

Chapter 3

The Measurement Problem

In Chap. 2, we reviewed the mathematical formalism of quantum mechanics and practiced applying it to a number of concrete examples. In the present chapter, we will begin the process of stepping back and turning a critical eye toward the theory. In particular, in this chapter, we will look carefully at the curiously central role that the theory gives to the process of “measurement” and discuss the network of related concerns, centering on the infamous example of Schrödinger’s Cat, that have come to be called “the measurement problem.”

3.1 The Quantum Description of Measurement

Our discussion of the examples in the previous Chapter focused on solving Schrödinger’s equation to understand how the quantum states of microscopic systems evolve in time, and then using Born’s rule to connect these quantum states to probabilities for various possible measurement outcomes. Here, we want to emphasize and develop two additional aspects of Born’s rule, and then step back and look at the quantum mechanical description of measurement and, really, the quantum mechanical description of the world as a whole.

The first new aspect of Born’s rule that we need to stress is the so-called “collapse postulate”. Recall that, according to Born’s rule, we calculate the probabilities of different measurement outcomes as follows: first, write the quantum state Ψ (of the system we are measuring) as a linear combination of eigenstates of the operator corresponding to the property we are measuring, as in

$$\Psi = \sum_i c_i \Psi_i \tag{3.1}$$

where Ψ_i is an eigenstate of the operator \hat{A} with eigenvalue A_i . Then the probability for the measurement to have outcome A_i is $P(A_i) = |c_i|^2$. That should be familiar and clear.

But it is an experimental fact that, if you *immediately repeat* a measurement – for example, you measure the energy of a particle and then immediately measure its energy again – you always get *the same result* for the second measurement that you got for the first. (Note how weird it would be if this weren't true. At very least, the word “measurement” would then seem quite inappropriate.) But then the consistent applicability of Born's rule to the two measurements implies that, by the time the second measurement occurs, the system must *be* in the eigenstate corresponding to the outcome of the first measurement. Only this (according to Born's rule) will ensure that the probability of seeing that same result again, in the second measurement, is 100%. This is the collapse postulate: when a measurement occurs, and has outcome A_n , the quantum state of the system being measured *ceases* to be whatever superposition it might have been previously, and “*collapses*” to the eigenstate Ψ_n whose eigenvalue is the realized outcome A_n .

Formally, for a measurement that begins at time t_1 and ends at time t_2 , one has

$$\Psi(t_1) = \sum_i c_i \Psi_i \quad \rightarrow \quad \Psi(t_2) = \Psi_n. \quad (3.2)$$

where the right arrow indicates the time-evolution. The crucial point – and the reason the “collapse postulate” is a *postulate* – is that this is a very different sort of time-evolution than wave functions normally undergo, when they are evolving according to Schrödinger's equation.

The second new aspect of the Born rule then has to do with the idea of experiments having definite outcomes. So far we have just used these words abstractly, without really thinking in concrete physical terms about what we mean. So consider, for example, the two-slit experiment described in the last Chapter. A single electron (or photon) propagates through the apparatus, and we describe its state with a wave function. But then, at some point, the particle interacts with the “detection screen” that is being used to measure its position. This (as we have just been discussing) apparently causes the wave function of the particle to “collapse” – to switch from something that is spread out across (say) nearly the whole screen, to something more like a δ -function spike at a particular spot.

But something important happens to the screen, as well! If, for example, we think of the screen as a piece of film, a certain little spot on the film *changes color*. Or if we think of it as a CCD array, an electrical signal is produced which (say) results in the coordinates of the particular pixel that the particle “hit” being printed on a computer screen. Whatever the details, exactly, the point is that the measuring device itself is a physical thing, which undergoes some kind of observable physical change, that is intimately coupled with the change (described by the collapse postulate) in the state of the particle that happens at that same moment. The “measurement”, in short, is a *physical interaction* between these two physical systems, which both change as a result.

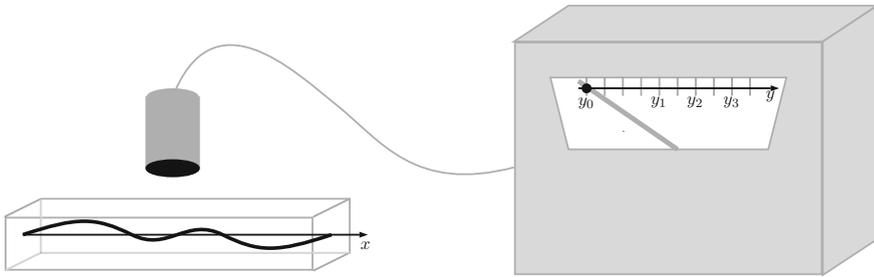


Fig. 3.1 The quantum particle-in-a-box (whose spatial degree of freedom is called x) is shown on the *left*; the curve is meant to indicate its wave function (though one should be careful not to take this picture too literally!). Then there is an energy-measurement device which will perform the measurement. The device has a macroscopic pointer, which we can idealize as a single, very heavy particle with horizontal coordinate y . Prior to the measurement-interaction, the pointer is sitting in its “ready” position (y_0); after the measurement interaction, the pointer will move to a new position which indicates the outcome of the measurement: y_1 will mean that the energy of the particle is E_1 , etc

Let us set up a semi-realistic concrete example to consider, involving the “pointer” concept introduced in Chap. 1. Suppose our microscopic quantum system is a particle-in-a-box, which has been prepared so that its state is a superposition of the first few energy eigenstates. Say:

$$\psi_0 = c_1\psi_1 + c_2\psi_2 + c_3\psi_3 \quad (3.3)$$

where ψ_1 is the $n = 1$ (i.e., ground state) particle-in-a-box wave function, etc. And suppose that we are planning to wait a few seconds and then measure the energy of the particle-in-a-box. This measurement will be performed using some kind of energy-measuring apparatus, which we will treat schematically as a black box (with internal workings whose details we need not worry about too much) with a macroscopic pointer whose position, after the interaction, will indicate the result of the energy measurement. See Fig. 3.1.

So then the quantum mechanical description of the measurement process goes like this. The particle-in-a-box is described by a wave function, which starts as ψ_0 and then evolves according to the Schrödinger equation until the measurement is made. The pointer, meanwhile, starts out at position $y = y_0$ and just sits there (obeying $F_{\text{net}} = ma$ with $F_{\text{net}} = 0$) until the measurement is made.

At some point, the measurement interaction occurs and when the dust settles, the situation is now as follows: the wave function of the particle-in-the-box has collapsed to one of the energy eigenfunctions ψ_n (with n being either 1, or 2, or 3); and the apparatus pointer has moved and is now pointing at the appropriate corresponding value, $y = y_n$. (See Fig. 3.2 for an illustration of one of the three possibilities.) With the interaction between the particle-in-the-box and the measuring apparatus now completed, each part – the particle-in-the-box and the apparatus pointer – continue evolving as normal (i.e., the wave function of the particle-in-the-box evolves

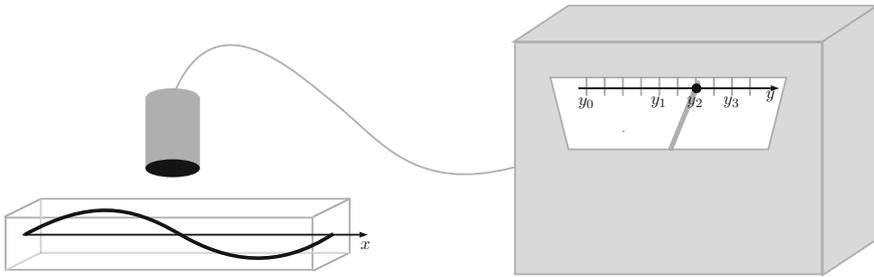


Fig. 3.2 One of the three possible post-measurement states of the particle-in-a-box and measurement apparatus pointer: the wave function of the PIB has “collapsed” to ψ_2 and the pointer has moved to position y_2 , indicating that the energy measurement had outcome E_2

according to the Schrödinger equation, and the apparatus pointer again obeys $F_{\text{net}} = ma$ with, presumably, $F_{\text{net}} = 0$ again).

Now there are two curious and connected things about all of this, one dynamical and one more ontological.

The curious thing, dynamically, is that the measurement interaction seems to involve a violation of the “usual” dynamics for each of the interacting objects. We have already pointed out that the “collapse” that occurs to the wave function of the particle-in-the-box is not (or at least is not obviously, not apparently) something that is described by Schrödinger’s equation. The wave function, during that brief time period, collapses *instead* of evolving in accordance with Schrödinger’s equation. And then something similar occurs with the pointer, although the schematic nature of the example makes this a little harder to see. The claim here is meant to be something like: “ $F_{\text{net}} = ma$ ceases to apply to the pointer when it suddenly swerves from y_0 to y_n for some particular value of n .” But of course, that cannot be exactly true. The pointer is, after all, just the last step in a long causal chain. The (say, magnetic) force on it – which results in it moving some particular distance to the right during the course of the interaction – can be perfectly explained in terms of (say) electrical currents that flow through some wires inside the black box part of the apparatus. But those currents can, in turn, be perfectly explained, in standard classical Maxwellian electro-dynamical terms, by (say) other electrical currents, in the cable, i.e., further upstream along the causal chain. And so on... but at *some* point, we must come to some change in some aspect of the physical state of the apparatus for which the usual classical dynamical rules do not provide an explanation. Otherwise, why would we need quantum mechanics at all?

This brings us to the other curious thing about this whole situation: the apparatus is being described in classical terms! We do not speak, for example, of “the wave function of the pointer,” but instead of the pointer’s *position*. The whole setup and description, that is, presupposes that the pointer (and presumably most of the contents of the entire black box, at least until we get far upstream into some murky meso- or micro-scopic stuff) is the kind of thing to which we can just unproblematically attribute definite, sharp, unambiguous, un-superposed, classical properties. The

picture, that is, seems to presume that quantum systems (described in terms of wave functions) exist *in addition to* classical systems (described, for example, in terms of particles with definite positions). If this kind of description is taken seriously, as faithfully capturing the true nature of the world being described, it implies for example that there are two fundamentally different types of particles in the world: those (like our particle-in-a-box here) which are “wavy” (i.e., which are properly described in terms of a spatially-spread-out wave function) and those (like the particle(s) composing the pointer here) which are “sharp” (i.e., which are properly described as having definite positions at all times).

At the broadest level, then, it seems like the implied quantum mechanical picture of the world goes something like the following. There is, to begin with, the familiar macroscopic “classical” realm in which things have definite properties and are described in clear, everyday terms. This macroscopic realm basically obeys the dynamical laws of classical mechanics. Then there is also a microscopic realm where our everyday classical intuitions don’t apply and we must instead describe things using quantum mechanical wave functions which, of course, do not necessarily attribute definite properties to things: energy, momentum, position, etc., can all be “smeared out” in quantum superpositions, i.e., these properties can fail to have definite values in the way we would have expected classically.

And then finally there are the special rules describing the *interaction* of the macroscopic and microscopic realms. In particular, during a *measurement*, the quantum mechanical wave function describing the microscopic system fails, momentarily, to evolve in accordance with Schrödinger’s equation, and instead *collapses* to one particular eigenstate of the operator corresponding to the physical property (energy or momentum or whatever) that is being measured. Which particular eigenstate? This is supposed to be irreducibly random and inexplicable, but the particular state that the quantum system collapses to is the one that corresponds to the particular measurement outcome that is displayed by the apparatus, as a result of its own process of jumping, inexplicably and in violation of the usual (here, classical) dynamical rules.

Stepping back, all of this should make one feel very uncomfortable. To begin with, there is a kind of schizophrenic division of the world into two “realms” (the microscopic quantum part, and the macroscopic classical part) which seem to have completely different ontologies and completely different dynamical laws. And then there are apparently special dynamical rules which come into play when the two realms interact, during a “measurement”. And the situation is then made worse by the *vagueness* of the concept “measurement”. If you say “during measurements, quantum wave functions momentarily cease to obey Schrödinger’s equation and instead collapse” that is already weird and troubling, but it becomes downright *meaningless* if you can’t specify *exactly* what kinds of physical processes count as “measurements”. Bell put this particular point very sharply as follows:

What exactly qualifies some physical systems to play the role of ‘measurer’? Was the wave-function of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system ... with a Ph.D.? If the theory is to apply to anything but highly idealised laboratory operations, are we not obliged to admit that more or less ‘measurement-like’ processes are

going on more or less all the time, more or less everywhere? Do we not have [quantum] jumping [i.e., collapse] then all the time? [1]

It indeed seems necessary to admit that “measurements” are ubiquitous, and occur even in places and times where there are no human experimenters. But it also seems hopeless to think that we will be able to give an appropriately sharp answer to the question of what, *exactly*, differentiates the ‘ordinary’ processes (where the usual dynamical rules apply) from the ‘measurement-like’ processes (where the rules momentarily change).

In an interview, Bell was once asked whether he thought the problems with quantum mechanics were philosophical or experimental. His answer is relevant here:

I think there are *professional* problems. That is to say, I’m a professional theoretical physicist and I would like to make a clean theory. And when I look at quantum mechanics I see that it’s a dirty theory. The formulations of quantum mechanics that you find in the books involve dividing the world into an observer and an observed, and you are not told where that division comes – on which side of my spectacles it comes, for example – or at which end of my optic nerve. You’re not told about this division between the observer and the observed. What you learn in the course of your apprenticeship is that for practical purposes it does not much matter where you put this division; that the ambiguity is at a level of precision far beyond human capability of testing. So you have a theory which is fundamentally ambiguous... [2].

Stepping back, it begins to seem like we must have taken the quantum mechanical “recipe” – Born’s rule and the collapse postulate in particular – too literally, somehow. Surely, for example, big macroscopic things like measuring devices and their pointers (not to mention cats and trees and planets) are just large collections of electrons and other microscopic particles. And so surely, if macroscopic stuff is literally made of lots and lots of little microscopic parts, then shouldn’t a fully microscopic description suffice, at least in principle even if not in practice? In other words, shouldn’t the familiar macroscopic world (meaning large objects with definite, classical properties, that at least approximately obey Newton’s laws of motion) somehow *emerge* from the more fundamental quantum mechanical description, as opposed to being postulated at the fundamental level?

3.2 Formal Treatment

Let us explore this possibility in a more formal way here. (Doing this will help us understand the point Schrödinger meant to be making with his cat example, when we turn to that shortly.)

Take the example from the previous section – measuring the energy of a particle-in-a-box – but now let us attempt to use the microscopic type of quantum description (in terms of wave functions obeying Schrödinger’s equation) for the entire setup, including the measuring apparatus. Suppose, as before, that the particle-in-a-box starts out in the state

$$\psi_0(x) = c_1\psi_1(x) + c_2\psi_2(x) + c_3\psi_3(x). \quad (3.4)$$

As for the pointer, previously we had been describing it classically and hence attributing to it some definite pre-measurement position y_0 . But now we want to instead describe the pointer quantum mechanically, with a wave function. Let's say that the pointer in its "ready position" can be described by a Gaussian wave packet centered on the position y_0 :

$$\phi(y) = N e^{-(y-y_0)^2/4\sigma^2}. \quad (3.5)$$

At the moment (call it $t = 0$) when the measurement interaction begins, the joint wave function of the particle and pointer will be

$$\Psi_0(x, y) = \psi_0(x)\phi(y). \quad (3.6)$$

This quantum system will then evolve in time in accordance with the Schrödinger equation,

$$i\hbar \frac{\partial \Psi(x, y, t)}{\partial t} = \hat{H}\Psi(x, y, t). \quad (3.7)$$

But what is the Hamiltonian, \hat{H} ? Evidently there will be three contributions. First, it will include the usual terms corresponding to the kinetic and potential energies of the particle-in-the-box, whose degree of freedom is "x":

$$\hat{H}_x = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x). \quad (3.8)$$

As we saw in Chap. 2, \hat{H}_x gives each term in Eq. (3.4) a time-dependent complex phase, with distinct frequencies for the different terms.

Second, the overall Hamiltonian should include the kinetic energy of the pointer, whose degree of freedom is "y":

$$\hat{H}_y = -\frac{\hbar^2}{2M} \frac{\partial^2}{\partial y^2}. \quad (3.9)$$

As we saw in Chap. 2, \hat{H}_y will tend to make the wave packet (describing here the state of the pointer) spread in time. But recall that our single pointer particle is just a schematic way of describing what is, in fact, some enormous macroscopic collection of elementary particles. We should thus probably attribute to our pointer particle a very large mass M . This means – recall Eq. 2.53 – that the packet will spread very slowly. Or even more simply, it warrants making the following approximation:

$$\hat{H}_y \approx 0. \quad (3.10)$$

The third (and here most important) contribution to the Hamiltonian will be the one describing the *interaction* of the particle and the pointer. Of course, in reality, this interaction would be quite indirect and quite complicated, mediated somehow

by all the *other* particles composing the apparatus. For our schematic treatment, though, all we really want to ensure is the following: *if* the particle were to start out definitely in an energy eigenstate ψ_n , then the pointer should move sideways by a distance proportional to E_n , the corresponding energy eigenvalue. That, after all, is what pointers on measuring devices *do*, rather by definition – they register the outcome of the measurement by their positions, and if the particle starts out in an energy eigenstate, then we know the outcome will definitely be that corresponding eigenvalue.

It turns out that an interaction Hamiltonian of the following form will achieve this:

$$\hat{H}_{int} = \lambda \hat{H}_x \hat{p}_y = -i\hbar\lambda \hat{H}_x \frac{\partial}{\partial y}. \quad (3.11)$$

Here λ is a constant; \hat{H}_x is the energy operator for the particle in the box, i.e., \hat{H}_x is the operator corresponding to the quantity that we are *measuring*; and \hat{p}_y is the momentum operator for the pointer.

Let's see why this works. Suppose again that the particle starts out in an energy eigenstate, so that the particle-pointer initial wave function is

$$\Psi(x, y, 0) = \psi_n(x)\phi(y). \quad (3.12)$$

Now for simplicity assume that λ (which describes the strength of the interaction) is very large, so that – during the period in which the interaction is occurring – we can ignore any other terms in the overall Hamiltonian. Then the Schrödinger equation reads

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}_{int} \Psi = -i\hbar\lambda \hat{H}_x \frac{\partial \Psi}{\partial y}. \quad (3.13)$$

Simplifying a bit gives

$$\frac{\partial \Psi}{\partial t} = -\lambda \hat{H}_x \frac{\partial \Psi}{\partial y}. \quad (3.14)$$

One sees that the t - and y -dependencies of Ψ are coupled, but the variable x is not involved. We may thus assume that $\Psi(x, y, t)$ remains proportional to $\psi_n(x)$ and hence remains an eigenstate of \hat{H}_x , so that the previous equation simplifies to

$$\frac{\partial \Psi}{\partial t} = -\lambda E_n \frac{\partial \Psi}{\partial y}. \quad (3.15)$$

It is then straightforward to check that the solution is

$$\Psi(x, y, t) = \psi_n(x)\phi(y - \lambda E_n t). \quad (3.16)$$

Suppose the interaction lasts until a time $t = T$. Then the quantum state of the particle-pointer system at the end of the interaction is evidently

$$\Psi(x, y, T) = \psi_n(x)\phi(y - \lambda E_n T) \quad (3.17)$$

which can be understood as follows: the particle-in-the-box remains in the n^{th} energy eigenstate, and the pointer remains a Gaussian wave packet but whose center has *moved*, to the right, a distance $\lambda E_n T$ so that it is now centered at $y_0 + \lambda E_n T$ which we can identify as y_n – the final location of the pointer when it registers the n^{th} energy value.

To summarize, the (admittedly weird-looking) interaction Hamiltonian, Eq. (3.11), does exactly the job we wanted: it makes the pointer *move* a distance proportional to the energy of the particle when the particle actually begins with a particular, definite energy. So it seems like this model – with this interaction Hamiltonian – provides a schematic, but still faithful, way of capturing all the complicated physical interactions that in fact couple the particle-in-the-box to the apparatus pointer in this kind of situation.

(A technical aside: if you are worried that we dropped something important, either by ignoring \hat{H}_x during the measurement interaction, or by setting the mass of the pointer to infinity and hence ignoring \hat{H}_y , you shouldn't be. Including \hat{H}_x would only have the effect of giving an extra factor $e^{-iE_n t/\hbar}$ in Eq. (3.17) – a meaningless overall phase. And if we had included a term corresponding to the kinetic energy of the pointer, this would only have the effect of making the pointer wave packet spread slightly during the course of the interaction. This also doesn't really change anything important, so the approximations made above really do seem to capture what is essential. You are invited to explore this in the Projects if you want to.)

So far so good. But of course we are not so interested in the special case where the particle starts out with a definite energy. We want to know what happens when the particle starts out in the superposition state, Eq. (3.4), and we try to treat the measurement interaction fully quantum mechanically – i.e., without bringing in any *ad hoc* extra *postulates* about exceptions to the Schrödinger evolution, the pointer always having a classical position, etc. Remember, the hope is that, if we just use the purely microscopic part of quantum mechanics – wave functions obeying Schrödinger's equation – to describe the entire interaction between the two systems, everything will work out the way we want it to: the final quantum state will attribute an approximately-definite position to the pointer, the wave function of the particle-in-a-box will be one of the energy eigenstates, etc.

But, sadly, our hope is immediately dashed. It is very easy to see – from the fact that the overall Schrödinger equation is *linear* – that with

$$\hat{H} = \hat{H}_{int} \quad (3.18)$$

and

$$\Psi(x, y, 0) = \left(\sum_i c_i \psi_i(x) \right) \phi(y) \quad (3.19)$$

the wave function at time T (when the measurement interaction ceases) is

$$\Psi(x, y, T) = \sum_i c_i \psi_i(x) \phi(y - \lambda E_i T). \quad (3.20)$$

This represents an *entangled superposition* of several states, in each of which the particle has a definite energy and the pointer's position is slightly fuzzy but centered on a definite position corresponding to the energy of the particle. But of course, the energy of the particle, and the post-measurement position of the pointer, are *different* in each of the superposed states. And that is seriously problematic. The particle-in-a-box does not end up in a particular energy eigenstate at all, and (worse!) the pointer is not localized around any particular one of (what we thought of previously as) its possible final positions.

To summarize: if you try to treat the measurement process using just the microscopic part of quantum mechanics, it simply doesn't give you what you want, which is some kind of explanation for the emergence of one definite outcome (pointer position). Instead of somehow *resolving* the initial ambiguity in the energy of the particle, and thereby causing the particle to end up with a definite energy and the pointer to end up in a definite place, the interaction between the particle-in-the-box and the pointer *infects* the pointer with its quantum ambiguity!

Now you might think that this result is simply a consequence of our having treated the measurement process so schematically. Perhaps a more detailed, more realistic, quantum mechanical description of the measuring apparatus and its interaction with the quantum system, would yield the desired result? Unfortunately it is easy to see that this cannot possibly work. You can include as many of those intermediate degrees of freedom as you want, making the description as realistic and as accurate as you like, and still it won't make any difference to the outcome. Here's why. Suppose we include another intermediate degree of freedom, called z , so that now the initial state is something like

$$\Psi(x, y, z, 0) = \left(\sum_i c_i \psi_i(x) \right) \phi(y) \chi_0(z). \quad (3.21)$$

Schrödinger's equation reads

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H} \Psi. \quad (3.22)$$

Since we mean to be describing a *measurement* of the particle's energy, we demand that the Hamiltonian \hat{H} have the property that if

$$\Psi(x, y, z, 0) = \psi_n(x) \phi(y) \chi_0(z) \quad (3.23)$$

then

$$\Psi(x, y, z, T) = \psi_n(x) \phi(y - \lambda E_n T) \chi_n(z). \quad (3.24)$$

That is, we demand that the position of the pointer should move, by an amount proportional to E_n , when the particle starts out in an energy eigenstate with energy

E_n . (And the intermediate degree of freedom ends up in some associated state.) But then it is obvious, again from the linearity of the Schrödinger equation, that in the general case where the particle starts out in a superposition of energies, as in Eq. (3.21), it ends up in the entangled superposition state

$$\Psi(x, y, z, T) = \sum_i c_i \psi_i(x) \phi(y - \lambda E_i T) \chi_i(z). \quad (3.25)$$

The intermediate degree of freedom just gets infected with the superposition, too. Including it – or, indeed, including as many such intermediate degrees of freedom as you might want and thereby making the description as complete and accurate and unschematic as you might want – changes nothing. Our simple schematic treatment brought out the essential and inevitable problem already.

3.3 Schrödinger’s Cat and Einstein’s Bomb

The most famous illustration of the problem described in the last section was presented by Schrödinger in the last section of his 1935 paper, “The present situation in quantum mechanics.” [3] Actually, in that paper, Schrödinger illustrates the problem several times, with several different examples. Here, for instance, is his discussion of the case of a radioactive nucleus emitting an alpha particle:

[the ψ -] function has provided quite intuitive and convenient ideas, for instance the ‘cloud of negative electricity’ around the nucleus, etc. But serious misgivings arise if one notices that the uncertainty affects macroscopically tangible and visible things, for which the term ‘blurring’ seems simply wrong. The state of a radioactive nucleus is presumably blurred in such degree and fashion that neither the instant of decay nor the direction, in which the emitted α -particle leaves the nucleus, is well-established. Inside the nucleus, blurring doesn’t bother us. The emerging particle is described, if one wants to explain intuitively, as a spherical wave that continuously emanates in all directions from the nucleus and that impinges continuously on a surrounding luminescent screen over its full expanse. The screen however does not show a more or less constant uniform surface glow, but rather lights up at *one* instant at *one* spot – or, to honor the truth, it lights up now here, now there, for it is impossible to do the experiment with only a single radioactive atom. If in place of the luminescent screen one uses a spatially extended detector, perhaps a gas that is ionised by the α -particles, one finds the ion pairs arranged along rectilinear columns, that project backwards on to the bit of radioactive matter from which the α -radiation comes (C.T.R. Wilson’s cloud chamber tracks, made visible by drops of moisture condensed on the ions) [3].

The idea here is illustrated in Fig. 3.3. Schrödinger’s point is that, according to quantum mechanics, the α particle emitted by a radioactive nucleus is not going in any particular direction. Instead, the theory describes it as coming out of the nucleus in a superposition of all possible directions, which is mathematically equivalent to a spherically symmetric, outward-propagating wave function. It is easy to understand how a flash – at a particular location on the screen – could be created by an α particle that had been emitted in a particular direction, namely, toward that particular spot.

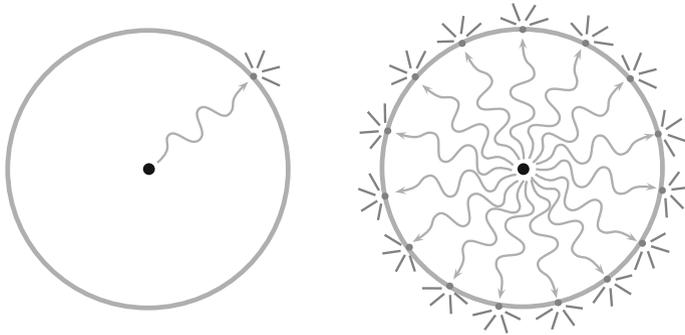


Fig. 3.3 A radioactive nucleus emits an alpha particle which then causes a visible scintillation on a surrounding circular detection screen. The left panel shows that, if the alpha particle is emitted in a particular direction, the scintillation will occur at a spot on the screen in that same direction. But if both the alpha particle and the detection screen are treated quantum mechanically, the linearity of Schrödinger's equation implies that if the alpha particle is emitted in a superposition of different directions (say, a spherically symmetric wave function propagating outward) the final quantum state will be an entangled superposition involving terms with scintillations in all directions on the screen. That is, the microscopic quantum dynamics cannot explain how it is that the flash is seen to occur at just some one particular spot on the screen

But if the emission of the alpha particle is really somehow spherically symmetric, there seems to be nothing in the microscopic part of quantum mechanics to break the symmetry and explain the flash occurring at a particular spot.

The situation here can be described quantum mechanically in the same way we described the measurement process in the last section. Take ψ_n to be the wave function of an α -particle that has a reasonably sharply-defined propagation direction θ_n , and take ϕ_0 to be the wave function of a photo-luminescent screen on which no flashes have yet appeared. Then the idea is that the overall wave function (for the α particle and screen jointly) would evolve, under Schrödinger's equation, as follows:

$$\psi_n \phi_0 \rightarrow \psi_n \phi_n \quad (3.26)$$

where ϕ_n is the wave function for the photo-luminescent screen with a bright flash at angle θ_n .

But then this immediately implies – from the linearity of Schrödinger's equation – that if (as is in fact the case in this kind of situation) the wave function of the α particle is a (say, spherically symmetric) superposition

$$\psi_{sph} \sim \sum_i \psi_i \quad (3.27)$$

the state will evolve, under Schrödinger's equation, as follows:

$$\psi_{sph}\phi_0 \rightarrow \sum_i \psi_i\phi_i. \quad (3.28)$$

This is of course an entangled superposition of states; each term in the superposition has the alpha particle being emitted in a particular direction and the screen flashing at a particular point, but the superposition as a whole includes such terms corresponding to all possible directions. No one particular direction is picked out, either for the alpha particle *or for the flash*. As Schrödinger points out, though, this democratic wave function does not seem to correspond appropriately to what we actually observe in this kind of case, which is a flash at some particular definite location: “The screen however does not show a more or less constant uniform surface glow, but rather lights up at *one* instant at *one* spot” [3].

Schrödinger immediately continues the discussion by describing the case of the famous cat:

One can even set up quite ridiculous cases. A cat is penned up in a steel chamber, along with the following diabolical device (which must be secured against direct interference by the cat): in a Geiger counter there is a tiny bit of radioactive substance, *so* small, that *perhaps* in the course of one hour one of the atoms decays, but also, with equal probability, perhaps none; if it happens, the counter tube discharges and through a relay releases a hammer which shatters a small flask of hydrocyanic acid. If one has left this entire system to itself for an hour, one would say that the cat still lives *if* meanwhile no atom has decayed. The first atomic decay would have poisoned it. The ψ -function of the entire system would express this by having in it the living and the dead cat (pardon the expression) mixed or smeared out in equal parts [3].

To put some formalism to this, we would say that, in the course of the hour, the single radioactive atom evolves (according to Schrödinger's equation) into an equally-weighted superposition of decayed and not-yet-decayed states:

$$\psi_0 \rightarrow \frac{1}{\sqrt{2}} \psi_{\text{decayed}} + \frac{1}{\sqrt{2}} \psi_{\text{not decayed}}. \quad (3.29)$$

If we treated the surrounding equipment *classically*, and applied the collapse postulate, we would say that when the atom interacts with (say) the Geiger counter, this interaction triggers a collapse and it thereby becomes unambiguously the case that *either* the atom has decayed and the Geiger counter has clicked, *or* the atom has not decayed and the Geiger counter has not clicked. Then, a classical interaction (mediated by the hammer and flask of acid) between the Geiger counter and the cat would result in the cat definitely being alive if the Geiger counter did not click, and the cat definitely being dead if the Geiger counter did click.

However, if we instead treat the surrounding apparatus quantum mechanically, we find that the final state is something like the following:

$$\Psi_f = \frac{1}{\sqrt{2}} \psi_{\text{decayed}} \phi_{\text{shattered}} \chi_{\text{dead}} + \frac{1}{\sqrt{2}} \psi_{\text{not decayed}} \phi_{\text{intact}} \chi_{\text{alive}}. \quad (3.30)$$

This is a superposition of two states: (i) a state in which the atom is decayed, the hammer is down and the flask of poison is shattered, and the cat is dead; and (ii) a state in which the atom is not decayed, the hammer is up and the flask of poison is intact, and the cat is alive. In particular, this is not a state in which there is any definite fact of the matter about whether the cat is dead or alive. The poor cat is, in Schrödinger's words, "mixed or smeared out" between living and dead.

As a bit of historical context, it is perhaps interesting to note that in the weeks leading up to Schrödinger's submitting the paper containing the cat example, he was exchanging letters with Einstein. And in one of his letters to Schrödinger (from August 8, 1935), Einstein suggested an example that illustrates the same basic point illustrated by the cat:

The system is a substance in chemically unstable equilibrium, perhaps a charge of gunpowder that, by means of intrinsic forces, can spontaneously combust, and where the average life span of the whole setup is a year. In principle this can quite easily be represented quantum-mechanically. In the beginning the ψ -function characterizes a reasonably well-defined macroscopic state. But, according to your equation, after the course of a year this is no longer the case at all. Rather, the ψ -function then describes a sort of blend of not-yet and of already-exploded systems. Through no art of interpretation can this ψ -function be turned into an adequate description of a real state of affairs; [for] in reality there is just no intermediary between exploded and not-exploded [4, p. 78].

So perhaps the basic idea of the "Schrödinger's cat" example actually started with Einstein? This is again suggested by a later letter in which Einstein seems to get slightly confused and mixes the two examples together in an amusing way:

Dear Schrödinger,

.... I am as convinced as ever that the wave representation of matter is an incomplete representation of the state of affairs, no matter how practically useful it has proved itself to be. The prettiest way to show this is by your example with the cat (radioactive decay with an explosion coupled to it). At a fixed time parts of the ψ -function correspond to the cat being alive and other parts to the cat being pulverized.

If one attempts to interpret the ψ -function as a complete description of a state, independent of whether or not it is observed, then this means that at the time in question the cat is neither alive nor pulverized. But one or the other situation would be realized by making an observation.

If one rejects this interpretation then one must assume that the ψ -function does not express the real situation but rather that it expresses the contents of our knowledge of the situation. This is Born's interpretation, which most theorists today probably share. But then the laws of nature that one can formulate do not apply to the change with time of something that exists, but rather to the time variation of the content of our legitimate expectations.

Both points of view are logically unobjectionable; but I cannot believe that either of these viewpoints will finally be established.

There is also the mystic, who forbids, as being unscientific, an inquiry about something that exists independently of whether or not it is observed, i.e., the question as to whether or not the cat is alive at a particular instant before an observation is made (Bohr). Then both interpretations fuse into a gentle fog, in which I feel no better than I do in either of the previously mentioned interpretations, which do take a position with respect to the concept of reality.

I am as convinced as ever that this most remarkable situation has come about because we have not yet achieved a complete description of the actual state of affairs.

Of course I admit that such a complete description would not be observable in its entirety in the individual case, but from a rational point of view one also could not require this....

Best regards from

Yours, A. Einstein [5, pp. 35-6]

In any case, whoever deserves credit for originating the example, it is nice to know that Schrödinger and Einstein agreed about what it established. Here, for example, is Einstein – again writing to Schrödinger – now on Sept 4, 1935, just after Schrödinger submitted his manuscript:

...your cat shows that we are in complete agreement concerning our assessment of the character of the current theory. A ψ -function that contains the living as well as the dead cat just cannot be taken as a description of a real state of affairs. To the contrary, this example shows exactly that it is reasonable to let the ψ -function correspond to a statistical ensemble that contains both systems with live cats and those with dead cats [4, p. 84].

For Einstein and Schrödinger, then, the cat/bomb example strongly suggested that, rather than providing direct, literal, complete descriptions of physical systems, quantum mechanical wave functions should instead be understood as describing our incomplete knowledge – our ignorance – about the physical states of these systems. As Einstein puts it here, the wave function (attributing, typically, a *range* of possible values to various system properties) should not be understood as a complete description of an individual system, but should instead be understood as characterizing a statistical ensemble of systems with some variation among the individual members of the ensemble. We will discuss this alternative viewpoint in more detail in the following section.

But, after all this talk of dead cats and exploding bombs, let's close this section with a slightly-happier image of Schrödinger's cat as re-envisaged by Bell. In Bell's re-telling – See Fig. 3.4 – the poison which is either released, or not, is replaced by

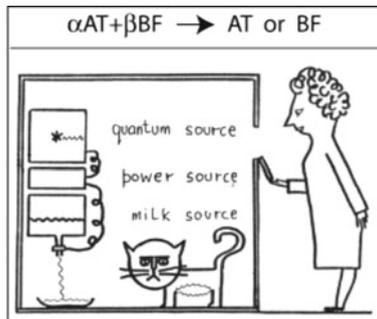


Fig. 3.4 Bell's version of Schrödinger's cat. The state of the radioactive nucleus ("A" for "not decayed" and "B" for "decayed") becomes entangled with the delivery (or not) of milk into the cat's dish and thereby also with the size of the cat's stomach ("T" for "thin" and "F" for "fat"). From Ref. [6]. Figure © IOP Publishing. Reproduced with permission. All rights reserved. <https://doi.org/10.1088/1751-8121/40/12/S02>

a portion of milk that is either released, or not, into a dish, for the cat to drink. If the nucleus decays, the cat gets fed and ends up fat – whereas if the nucleus does not decay, the cat does not get fed and ends up thin. But in either case he survives!

3.4 Hidden Variables and the Ignorance Interpretation

In recent popular culture, Schrödinger’s cat has become a kind of symbol or emblem of the weirdness of quantum mechanics. Many people therefore have the impression that Schrödinger thought the cat really *would* end up “mixed or smeared out” between live and dead, and that one simply had to accept this as true despite its incomprehensibility. But that is completely wrong. In fact, Schrödinger intended the cat thought experiment as a *reductio ad absurdum* of the idea that quantum mechanical wave functions provide *complete* descriptions of physical systems. The idea was that, in this kind of case, quantum mechanics implies something that is (at least in the opinion of Schrödinger and Einstein) *obviously wrong*. The pointer does *not* in fact end up in some kind of superposition of different locations, but rather points to one particular spot. The screen around the radioactive nucleus “does not show a more or less constant uniform surface glow, but rather lights up at *one* instant at *one* spot”. And the cat is most certainly either (fully, definitely) alive or (fully, definitely) dead – not both alive and dead, “mixed or smeared out in equal parts.” [3]

Schrödinger summarizes his point as follows just after presenting the cat example:

It is typical of these cases that an indeterminacy originally restricted to the atomic domain becomes transformed into macroscopic indeterminacy, which can then be *resolved* by direct observation. That prevents us from so naively accepting as valid a ‘blurred model’ for representing reality. In itself it would not embody anything unclear or contradictory. There is a difference between a shaky or out-of-focus photograph and a snapshot of clouds and fog banks [3].

He continues shortly after by noting that, with the cat example,

we saw that the indeterminacy is not even an actual blurring, for there are always cases where an easily executed observation provides the missing knowledge [3].

I would summarize Schrödinger’s point here this way. If we describe the entire measurement process using the microscopic part of quantum mechanics, the theory tells us that the measuring apparatus (or some other macroscopic object like the cat) ends up in its own ambiguous, superposed state. But *we know this cannot be a complete description of the state of such things* since direct observation reveals that such objects are always in perfectly definite states. Therefore, at least when it is used to describe the state of macroscopic things, the quantum mechanical description *cannot be complete*: the ambiguity of quantum superposition must (as Einstein also remarked in the letter quoted in the previous section) refer to *our ignorance* about which of several possibilities is in fact realized, as opposed to describing an objective blurring in the physical state of the object itself. This, I think, is the point of the intriguing sentence “There is a difference between a shaky or out-of-focus photograph and a snapshot of

clouds and fog banks.” He is suggesting that quantum mechanical wave functions, thought of as describing or depicting the objective physical states of things, are *not* like (sharp, in-focus) photographs of clouds – i.e., faithful reproductions of things which are themselves, objectively, smeared out and fuzzy. Instead, he means to suggest, we should understand quantum mechanical wave functions as like “shaky or out-of-focus photograph[s]” of objects that are, in themselves, perfectly sharp. In this kind of case, the smeared out or fuzzy character does not pertain to the object described, but is instead a kind of failing or imperfection in the reproduction process.

But surely there is no fundamental distinction between microscopic and macroscopic systems (the latter, after all, literally being made of the former). This, I think, is the point of including, in the cat example, the detailed description of the intermediate parts of the mechanism – the causal chain – whereby the state of the nucleus is coupled to the state of the cat. Surely, Schrödinger invites us to think, there is no particular spot along this continuous chain between micro and macro where it would make sense to draw a sharp line and say “different dynamical laws start applying *here*”. But if the micro and the macro must be treated uniformly, and if the quantum mechanical description (in terms of wave functions) of macroscopic systems (like cats and pointers) is not complete, then surely this is also the case for microscopic systems.

If that’s right, then, for example, when we say that the particle-in-the-box is in a superposition of several different energy eigenstates,

$$\psi_0 = c_1\psi_1(x) + c_2\psi_2(x) + c_3\psi_3(x), \quad (3.31)$$

what this must *mean* is that the particle is *either* in the state $\psi_1(x)$ *or* the state $\psi_2(x)$ *or* the state $\psi_3(x)$... we’re just not sure which one! It’s not that the energy of the particle is somehow blurred or indefinite – rather, it’s only our *knowledge* which is blurred or indefinite. The energy is *uncertain* (in the literal sense, meaning “unknown to us”) but it is perfectly sharp, some one definite value or another, in reality. And then, if the energy is measured, we simply *find out* what the energy was all along.

Or similarly, when we say that the wave function of the emitted alpha particle is spherically symmetric, what this *means*, according to this viewpoint, is just that we have no idea which direction the alpha particle is going. The subjective probability distribution we would assign to its direction is spherically symmetric, but the thing itself isn’t! The alpha particle itself, on this view, is already moving in some one particular direction – we don’t know which one, but it is perfectly definite in reality all the same. Seeing a flash at some particular spot on the surrounding screen is then not a big mystery and not a proof that the microscopic quantum dynamics is wrong... it’s simply the way we *find out* which direction the alpha particle was going all along.

In its simplest (or, one might say, most naive possible) form, this view might be called the “ignorance interpretation of superposition”. I think it should be admitted that it has a certain alluring reasonableness. Indeed, for some people reading this, it may be the view that you have had in mind all along! By getting rid of any “quantum fuzziness” at the root, down at the microscopic scale, the “ignorance interpretation of superposition” totally pre-empts the difficulty, illustrated by Schrödinger’s cat, of

amplifying the fuzziness up to a macroscopic scale where, apparently, it conflicts with our direct experience of the world.

Notice also that the ignorance interpretation provides a beautifully simple resolution of our earlier worries about the collapse postulate: if wave functions are not really descriptions of the physical states of systems at all, but instead descriptions of the state of our *knowledge* of those systems, then there is nothing remotely problematic about wave functions collapsing. The collapse is simply an updating of our knowledge, when we get new data! So for example when we measure the energy of the particle-in-the-box and its wave function collapses from $\psi_0 = c_1\psi_1 + c_2\psi_2 + c_3\psi_3$ to, say, ψ_2 , this was not a dynamical process, a change in the physical state of the particle, at all. The measurement simply *reveals* something that was there all along but unknown to us. We simply learn the value of the particle's energy, which we did not know before, and so update ψ – our “knowledge catalog” – accordingly. All the worries about inconsistent dynamical rules, and ambiguities about which ones should apply when, and so on – all of these simply evaporate if we adopt the ignorance interpretation.

This was essentially the view put forward by Max Born in 1926. Born's view morphed significantly as it became bound up with the Copenhagen Interpretation that we will discuss in Chap. 6, but the *original* “Born interpretation” was nothing but the ignorance interpretation we have discussed here. (See Sect. 2.4 of Ref. [7] for a nice overview.) And this view continued to enjoy support (from those who resisted the Copenhagen orthodoxy) in subsequent decades. For example, in an essay written in 1949, just a few years before his death in 1955, Einstein seemed to again advocate something along these lines:

Within the framework of statistical quantum theory there is no such thing as a complete description of the individual system. More cautiously it might be put as follows: The attempt to conceive the quantum-theoretical description as the complete description of the individual systems leads to unnatural theoretical interpretations, which become immediately unnecessary if one accepts the interpretation that the description refers to ensembles of systems and not to individual systems. In that case the whole ‘egg-walking’ performed in order to avoid the ‘physically real’ becomes superfluous. There exists, however, a simple psychological reason for the fact that this most nearly obvious interpretation is being shunned. For if the statistical quantum theory does not pretend to describe the individual system (and its development in time) completely, it appears unavoidable to look elsewhere for a complete description of the individual system; in doing so it would be clear from the very beginning that the elements of such a description are not contained within the conceptual scheme of the statistical quantum theory. With this one would admit that, in principle, this scheme could not serve as the basis of theoretical physics. Assuming the success of efforts to accomplish a complete physical description, the statistical quantum theory would, within the framework of future physics, take an approximately analogous position to the statistical mechanics within the framework of classical mechanics. I am rather firmly convinced that the development of theoretical physics will be of this type; but the path will be lengthy and difficult [8, p. 671].

So this point of view not only seems quite sensible but also seems to have a strong pedigree.

But, unfortunately, the ignorance interpretation – at least in its simplest form – cannot possibly be right. It would seem to imply, for example, that in the double

slit experiment described in the last chapter, in the middle of which the particle's quantum state is a superposition of "going through slit 1" and "going through slit 2"

$$\psi = \frac{1}{\sqrt{2}} [\psi_{\text{slit 1}} + \psi_{\text{slit 2}}], \quad (3.32)$$

there is nevertheless some fact of the matter about which slit the particle *really* went through: *either* slit 1 *or* slit 2. But if each particle is really just a literal particle which has a perfectly definite (if sometimes unknown) position, it is (to understate it) very difficult to understand how the subsequent particle locations could form an *interference pattern*. The interference pattern strongly – indeed, I think, conclusively – establishes that the quantum wave function is really something physical, something real, not just a description of our incomplete state of knowledge.

In addition, there are a number of rigorous mathematical theorems proving that it is impossible to assign definite values to quantum properties in, at least, the naive way suggested by the ignorance interpretation. Here we briefly indicate the flavor of these so-called "no hidden variable theorems" with a simple example. (See the Projects for two additional examples discussed already by Schrödinger in his 1935 paper.)

Consider the case of a single spin-1/2 particle whose spin might be measured along the \hat{z} , \hat{x} , or \hat{n} directions (where \hat{n} is in the $x - z$ plane and halfway between \hat{x} and \hat{z} , i.e., 45° away from both). The operators corresponding to the particle's spin along these three directions are

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (3.33)$$

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (3.34)$$

and

$$\hat{\sigma}_n = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{\sqrt{2}} [\hat{\sigma}_x + \hat{\sigma}_z]. \quad (3.35)$$

Now if the particle has a definite spin along one of these directions (i.e., if its wave function is an eigenstate of one of the three operators) it will be in a superposition of "spin-up" and "spin-down" with respect to the other two directions. Thus, according to the idea that the wave function provides a complete description of the physical state of the particle, a particle can never possess a definite value of spin along all three of these directions at once.

But, according to the ignorance interpretation of superposition, the fact that the quantum state is a superposition does not mean that the particle doesn't have a definite value of spin. So we contemplate the possibility that particles have definite spin values, s_x , s_y , and s_n , along all three directions at once. These values would in general not all be known at once, although presumably the idea is that a measurement

of the corresponding quantity will simply reveal the value in question. Since (let's say) such measurements have outcomes $+1$ (meaning “spin-up”) or -1 (meaning “spin-down”), we should assign either the value $+1$ or -1 to each of s_x , s_y , and s_n .

Furthermore, there is some reason to expect that the values of the spin along these three directions should obey the same mathematical relationships that are obeyed by the corresponding operators: the quantum mechanical expectation values, for example, will have this relationship, and it is plausible to suspect that, in the context of a “hidden variable” theory like we are contemplating here, the average values will have this relationship because each individual set of possible values has this relationship. Equation (3.35) would then lead us to demand that

$$s_n = \frac{1}{\sqrt{2}} [s_x + s_z]. \quad (3.36)$$

But it is immediately obvious that it will be mathematically impossible to assign values $+1$ or -1 to s_x , s_z , and s_n in accordance with Eq. (3.36): the quantity in square brackets on the right will be -2 , 0 , or $+2$, and none of these, divided by $\sqrt{2}$ is $+1$ or -1 .

The argument just sketched is closely related to the first “no hidden variables” proof, given by John von Neumann in 1932. Historically, von Neumann’s argument convinced many people that the type of ignorance interpretation favored by Schrödinger and Einstein was untenable, and thus provided a nudge in the direction of the Copenhagen interpretation (which we will study in greater depth in Chap. 6). As it turns out, though, von Neumann’s argument is rather impotent. This was first pointed out in 1935 by Grete Hermann, but her critique tragically failed to gain any traction in the physics community. [7] Several decades later, John Stewart Bell independently tackled the question of whether “hidden variable” theories had been mathematically refuted; he would later describe von Neumann’s proof as “not merely false but *foolish!*” [9].

The reason for this harsh assessment has to do with the requirement that the “hidden variables” should obey the same mathematical relationships as their corresponding quantum mechanical operators – a requirement which perhaps makes sense for *commuting* operators, but which is completely unmotivated for sets of *non-commuting* operators like $\hat{\sigma}_x$, $\hat{\sigma}_z$, and $\hat{\sigma}_n$.

In more recent decades, “no hidden variables” proofs have been found whose assumptions are somewhat more reasonable. Getting into the details here would take us too far afield, but suffice it to say it is seriously problematic to think that one can understand all measurements as simply revealing the values of properties that are, while unknown, perfectly definite and independent of the measurement procedure itself. The naive ignorance interpretation, that is, is really not tenable. (Crucial references here include Bell’s paper “On the problem of hidden variables in quantum mechanics” [10] and Mermin’s review article [9].) As we will see in Chap. 7, however, there does exist a perfectly viable “hidden variable” theory with

a crucial property called “contextuality” (meaning, in a nutshell, that some *but not all* measurements simply reveal the pre-existing values of properties) that allows it to elude all of these impossibility theorems!

3.5 Wrap-Up

For now, let us try to recap, summarize, and package what we’ve seen in this Chapter. It will probably be helpful, for example, to step back and try to get clear on the answer to the following question:

What, exactly, is “the measurement problem”?

It is admittedly a little confusing, because this phrase is used to refer to several inter-related things, all of which we have touched on here.

To begin with, sometimes “the measurement problem” refers to the fact that the postulates of textbook quantum mechanics include statements about “measurements” (and their outcomes), even though “measurement” is a very fuzzy and human concept. That is, it simply is not clear exactly which set of physical interactions or processes in nature should count as “measurements”. And so, until or unless this is somehow clarified, it simply isn’t clear exactly what the theory is even saying. This is the point Bell had in mind when he wrote that “[t]he concept of ‘measurement’ becomes so fuzzy on reflection that it is quite surprising to have it appearing in physical theory *at the most fundamental level.*” [11] Or, as he put it elsewhere – somewhat less diplomatically – “conventional formulations of quantum theory, and of quantum field theory in particular, are unprofessionally vague and ambiguous” [12].

Then there is a closely-related, second, meaning to (or aspect of) “the measurement problem”. Even if the notion of “measurement” were somehow given a clear and precise meaning – even if, that is, a sharp boundary were somehow drawn between “measurements” and “non-measurements” so that it became unambiguous when to apply which part of the quantum formalism – there would still be something unbelievable about the idea that there are these two fundamentally distinct types of processes. Equivalently (since the difference between the two supposedly distinct types of processes has to do with whether a microscopic system is, or is not, interacting appropriately with something from the other, macroscopic, “realm”) there is something unbelievable about the idea that the world is fundamentally “split” into these two distinct “realms”. Surely a proper fundamental theory should describe the *entire* universe in a coherent, unified way. And so, in this aspect, “the measurement problem” refers to the failure of standard quantum theory to provide such a unified description. Bell often remarked that quantum mechanics involved what he described as a “shifty split”. For example:

There can be no question then of identifying the quantum system S with the whole world W . There can be no question – without changing the axioms – of getting rid of the shifty split. Sometimes some authors of ‘quantum measurement’ theories seem to be trying to do just that. It is like a snake trying to swallow itself by the tail. It can be done – up to a point.

But it becomes embarrassing for the spectators even before it becomes uncomfortable for the snake [13].

And Schrödinger as well had already criticized the idea that “measurement” was somehow a special, dynamically distinct kind of process, with its own special dynamical rules. In quantum mechanics, he wrote,

any *measurement* suspends the law that otherwise governs continuous time-dependence of the ψ -function and brings about in it a quite different change, not governed by any law but rather dictated by the result of the measurement. But laws of nature differing from the usual ones cannot apply during a measurement, for objectively viewed it is a natural process like any other, and it cannot interrupt the orderly course of natural events. Since it does interrupt that of the ψ -function, the latter ... can *not* serve ... as an experimentally verifiable representation of an objective reality [3].

The idea that, “objectively viewed”, “measurement ... is a natural process like any other” perfectly captures this second aspect of “the measurement problem.”

And then, finally, “the measurement problem” also sometimes denotes the theory’s apparent inability to provide sensible results when it is modified in the obvious way in response to the criticisms of the previous paragraphs. This is the aspect that is illustrated by Schrödinger’s cat. If we refuse to accept the “fractured universe” implied by the most straightforward reading of the quantum formalism, the easiest way to try to solve *that* problem is to simply get rid of the postulates about “measurement” (and the separately-presupposed classical “realm”) and retain just the microscopic part of the theory. In this modified understanding of the theory, *everything* will be described in terms of wave functions obeying Schrödinger’s equation *always*, and so, to be sure, we have a coherent, unified worldview. But the problem – as we saw – is that this worldview simply doesn’t seem to be right. Detection screens do not “show a more or less constant uniform surface glow”, pointers on measuring devices are never blurry, and cats are never observed to be “mixed or smeared out in equal parts” of living and dead.

Sometimes this last aspect of “the measurement problem” is expressed by noting that the theory does not seem to be able to explain the occurrence of definite measurement results. That is fine as far as it goes, but it can also be confusing or misleading. The full, original theory – with “collapse postulate” and all – certainly has no difficulty explaining the occurrence of definite measurement results! Indeed, they are right there in the postulates of that version of the theory! But that is precisely the problem: those rules appear to have been implausibly put in by hand, to avoid embarrassment, and (for the reasons I summarized in what I described as the first two aspects of “the measurement problem”) it seems impossible to take them seriously as fundamental physical laws.

At the end of the day, the measurement problem is probably best understood as the problem of understanding the seemingly paradoxical relation of the collapse postulate to the rest of quantum theory. On the one hand, it seems impossible to include the collapse postulate in the axioms of the theory and still regard the theory as providing a fundamental account of the microscopic world. On the other hand, it seems impossible to eliminate the collapse postulate from the dynamical axioms of

the theory, either by simply jettisoning it (and letting wave functions evolve according to the Schrödinger equation all the time) or by interpreting it (and the wave function generally) as pertaining not to reality itself but only to our knowledge of reality.

In later Chapters, we will explore some concrete proposals for resolving (or dissolving) the measurement problem. One of these – Everett’s “Many Worlds Interpretation”, the subject of Chap. 10 – attempts to retain the idea of the quantum descriptions of reality (in terms of wave functions alone, obeying Schrödinger’s equation always) being complete. Another – the “Pilot-Wave Theory” of de Broglie and Bohm, the subject of Chap. 7 – proposes to supplement the wave function with additional (“hidden”) variables that resolve the dilemmas posed by Schrödinger, but in a way that avoids the “no hidden variable” theorems discussed here. A third proposal – the “Spontaneous Collapse” theory discussed in Chap. 9 – attempts to unify the Schrödinger equation and the collapse postulate to give a single uniform dynamical description that can be applied coherently at all scales. And then there is also a philosophical perspective – the “Copenhagen Interpretation” of Bohr and Heisenberg, discussed in Chap. 6 – that urges us to reject Schrödinger’s worries as somehow baseless and meaningless.

Before turning to these proposals, however, we explore in the following two Chapters two additional problems that seem to afflict textbook quantum theory.

Projects:

- 3.1 Show explicitly that Eq. (3.20) indeed satisfies the Schrödinger equation with $\hat{H} = \hat{H}_{int} = \lambda \hat{H}_x \hat{p}_y$.
- 3.2 Suppose, in our schematic formal treatment of the measurement of the energy of a particle-in-a-box, we use the more complete Hamiltonian operator

$$\hat{H} = \hat{H}_x + \hat{H}_y + \hat{H}_{int} \quad (3.37)$$

(with M finite so \hat{H}_y cannot just be ignored). What now is the solution to the Schrödinger equation with $\Psi(x, y, 0)$ still given by Eq. (3.6)?

- 3.3 Sketch some configuration space cartoons – in the style of Figs. 2.10 and 2.11 – to illustrate the evolution of the wave function (for the particle-in-a-box + pointer system) from Sect. 3.2.
- 3.4 In Chap. 2, we saw a simple example of measuring the momentum of a particle whose wave function was $\psi_0(x) = \sqrt{2} \sin(kx)$. Set up a formal (purely microscopic) quantum description of the measurement process: assume a “pointer” degree of freedom y , which starts in a Gaussian state centered at $y = 0$. What interaction Hamiltonian is appropriate for coupling the post-interaction pointer position to the particle’s momentum? What is the final quantum state $\Psi(x, y, T)$ at the end of the interaction?
- 3.5 A particle is in the following superposition of position eigenstates:

$$\psi(x) = \frac{1}{\sqrt{2}} [\delta(x - a) + \delta(x + a)] \quad (3.38)$$

where the “ δ ”s are Dirac delta functions. The position of this particle is to be measured by a position measuring apparatus (also described quantum mechanically) whose pointer (with degree of freedom “ y ”) should indicate the particle’s position, $x = +a$ or $x = -a$, by moving to the right or to the left, respectively. What interaction Hamiltonian \hat{H}_{int} will accomplish this? If the interaction turns on at $t = 0$ (and any contributions to the total Hamiltonian other than \hat{H}_{int} are negligible) what is the wave function $\Psi(x, y, t)$ at time t ? You should show explicitly that your answer is a solution of Schrödinger’s equation. Finally, sketch/indicate the time-evolution of $\Psi(x, y, t)$ in configuration space.

- 3.6 Schrödinger says that “[i]nside the nucleus, blurring doesn’t bother us” [3]. Why not? Why is “blurring” a problem for macroscopic things like pointers, but not a problem for microscopic things like nuclei?
- 3.7 Here is another simple “no hidden variables” argument that Schrödinger gives in Ref. [3]. Suppose that a quantum mechanical particle actually has a definite position \vec{r} and a definite momentum \vec{p} (even though these two quantities cannot be simultaneously known). Then it will have an angular momentum of magnitude $|\vec{L}| = |\vec{r} \times \vec{p}|$. Suppose that a measurement of the angular momentum magnitude simply reveals this pre-existing value. Now note that, by varying the origin with respect to which we measure the position, \vec{r} – and therefore also $|\vec{L}|$ – can take on any value in a whole continuous spectrum. Explain how this is inconsistent with the (quantized!) measurement outcomes for angular momentum measurements, and therefore why the quantities \vec{r} , \vec{p} , and \vec{L} cannot possess pre-existing definite values (satisfying the relation $\vec{L} = \vec{r} \times \vec{p}$) which are simply revealed by measurements.
- 3.8 Here is yet another simple “no hidden variables” argument from Ref. [3]. Consider a quantum mechanical simple harmonic oscillator, with energy operator (Hamiltonian) $\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2$. Suppose the oscillator is in its ground state with energy $E = \frac{1}{2}\hbar\omega$. The naive sort of “hidden variable” theory (associated with the ignorance interpretation of superposition) would say that this state describes an ensemble of individual systems, all with energy E , but different values of x and p satisfying $E = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2$. Explain why this is not straightforwardly possible. (Hint: if the assumption is that position measurements simply reveal the actual pre-existing value of x , the Born rule implies that arbitrarily large values of $|x|$ are represented in the ensemble.)
- 3.9 One of the main ideas of this chapter is that there is no hope of introducing a sharply defined notion of dynamical collapse, such that the Schrödinger evolution and the other kind of evolution each apply in their own well-defined and non-overlapping spheres. But there is one idea for sharply defining such a boundary; it was proposed (or at least considered) by Eugene Wigner (and is perhaps somewhat widespread in more popular accounts of QM, for example the weird movie “What the bleep do we know”). The idea is that wave function collapse happens when physical matter interacts with *mind*. So, for example, in the Schrödinger’s cat case, the wave function obeys the linear Schrödinger

equation when the radioactive atom is in the process of decaying, and when it is interacting with the Geiger counter, which in turn interacts with the hammer which interacts with the vial of poison which interacts with the cat ... all of this ends up in the superposition described in the text ... until the moment some human opens the box and becomes *consciously aware* of the result, at which point the involvement of her *mind* (presumably, to be specific, the interaction of her mind with her brain) collapses the wave function, for the whole physical system up through and including the brain, down to one or the other of the definite results. What do you think of this idea? Is it a good possible solution to the measurement problem, or utter nonsense, or what? (See Wigner's essay "Remarks On the Mind-Body Question" [14].)

- 3.10 In his "Reply to criticisms" Einstein gives a nice one-particle version of a Schrödinger's Cat type argument:

If our concern is with macroscopic masses (billiard balls or stars), we are operating with very short de Broglie waves, which are determinative for the behavior of the center of gravity of such masses. This is the reason why it is possible to arrange the quantum-theoretical description for a reasonable time in such a manner that for the macroscopic way of viewing things, it becomes sufficiently precise in position as well as in momentum. It is true also that this sharpness remains for a long time and that the quasi-points thus represented behave just like the mass-points of classical mechanics. However, the theory shows also that, after a sufficiently long time, the point-like character of the ψ -function is completely lost to the center of gravity-co-ordinates, so that one can no longer speak of any quasi-localisation of the centers of gravity. The picture then becomes, for example in the case of a single macro-mass-point, quite similar to that involved in a single free electron.

If now, in accordance with the orthodox position, I view the ψ -function as the complete description of a real matter of fact for the individual case, I cannot but consider the essentially unlimited lack of sharpness of the position of the (macroscopic) body as *real*. On the other hand, however, we know that, by illuminating the body by means of a lantern ... we get a (macroscopically judged) sharp determination of position. In order to comprehend this I must assume that the sharply defined position is determined not merely by the real situation of the observed body, but also by the act of illumination. This is again a paradox.... The spook disappears only if one relinquishes the orthodox standpoint, according to which the ψ -function is accepted as a complete description of the single system [8].

Work out some quantitative estimates of the time durations involved in this kind of case. For example, consider the center-of-mass coordinate of a billiard ball. Suppose, at $t = 0$, it is described quantum mechanically by a Gaussian wave function of width one nanometer. How long would it take for the wave function to spread to a width of order, say, a meter? How long would it take the position of, say, a planet to become implausibly fuzzy?

- 3.11 Read Schrödinger's cat paper, Ref. [3], and report on anything you find interesting that wasn't already covered here.
- 3.12 In a Stern–Gerlach experiment, one can think of the *position* of the particle as the "pointer" that indicates the outcome of the spin measurement. Suppose, for example, a spin 1/2 particle begins in the product state

$$\Psi_0 = \psi_{+z} \phi(z) \quad (3.39)$$

where ψ_{+z} is the spin eigenstate (“spin-up along the z -axis”) and $\phi(z)$ is a Gaussian wave packet. (The z axis here is the one along which the Stern–Gerlach apparatus has a non-uniform magnetic field, i.e., the direction along which the beam of incoming particles will be split.) What does the wave function evolve into during the course of the experiment? Sketch a diagram. What if the initial state is instead

$$\Psi_0 = \psi_{+x} \phi(z) \quad (3.40)$$

and it is still the z -component of the spin that is being measured? Sketch another diagram and discuss the relationship to the examples discussed in the Chapter. (Could the role of the two properties be reversed? That is, could the spin be considered as a pointer indicating the position along the z -axis? Discuss.)

- 3.13 In his essay “The Problem of Measurement” [15], Wigner discusses an example that is now part of many introductory textbook explanations of spin: a beam of particles (with, say, initial spin state ψ_{+x}) is sent through a Stern–Gerlach device to measure the z -spin. As discussed in the previous Project, this results in two sub-beams that are spatially separated (transverse to the direction of propagation of the particles). But now suppose some additional magnets are added, which have the effect of re-combining the two beams. The recombined beam is then sent through another Stern–Gerlach device, this time oriented in the x -direction. (Draw a picture to keep track of all this!) If you think the particle’s passage through the z -oriented S-G device constitutes a *measurement* of the particle’s z -spin, you would say that the particle’s wave function collapses in this intermediate stage. Discuss what you would then expect to see in the subsequent x -spin measurement. In fact, *all* particles in this kind of situation are observed to emerge from the final x -spin measurement as spin-up along x . Discuss the implications of this and relate it to the other examples from the Chapter.
- 3.14 A beam of spin-1/2 particles is sent through a Stern–Gerlach device aligned along the x -axis. Those particles which emerge spin-up along the x -axis then enter another Stern–Gerlach device aligned along the z -axis. What happens, and how would an advocate of the ignorance interpretation explain the results? Now suppose we allow particles emerging from the second S-G device as spin-up along the z -axis to enter a third S-G device, oriented parallel to the x -axis. What happens? Can an advocate of the ignorance interpretation explain these results? How?
- 3.15 In Chap. 2 I described – as something that should “kind of blow your mind” – a two-particle entangled state in which neither particle has a definite energy, but the two-particle system does have a definite total energy. The discussion in Chap. 3 should help you understand better exactly how one needs to be understanding quantum descriptions in order for this kind of situation to be

interesting. Would this sort of entangled state be at all mind-blowing to someone who adopted the ignorance interpretation of superposition?

- 3.16 One difference between the Schrödinger evolution of the wave function, and the collapse of the wave function, is that the former is deterministic while the second is supposed to be irreducibly random. Sometimes it is claimed that people (like Schrödinger and Einstein) who had problems with quantum mechanics really just had problems with accepting irreducible randomness, i.e., the failure of determinism. (Think here, for example, of Einstein's famous and oft-quoted remark "God does not play dice".) To what extent do you think it is accurate to say that the (supposed) "measurement problem" is really just based on a philosophical insistence on pure determinism?

References

1. J.S. Bell, Against 'Measurement', reprinted in *Speakable and Unsayable in Quantum Mechanics*, 2nd edn. (Cambridge University Press, Cambridge, 2004)
2. P.C.W. Davies, R. Brown (eds.), *The Ghost in the Atom*, interview with J.S. Bell (Cambridge University Press, Cambridge, 1986)
3. E. Schrödinger, The present situation in quantum mechanics. *Naturwissenschaften* **23** (1935), translated by J. Trimmer, in *Proceedings of the American Philosophical Society*, vol. 124, 10 October 1980 (1980), pp. 323–338
4. A. Fine, *The Shaky Game* (University of Chicago Press, Chicago, 1996)
5. I. Born, trans., *The Born–Einstein Letters* (Walker and Company, New York, 1971)
6. J.S. Bell, The trieste lecture of John Stewart Bell, transcribed by A. Bassi, G.C. Ghirardi. *J. Phys. A: Math. Theor.* **40**, 2919–2933 (2007)
7. M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974)
8. A. Einstein, Reply to criticisms, in *Albert Einstein: Philosopher-Scientist*, ed. by P.A. Schilpp (1949)
9. N.D. Mermin, Hidden variables and the two theorems of John Bell. *Rev. Mod. Phys.* **65**, 803–815
10. J.S. Bell, On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447–452 (1966); reprinted in *Speakable and Unsayable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
11. J.S. Bell, Quantum mechanics for cosmologists, in *Quantum Gravity 2*, ed. by C. Isham, R. Penrose, D. Sciama (Clarendon Press, Oxford, 1981), pp. 611–637; reprinted in *Speakable and Unsayable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
12. J.S. Bell, Beables for QFT, *Speakable and Unsayable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
13. J.S. Bell, Against 'measurement', in *62 Years of Uncertainty: Erice, 5–14 August 1989* (Plenum Publishers, New York); reprinted in *Speakable and Unsayable in Quantum Mechanics* (Cambridge University Press, Cambridge, 2004)
14. E. Wigner, Remarks on the mind-body question, in *Symmetries and Reflections: Scientific Essays*, ed. by W. Moore, M. Scriven (Indiana University Press, Indiana, 1967)
15. E. Wigner, The problem of measurement. *Am. J. Phys.* **31**, 6 (1963)