1

Quantum measurement theory

1.1 Classical measurement theory

1.1.1 Basic concepts

Although this chapter, and indeed this book, is concerned with quantum measurements, we begin with a discussion of some elementary notions of measurement for classical systems. These are systems that operate at a level where quantum effects are not apparent. The purpose of this discussion is to introduce some ideas, which carry over to quantum theory, concerning states, conditional and non-conditional distributions, and stochastic processes. It will also make the distinct features of quantum measurements plainer.

A classical system can be described by a set of *system variables*, which we will call the system *configuration*. For example, for a system of N interacting particles these could be the N position and momentum vectors of the particles. The possible values of these variables form the *configuration space* $\mathbb{S}$ for the system. In the above dynamical example, the configuration space would be $\mathbb{R}^{6N}$, where $\mathbb{R}$ is the real line.[1] Alternatively, to take the simplest possible example, there may be a single system variable X that takes just two values, $X = 0$ or $X = 1$, so that the configuration space would be $\{0, 1\}$. Physically, this binary variable could represent a coin on a table, with $X = 0$ and $X = 1$ corresponding to heads and tails, respectively.

We define the *state* of a classical system to be a probability distribution on configuration space. Say (as in the example of the coin) that there is a single system variable $X \in \mathbb{S}$ that is discrete. Then we write the probability that X has the value x as $\Pr[X = x]$. Here, in general, $\Pr[E]$ is the probability of an event E. When no confusion is likely to arise, we write $\Pr[X = x]$ simply as $\wp(x)$. Here we are following the convention of representing variables by upper-case letters and the corresponding arguments in probability distributions by the corresponding lower-case letters. If X is a continuous variable, then we define a probability density $\wp(x)$ by $\wp(x)\mathrm{d}x = \Pr[X \in (x, x + \mathrm{d}x)]$. In either case, the state of the system is represented by the function $\wp(x)$ for all values of x. When we choose to be more careful, we write this as $\{\wp(x)\colon x \in \mathbb{S}\}$, or as $\{\wp(x)\colon x\}$. We will use these conventions

[1] This space is often called 'phase space', with 'configuration space' referring only to the space of positions. We will not use 'configuration space' with this meaning.

conscientiously for the first two chapters, but in subsequent chapters we will become more relaxed about such issues in order to avoid undue notational complexity.

The system state, as we have defined it, represents an observer's knowledge about the system variables. Unless the probability distribution is non-zero only for a single configuration, we say that it represents a state of *uncertainty* or incomplete knowledge. That is, in this book we adopt the position that probabilities are *subjective*: they represent degrees of certainty rather than objective properties of the world. This point of view may be unfamiliar and lead to uncomfortable ideas. For example, different observers, with different knowledge about a system, would in general assign different states to the same system. This is not a problem for these observers, as long as the different states are *consistent*. This is the case as long as their supports on configuration space are not disjoint (that is, as long as they all assign a non-zero probability to at least one set of values for the system variables). This guarantees that there is at least one state of complete knowledge (that is, one configuration) that all observers agree is a possible state.

We now consider measurement of a classical system. With a perfect measurement of X, the observer would simply find out its value, say x'. The system state would then be a state of complete knowledge about this variable. For discrete variables this is represented by the Kronecker δ-function $\wp(x) = \delta_{x,x'}$, whereas for a continuous variable it is represented by the Dirac δ-function $\wp(x) = \delta(x - x')$. For comparison with the quantum case (in following sections), it is more enlightening to consider imperfect measurements. Suppose that one only has access to the values of the system variables indirectly, through an *apparatus* variable Y. The state of the apparatus is also specified by a probability distribution $\wp(y)$. By some physical process, the apparatus variable becomes statistically dependent on the system variable. That is, the configuration of the apparatus is correlated (perhaps imperfectly) with the configuration of the system. If the apparatus variable is observed, $\{\wp(y)\colon y\}$ is simply the probability distribution of measurement outcomes.

One way of thinking about the system–apparatus correlation is illustrated in Fig. 1.1. The correlation is defined by a functional relationship among the readout variable, Y, the system variable, X, before the measurement, and a random variable, Ξ, which represents extra noise in the measurement outcome. We can specify this by a function

$$Y = G(X, \Xi), \tag{1.1}$$

together with a probability distribution $\wp(\xi)$ for the noise. Here, the noise is assumed to be independent of the system, and is assumed not to affect the system. That is, we restrict our consideration for the moment to *non-disturbing measurements*, for which X after the measurement is the same as X before the measurement.

1.1.2 Example: binary variables

To illustrate the above theory, consider the case of binary variables. As we will see, this is relevant in the quantum setting also. For this case, the state of the system $\wp(x)$ is completely specified by the probability $\wp(x := 0)$, since $\wp(x := 1) = 1 - \wp(x := 0)$. Here

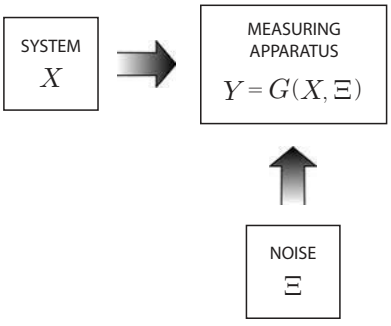

Fig. 1.1 System–apparatus correlation in a typical classical measurement.

we have introduced another abuse of notation, namely that $\wp(x := a)$ means $\wp(x)$ evaluated at $x = a$, where a is any number or variable. In other words, it is another way of writing $\Pr[X = a]$ (for the case of discrete variables). The convenience of this notation will become evident.

We assume that the apparatus and noise are also described by binary variables with values 0 and 1. We take the output variable Y to be the *binary addition* (that is, addition modulo 2) of the system variable X and the noise variable Ξ. In the language of binary logic, this is called the 'exclusive or' (XOR) of these two variables, and is written as

$$Y = X \oplus \Xi. \tag{1.2}$$

We specify the noise by $\wp(\xi := 0) = \mu$.

Equation 1.2 implies that the readout variable Y will reproduce the system variable X if $\Xi = 0$. If $\Xi = 1$, the readout variable is (in the language of logic) the negation of the system variable. That is to say, the readout has undergone a *bit-flip* error, so that $Y = 1$ when $X = 0$ and vice versa. We can thus interpret μ as the probability that the readout variable is 'correct'. If no noise is added by the measurement apparatus, so that $\wp(\xi := 0) = 1$, we call the measurement *ideal*.

It should be intuitively clear that the apparatus state (i.e. the readout distribution) $\wp(y)$ is determined by the function G together with the noise probability $\wp(\xi)$ and the system state before the measurement, $\wp(x)$. This last state is called the a-priori state, or prior state. In the example above we find that

$$\wp(y := 1) = \mu\wp(x := 1) + (1 - \mu)\wp(x := 0), \tag{1.3}$$
$$\wp(y := 0) = \mu\wp(x := 0) + (1 - \mu)\wp(x := 1). \tag{1.4}$$

This may be written more succinctly by inverting Eq. (1.2) to obtain $\Xi = X \oplus Y$ and writing

$$\wp(y) = \sum_{x=0}^{1} \wp(\xi := x \oplus y)\wp(x). \tag{1.5}$$

In the case of a binary variable X with distribution $\wp(x)$, it is easy to verify that the mean is given by $\mathrm{E}[X] = \wp(x := 1)$. Here we are using E to represent 'expectation of'. That is, in the general case,

$$\mathrm{E}[X] = \sum_x x \Pr[X = x] = \sum_x x\wp(x). \tag{1.6}$$

More generally,

$$\mathrm{E}[f(X)] = \sum_x f(x)\Pr[X = x] = \sum_x f(x)\wp(x). \tag{1.7}$$

Using this notation, we define the variance of a variable as

$$\mathrm{Var}[X] \equiv \mathrm{E}[X^2] - (\mathrm{E}[X])^2. \tag{1.8}$$

Exercise 1.1 *Show that*

$$\mathrm{E}[Y] = (1 - \mu) + (2\mu - 1)\mathrm{E}[X], \tag{1.9}$$

$$\mathrm{Var}[Y] = \mu(1 - \mu) + (2\mu - 1)^2\mathrm{Var}[X]. \tag{1.10}$$

Equation (1.9) shows that the average measurement result is the system variable mean, scaled by a factor of $2\mu - 1$, plus a constant off-set of $1 - \mu$. The scaling factor also appears in the variance equation (1.10), together with a constant (the first term) due to the noise added by the measurement process. When the measurement is ideal ($\mu = 1$), the mean and variance of the readout variable directly reflect the statistics of the measured system state.

1.1.3 Bayesian inference

We stated above that we are considering, at present, non-disturbing measurements, in which the system variable X is unaffected by the measurement. However, this does not mean that the system *state* is unaffected by the measurement. Recall that the state represents the observer's incomplete knowledge of the system, and the point of making a measurement is (usually) to obtain more knowledge. Thus we *should* expect the state to change *given* that a certain readout is obtained.

The concept we are introducing here is the *conditional* state of the system, also known as the state conditioned on the readout. This state is sometimes called the a-posteriori state, or posterior state. The key to finding the conditioned state is to use Bayesian inference. Here, inference means that one infers information about the system from the readout, and Bayesian inference means doing this using Bayes' theorem. This theorem is an elementary consequence of basic probability theory, via the double application of the conditional-probability definition

$$\Pr(A|B) = \Pr(A \cap B)/\Pr(B), \tag{1.11}$$

where A and B are events, $A \cap B$ is their intersection and $A|B$ is to be read as 'A given B'. In an obvious generalization of this notation from events to the values of system variables,

Bayes' theorem says that the conditional system state may be written in terms of the a-priori (or prior) system state $\wp(x)$ as

$$\wp'(x|y) = \frac{\wp(y|x)\wp(x)}{\wp(y)}. \tag{1.12}$$

Here the prime emphasizes that this is an a-posteriori state, and (sticking to the discrete case as usual)

$$\wp(y) = \sum_x \wp(y|x)\wp(x), \tag{1.13}$$

as required for the conditional state to be normalized.

The crucial part of Bayesian inference is the known conditional probability $\wp(y|x)$, also known as the 'forward probability'. This is related to the measurement noise and the function G as follows:

$$\wp(y|x) = \sum_\xi \wp(y|x,\xi)\wp(\xi) = \sum_\xi \delta_{y,G(x,\xi)}\wp(\xi). \tag{1.14}$$

Here $\wp(y|x,\xi)$ means the state of y given the values of x and ξ. If the output function $Y = G(X, \Xi)$ is invertible in the sense that there is a function G^{-1} such that $\Xi = G^{-1}(X, Y)$, then we can further simplify this as

$$\wp(y|x) = \sum_\xi \delta_{\xi,G^{-1}(x,y)}\wp(\xi) = \wp(\xi := G^{-1}(x,y)). \tag{1.15}$$

Thus we obtain finally for the conditional system state

$$\wp'(x|y) = \frac{\wp(\xi := G^{-1}(x,y))\wp(x)}{\wp(y)}. \tag{1.16}$$

Exercise 1.2 *If you are unfamiliar with probability theory, derive the first equality in Eq. (1.14).*

As well as defining the conditional post-measurement system state, we can define an unconditional posterior state by averaging over the possible measurement results:

$$\wp'(x) = \sum_y \wp'(x|y)\wp(y) = \sum_y \wp(\xi := G^{-1}(x,y))\wp(x). \tag{1.17}$$

The terms conditional and unconditional are sometimes replaced by the terms selective and non-selective, respectively. In this case of a non-disturbing measurement, it is clear that

$$\wp'(x) = \wp(x). \tag{1.18}$$

That is, the unconditional posterior state is always the same as the prior state. This is the counterpart of the statement that the system variable X is unaffected by the measurement.

Exercise 1.3 *Determine the posterior conditional states $\wp'(x|y)$ in the above binary example for the two cases $y = 0$ and $y = 1$. Show that, in the limit $\mu \to 1$, $\wp'(x|y) \to \delta_{x,y}$ (assuming that $\wp(x := y) \neq 0$), whereas, in the case $\mu = 1/2$, $\wp'(x|y) = \wp(x)$. Interpret these results.*

1.1.4 Example: continuous variables

We now turn to the case of continuous state variables. Suppose one is interested in determining the position of a particle on the real line. Let the a-priori state of this system be the probability density $\wp(x)$. As explained earlier, this means that, if an ideal measurement of the position X is made, then the probability that a value between x and $x + \mathrm{d}x$ will be obtained is $\wp(x)\mathrm{d}x$.

As in the binary case, we introduce an apparatus with configuration Y and a noise variable Ξ, both real numbers. To define the measurement we specify the output function, $G(X, \Xi)$, for example

$$Y = X + \Xi, \tag{1.19}$$

so that $\Xi = G^{-1}(X, Y) = Y - X$. We must also specify the probability density for the noise variable ξ, and a common choice is a zero-mean Gaussian with a variance Δ^2:

$$\wp(\xi) = (2\pi\Delta^2)^{-1/2}\mathrm{e}^{-\xi^2/(2\Delta^2)}. \tag{1.20}$$

The post-measurement apparatus state is given by the continuous analogue of Eq. (1.13),

$$\wp(y) = \int_{-\infty}^{\infty} \wp(y|x)\wp(x)\mathrm{d}x. \tag{1.21}$$

Exercise 1.4 *Show that the mean and variance of the state $\wp(y)$ are $\mathrm{E}[X]$ and $\mathrm{Var}[X] + \Delta^2$, respectively. This clearly shows the effect of the noise.*

Finding the conditional states in this case is difficult in general. However, it is greatly simplified if the a-priori system state is Gaussian:

$$\wp(x) = (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), \tag{1.22}$$

because then the conditional states are still Gaussian.

Exercise 1.5 *Verify this, and show that the conditional mean and variance given a result y are, respectively,*

$$\bar{x}' = \frac{\sigma^2 y + \Delta^2\bar{x}}{\Delta^2 + \sigma^2}, \qquad (\sigma')^2 = \frac{\sigma^2\Delta^2}{\Delta^2 + \sigma^2}. \tag{1.23}$$

Hence show that, in the limit $\Delta \to 0$, the conditional state $\wp'(x|y)$ converges to $\delta(x - y)$, and an ideal measurement is recovered.

1.1.5 Most general formulation of classical measurements

As stated above, so far we have considered only non-disturbing classical measurements; that is, measurements with no *back-action* on the system. However, it is easy to consider classical measurements that do have a back-action on the system. For example, one could measure whether or not a can has petrol fumes in it by dropping a lit match inside. The

result (nothing, or flames) will certainly reveal whether or not there *was* petrol inside the can, but the final state of the system after the measurement will have no petrol fumes inside in either case.

We can generalize Bayes' theorem to deal with this case by allowing a state-changing operation to act upon the state after applying Bayes' theorem. Say the system state is $\wp(x)$. For simplicity we will take X to be a discrete random variable, with the configuration space being $\{0, 1, \ldots, n-1\}$. Say Y is the result of the measurement as usual. Then this state-changing operation is described by an $n \times n$ matrix $\mathcal{B}_y$, whose element $\mathcal{B}_y(x|x')$ is the probability that the measurement will cause the system to make a transition, from the state in which $X = x'$ to the state in which $X = x$, given that the result $Y = y$ was obtained. Thus, for all x' and all y,

$$\mathcal{B}_y(x|x') \geq 0, \qquad \sum_x \mathcal{B}_y(x|x') = 1. \tag{1.24}$$

The posterior system state is then given by

$$\wp'(x|y) = \frac{\sum_{x'} \mathcal{B}_y(x|x')\wp(y|x')\wp(x')}{\wp(y)}, \tag{1.25}$$

where the expression for $\wp(y)$ is unchanged from before.

We can unify the Bayesian part and the back-action part of the above expression by defining a new $n \times n$ matrix $\mathcal{O}_y$ with elements

$$\mathcal{O}_y(x|x') = \mathcal{B}_y(x|x')\wp(y|x'), \tag{1.26}$$

which maps a normalized probability distribution $\wp(x)$ onto an unnormalized probability distribution:

$$\tilde{\wp}'(x|y) = \sum_{x'} \mathcal{O}_y(x|x')\wp(x'). \tag{1.27}$$

Here we are introducing the convention of using a tilde to indicate an unnormalized state, with a norm of less than unity. This norm is equal to

$$\wp(y) = \sum_x \sum_{x'} \mathcal{O}_y(x|x')\wp(x'), \tag{1.28}$$

the probability of obtaining the result $Y = y$. Maps that take states to (possibly unnormalized) states are known as *positive maps*. The normalized conditional system state is

$$\wp'(x|y) = \sum_{x'} \mathcal{O}_y(x|x')\wp(x')/\wp(y). \tag{1.29}$$

From the properties of $\mathcal{O}_y$, it follows that it is possible to find an n-vector E_y with positive elements $E_y(x)$, such that the probability formula simplifies:

$$\sum_x \sum_{x'} \mathcal{O}_y(x|x')\wp(x') = \sum_x E_y(x)\wp(x). \tag{1.30}$$

Specifically, in terms of Eq. (1.26),

$$E_y(x) = \wp(y|x). \tag{1.31}$$

This satisfies the completeness condition

$$\forall x, \sum_y E_y(x) = 1. \tag{1.32}$$

This is the only mathematical restriction on $\{\mathcal{O}_y: y\}$ (apart from requiring that it be a positive map).

Exercise 1.6 *Formulate the match-in-the-tin measurement technique described above. Let the states $X = 1$ and $X = 0$ correspond to petrol fumes and no petrol fumes, respectively. Let the results $Y = 1$ and $Y = 0$ correspond to flames and no flames, respectively. Determine the two matrices $\mathcal{O}_{y:=0}$ and $\mathcal{O}_{y:=1}$ (each of which is a 2×2 matrix).*

The unconditional system state after the measurement is

$$\wp'(x) = \sum_y \sum_{x'} \mathcal{O}_y(x|x')\wp(x') = \sum_{x'} \mathcal{O}(x|x')\wp(x'). \tag{1.33}$$

Here the unconditional evolution map $\mathcal{O}$ is

$$\mathcal{O} = \sum_y \mathcal{O}_y. \tag{1.34}$$

Exercise 1.7 *Show that $\mathcal{O}$ is the identity if and only if there is no back-action.*

1.2 Quantum measurement theory

1.2.1 Probability and quantum mechanics

As we have discussed, with a classical system an ideal measurement can determine with certainty the values of all of the system variables. In this situation of complete knowledge, all subsequent ideal measurement results are determined with certainty. In consequence, measurement and probability do not play a significant role in the foundation of classical mechanics (although they do play a very significant role in practical applications of classical mechanics, where noise is inevitable [Whi96]).

The situation is very different in quantum mechanics. Here, for any sort of measurement, there are systems about which one has maximal knowledge, but for which the result of the measurement is not determined. The best one can do is to give the probability distributions for measurement outcomes. From this it might be inferred that a state of maximal knowledge about a quantum system is not a state of complete knowledge. That is, that there are 'hidden' variables about which one has incomplete knowledge, even when one has maximal knowledge, and these hidden variables determine the measurement outcomes.

Although it is possible to build a perfectly consistent interpretation of quantum mechanics based on this idea (see for example Ref. [BH93]), most physicists reject the idea. Probably the chief reason for this rejection is that in 1964 John Bell showed that any such deterministic hidden-variables theory must be nonlocal (that is, it must violate local causality) [Bel64]. That is, in such a theory, an agent with control over some local macroscopic parameters (such as the orientation of a magnet) can, under particular circumstances, instantaneously affect the hidden variables at an arbitrarily distant point in space. It can be shown that this effect cannot allow faster-than-light signalling, and hence does not lead to causal paradoxes within Einstein's theory of relativity. Nevertheless, it is clearly against the spirit of relativity theory. It should be noted, however, that this result is not restricted to hidden-variable interpretations; any interpretation of quantum mechanics that allows the concept of local causality to be formulated will be found to violate it – see Refs. [Bel87] (p. 172) and [Wis06].

Another, perhaps better, justification for ignoring hidden-variables theories is that there are infinitely many of them (see Ref. [GW04] and references therein). While some are more natural than others [Wis07], there is, at this stage, no compelling reason to choose one over all of the others. Thus one would be forced to make a somewhat arbitrary choice as to which hidden-variables interpretation to adopt, and each interpretation would have its own unique explanation as to the nature of quantum-mechanical uncertainty.

Rather than grappling with these difficulties, in this book we take an *operational* approach. That is, we treat quantum mechanics as simply an algorithm for calculating what one expects to happen when one performs a measurement. We treat uncertainty about future measurement outcomes as a primitive in the theory, rather than ascribing it to lack of knowledge about existing hidden variables.

We will still talk of a quantum state as representing our knowledge about a system, even though strictly it is our knowledge about the outcomes of our future measurements on that system. Also it is still useful, in many cases, to think of a quantum state of maximal knowledge as being like a classical state of incomplete knowledge about a system. Very crudely, this is the idea of 'quantum noise'. The reader who is not familiar with basic quantum mechanics (pure states, mixed states, time-evolution, entanglement etc.) should consult Appendix A for a summary of this material. However, before moving to quantum measurements, we note an important point of terminology. The matrix representation ρ of a mixed quantum state is usually called (for historical reasons) the density operator, or density matrix. We will call it the *state matrix*, because it generalizes the *state vector* for pure states.

Finally, just as in the classical case, observers with different knowledge may assign different states simultaneously to a single system. The most natural way to extend the concept of consistency to the quantum case is to replace the common state of maximal knowledge with a common pure state. That is, the condition for the consistency of a collection of states $\{\rho_j\}$ from different observers is that there exists a positive ϵ and a ket $|\psi\rangle$ such that, for all j, $\rho_j - \epsilon|\psi\rangle\langle\psi|$ is a positive operator. In other words, each observer's state ρ_j can be written as a mixture of the pure state $|\psi\rangle\langle\psi|$ and some other states.

Exercise 1.8 *Show, from this definition, that two different pure states $\hat{\pi}_1$ and $\hat{\pi}_2$ cannot be consistent states for any system.*
Hint: *Consider the operator $\hat{\sigma}_j = \hat{\pi}_j - \epsilon|\psi\rangle\langle\psi|$. Assuming $\hat{\pi}_j \neq |\psi\rangle\langle\psi|$, show that $\mathrm{Tr}[\sigma_j^2] > (\mathrm{Tr}[\sigma_j])^2$ and hence deduce the result.*

1.2.2 Projective measurements

The traditional description of measurement in quantum mechanics is in terms of projective measurements, as follows. Consider a measurement of the physical quantity Λ. First we note that the associated operator $\hat{\Lambda}$ (often called an *observable*) can be diagonalized as

$$\hat{\Lambda} = \sum_{\lambda} \lambda \hat{\Pi}_\lambda, \tag{1.35}$$

where $\{\lambda\}$ are the eigenvalues of $\hat{\Lambda}$ which are real and which we have assumed for convenience are discrete. $\hat{\Pi}_\lambda$ is called the projection operator, or projector, onto the subspace of eigenstates of $\hat{\Lambda}$ with eigenvalue λ. If the spectrum (set of eigenvalues $\{\lambda\}$) is non-degenerate, then the projector would simply be the rank-1 projector $\hat{\pi}_\lambda = |\lambda\rangle\langle\lambda|$. We will call this special case *von Neumann measurements*.

In the more general case, where the eigenvalues of $\hat{\Lambda}$ are N_λ-fold degenerate, $\hat{\Pi}_\lambda$ is a rank-N_λ projector, and can be written as $\sum_{j=1}^{N_\lambda} |\lambda, j\rangle\langle\lambda, j|$. For example, in the simplest model of the hydrogen atom, if Λ is the energy then λ would be the principal quantum number n and j would code for the angular-momentum and spin quantum numbers l, m and s of states with the same energy. The projectors are orthonormal, obeying

$$\hat{\Pi}_\lambda \hat{\Pi}_{\lambda'} = \delta_{\lambda,\lambda'} \hat{\Pi}_\lambda. \tag{1.36}$$

The existence of this orthonormal basis is a consequence of the spectral theorem (see Box 1.1).

When one measures Λ, the result one obtains is one of the eigenvalues λ. Say the measurement begins at time t and takes a time T. Assuming that the system does not evolve significantly from other causes during the measurement, the probability for obtaining that particular eigenvalue is

$$\Pr[\Lambda(t) = \lambda] = \wp_\lambda = \mathrm{Tr}[\rho(t)\hat{\Pi}_\lambda]. \tag{1.37}$$

After the measurement, the conditional (a-posteriori) state of the system given the result λ is

$$\rho_\lambda(t+T) = \frac{\hat{\Pi}_\lambda \rho(t) \hat{\Pi}_\lambda}{\Pr[\Lambda(t) = \lambda]}. \tag{1.38}$$

That is to say, the final state has been projected by $\hat{\Pi}_\lambda$ into the corresponding subspace of the total Hilbert space. This is known as the *projection postulate*, or sometimes as *state collapse*, or *state reduction*. The last term is best avoided, since it invites confusion with the reduced state of a bipartite system as discussed in Section A.2.2 of Appendix A. This process