1 Work carefully through all the steps to convince yourself that Eqs. (9.16) and (9.17) are correct.

9.2 Suppose a particle with a Gaussian wave function $\psi(x) \sim e^{-x^2/4w_0^2}$ suffers a spontaneous collapse centered at $x = r$. Show that the post-collapse wave function remains Gaussian, and find a formula for its width w in terms of w_0 and σ. (Confirm that your expression for w implies that $w \approx w_0$ if $\sigma \gg w_0$, and implies that $w \approx \sigma$ if $w_0 \gg \sigma$.)

9.3 Suppose a particle has a Gaussian wave function $\psi(x) \sim e^{-x^2/4w_0^2}$ at the moment just before it suffers a spontaneous collapse. What is the probability density $P(r)$ for the collapse to be centered at $x = r$?

9.4 Argue that the probability distribution $P(r)$ defined in Eq.(9.11) is indeed a legitimate probability distribution since $P(r) > 0$ and $\int P(r)\, dr = 1$.

9.5 The discussion in Sect. 9.2 suggests that whereas for a single particle the localization rate is λ, for a collection of N particles the localization rate is $N\lambda$. This is basically equivalent to saying that collections of particles should have an overall localization rate that is proportional to the total mass of the collection – an idea that can and probably should be instituted as part of the formulation of the theory at the fundamental level: different particle species (electrons and protons, for example) may have different fundamental localization rates, with the rates being proportional to the mass of the particle. Assuming such a modification of the theory, is it nucleons (neutrons and protons) or electrons that suffer most of the localizations associated with ordinary matter?

9.6 It may appear puzzling that the spatial probability distribution for the point r at which a localization is centered, is given by $P(r) = N(r)^2$ rather than the seemingly simpler and approximately equivalent alternative $P(r) = |\psi(r, t^-)|^2$. The reason for this has to do with the requirement that non-local signaling (i.e., instantaneous communication across arbitrary distances) should be impossible. Consider a situation involving two entangled and spatially-separated particles, 1 and 2, and suppose that particle 1 suffers a spontaneous collapse centered at $x_1 = r$ at time t. Show that the pre-collapse marginal distribution for particle 2 to be observed at position x_2, namely

$$P(x_2, t^-) = \int \left|\psi(x_1, x_2, t^-)\right|^2 \, dx_1 \tag{9.38}$$

is the same as the post-collapse marginal distribution (averaged over all the points r at which the collapse might have been centered)

$$P(x_2, t^+) = \int \int \left|\psi(x_1, x_2, t^+)\right|^2 \, P(r)\, dx_1\, dr \tag{9.39}$$

provided $P(r) = N(r)^2$. (This means, for example, that Alice cannot tell, by measurements made on her particle, whether a distant entangled particle has suffered a collapse. This in turn prevents Bob from sending her a message, by for example choosing whether or not to allow his particle – entangled with her distant one – to interact with a macroscopic object such as a measuring device and thereby trigger a collapse.)

9.7 Consider a one gram pointer. In GRW with the "flash" ontology, approximately how many flashes occur, associated with the pointer, per second, if the flash rate f is as given in the text? (Assume for simplicity that only the nucleons are hit by spontaneous localizations.)

9.8 Approximately what fraction of the particles composing your body will pop briefly into existence (in a "flash") at least once during your lifetime, according to GRWf?

9.9 Consider the conduction electrons in a macroscopic piece of metal. These can be thought of as having wave functions that spread out over the entire, macroscopic extent of the metal. For such an electron with essentially zero momentum, its kinetic energy will also be approximately zero. However, if it happens to suffer a spontaneous localization its wave function will subsequently be a width-σ Gaussian. Estimate the increase in the particle's kinetic energy that results from this spontaneous localization, and use this to estimate the rate at which the temperature of a thermally isolated piece of metal should increase according to the spontaneous collapse theory. Would this "anomalous heating" be easy to detect, experimentally?

9.10 Use Bell's formulation of locality (and/or one of our modified versions from Chap. 1 or Chap. 5) to more formally diagnose GRWm as a non-local theory, using the example displayed in and discussed around Fig. 9.8. (Note: you will need to think carefully about which formulation of locality it is possible and appropriate to use here.)

9.11 In Sect. 9.3, we discussed the non-local character of both GRWm and GRWf in terms of a "double Einstein's boxes situation, in which two particles are each split between two pairs of half-boxes." This example is nice because it provides another opportunity to think about how the spontaneous collapses function in the presence of entanglement. But it is really more complicated than is minimally necessary to establish the non-locality of the theory. Show and explain how a "[single] Einstein's boxes" situation, like that discussed in Sect. 4.1 of Chap. 4, can already be used to diagnose the spontaneous collapse theories as non-local. (Note that this means, interestingly, that there are situations whose explanation involves non-locality in GRW, but is local in the pilot-wave theory.)

9.12 A single character printed in ink contains something of order 10^{17} carbon atoms or roughly 10^{18} nucleons. In GRWf, how many flashes per second (associated with that small amount of ink) do you think are sufficient to say that the ink drop is really there, with the particular shape we see? (Hint: human visual perception can be modeled as something like a digital camera which captures roughly 30 frames per second. Consistency with perceptual experience would seem to require that typical frames contain enough flashes to construct the shape of the appropriate letter unambiguously.) Use your estimate to calculate the minimum localization rate λ compatible with "perceptual acceptability", and compare your calculated value to Fig. 9.10.

9.13 In Ref. [8], Tumulka and Feldmann raise an interesting question: what if some future experiment demonstrates violations of ordinary QM and confirms the empirical predictions of GRW/CSL, but for parameter values which lie in the "perceptually unsatisfactory region" (PUR) of Fig. 9.10. What would you say/conclude in such a situation?

9.14 True of false: according to the spontaneous collapse theories, matter is made of particles. Explain.

9.15 There are a lot of things to like about the spontaneous collapse theories: they sharpen, with precise mathematics, the "loose talk" of Copenhagen QM; they provide comprehensible (if unanticipated) ontologies; and they make empirically testable predictions that differ from other versions of QM. But it is also possible to find the spontaneous collapse theories somewhat contrived and *ad hoc*. Explain why, by listing and discussing some of the details about the theory's formulation which seem arbitrary and/or which could easily be changed without dramatically affecting the theory's structure or predictions.

References

1. J.S. Bell, Are there quantum jumps? in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, 2004)
2. G.C. Ghirardi, A. Rimini, T. Weber, Unified dynamics for microscopic and macroscopic systems. Phys. Rev. D **34**, 470 (1986)
3. J.S. Bell, quoted by G.C. Ghirardi in *Sneaking a Look at God's Cards*, Gerald Malsbary, trans., Revised Edition (Princeton University Press, Princeton, 2005), p. 415
4. R. Tumulka, A relativistic version of the Ghirardi-Rimini-Weber model. J. Stat. Phys. **125**, 821–840 (2006)
5. D. Dürr, D.J. Bedingham, G. Ghirardi, S. Goldstein, R. Tumulka, N. Zanghi, Matter density and relativistic models of wave function collapse. J. Stat. Phys. **154**, 623–631 (2014)
6. M. Arndt, O. Nairz, J. Vos-Andreae, C. Keller, G. van der Zouw, A. Zeilinger, Waveparticle duality of C60 molecules. Nature **401**, 680 (1999)
7. M. Arndt, K. Hornberger, Testing the limits of quantum mechanical superpositions. Nat. Phys. **10**, 271–277 (2014)
8. W. Feldmann, R. Tumulka, Parameter diagrams of the GRW and CSL theories of wave function collapse. J. Phys. A: Math. Theor. **45**, 065304 (2012)