# Chapter 9

# The Born Rule

## 9.1 Classical Random Walk

The previous chapter showed how to construct sample spaces of histories for both classical and quantum systems. Now we shall see how to use dynamical laws in order to assign probabilities to these histories. It is useful to begin with a classical random walk of a particle in one dimension, as it provides a helpful guide for quantum systems, which are discussed beginning in Sec. 9.3, as well as in the next chapter. The sample space of random walks, Sec. 8.2, consists of all sequences of the form

$$\boldsymbol{s} = (s_0, s_1, s_2, \dots s_f), \tag{9.1}$$

where $s_j$, an integer in the range

$$-M_a \leq s_j \leq M_b, \tag{9.2}$$

is the position of the particle or random walker at time $t = j$.

We shall assume that the *dynamical law* for the particle's motion is that when the time changes from $t$ to $t+1$, the particle can take one step to the left, from $s$ to $s-1$, with probability $p$, remain where it is with probability $q$, or take one step to the right, from $s$ to $s+1$, with probability $r$, where

$$p + q + r = 1. \tag{9.3}$$

The probability for hops in which $s$ changes by 2 or more is zero. The endpoints of the interval (9.1) are thought of as connected by a periodic boundary condition, so that $M_b$ is one step to the left of $-M_a$, which in turn is one step to the right of $M_b$. The dynamical law can be used to generate a probability distribution on the sample space of histories in the following way. We begin by assigning to each history a *weight*

$$W(\boldsymbol{s}) = \prod_{j=1}^{n} w(s_j - s_{j-1}), \tag{9.4}$$

where the hopping probabilities

$$w(-1) = p, \quad w(0) = q, \quad w(+1) = r \tag{9.5}$$

were introduced earlier, and $w(\Delta s) = 0$ for $|\Delta s| \geq 2$.

The weights by themselves do not determine a probability. Instead, they must be combined with other information, such as the starting point of the particle at $t = 0$, or a probability distribution for this starting point, or perhaps information about where the particle is located at some later time(s). This information is not contained in the dynamical laws themselves, so we shall refer to it as *contingent information* or *initial data*. The "initial" in initial data refers to the beginning of an argument or calculation, and not necessarily to the earliest time in the random walk. The single contingent piece of information "$s = 3$ at $t = 2$" can be the initial datum used to generate a probability distribution on the space of all histories of the form (9.2). Contingent information is also needed for deterministic processes. The orbit of the planet Mars can be calculated using the laws of classical mechanics, but to get the calculation started one needs to provide its position and velocity at some particular time. These data are contingent in the sense that they are not determined by the laws of mechanics, but must be obtained from observations. Once they are given, the position of Mars can be calculated at earlier as well as later times.

Contingent information in the case of a random walk is often expressed as a probability distribution $p_0(s_0)$ on the coarse sample space of positions at $t = 0$. (If the particle starts at a definite location, the distribution $p_0$ assigns the value 1 to this position and 0 to all others.) The probability distribution on the refined sample space of histories is then determined by a *refinement rule* that says, in essence, that for each $s_0$, the probability $p_0(s_0)$ is to be divided up among all the different histories which start at this point at $t = 0$, with history $\boldsymbol{s}$ assigned a fraction of $p_0(s_0)$ proportional to its weight $W(\boldsymbol{s})$. One could also use a refinement rule if the contingent data were in the form of a position or a probability distribution at some later time, say $t = 2$ or $t = f$, or if positions were given at two or more different times. Things are more complicated when probability distributions are specified at two or more times.

In order to turn the refinement rule for a probability distribution at $t = 0$ into a formula, let $J(s_0)$ be the set of all histories which begin at $s_0$, and

$$N(s_0) := \sum_{\boldsymbol{s} \in J(s_0)} W(\boldsymbol{s}) \tag{9.6}$$

the sum of their weights. The probability of a particular history is then given by the formula

$$\Pr(\boldsymbol{s}) = p_0(s_0) W(\boldsymbol{s}) / N(s_0). \tag{9.7}$$

These probabilities sum to 1 because the initial probabilities $p_0(s_0)$ sum to one, and because the weights have been suitably normalized by dividing by the normalization factor $N(s_0)$. In fact, for the weights defined by (9.4) and (9.5) using hopping probabilities which satisfy (9.3), it is not hard to show that the sum in (9.6) is equal to 1, so that in this particular case the normalization can be omitted from (9.7). However, it is sometimes convenient to work with weights which are not normalized, and then the factor of $1/N(s_0)$ is needed.

Suppose the particle starts at $s_0 = 2$, so that $p_0(2) = 1$. Then the histories $(2, 1)$, $(2, 2)$, $(2, 3)$, and $(2, 4)$, which are compound histories for $f \geq 2$, have probabilities $p$, $q$, $r$, and 0, respectively. Likewise, the histories $(2, 2, 2)$, $(2, 2, 3)$, $(2, 3, 4)$, $(2, 4, 3)$ have probabilities of $q^2$, $qr$, $r^2$ and 0. Any history in which the particle hops by a distance of two or more in a single time step has zero probability, i.e., it is impossible. One could reduce the size of the sample space by eliminating impossible histories, but in practice it is more convenient to use the larger sample space.

As another example, suppose that $p_0(0) = p_0(1) = p_0(2) = 1/3$. What is the probability that $s_1 = 2$ at time $t = 1$? Think of $s_1 = 2$ as a compound history given by the collection of all histories which pass through $s = 2$ when $t = 1$, so that its probability is the sum of probabilities of the histories in this collection. Clearly histories with zero probability can be ignored, and this leaves only three two-time histories: $(1, 2)$, $(2, 2)$ and $(3, 2)$. In the case $f = 1$, formula (9.7) assigns them probabilities $r/3$, $q/3$, and 0, so the answer to the question is $(q + r)/3$. This answer is also correct for $f \geq 2$, but then it is not quite so obvious. The reader may find it a useful exercise to work out the case $f = 2$, in which there are 9 histories of non-zero weight passing through $s = 2$ at $t = 1$.

Once probabilities have been assigned on the sample space, one can answer questions such as: "What is the probability that the particle was at $s = 2$ at time $t = 3$, given that it arrived at $s = 4$ at time $t = 5$?" by means of conditional probabilities:

$$\Pr(s_3 = 2 \,|\, s_5 = 4) = \Pr\bigl[(s_3 = 2) \wedge (s_5 = 4)\bigr] / \Pr(s_5 = 4). \tag{9.8}$$

Here the event $(s_3 = 2) \wedge (s_5 = 4)$ is the compound history consisting of all elementary histories which pass through $s = 2$ at time $t = 3$ and $s = 4$ at time $t = 5$. Such conditional probabilities depend, in general, both on the initial data and the weights. However, *if a value of* $s_0$ *is one of the conditions*, then the conditional probability does not depend upon $p_0(s_0)$ (assuming $p_0(s_0) > 0$, so that the conditional probability is defined). In particular,

$$\Pr(\boldsymbol{s} \,|\, s_0 = m) = \delta_{ms_0} W(\boldsymbol{s}) / N(s_0). \tag{9.9}$$

To obtain similar formulas in other cases, it is convenient to extend the definition of weights to include compound histories in the event algebra using the formula

$$W(E) = \sum_{\boldsymbol{s} \in E} W(\boldsymbol{s}). \tag{9.10}$$

Defining $W(E)$ for the compound event $E$ in this way makes it an "additive set function" or "measure" in the sense that if $E$ and $F$ are disjoint (they have no elementary histories in common) members of the event algebra of histories, then

$$W(E \cup F) = W(E) + W(F). \tag{9.11}$$

Using this extended definition of $W$, one can, for example, write

$$\Pr\bigl[s_3 = 2 \,|\, (s_0 = 1) \wedge (s_5 = 4)\bigr] = \frac{W\bigl[(s_0 = 1) \wedge (s_3 = 2) \wedge (s_5 = 4)\bigr]}{W\bigl[(s_0 = 1) \wedge (s_5 = 4)\bigr]}. \tag{9.12}$$

That is, take the total weight of all the histories which satisfy the conditions $s_0 = 1$ and $s_5 = 4$, and find what fraction of it corresponds to histories which also have $s_3 = 2$.

## 9.2 Single-Time Probabilities

The probability that at time $t$ the random walker of Sec. 9.1 will be located at $s$ is given by the *single-time probability distribution*[1]

$$\rho_t(s) = \sum_{\boldsymbol{s}\in J_t(s)} \Pr(\boldsymbol{s}), \tag{9.13}$$

where the sum is over the collection $J_t(s)$ of all histories which pass through $s$ at time $t$. Because the particle must be somewhere at time $t$, it follows that

$$\sum_s \rho_t(s) = 1. \tag{9.14}$$

It is easy to show that the dynamical law used in Sec. 9.1 implies that $\rho_t(s)$ satisfies the difference equation

$$\rho_{t+1}(s) = p\,\rho_t(s+1) + q\,\rho_t(s) + r\,\rho_t(s-1). \tag{9.15}$$

In particular, if the contingent information is given by a probability distribution at $t=0$, so that $\rho_0(s) = p_0(s)$, (9.15) can be used to calculate $\rho_t(s)$ at any later time $t$. For example, if the random walker starts off at $s=0$ when $t=0$, and $p=q=r=1/3$, then $\rho_0(0)=1$, while

$$\begin{gathered} \rho_1(-1) = \rho_1(0) = \rho_1(1) = 1/3, \\ \rho_2(-2) = \rho_2(2) = 1/9,\ \ \rho_2(-1) = \rho_2(1) = 2/9,\ \ \rho_2(0) = 1/3 \end{gathered} \tag{9.16}$$

are the non-zero values of $\rho_t(s)$ for $t = 1$ and 2.

The single-time distribution $\rho_t(s)$ is a marginal probability distribution and contains less information than the full probability distribution $\Pr(\boldsymbol{s})$ on the set of all random walks. This is so even if one knows $\rho_t(s)$ for every value of $t$. In particular, $\rho_t(s)$ does not tell one how the particle's position is correlated at successive times. For example, given $\Pr(\boldsymbol{s})$, one can show that the conditional probability $\Pr(s_{t+1} \mid s_t)$ is zero whenever $|s_{t+1} - s_t|$ is larger than 1, whereas the values of $\rho_1(s)$ and $\rho_2(s)$ in (9.16) are consistent with the possibility of the particle hopping from $s=1$ at $t=1$ to $s=-2$ at $t=2$. It is not a defect of $\rho_t(s)$ that it contains less information than the total probability distribution $\Pr(\boldsymbol{s})$. Less detailed descriptions are often very useful in helping one see the forest and not just the trees. But one needs to be aware of the fact that the single-time distribution as a function of time is far from being the full story.

For a Brownian particle the analog of $\rho_t(s)$ for the random walker is the *single-time probability distribution density* $\rho_t(\boldsymbol{r})$, defined in such a way that the integral

$$\int_R \rho_t(\boldsymbol{r})\,d\boldsymbol{r} \tag{9.17}$$

over a region $R$ in three-dimensional space is the probability that the particle will lie in this region at time $t$. In the simplest theory of Brownian motion, $\rho_t(\boldsymbol{r})$ satisfies a partial differential equation

$$\partial\rho/\partial t = D\nabla^2\rho, \tag{9.18}$$

[1] The term *one-dimensional distribution* is often used, but in the present context "one-dimensional" would be misleading.

where $D$ is the diffusion constant and $\nabla^2$ is the Laplacian. If the particle starts off at $\boldsymbol{r} = 0$ when $t = 0$, the solution is

$$\rho_t(\boldsymbol{r}) = (4\pi Dt)^{-3/2} \exp[-r^2/4Dt], \tag{9.19}$$

where $r$ is the magnitude of $\boldsymbol{r}$.

Just as for $\rho_t(s)$ in the case of a random walk, $\rho_t(\boldsymbol{r})$ lacks information about the correlation between positions of the Brownian particle at successive times. Suppose, for example, that a particle starting at $\boldsymbol{r} = 0$ at time $t = 0$ is at $\boldsymbol{r}_1$ at a time $t_1 > 0$. Then at a time $t_2 = t_1 + \epsilon$, where $\epsilon$ is small compared to $t_1$, there is a high probability that the particle will still be quite close to $\boldsymbol{r}_1$. This fact is not, however, reflected in $\rho_{t_2}(\boldsymbol{r})$, as (9.19) gives the probability density for the particle to be at $\boldsymbol{r}$ using no information beyond the fact that it was at the origin at $t = 0$.

## 9.3 The Born Rule

As in Ch. 7, we shall consider an *isolated system* which does not interact with its environment, so that one can define unitary time-development operators of the form $T(t', t)$. To describe its stochastic time development one must assign probabilities to histories forming a suitable sample space of the type discussed in Sec. 8.5. Just as in the case of the random walk considered above in Sec. 9.1, these probabilities are determined both by the contingent information contained in initial data, and by a set of weights. The weights are given by the laws of quantum mechanics, and for an isolated system they can be computed using the time-development operators.

In this section we consider a very simple situation in which the initial datum is a normalized state $|\psi_0\rangle$ at time $t_0$, and the histories involve only two times, $t_0$ and a later time $t_1$ at which there is a decomposition of the identity corresponding to an orthonormal basis $\{|\phi_1^k\rangle, k = 1, 2 \ldots\}$. Histories of the form

$$Y^k = [\psi_0] \odot [\phi_1^k], \tag{9.20}$$

together with a history

$$Z = (I - [\psi_0]) \odot I \tag{9.21}$$

constitute a decomposition of the history identity $\breve{I}$, and thus a sample space of histories based upon the initial state $[\psi_0]$, to use the terminology of Sec. 8.5. We assign initial probabilities $p_0(I - [\psi_0]) = 0$ and $p_0([\psi_0]) = 1$, in the notation of Sec. 9.1.

The *Born rule* assigns a weight

$$W(Y^k) = |\langle\phi_1^k|T(t_1, t_0)|\psi_0\rangle|^2 \tag{9.22}$$

to the history $Y^k$. These weights sum to 1,

$$\begin{aligned}\sum_{k>0} W(Y^k) &= \sum_k \langle\psi_0|T(t_0, t_1)|\phi_1^k\rangle\langle\phi_1^k|T(t_1, t_0)|\psi_0\rangle \\ &= \langle\psi_0|T(t_0, t_1)T(t_1, t_0)|\psi_0\rangle = \langle\psi_0|I|\psi_0\rangle = \langle\psi_0|\psi_0\rangle = 1,\end{aligned} \tag{9.23}$$

because $|\psi_0\rangle$ is normalized and the $\{|\phi_1^k\rangle\}$ are an orthonormal basis. It is important to notice that the Born rule does not follow from any other principle of quantum mechanics. It is a fundamental postulate or axiom, the same as Schrödinger's equation. The weights can be used to assign probabilities to histories using the obvious analog of (9.7), with the normalization $N$ equal to 1 because of (9.23):

$$\Pr(\phi_1^k) = \Pr(Y^k) = W(Y^k) = |\langle\phi_1^k|T(t_1,t_0)|\psi_0\rangle|^2 \tag{9.24}$$

where $\Pr(\phi_1^k)$, which could also be written as $\Pr(\phi_1^k \,|\, \psi_0)$, is the probability of the event $\phi_1^k$ at time $t_1$. The square brackets around $\phi_1^k$ have been omitted where these dyads appear as arguments of probabilities, since this makes the notation less awkward, and there is no risk of confusion. Given an observable of the form

$$V = V^\dagger = \sum_k v_k[\phi_1^k] = \sum_k v_k|\phi_1^k\rangle\langle\phi_1^k|, \tag{9.25}$$

one can compute its average, see (5.42), at the time $t_1$ using the probability distribution $\Pr(\phi_1^k)$:

$$\langle V\rangle = \sum_k v_k \Pr(\phi_1^k) = \langle\psi_0|T(t_0,t_1)VT(t_1,t_0)|\psi_0\rangle. \tag{9.26}$$

The validity of the right side becomes obvious when $V$ is replaced by the right side of (9.25).

Let us analyze two simple but instructive examples. Consider a spin-half particle in zero magnetic field, so that the spin dynamics is trivial: $H = 0$ and $T(t',t) = I$. Let the initial state be

$$|\psi_0\rangle = |z^+\rangle. \tag{9.27}$$

For the first example use

$$|\phi_1^1\rangle = |z^+\rangle, \quad |\phi_1^2\rangle = |z^-\rangle. \tag{9.28}$$

as the orthonormal basis at $t_1$. Then (9.24) results in:

$$\Pr(\phi_1^1) = \Pr(z^+) = 1, \quad \Pr(\phi_1^2) = \Pr(z^-) = 0. \tag{9.29}$$

We have here an example of a unitary family of histories as defined in Sec. 8.7. Since the ket $T(t_1,t_0)|\psi_0\rangle$ is equal to one of the basis vectors at $t_1$, it is necessarily orthogonal to the other basis vector. Thus the unitary history $[\psi_0]\odot[\phi_1^1]$ has probability one, whereas the other history $[\psi_0]\odot[\phi_1^2]$ which begins with $[\psi_0]$ has probability zero. It follows from (9.29) that

$$\langle S_z\rangle = 1/2, \tag{9.30}$$

where $S_z = \frac{1}{2}\big([z^+]-[z^-]\big)$ is the operator for the $z$ component of spin angular momentum in units of $\hbar$—see (5.30).

The second example uses the same initial state (9.27), but at $t_1$ an orthonormal basis

$$|\bar\phi_1^1\rangle = |x^+\rangle, \quad |\bar\phi_1^2\rangle = |x^-\rangle, \tag{9.31}$$

where bars have been added to distinguish these kets from those in (9.28). A straightforward calculation yields

$$\Pr(x^+) = 1/2 = \Pr(x^-). \tag{9.32}$$

Stated in words, if $S_z = +1/2$ at $t_0$, the probability is 1/2 that $S_x = +1/2$ at $t_1$, and 1/2 that $S_x = -1/2$. Consequently, the average of the $x$ component of angular momentum is

$$\langle S_x\rangle = 0. \tag{9.33}$$

The second example may seem counterintuitive for the following reason. The unitary quantum dynamics is trivial: nothing at all is happening to this spin half particle. It is not in a magnetic field, and therefore there is no reason why the spin should precess. Nonetheless, it might seem as if the spin orientation has managed to "jump" from being along the positive $z$ axis at time $t_0$ to an orientation either along or opposite to the positive $x$ axis at $t_1$. However, the idea that something is "jumping" comes from a misleading mental picture of a spin-half particle. To better understand the situation, imagine a classical object spinning in free space and not subject to any torques, so that its angular momentum is conserved. Suppose we know the $z$ component of its angular momentum at $t_0$, and for some reason want to discuss the $x$ component at a later time $t_1$. The fact that two different components of angular momentum are considered at the two different times does not mean there has been a change in the angular momentum of the object between $t_0$ and $t_1$. This analogy, like all classical analogies, is far from perfect, but in the present context it is less misleading than thinking of $S_z = +1/2$ for a spin-half particle as corresponding to a classical object with its total angular momentum in the $+z$ direction. Applying this analogy to the quantum case, we see that the probabilities in (9.32) are not unreasonable, given that we have adopted a sample space in which values of $S_x$ occur at $t_1$, rather than values of $S_z$, as in the first example.

The odd thing about quantum theory is the fact that one cannot combine the conclusions in (9.29) and (9.32) to form a single description of the time development of the particle, whereas it would be perfectly reasonable to do so for a classical spinning object. It is incorrect to conclude from (9.29) and (9.32) that at $t_1$ either it is the case that $S_z = +1/2$ AND $S_x = +1/2$, or else it is the case that $S_z = +1/2$ AND $S_x = -1/2$. Both of the statements connected by AND are quantum nonsense, as they do not correspond to anything in the quantum Hilbert space; see Sec. 4.6. For the same reason the two averages (9.30) and (9.33) cannot be thought of as applying simultaneously to the same system, since the observables $S_z$ and $S_x$ do not commute with each other, and hence correspond to incompatible sample spaces. It is always possible to apply the Born rule in a large number of different ways by using different orthonormal bases at $t_1$, but these different results cannot be combined in a single sensible quantum description of the system. Attempting to do so violates the single framework rule (to be discussed in Sec. 16.1) and leads to confusion.

The Born rule is often discussed in the context of *measurements*, as a formula to compute the probabilities of various outcomes of a measurement carried out by an apparatus $\mathcal{A}$ on a system $\mathcal{S}$. Hence it is worth emphasizing that the probabilities in (9.24) refer to an *isolated* system $\mathcal{S}$ which is *not* interacting with a separate measurement device. Indeed, our discussion of the Born rule has made no reference whatsoever to measurements of any sort. Measurements will be taken up in Chs. 17 and 18 below, where the usual formulas for the probabilities of different measurement outcomes will be derived by applying general quantum principles to the combined apparatus and measured system thought of as constituting a single, isolated system.

## 9.4 Wave Function as a Pre-Probability

The basic formula (9.24) which expresses the Born rule can be rewritten in various ways. One rather common form is the following. Let

$$|\psi_1\rangle = T(t_1, t_0)|\psi_0\rangle \tag{9.34}$$

be the wave function obtained by integrating Schrödinger's equation from $t_0$ to $t_1$. Then (9.24) can be written in the compact form

$$\Pr(\phi_1^k) = |\langle\phi_1^k|\psi_1\rangle|^2. \tag{9.35}$$

Note that $|\psi_1\rangle$ or $[\psi_1]$, regarded as a quantum property at time $t_1$, is *incompatible* with the collection of properties $\{[\phi_1^k]\}$ if at least two of the probabilities in (9.35) are non-zero, that is, if one is not dealing with a unitary family. Thus in the second spin-half example considered above, $|\psi_1\rangle = |z^+\rangle$ is incompatible with both $|x^+\rangle$ and $|x^-\rangle$. Therefore, in the context of the family based on (9.20) and (9.21) it does not make sense to suppose that at $t_1$ the system possesses the *physical property* $|\psi_1\rangle$. Instead, $|\psi_1\rangle$ must be thought of as a *mathematical* construct suitable for calculating certain probabilities. We shall refer to $|\psi_1\rangle$ understood in this way as a *pre-probability*, since it is (obviously) not a probability, nor a property of the physical system, but instead something which is used to calculate probabilities. In addition to wave functions obtained by unitary time development, density matrices are often employed in quantum theory as pre-probabilities; see Ch. 15. The pre-probability $|\psi_1\rangle$ is very convenient for calculations because it does not depend upon which orthonormal basis $\{|\phi_1^k\rangle\}$ is employed at $t_1$. The theoretical physicist may want to compute probabilities for various different bases, that is, for various different families of histories, and $|\psi_1\rangle$ is a convenient tool for doing this. There is no harm in carrying out such calculations as long as one does not try to combine the results for incompatible bases into a single description of the quantum system.

Another way to see that $|\psi_1\rangle$ on the right side of (9.35) is a calculational device and not a physical property is to note that these probabilities can be computed equally well by an alternative procedure. For each $k$, let

$$|\phi_0^k\rangle = T(t_0, t_1)|\phi_1^k\rangle \tag{9.36}$$

be the ket obtained by integrating Schrödinger's equation backwards in time from the final state $|\phi_1^k\rangle$. It is then obvious, see (9.24), that

$$\Pr(\phi_1^k) = |\langle\phi_0^k|\psi_0\rangle|^2. \tag{9.37}$$

There is no reason in principle to prefer (9.35) to (9.37) as a method of calculating these probabilities, and in fact there are a lot of other methods of obtaining the same answer. For example, one can integrate $|\psi_0\rangle$ forwards in time and each $|\phi_1^k\rangle$ backwards in time until they meet at some intermediate time, and then evaluate the absolute square of the inner product. To be sure, the most efficient procedure for calculating $\Pr(\phi_1^k \mid \psi_0)$ for all values of $k$ is likely to be (9.35): one only has to do one time integration, and then evaluate a number of inner products. But the fact that other procedures are equally valid, and can give very different "pictures" of what is going on at intermediate times if one takes them literally, is a warning that one has no more justification for identifying $|\psi_1\rangle$, as defined in (9.34), as "the real state of the system" at time $t_1$ than one has for identifying one or more of the $|\phi_0^k\rangle$, as defined in (9.36), with the "the real state of the system" at time $t_0$. Instead, both $|\psi_1\rangle$ and the $|\phi_0^k\rangle$ are functioning as pre-probabilities.

It is evident from (9.26) and (9.34) that the average of an observable $V$ at time $t_1$ can be written in the compact and convenient form

$$\langle V\rangle = \langle\psi_1|V|\psi_1\rangle, \tag{9.38}$$

where $|\psi_1\rangle$ is again functioning as a pre-probability. A similar expression holds for any other observable $W$, and there is no harm in simultaneously calculating averages for $\langle V\rangle$, $\langle W\rangle$ provided one keeps in mind the fact that when $V$ and $W$ do not commute with each other, one cannot regard $\langle V\rangle$ and $\langle W\rangle$ as belonging to a single (stochastic) description of a quantum system, for the two averages are necessarily based on incompatible sample spaces that cannot be combined. See the comments towards the end of Sec. 9.3 in connection with the example of a spin-half particle. Any time the symbol $\langle V\rangle$ is used with reference to the physical properties of a quantum system there is an implicit reference to a sample space, and ignoring this fact can lead to serious misunderstanding.

It is important to remember when applying the Born formula that a family of histories involving two times *tells us nothing at all about what happens at intermediate times.* Such times can, of course, be introduced formally by extending the history, in the manner indicated in Sec. 8.4,

$$Y^k = [\psi_0] \odot [\phi_1^k] = [\psi_0] \odot I \odot I \odot \cdots \odot I \odot [\phi_1^k], \tag{9.39}$$

for as many intermediate times as one wants. But each $I$ at the intermediate time tells us nothing at all about what actually happens at this time. Imagine being outdoors on a dark night during a thunder storm. Each time the lightning flashes you can see the world around you. Between flashes, you cannot tell what is going on. To be sure, if we are curious about what is going on at intermediate times in a quantum history of the form (9.39), we can *refine* the history in the manner indicated in Sec. 8.6, by writing the projector as a sum of history projectors which include non-trivial information about the intermediate times, and then compute probabilities for these different possibilities. That, however, *cannot* be done by means of the Born formula (9.22), and requires an extension of this formula which will be introduced in the next chapter.

A similar restriction applies to a wave function understood as a pre-probability. Even if

$$|\psi_t\rangle = T(t, t_0)|\psi_0\rangle \tag{9.40}$$

is known for all values of the time $t$, it can only be used to compute probabilities of histories involving just two times, $t_0$ and $t$. These probabilities are the quantum analogs of the single-time probabilities $\rho_t(\boldsymbol{r})$ for a classical Brownian particle which started off at a definite location at the initial time $t_0$. As discussed in Sec. 9.2, $\rho_t(\boldsymbol{r})$ does not contain probabilistic information about correlations between particle positions at intermediate times, and in the same way correlations between quantum properties at different times cannot be computed from $|\psi_t\rangle$. Instead, one must use the procedures discussed in the next chapter.

## 9.5 Application: Alpha Decay

A toy model of alpha decay was introduced in Sec. 7.4, see Fig. 7.2, as an example of unitary time evolution. In this section we shall apply the Born formula in order to calculate some of the associated probabilities, but before doing so it will be convenient to add a toy detector of the sort shown in Fig. 7.1, in order to detect the alpha particle after it leaves the nucleus, see Fig. 9.1. Let $\mathcal{M}$ be the Hilbert space of the particle, and $\mathcal{N}$ that of the detector. For the combined system $\mathcal{M} \otimes \mathcal{N}$ we define the time development operator to be

$$T = S_a R, \tag{9.41}$$

where $S_a$ is defined in (7.56) and $R$ in (7.53). Note the similarity with (7.52), which means that the discussion of the operation of the detector found in Sec. 7.4, see Fig. 7.1, applies to the arrangement in Fig. 9.1, with a few obvious modifications.

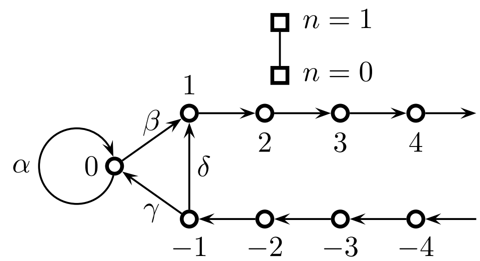


Figure 9.1: Toy model of alpha decay (Fig. 7.2) plus a detector.

Assume that at $t = 0$ the alpha particle is at $m = 0$ inside the nucleus, which has not yet decayed, and the detector is in its ready state $n = 0$, so the wave function for the total system is

$$|\Psi_0\rangle = |m = 0\rangle \otimes |n = 0\rangle = |0, 0\rangle. \tag{9.42}$$

Unitary time evolution using (9.41) results in

$$|\Psi_t\rangle = T^t|\Psi_0\rangle = |\chi_t\rangle \otimes |0\rangle + |\omega_t\rangle \otimes |1\rangle, \tag{9.43}$$

where

$$\begin{aligned} |\chi_1\rangle &= \alpha|0\rangle + \beta|1\rangle, & |\omega_1\rangle &= 0,\\ |\chi_2\rangle &= \alpha^2|0\rangle + \alpha\beta|1\rangle + \beta|2\rangle, & |\omega_2\rangle &= 0,\\ |\chi_3\rangle &= \alpha^3|0\rangle + \alpha^2\beta|1\rangle + \alpha\beta|2\rangle, & |\omega_3\rangle &= \beta|3\rangle, \end{aligned} \tag{9.44}$$

and for $t \geq 4$

$$\begin{aligned} |\chi_t\rangle &= \alpha^t|0\rangle + \alpha^{t-1}\beta|1\rangle + \alpha^{t-2}\beta|2\rangle,\\ |\omega_t\rangle &= \alpha^{t-3}\beta|3\rangle + \alpha^{t-4}\beta|4\rangle + \cdots + \beta|t\rangle. \end{aligned} \tag{9.45}$$

Let us apply the Born rule with $t_0 = 0$, $t_1 = t$ for some integer $t > 0$, using $|\Psi_0\rangle$ as the initial state at time $t_0$, and at time $t_1$ the orthonormal basis $\{|m, n\rangle\}$, in which the alpha particle has a definite position $m$ and the detector either has or has not detected the particle. The joint probability distribution of $m$ and $n$ at time $t$,

$$p_t(m, n) := \Pr([m, n]_t), \tag{9.46}$$

is easily computed by regarding $|\Psi_t\rangle$ in (9.43) as a pre-probability: $p_t(m, n)$ is the absolute square of the coefficient of $|m\rangle$ in $|\chi_t\rangle$ if $n = 0$, or in $|\omega_t\rangle$ if $n = 1$. These probabilities vanish except for the cases

$$p_t(0, 0) = e^{-t/\tau}, \tag{9.47}$$

$$p_t(m, 0) = \kappa e^{-(t-m)/\tau} \text{ for } m = 1, 2 \text{ and } m \leq t, \tag{9.48}$$

$$p_t(m, 1) = \kappa e^{-(t-m)/\tau} \text{ for } 3 \leq m \leq t. \tag{9.49}$$

The positive constants $\kappa$ and $\tau$ are defined by

$$e^{-1/\tau} = |\alpha|^2, \quad \kappa = |\beta|^2 = 1 - |\alpha|^2. \tag{9.50}$$

The probabilities in (9.47) to (9.50) make good physical sense. The probability (9.47) that the alpha particle is still in the nucleus decreases exponentially with time, in agreement with the well-known exponential decay law for radioactive nuclei. That $p_t(m,n)$ vanishes for $m$ larger than $t$ reflects the fact that the alpha particle was (by assumption) inside the nucleus at $t = 0$ and, since it hops at most one step during any time interval, cannot arrive at $m$ earlier than $t = m$. Finally, if the alpha particle is at $m = 0$, 1 or 2, the detector is still in its ready state $n = 0$, whereas for $m = 3$ or larger the detector will be in the state $n = 1$, indicating that it has detected the particle. This is just what one would expect for a detector designed to detect the particle as it hops from $m = 2$ to $m = 3$ (see the discussion in Sec. 7.4).

It is worth emphasizing once again that $p_t(m,n)$ is the quantum analog of the single-time probability $\rho_t(s)$ for the random walk discussed in Sec. 9.2. The reason is that the histories to which the Born rule applies involve only two times, $t_0$ and $t_1$ in the notation of Sec. 9.3, and thus no information is available as to what happens between these times. Consequently, just as $\rho_t(s)$ does not tell us all there is to be said about the stochastic behavior of a random walker, there is also more to the story of (toy) alpha decay and its detection than is contained in $p_t(m,n)$. However, providing a more detailed description of what is going on requires the additional mathematical tools introduced in the next chapter, and we shall return to the problem of alpha decay using more sophisticated methods (and a better detector) in Sec. 12.4.

It is not necessary to employ the basis $\{|m,n\rangle\}$ in order to apply the Born rule; one could use any other orthonormal basis of $\mathcal{M} \otimes \mathcal{N}$, and there are many possibilities. However, the physical properties which can be described by the resulting probabilities depend upon which basis is used, and not every choice of basis at time $t$ (an example will be considered in the next section) allows one to say whether $n = 0$ or 1, that is, whether the detector has detected the particle. It is customary to use the term *pointer basis* for an orthonormal basis, or more generally a decomposition of the identity such as employed in the generalized Born rule defined in Sec. 10.3 below, that allows one to discuss the outcomes of a measurement in a sensible way. (The term arises from a mental picture of a measuring device equipped with a visible pointer whose position indicates the outcome after the measurement is over.) Thus $\{|m,n\rangle\}$ is a pointer basis, but so is any basis of the form $\{|\xi^j,n\rangle\}$, where $\{|\xi^j\rangle\}$, $j = 1,2,\cdots$, is some orthonormal basis of $\mathcal{M}$. While quantum calculations which are to be compared with experiments usually employ a pointer basis for calculating probabilities, for obvious reasons, there is no fundamental principle of quantum theory which restricts the Born rule to bases of this type.

## 9.6 Schrödinger's Cat

What is the physical significance of the state $|\Psi_t\rangle$ which evolves unitarily from $|\Psi_0\rangle$ in the toy model discussed in the preceding section? This is a difficult question to answer, because for $t \geq 3$ $|\Psi_t\rangle$ is of the form $|A\rangle + |B\rangle$, (9.43), where $|A\rangle = |\chi_t\rangle \otimes |0\rangle$ has the significance that the alpha particle is inside or very close to the nucleus and the detector is ready, whereas $|B\rangle = |\omega_t\rangle \otimes |1\rangle$ means that the detector has triggered and the nucleus has decayed. What can be the significance of a linear combination $|A\rangle + |B\rangle$ of states with quite distinct physical meanings? Could it signify that the detector both has and has not detected the particle?

The difficulty of interpreting such wave functions is often referred to as the problem or paradox of *Schrödinger's cat*. In a famous paper Schrödinger pointed out that in the case of alpha decay, unitary time evolution applied to the system consisting of a decaying nucleus plus a detector will quite generally lead to a superposition state $|S\rangle = |A\rangle + |B\rangle$, where the (macroscopic) detector either has, state $|B\rangle$, or has not, state $|A\rangle$, detected the alpha particle. To dramatize the conceptual difficulty Schrödinger imagined the detector hooked up to a device which would kill a live cat upon detection of an alpha particle, thus raising the problem of interpreting $|A\rangle + |B\rangle$ when $|A\rangle$ corresponds to an undecayed nucleus, untriggered detector, and live cat, and $|B\rangle$ to a nucleus which has decayed, a triggered detector, and a dead cat. We shall call $|A\rangle + |B\rangle$ a *macroscopic quantum superposition* or MQS state when $|A\rangle$ and $|B\rangle$ correspond to situations which are macroscopically distinct, and use the same terminology for a superposition of three or more macroscopically distinct states. In the literature MQS states are often called *Schrödinger cat states*.

Rather than addressing the general problem of MQS states, let us return to the toy model with its toy example of such a state and, to be specific, consider $|\Psi_5\rangle$ at $t = 5$ under the assumption that $\alpha$ and $\beta$ have been chosen so that $\langle\chi_5|\chi_5\rangle$ and $\langle\omega_5|\omega_5\rangle$ are of the same order of magnitude, which will prevent us from escaping the problem of interpretation by supposing that either $|A\rangle$ or $|B\rangle$ is very small and can be ignored. It is easy to show that $[\Psi_5] = |\Psi_5\rangle\langle\Psi_5|$ does not commute with either of the projectors $[n = 0]$ or $[n = 1]$. Nor does it commute with a projector $[\hat{n}]$, where $|\hat{n}\rangle$ is some linear combination of $|n = 0\rangle$ and $|n = 1\rangle$. This means that it makes no sense to say that the combined system has the property $[\Psi_5]$, whatever its physical significance might be, while at the very same time the detector has or has not detected the particle (or has some other physical property). See the discussion in Sec. 4.6. Saying that the system is in the state $[\Psi_5]$ and then ascribing a property to the detector is no more meaningful than assigning simultaneous values to $S_x$ and $S_z$ for a spin-half particle. The converse is also true: if it makes sense (using an appropriate quantum description) to say that the detector is either ready or triggered at $t = 5$, then one cannot say that the combined system has the property $[\Psi_5]$, because that would be nonsense.

Note that these considerations cause no problem for the analysis in Sec. 9.5, because in applying the Born rule to the basis $\{|m, n\rangle\}$, $|\Psi_t\rangle$ is employed as a pre-probability, Sec. 9.4, a convenient mathematical tool for calculating probabilities which could also be computed by other methods. When it is used in this way there is obviously no need to ascribe some physical significance to $|\Psi_5\rangle$, nor is there any motivation for doing so, since $[\Psi_5]$ must in any case be excluded from any meaningful quantum description based upon $\{|m, n\rangle\}$.

Very similar considerations apply to the situation considered by Schrödinger, although analyzing it carefully requires a model of macroscopic measurement, see Secs. 17.3 and 17.4. The question of whether the cat is dead or alive can be addressed by using the Born rule with an appropriate pointer basis (as defined at the end of Sec. 9.5), and one never has to give a physical interpretation to Schrödinger's MQS state $|S\rangle$, since it only enters the calculations as a pre-probability. In any case, treating $[S]$ as a physical property is meaningless when one uses a pointer basis. To be sure, this does not prevent one from asking whether $|S\rangle$ by itself has some intuitive physical meaning. What the preceding discussion shows is that whatever that meaning may be, it cannot possibly have anything to do with whether the cat is dead or alive, as these properties will be incompatible with $[S]$. Indeed, it is probably the case that the very concept of a "cat" (small furry animal, etc.)

cannot be meaningfully formulated in a way which is compatible with $[S]$.

Quite apart from MQS states, it is in general a mistake to associate a physical meaning with a linear combination $|C\rangle = |A\rangle + |B\rangle$ by referring to the properties of the separate states $|A\rangle$ and $|B\rangle$. For example, the state $|x^+\rangle$ for a spin half particle is a linear combination of $|z^+\rangle$ and $|z^-\rangle$, but its physical signficance of $S_x = 1/2$ is unrelated to $S_z = \pm 1/2$. For another example, see the discussion of (2.27) in Sec. 2.5. In addition, there is the problem that for a given $|C\rangle = |A\rangle + |B\rangle$, the choice of $|A\rangle$ and $|B\rangle$ is far from unique. Think of an ordinary vector in three dimensions: there are lots of ways of writing it as the sum of two other vectors, even if one requires that these be mutually perpendicular, corresponding to the not-unreasonable orthogonality condition $\langle A|B\rangle = 0$. But if $|C\rangle$ is equal to $|A'\rangle + |B'\rangle$ as well as to $|A\rangle + |B\rangle$, why base a physical interpretation upon $|A\rangle$ rather than $|A'\rangle$? See the discussion of (2.28) in Sec. 2.5.

Returning once again to the toy model, it is worth emphasizing that $|\Psi_5\rangle$ is a perfectly good element of the Hilbert space, and enters fundamental quantum theory on precisely the same footing as all other states, despite our difficulty in assigning it a simple intuitive meaning. In particular, we can choose an orthonormal basis at $t_1 = 5$ which contains $|\Psi_5\rangle$ as one of its members, and apply the Born rule. The result is that a weight of 1 is assigned to the unitary history $[\Psi_0] \odot [\Psi_5]$, and 0 to all other histories in the family with initial state $[\Psi_0]$. This means that the state $[\Psi_5]$ will certainly occur (probability 1) at $t = 5$ given the initial state $[\Psi_0]$ at $t = 0$.

But if $[\Psi_5]$ occurs with certainty, how is it possible for there be a different quantum description in which $[n = 0]$ occurs with a finite probability, when we know the two properties $[\Psi_5]$ and $[n = 0]$ cannot consistently enter the same quantum description at the same time? The brief answer is that quantum probabilities only have meaning within specific families, and those from incompatible families—the term will be defined in Sec. 10.4, but we have here a particular instance—cannot be combined. Going beyond the brief answer to a more detailed discussion requires the material in the next chapter and its application to some additional examples. The general principle which emerges is called the *single family* or *single framework* rule, and is discussed in Sec. 16.1.