On the Postulates of Quantum Mechanics
and their Interpretation.

Marcel Nooijen

University of Waterloo

One arrives at very implausible theoretical conceptions, if one attempts to maintain the thesis that the statistical quantum theory is in principle capable of producing a complete description of an individual physical system. On the other hand, those difficulties of theoretical interpretation disappear, if one views the quantum mechanical description as the description of ensembles of systems. Albert Einstein (1949).

Quantum mechanics has given rise to a lot of discussion since its conception, as some things are so weird. I will try to clarify some of the issues here. One thing is very important from the outset. Quantum mechanics is a statistical theory. It tells us the various possible outcomes of experiments and the corresponding probabilities if we would do *a large number of identical experiments on individual quantum systems*. Identical experiments are necessarily idealizations, but this is not much of a restriction in practice, as many variables (e.g. what's is going on in Sidney or on the next bench in the lab) are irrelevant. In this context taking a spectrum of a sample in the gas phase appears to be a single experiment but it really amounts to doing measurements on many individual quantum systems. The systems are not all identical but this is the same type of fluctuation that occurs in classical statistical descriptions. At first sight the situation may not appear very different therefore from the description provided by classical statistical mechanics. In that case however, we have an underlying description (classical mechanics) that provides a complete (i.e. non-statistical) description of the world, which in general is far too complex, however, to be of use. Throwing a dice is a good example. Here we find the probabilities to obtain either of the outcomes 1, 2, 3, 4, 5, 6 equally likely (1/6). However, in principle it is possible to throw the dice always in precisely the

same way such that it would yield a 6 say every time. It is unlikely that you would master this skill yourself but it appears well possible to built a device that would accomplish this. Classical mechanics says that in principle it can be done. The dice in the quantum world is different. It is *intrinsically impossible to predict the outcome of a single event* (unless the probability is unity). In fact this type of 'single event' experiment is not even the subject of quantum theory. We can only predict statistics and this *requires* that one does many identical experiments. Let me emphasize this aspect as it is easily to get confused on this issue. For example in every day language we might hear that there is a 70% chance that it will rain tomorrow. What does this mean? How can we verify that this 70% is a valid prediction? These kind of predictions only have a meaning when evaluated over many tomorrows.

Every description of an experiment on a microscopic system (even single molecule spectroscopy) is essentially statistical. Typically one performs an experiment on a sample consisting of similar microscopic systems. In an idealized theoretical description we view such an experiment as equivalent to performing a sequence of measurements on each (now supposedly identical) microscopic system in isolation. This generates a definite result for each individual experiment, and *the statistics of the distribution of results is described by quantum mechanics*. This theoretically predicted distribution of results may not quite agree with the experimental result for a variety of reasons. The actual experimental sample will contain a distribution of different microsystems; it will involve some interaction between different microsystems or between microsystems and the environment, and so forth. Some of these effects can be taken into account by using statistical mechanics. This is beyond where I want to go, however. Let us simply assume that the quantum description would agree very well with the experimental result, and analyse what it says and what it means.

The above abstraction of an experiment shows that quantum mechanics describes the statistical outcome of an experiment performed on an *ensemble* of identical microsystems. *The wave function in quantum mechanics describes the properties of the ensemble.* This is very different from classical theory, where the laws of physics describe

individual systems, and we can imagine how they relate to 'reality'. We cannot claim that this is true in quantum mechanics. The wave function relates to an ensemble of identical systems. In most textbooks on quantum mechanics the wave function is said to describe the 'system'. This is slightly inaccurate and it is analogous to the fact that the weather forecast does not describe tomorrow but an 'average of possible tomorrows' given the current situation in the atmosphere. The reason that I emphasize this seemingly minor point is that it helps to explain many of the 'strange' results of quantum mechanics. As long as we strictly adhere to the so-called ensemble interpretation of quantum mechanics we seldom run into difficulties. It is unsatisfactory perhaps that this is all we (can?) know about the microscopic world, but this is the (current) state of affairs.

Below I will discuss the postulates of quantum mechanics. We will phrase these postulates using just the *x*-coordinate of a single particle. The generalization to many-particles in 3d space is straightforward, but leads to less transparent equations unless some new notation would be introduced. Also we will discuss everything in the so-called coordinate representation. A more general formulation is not needed for our present purposes.

1. <u>**Measurements in Quantum Mechanics.**</u>

In MS (McQuarrie and Simon) mathchapter 2 you find a discussion on the statistical analysis of a repeated measurement of the quantity A on a given ensemble. Let us assume that if we measure the quantity A, we have possible outcomes $a_1, a_2, ...,$. The probability $P(a_i) = P_i$ to find the value a_i for our given ensemble is N_i / N_{tot}, where N_i is the number of times we obtain a_i out of a total number of measurements N_{tot}. In the limit of a very large set of measurements these probabilities will converge to definite numbers $0 \le P(a_i) \le 1$. This leads to the purely classical definitions of the average value $\overline{A}$ or $\langle A \rangle$

$$\overline{A} = P_1 a_1 + P_2 a_2 + ... = \sum_k P_k a_k \qquad (4.1)$$

and the standard deviation $\sigma(A)$, or similarly its square, the so-called variance:

$$\sigma(A)^2 = P_1(a_1 - \overline{A})^2 + P_2(a_2 - \overline{A})^2 + ... = \sum_k P_k(a_k - \overline{A})^2 \geq 0 \quad (4.2)$$

The variance can be written alternatively as

$$\sum_k P_k(a_k - \overline{A})^2 = \sum_k P_k(a_k^2 - 2a_k\overline{A} + (\overline{A})^2)$$

$$= \sum_k P_k a_k^2 - 2\langle A \rangle \sum_k P_k a_k + \langle A \rangle^2 \sum_k P_k \quad (4.3)$$

$$= \langle A^2 \rangle - \langle A \rangle^2 \geq 0$$

and this is the form that turns out to be very convenient in QM. It is seen that the variance is 0 only if we only have a single possible outcome of the experiment with unit probability. This will be an intuitively obvious observation.

In Quantum Mechanics we have a set of strict rules for discussing measurements. The prime role is played by the quantity being measured, the so-called observable, which is analogous to the 'measuring apparatus'. In Quantum Mechanics observables are described by so-called Hermitian operators, which describe the possible outcomes of experiments. The second ingredient in a measurement is the sample, or 'ensemble'. It is a collection of 'identical' microsystems, and this ensemble is described by the so-called (normalized) wave function. The wave function determines the probabilities for the outcome of the experiment. There are generalizations of the wave function, involving the so-called density matrix, that I will only briefly allude to in these notes. Let us proceed with a brief description of the quantum machinery.

An observable classical quantity A corresponds to a *linear Hermitian operator* $\hat{A}$ in quantum mechanics (Operators carry hats to distinguish them from numbers. They act on functions.). Classical quantities are functions of the basic variables, the positions and momenta of the particles (energy, angular momentum are some examples). In the classical theory the basic variables depend on time, and the laws of physics describe how they involve in time, given initial conditions, and given the forces between the particles. In Quantum Mechanics position and momentum become abstract operators, that themselves do not depend on time. There is a translation table to go from the classical quantities to the quantum operators. Some examples:

Classical	Quantum
$x(t), y(t), z(t)$	$\hat{X}, \hat{Y}, \hat{Z}$
$p_x(t), p_y(t), p_z(t)$	$\hat{P}_x = -i\hbar\dfrac{\partial}{\partial x}, \hat{P}_y = -i\hbar\dfrac{\partial}{\partial y}, \ldots$
$E = T + V = \dfrac{p^2}{2m} + V(x,y,z)$	$\hat{H} = -\dfrac{\hbar^2}{2m}(\dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2} + \dfrac{\partial^2}{\partial z^2}) + V(x,y,z)$
$L_x = yp_z - zp_y$	$\hat{L}_x = \hat{Y}\hat{P}_z - \hat{Z}\hat{P}_y \rightarrow -i\hbar(y\dfrac{\partial}{\partial z} - z\dfrac{\partial}{\partial y})$
$L_y = zp_x - xp_z$	$\hat{L}_y = \hat{Z}\hat{P}_x - \hat{X}\hat{P}_z \rightarrow -i\hbar(z\dfrac{\partial}{\partial x} - x\dfrac{\partial}{\partial z})$
$L_z = xp_y - yp_x$	$\hat{L}_z = \hat{X}\hat{P}_y - \hat{Y}\hat{P}_x \rightarrow -i\hbar(x\dfrac{\partial}{\partial y} - y\dfrac{\partial}{\partial x})$
	$\hat{S}_x, \hat{S}_y, \hat{S}_z$

This allows us to make the corresponding operators in QM: Replace x, y, z by the multiplication operators $\hat{X}, \hat{Y}, \hat{Z}$, and replace the momenta by differential operators: $p_x \rightarrow \hat{P}_x = -i\hbar\dfrac{\partial}{\partial x}$ and so forth. This allows you to construct most operators that occur in QM. In addition there are three operators with no classical analog. They are the spin operators $\hat{S}_x, \hat{S}_y$ and $\hat{S}_z$. These will be discussed later on.

An operator $\hat{A}$ is linear if for any linear combination of functions (in its domain)
$$\hat{A}(c_1 f(x) + c_2 g(x)) = c_1 \hat{A}(f(x)) + c_2 \hat{A}(g(x)) \qquad (4.4)$$
while it is called Hermitian if for any two functions (in its domain)
$$\int_{-\infty}^{\infty} g^*(x)\hat{A}(f(x))dx = \int_{-\infty}^{\infty} f(x)(\hat{A}g(x))^* dx = \int_{-\infty}^{\infty} (\hat{A}g(x))^* f(x)dx (4.5)$$
Hence instead of acting on $f(x)$ we can act on $g(x)$ and take the complex conjugate of the result. The value for the integral in either form is the same irrepective of the actual functions f and g. Therefore, Hermiticity is a property of the operator. The Hermiticity of an operator is connected to the precise definition of the integration. Typically operators

are Hermitean only if they act on functions that satisfy the boundary conditions in the problem (This is again tied to the integration. Think of the particle in the box, or a particle on the ring). Such functions that satisfy the boundary conditions are in the so-called domain of the operator. Hermitian operators have some very important properties that make them suitable for the representation of observables which play a crucial role in QM. In particular they have real eigenvalues a_k, and their eigenfunctions can always be chosen to be orthonormal:

Hermitian operator $\hat{A}$:   orthonormal eigenfunctions $\phi_k(x)$, real eigenvalues a_k:

$$\hat{A}\phi_k(x) = a_k\phi_k(x) \tag{4.6}$$

$$\int_{-\infty}^{\infty} \phi_k^*(x)\phi_l(x)dx = \delta_{kl} \tag{4.7}$$

where $\delta_{kl} = 1, k = l$; $\delta_{kl} = 0, k \neq l$; is the the Kronecker delta symbol.

Moreover <u>any</u> function $\Psi(x)$ (that satisfies the boundary conditions) can be expressed as a linear combination of the eigenfunctions of a Hermitian operator $\hat{A}$ (this is the the so-called completeness property)

$$\Psi(x) = c_1\phi_1(x) + c_2\phi_2(x) + ... \tag{4.8}$$

where the expansion coefficients are given by

$$c_k = \int_{-\infty}^{\infty} \phi_k^*(x)\Psi(x)dx \tag{4.9}$$

These mathematical definitions lead to a very precise description of measurements on an ensemble of microscopic systems. The statements below are a concise form of the postulates of QM related to measurements. The ensemble in QM is described by a normalized wave function $\Psi(x)$: $\int_{-\infty}^{\infty} \Psi(x)^*\Psi(x)dx = 1$, and the results of measurement of a quantity A for the entire ensemble can be described as follows:

- Only eigenvalues a_k can be obtained, one of them for 'each measurement on an individual system in the ensemble'. Collecting the data on the individual measurements yields the statistical information.

- Each possible value a_k shows up with probability $P_k = |c_k|^2 = c_k^* c_k$, where the (in general complex) coefficient c_k is defined in Eqn. (4.9). This probability depends on the sample, described by the normalized wave function $\Psi(x)$, and it depends on the observable being measured (through its eigenfunctions $\phi_k(x)$).

- After measuring A the initial ensemble is split up in subensembles, one corresponding to each of the possible eigenvalues a_k. The wave function that describes the subensemble corresponding to a_k is the corresponding (normalized) eigenstate $\phi_k(x)$.

- If we would continue to pursue with some further measurements on the complete ensemble, we would work with each of the subensembles described by $\phi_k(x)$, and collect the results by summing over each subensemble and multiplying this subresult with the respective probability P_k.

In the above description we have assumed that each of the eigenvalues of the operator $\hat{A}$ is non-degenerate. If there is more than one independent eigenfunction, say $\phi_{k_1}(x), \phi_{k_2}(x)$ corresponding to the same eigenvalue a_k the rules are slightly more complicated. We will outline this, after we have introduced some more notation.

============ *Average values, expectation values and variances in QM.*

The above is a fairly precise description of the information content of QM. In particular we can obtain information on the averages and variances discussed above:

$$\overline{A} = \langle A \rangle = \langle \hat{A} \rangle \equiv \int_{-\infty}^{\infty} \Psi^*(x) \hat{A} \Psi(x) dx \qquad (4.10)$$

This average quantity is also called the expectation value. For the variance we have

$$\sigma^2(A) = \langle A^2 \rangle - \langle A \rangle^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2 = \langle (\hat{A} - \langle \hat{A} \rangle)^2 \rangle$$
$$= \int_{-\infty}^{\infty} \Psi^*(x) \left(\hat{A} - \langle \hat{A} \rangle \right)^2 \Psi(x) dx \qquad (4.11)$$

These relations are not extra postulates. They follow from the previous statements and their classical definitions as can be verified by expanding $\Psi(x)$ in terms of eigenfunctions of $\hat{A}$. e.g.

$$\left\langle \hat{A} \right\rangle = \int_{-\infty}^{\infty} \Psi^*(x) \hat{A} \Psi\ (x) dx$$

$$= \int_{-\infty}^{\infty} \sum_k c_k^* \phi_k^*(x) \hat{A} \sum_l c_l \phi_l(x) dx$$

$$= \sum_k c_k^* \sum_l c_l \int_{-\infty}^{\infty} \phi_k^*(x) \hat{A} \phi_l(x) dx \qquad\qquad (4.12)$$

$$= \sum_k c_k^* \sum_l c_l a_l \int_{-\infty}^{\infty} \phi_k^*(x) \phi_l(x) dx$$

$$= \sum_{k,l} c_k^* c_l a_l \delta_{kl} = \sum_k c_k^* c_k a_k = \sum_k P_k a_k = \left\langle A \right\rangle$$

There are important general aspects of this proof that you are expected to know. In the first step we substitute the expansion for $\Psi(x)$. Please note that we use a different name for the summation indices to represent $\Psi(x)$ and its complex conjugate. This is completely general technique: never use the same summation index twice. They are always independent. To emphasize this: an index may appear multiple times, but not under the summation sign. In the second step we use linearity, both of the operator $\hat{A}$ and of performing integration. In going from the third to the fourth line we use that the expansion functions are eigenfunctions of $\hat{A}$, and finally we use that the eigenfunctions are orthonormal. The integral in the fourth line is non-zero (unity in fact) only if $l = k$, which gets rid of the sum over l.

Similarly the QM variance can be shown to reduce to $\sigma^2(A) = \sum_k P_k (a_k - \overline{A})^2$, which is the classical expression. The variance vanishes if and only if only one term in the sum contributes, in which case the average equals this particular eigenvalue a_l say. It follows therefore that the variance vanishes only if the wave function $\Psi(x)$ that describes the ensemble is an eigenfunction of $\hat{A}$. One of the postulates was that the subensemble obtained by measuring an eigenvalue a_i was described by the eigenfunction $\phi_i(x)$. Here it is seen that if we measure A again on this particular subensemble we find the eigenvalue a_i with absolute certainty. This is what we would expect classically. Measuring a quantity is therefore a good way to prepare an ensemble of identical microsystems(!), as the subensemble corresponding to a particular eigenvalue is collectively described by the same wave function after the measurement.

For those of you who are familiar with linear algebra it may be illustrative to point out the analogy of the orthonormal eigenfunctions of an operator $\hat{A}$ with an orthonormal basis for a vector space. The functions $\phi_1(x), \phi_2(x), \ldots$ can be thought to form orthonormal basis vectors (coordinate system, or axes). $\Psi(x)$ represents an arbitrary vector in the space, with coordinates $(c_1, c_2, \ldots)$ with respect to this basis. The inner product between two vectors (needed to define the vector space) is defined as :

$$\Psi_a(x) = a_1\phi_1(x) + a_2\phi_2(x) + \ldots$$
$$\Psi_b(x) = b_1\phi_1(x) + b_2\phi_2(x) + \ldots$$

(4.13)

$$"\Psi_a \cdot \Psi_b" = \int_{-\infty}^{\infty} \Psi_a^*(x)\Psi_b(x)dx = a_1^*b_1 + a_2^*b_2 + \ldots$$

The act of measurement is then described as decomposing the full ensemble corresponding to eigenfunction $\Psi(x) = \sum_k c_k \phi_k(x)$ into subensembles corresponding to eigenvalues a_i that are each described by a particular basis vector $\phi_i(x)$. The probability to find eigenvalue a_i equals $|c_i|^2$. It is not correct to think of the ensemble to consist of subensembles in the first place (this many particles in state $\phi_1(x)$, that many in state $\phi_2(x)$, etc.). This would give the wrong result whenever so-called interference effects play a role. Using this vector analogy different measurements A, B correspond to different sets of coordinate axes (different eigenfunctions in general), and a different decomposition of the wave function $\Psi(x)$ that describes the ensemble. The vector analog is an intuitive (*and rigorously valid*) picture that may help you to understand this somewhat abstract chapter.

====== Dirac notation.

The analogy between complete sets of orthonormal functions and that of orthonormal vector space can be made very naturally by using a clever notation invented by Dirac, called Dirac bra-ket notation. I will list a little dictionary of the translation, as this clarifies the ideas and the power of Dirac notation. It will be used extensively throughout the course.

Functions	Dirac notation	Comments
$\Psi(x)$	$\lvert\Psi\rangle$	Arbitrary state, "ket"
$\phi_i(x)$	$\lvert i\rangle$, or $\lvert a_i\rangle$	Orthonormal basis function
$\int\Psi_a^*(x)\Psi_b(x)dx$	$\langle\Psi_a\lvert\Psi_b\rangle=\langle\Psi_b\lvert\Psi_a\rangle^*$	Inner product of two vectors: bracket
$\int\phi_i^*(x)\phi_j(x)dx=\delta_{ij}$	$\langle i\lvert j\rangle=\delta_{ij}$	Orthonormality of the basis.
$\int\phi_i^*(x)....dx$	$\langle i\rvert$	"Bra": (anti-)linear Functional:

associates a number when acting on arbitrary vector…

$$\Psi(x)=\sum_i\phi_i(x)c_i \qquad \lvert\Psi\rangle=\sum_i\lvert i\rangle\langle i\lvert\Psi\rangle$$
$$=\sum_i\phi_i(x)\int\phi_i^*(x)\Psi(x)dx \qquad \hat{\mathbf{1}}=\sum_i\lvert i\rangle\langle i\rvert$$

Completeness relation

The Dirac notation has the enormous advantage that it is more compact than the explicit notation using functions. This is true in particular if the coordinate space extends over many particles and 3 dimensions! All of the variables are implicit. This was of great importance when Dirac introduced this notation, and he could relate different versions of quantum mechanics using his so-call transformation theory. A very nice feature of the notation is that the resolution of the identity, $\hat{\mathbf{1}}=\sum_i\lvert i\rangle\langle i\rvert$, takes on a convenient form. Many useful manipulations will just be performed by inserting this resolution of the identity. Always bear in mind that summation indices are independent and should get different names. Some examples of using Dirac notation:

$$\langle\Psi\lvert\Psi\rangle=\sum_i\langle\Psi\lvert i\rangle\langle i\lvert\Psi\rangle=\sum_i(\langle i\lvert\Psi\rangle)^*\langle i\lvert\Psi\rangle=\sum_ic_i^*c_i=\sum_i\lvert c_i\rvert^2=1$$

$$\hat{O}\lvert\Psi\rangle=\sum_i\hat{O}\lvert i\rangle\langle i\lvert\Psi\rangle=\sum_{i,j}\lvert j\rangle\langle j\lvert\hat{O}\lvert i\rangle\langle i\lvert\Psi\rangle=\sum_{i,j}\lvert j\rangle O_{ji}c_i$$

Here $\langle j|\hat{O}|i\rangle = \int \phi_j^*(x)\hat{O}\phi_i(x)dx$ are the matrix elements of the operator $\hat{O}$ in the basis $|i\rangle$. We just inserted the resolution of the identity twice. The final result expresses the result $\hat{O}|\Psi\rangle$ in the basis set $|j\rangle$.

An operator is Hermitian if $\langle j|\hat{O}|i\rangle = \langle i|\hat{O}|j\rangle^*$ $\forall i,j$

======= *Commutators and the uncertainty principle.*

Above we discussed the variance σ_A^2 of measuring the quantity A corresponding to an operator $\hat{A}$. The variance depends on the ensemble (of course), hence on the wave function $\Psi(x)$, and we found that $\sigma_A = 0$ if and only if the wave function $\Psi(x)$ that describes the ensemble is actually an eigenfunction of $\hat{A}$. The easiest way to prepare an ensemble in such a state is by performing the measurement (as discussed). Now we can measure a quantity B after measuring A and ask the question: "Can we prepare an ensemble such that both σ_A and σ_B are zero?" This would mean that both the quantity A and B are uniquely specified for each element of the ensemble. Following the postulates this would be the case if the wave function $\Psi(x)$ is an eigenfunction of both the operators $\hat{A}$ *and* $\hat{B}$. A more elaborate question would be "Can we create an ensemble that has sharp values for A and B for *all possible* outcomes of measuring A then B?". This would imply that there is a function for *each possible pair* of eigenvalues (a_i, b_j) that is an eigenfunction of both $\hat{A}$ and $\hat{B}$. The operators $\hat{A}$ and $\hat{B}$ in such a case should have a complete set of *common* eigenfunctions

$$\phi_{a_1,b_1}(x), \phi_{a_2,b_2}(x),....:$$

where

$$\hat{A}\phi_{a_ib_j}(x) = a_i\phi_{a_ib_j}(x)$$
$$\hat{B}\phi_{a_ib_j}(x) = b_j\phi_{a_ib_j}(x)$$

(4.14)

From this it follows that

$$\hat{A}\hat{B}\phi_{a_i b_j}(x) = \hat{B}\hat{A}\phi_{a_i b_j}(x) = a_i b_j \phi_{a_i b_j}(x)$$

and since any $\Psi(x)$ can be expanded in this set of eigenfunctions (the completeness property) it follows that

$$\hat{A}\hat{B}\Psi(x) = \hat{B}\hat{A}\Psi(x) \text{ for any } \Psi(x) \tag{4.15}$$

or
$$\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A} = 0 \tag{4.16}$$

Here we have shown that if it is in principle possible to create an ensemble that has sharp values for A and B for any permissable pair of values of these observables only if $\hat{A}$ and $\hat{B}$ commute. The eigenvalues come in specific pairs, as they have to correspond to specific common eigenfunctions. How can we prepare such an ensemble? Simple. Measure a quantity A, take the subensemble that corresponds to eigenvalue a_i and go on to measure the quantity B. If we now pick the subensemble corresponding to b_j it would be described by a common eigenfunction of $\hat{A}$ and $\hat{B}$, $\phi_{a_i, b_j}(x)$. We can reverse the order of the measurements to create the specific ensemble, and we can do it for any pair of commensurate eigenvalues a_i and b_j. This also works if $\hat{A}$ and $\hat{B}$ have degenerate eigenvalues, although I will not venture a derivation of this result here. Finally the converse is also true: if $\left[\hat{A}, \hat{B}\right] = 0$ then A and B can be measured to indefinite precision. This can be more accurately phrased as: It is possible to prepare specific ensembles that will yield definite results for the (sequential) measurement of A and B.

From the postulates it also transpires what happens if we subsequently measure observables A and B whose corresponding operators $\hat{A}$ and $\hat{B}$ do not have common eigenfunctions. If we measure $\hat{A}$ and continue with the subensemble corresponding to eigenvalue a_i, which is described by the eigenfunction $\phi_i(x)$ we can continue to measure $\hat{B}$. However by assumption $\phi_i(x)$ is not an eigenfunction of $\hat{B}$ and we will therefore <u>not</u> get a sharp value for B. There is necessarily a spread. After measuring $\hat{B}$ and continuing with the subensemble corresponding to b_i described by eigenfunction $\psi_j(x)$ (eigenfunction of $\hat{B}$ but not of $\hat{A}$) we could decide to measure $\hat{A}$ again. However, the

new wave function $\psi_j(x)$ is not an eigenfunction of $\hat{A}$ and we would get a distribution of values for the observable A. This means that measuring B has destroyed the sharp value for A we created by the very act of measuring A. If the operators $\hat{A}$ and $\hat{B}$ do not commute in general the measurement of $\hat{B}$ destroys the information we obtained by measuring $\hat{A}$. This is the content of the uncertainty principle. I might add that it is possible that $\hat{A}$ and $\hat{B}$ do not commute, but still have *some* eigenfunctions in common. In such a case the corresponding eigenvalues can be specified sharply and are not disturbed by subsequent measurements of $\hat{A}$ or $\hat{B}$. Let me also note that the notion of *simultaneous* measurement does not enter the discussion. Using our ensemble interpretation the quest is for an ensemble that would yield a sharp value for the observables upon measurement. Be it in sequence or simultaneous or in combination. None of this matters if $\hat{A}$ and $\hat{B}$ commute. (This is strictly true only if $\hat{A}$ and $\hat{B}$ also commute with the Hamiltonian, see below).

For your information I will quote the general result about the product of standard deviations for operators that do not commute. The minimum spread depends on the initial wave function that describes the ensemble and the general result reads

$$\Delta A \Delta B \geq \frac{1}{2}\left|\left\langle\left[\hat{A},\hat{B}\right]\right\rangle\right| \qquad \text{(for given state } \Psi(x)) \qquad (4.17)$$

For the quantum aficionados in the class the proof follows by defining

$$\phi_\lambda(x) = \{(\hat{A}-\langle\hat{A}\rangle)+\lambda(\hat{B}-\langle\hat{B}\rangle)\}\Psi(x); \qquad (4.18)$$

Then using that $\int_{-\infty}^{\infty}\phi_\lambda^*(x)\phi_\lambda(x)dx \geq 0 \; \forall \; \lambda$ (this depends on the Hermiticity of $\hat{A}$ and $\hat{B}$) you can derive the general result (4.17). Give it a try!

Finally we have the special case: $\left[\hat{A},\hat{B}\right]=c$ (a constant), which implies $\Delta A \Delta B \geq \frac{1}{2}|c|$:
This leads to a lower bound on the precision of measuring A and B that is *independent of the initial ensemble (wave function)*. The most famous example is of course:

$\Delta x \Delta p_x \geq \frac{1}{2}|i\hbar| = \frac{\hbar}{2}$, Heisenbergs uncertainty principle.

==== *Some tricks when evaluating Commutators:*

$$(\hat{A}\hat{B} - \hat{B}\hat{A})\Psi(x,...) = ?? = \hat{C}\Psi(x,...), \text{ for any } \Psi(x) \quad \rightarrow \left[\hat{A}, \hat{B}\right] = \hat{C}$$

You need to act on an arbitrary function to get the desired result.

Examples: $\left[\hat{p}_x, \hat{p}_y\right] = 0;\ \left[\hat{x}, \hat{y}\right] = 0;\ \left[\hat{x}, \hat{p}_y\right] = 0;\ \left[\hat{x}, \hat{p}_x\right] = i\hbar$

$$\left[\hat{A}, \hat{B} + \hat{C}\right] = \left[\hat{A}, \hat{B}\right] + \left[\hat{A}, \hat{C}\right];\quad \left[\hat{A}, \hat{B}\hat{C}\right] = \left[\hat{A}, \hat{B}\right]\hat{C} + \hat{B}\left[\hat{A}, \hat{C}\right]$$

e.g. $\left[\hat{x}, \hat{p}_x^2\right] = \left[\hat{x}, \hat{p}_x\right]\hat{p}_x + \hat{p}_x\left[\hat{x}, \hat{p}_x\right] = 2i\hbar\hat{p}_x$

2. The Time-Dependent Schrödinger equation.

Another important postulate in quantum mechanics concerns the time-dependence of the wave function. This is governed by the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} = \hat{H}\Psi(x,t) \tag{4.19}$$

where $\hat{H}$ is the Hamiltonian operator of the system (the operator corresponding to the classical expression for the energy). This is a first order differential equation in t, which means that if we specify the wave function at an initial time t_0, the wave function is determined at all later times. Let me emphasize that this means the wave function has to be specified for *all* x at initial time t_0. These initial conditions are familiar from wave equations as discussed in MS Chapter 2. In classical physics we often deal with second order differential equations and in addition the time derivative $\partial\Psi(x,t)/\partial t$ would then need to be specified for all x. Let me emphasize here that although the experimental results that can be predicted from QM are statistical in nature, *the Schrödinger equation that determines the wave function as a function of time is completely deterministic.*

======= *Special solutions: Stationary States* (only if $\hat{H}$ is time-independent).

If we assume that the wave function can be written as a product: $\Psi(x,t) = \phi(x)\gamma(t)$ we can seperate the time dependence from the spatial dependence of the wave function in the usual way. The separation constant is called E and will turn out to be the energy of the system for such solutions

$$\rightarrow \dots \, i\hbar\frac{d\gamma(t)}{dt} = E\gamma(t) \rightarrow \gamma(t) = e^{-iE(t-t_0)/\hbar} \qquad (4.20)$$

$$\hat{H}\phi(x) = E\phi(x) \rightarrow \hat{H}\phi_n(x) = E_n\phi_n(x) \qquad (4.21)$$

equation (4.21) is called the time-independent Schrödinger equation and plays a central role in all of chemistry. Since the operator $\hat{H}$ is Hermitian the eigenfunctions form a complete and (can be chosen to be an) orthonormal set of functions. Using these eigenfunctions of $\hat{H}$ special solutions to the time-dependent Schrödinger equation can be expressed as

$$\Psi(x,t) = \phi_n(x)e^{-iE_n(t-t_0)/\hbar}; \qquad \Psi(x,t_0) = \phi_n(x)$$

For these special solutions of the Schrödinger equation, all measurable properties are independent of time. For this reason they are called stationary states. For example the probability distribution

$$\left|\Psi(x,t)\right|^2 = \left|\Psi(x,t_0)\right|^2 = \left|\phi_n(x)\right|^2,$$

but also

$$\left\langle \hat{A} \right\rangle_t = \left\langle \hat{A} \right\rangle_{t_0} \quad \forall \, \hat{A},$$

as is easily verified, by substituting the product form of the wave function. Also the probabilities to measure an eigenvalue a_k are independent of time, as seen below

$$\hat{A}\varphi_k(x) = a_k\varphi_k(x) \rightarrow P_k(t) = \left|\int_{-\infty}^{\infty}\varphi_k^*(x)\Psi(x,t)dx\right|^2 = P_k(t_0)$$

The common element in each of these proofs is that the time-dependent phase factor cancels because we have both $\Psi(x,t)$ and $\Psi^*(x,t)$ in each expression, while the time-dependent phase can be taken outside of the integration. Let me note that the stationary

solutions are determined by the initial condition. If you start off with a stationary state at t_0, the wave function is a stationary state for all time.

The general solution of the time-dependent Schrödinger equation (TDSE) can be written as a time-dependent linear combination of stationary states. If we assume that the initial state is given by

$$\Psi(x,t_0) = c_1\phi_1(x) + c_2\phi_2(x) + ...$$

(it can always be written in this fashion as the eigenfunctions of $\hat{H}$ form a complete set), then it is easily verified that

$$\Psi(x,t) = c_1 e^{-iE_1(t-t_0)/\hbar}\phi_1(x) + c_2 e^{-iE_2(t-t_0)/\hbar}\phi_2(x) + ...$$

satisfies the TDSE and the initial condition. For this general linear combination of eigenstates of $\hat{H}$ (the general case) properties *do* depend on time. This is true for expectation values and probabilities, and is due to the fact that different 'components' in the wave function oscillate with different time factors. In calculating expectation values we get cross terms and the time-dependent phase factors do not cancel out. This situation is part of a homework problem, to help you digest the material.

Independent of the initial wave function, energy is always conserved (as would be expected from classical physics), and also the probabilities to measure a particular energy E_k:

$$\left| \int_{-\infty}^{\infty} \phi_k^*(x)\Psi(x,t)dx \right|^2 = \left| c_k e^{-iE_k(t-t_0)/\hbar} \right|^2 = \left| c_k \right|^2$$

This is true essentially because each component in the wave function that corresponds to an eigenstate oscillates with its own time constant. But the length of this component, which is the relevant quantity upon measurement of the energy is not affected. There are no cross terms in assigning the probabilities.

Further Remarks:

-   All depends on the initial wave function, which is arbitrary in principle. The most common way is to specify it by means of a measurement!

-   Stationary states: the inital state is defined to be an eigenfunction of $\hat{H}$. In this case nothing moves except the phase of the wave function. In our current version of QM we would find infinite lifetimes of excited states! (This is because the electro-magnetic field is missing from our treatment and we have assumed a time-independent $\hat{H}$). However, the fact that in stationary states there are no moving charges explains why electrons in stationary states do not fall back into the nuclei while emitting radiation. There *is* no oscillating charge in these states. Electrons are described by wave functions, not as classically moving particles.

-   In the most general case (if the initial state is not a stationary state) properties oscillate in time (e.g. electron density) $\rightarrow$ radiation!? Again there is a need to include e.m. field. In the real world, sytems do not satisfy our time-dependent Schrödinger equation indefinitely. The system interacts with the electromagnetic field, and in this way makes a transition to the ground state (eventually). This occurs even if no radiation field is present (spontaneous emission). These are the reasons that stationary states and in particular the ground state is so important.

Discussion of postulates using density matrices and projectors.

The formulation of quantum mechanics can be phrased a little more compactly and elegantly using density matrices and projectors. This avoids distinguishing between cases where eigenvalues are degenerate and also the overall phase of the wave function is irrelevant. Density matrices are used extensively in both statistical mechanics and in advanced treatments of spectroscopy. They generalize the wave function picture. We will not pursue this in great depth, but it is useful to be acquainted with it.

Degeneracies:

Let us denote by $\left|a_i,t\right\rangle \equiv \left|i,t\right\rangle$, $t = 1, n_i$ a set of orthonormal eigenvectors of the operator $\hat{A}$ that correspond to the *same* n_i-fold degenerate eigenvalue a_i that hence span the corresponding degenerate subspace. Then we can define the orthogonal projector on the eigenspace by $\hat{P}(a_i) = \sum\limits_{t=1}^{n_i} \left|i,t\right\rangle\left\langle i,t\right|$, with general matrix representation

$P_{pq}(a_i) = \sum\limits_{t=1}^{n_i} \left\langle p | i,t\right\rangle\left\langle i,t | q\right\rangle$. An orthogonal projector has the important properties $\hat{P}^\dagger = \hat{P}, \hat{P}^2 = \hat{P}$ (idempotency) as you can verify for yourself. The latter property is intuitively essential for a projector. If you project a vector on a subspace, the result lies in the subspace. If you then project again the same result is obtained. This is expressed by the idempotency condition $\hat{P}^2 = \hat{P}$. Moreover the only eigenvalues of $\hat{P}(a_i)$ are 0 or 1. Any vector (completely) within the subspace corresponds to eigenvalue 1, while vectors orthogonal to the subspace have eigenvalue 0. $\hat{P}(a_i)$ acts as the identity operator within the subspace. The operator $\hat{P}(a_i)$ is independent of the precise definition of the eigenvectors $\left|i,t\right\rangle$. Another orthonormal set of vectors $\left|i,x\right\rangle$ that span the subspace would do just as well: $\hat{P}(a_i) = \sum\limits_{x=1}^{n_i} \left|i,x\right\rangle\left\langle i,x\right|$ would give the same matrix repesentation $P_{pq}(a_i)$ (verify). Finally $\hat{P}(a_i)\hat{P}(a_j) = 0$, $a_i \neq a_j$, because the eigenvectors corresponding to different eigenvalues are orthogonal. We can hence write $\hat{P}(a_i)\hat{P}(a_j) = \hat{P}(a_i)\delta_{ij}$.

The operator $\hat{A}$ can be represented as $\hat{A} = \sum\limits_i a_i \hat{P}(a_i)$, from which follows immediately

$$\hat{A}^2 = \sum\limits_i a_i \hat{P}(a_i) \sum\limits_j a_j \hat{P}(a_j) = \sum\limits_{i,j} a_i a_j \hat{P}(a_i)\hat{P}(a_j) = \sum\limits_i a_i^2 \hat{P}(a_i)$$
$$f(\hat{A}) = \sum\limits_i f(a_i)\hat{P}(a_i)$$

The probability to measure an eigenvalue a_i in a state $\left|\Psi\right\rangle$ is given by

$\sum_{t=1}^{n_i} \langle \Psi | i,t \rangle \langle i,t | \Psi \rangle = \langle \Psi | \hat{P}(a_i) | \Psi \rangle$. Moreover the (unnormalized) state after the measurement is given by $\hat{P}(a_i) | \Psi \rangle = \sum_{t=1}^{n_i} | i,t \rangle \langle i,t | \Psi \rangle$. In short projectors are a convenient way to deal with degenerate states. Only the subspaces are relevant and this is precisely the focus of the projectors. Compared to the formalism discussed before, where all eigenstates were non-degenerate, we now have to include a sum over the set of degenerate eigenstates. Since degenerate eigenstates are not unique, it is more convenient (and insightful) to work with the projectors instead.

We can go one step further and associate a projector with the state $|\Psi\rangle$ itself. This is called the density operator and is denoted $\hat{D} = |\Psi\rangle\langle\Psi|$. In this case the density operator corresponds *to a pure state and is a projector*. This operator has one eigenvalue 1, with corresponding eigenstate $|\Psi\rangle$, while a state that is orthogonal to $|\Psi\rangle$ will be an eigenstate of $\hat{D}$ with eigenvalue 0. Moreover

$$Tr(\hat{D}) = \sum_p \langle p | \Psi \rangle \langle \Psi | p \rangle = \sum_p \langle \Psi | p \rangle \langle p | \Psi \rangle = \langle \Psi | \Psi \rangle = 1$$

The density $\hat{D}$ completely characterizes the system and is independent of the overall phase of $|\Psi\rangle$. You may have noticed before, that this overall phase does never play a role in the theory. It is only when different components in the wave function get different phases, that phase is important, as it truly changes the direction of the vector in Hilbert space. The probability to measure a_i on a system described by $\hat{D}$ is given by

$$p(a_i) = Tr(\hat{P}(a_i)\hat{D}) = \sum_p \sum_{t=1}^{n_i} \langle p | i,t \rangle \langle i,t | \Psi \rangle \langle \Psi | p \rangle =$$

$$= \sum_p \sum_{t=1}^{n_i} \langle \Psi | p \rangle \langle p | i,t \rangle \langle i,t | \Psi \rangle = \sum_{t=1}^{n_i} \langle \Psi | i,t \rangle \langle i,t | \Psi \rangle$$

which agrees with the postulates. The system after measurement of eigenvalue a_i (without normalization) would be given by

$$\hat{P}(a_i)\hat{D}\hat{P}(a_i) = \sum_{t,s} | i,t \rangle \langle i,t | \Psi \rangle \langle \Psi | i,s \rangle \langle i,s |$$

The density as given above is normalized to

$Tr(\hat{P}(a_i)\hat{D}P(a_i)) = Tr(P(a_i)\hat{P}(a_i)\hat{D}) = Tr(\hat{P}(a_i)\hat{D}) = p(a_i)$, where we use that the trace is invariant under cyclic permutations of operators, and the property of idempotency.

The *complete* ensemble after the measurement of $\hat{A}$ (at time t_a) can be represented as $\hat{D}(t_a) = \sum_i \hat{P}(a_i)\hat{D}\hat{P}(a_i)$ with normalization $\sum_i p(a_i) = 1$. This density is not idempotent in general and is not a projector. It would not correspond to a pure state but to a mixture $\hat{D}(t_a) = \sum_i p(a_i)|\Psi_i\rangle\langle\Psi_i|$. This will be the general situation after a measurement, if the system is not separated into individual sub ensembles. In this sense the density formulation goes beyond the conventional quantum mechanical formulation. In actual experiments, and in statistical mechanics we are almost always dealing with statistical mixtures. The key idea is that every subensemble, describe by one component of the density matrix can be treated as an individual quantum systems, and the resulting probabilities can be averaged in the classical sense. So the sum over the subsystems has the same role as in classical statistics, while within a sub-branch $|\Psi_i\rangle$ the laws of quantum mechanics would be fully applicable. The density formulation unifies these two descriptions, but we will not pursue this further at this point.

For completeness he time dependence of the density operator (in general, pure state *or* mixture) is given by $-i\hbar\frac{\partial\hat{D}}{\partial t} = [\hat{D}, \hat{H}]$. This is discussed on pages 295-307 of Cohen-Tannoudji

Let us finally consider two Hermitian operators $\hat{A}$ and $\hat{B}$ having the respective eigenspace projectors $\hat{P}(a_i)$ and $\hat{P}(b_j)$. If $\hat{A}$ and $\hat{B}$ commute they have a complete set of common eigenvectors. It can be shown that in this case the projectors on the respective eigenspaces commute $\left[\hat{P}(a_i), \hat{P}(b_j)\right] = 0$ $\forall i, j$. The proof runs as follows. Let

$$\hat{A} = \sum_i a_i \hat{P}(a_i); \quad \hat{B} = \sum_j b_j \hat{P}(b_j)$$

and

$$\left[\hat{A},\hat{B}\right]=\sum_{i,j}a_{i}b_{j}\left[\hat{P}(a_{i}),\hat{P}(b_{j})\right]=0$$

Each individual term in the sum should equal zero as the operator parts are independent (projectors on different subspaces). Therefore either $a_{i}=0$, $b_{j}=0$ or the individual projectors commute. The special cases require some extra work. Since all projectors *not* corresponding to zero eigenvalues necessarily commute, we know that

$$\left[\sum_{a_{i}\neq0}\hat{P}(a_{i}),\hat{B}\right]=0=\left[(\hat{1}-\hat{P}(a_{0}=0)),\hat{B}\right]=-\left[\hat{P}(a_{0}=0),\sum_{j}b_{j}\hat{P}(b_{j})\right]=0$$

Hence the "null-projector" for $\hat{A}$ commutes with all non-null-projectors for $\hat{B}$ and therefore also with $(\hat{1}-\sum_{b_{j}\neq0}\hat{P}(b_{j}))=\hat{P}(b_{j}=0)$, which completes the proof. This result is completely equivalent to the statement that $\hat{A}$ and $\hat{B}$ have a complete set of common eigenfunctions. The projectors $\hat{P}(a_{i})\hat{P}(b_{j})$ would project on the subspace spanned by those eigenvectors $|a_{i},b_{j},t\rangle$ that all have the same eigenvalues a_{i} and b_{j}.

Tale on measuring non-commuting variables.

The most famous example of two non-commuting observables are position and momentum. The properties of these operators are a little complicated because their spectra are continous. It is easier to consider the case of measuring angular momentum or even better the spin of an S=1/2 system. The three cartesian components of $\hat{S}$ do not commute and we have the commutation relations $\left[\hat{S}_{x},\hat{S}_{y}\right]=i\hbar\hat{S}_{z}$. However we can very well measure any of these individual quantities and we can also perform a sequence of measurements and analyse the results. In the absence of magnetic interactions in the hamiltonian the resulting state vectors after the measurement are independent of time, which is another simplification. In fact to discuss the results of quantum mechanics let us not use any mathematics at all. Let us analyse the real content first and then venture into mathematical formulations.

As our ensemble we take a class of schoolkids. Each of these kids has a lunchpacket that consists of three items. They all have a turkey or roast beef sandwich (t or r), a coke or a sprite to drink (c or s) and an apple or an orange (a or o) for desert. Our measurement consists of asking a kid what is in the lunch bag, and getting statistics on the ensemble (the class). However, we can ask only one question at a time. For example: "everybody with a turkey sandwich stand to the right". But not: "All that have an orange and a coke please stand on the left". That is asking two questions at once, and in the anology with the spin system reflects the impossibility to simultaneously measure non-commuting observables. In fact any 'measurement' we do should obey the laws of quantum mechanics. Our goal is to characterize the distribution of lunch bags (e.g how many $tca, rca, tsa, rco,...$) etc, are there. Can we do this? If things behaved classically, easily. But not in the quantum world. Let us try. We would first ask all kids who has turkey and who has roast beef, and partition them into two groups. Then we would ask the turkey group who has a coke and set them apart. Fine. we already have an ensemble that has both a turkey and a coke, right? Let us check, and ask again. Who has a coke? Everybody has a coke. Now, who has a turkey sandwich? Oops. This doesn't work. Only about half of them has turkey. Asking the coke question destroyed the information we had on the turkey. In the quantum world it is impossible to isolate a group where everybody has both a coke and turkey. Asking the question changes the ensemble. This is fairly easy to understand mathematically, describing an ensemble as a vector in Hilbert space, that rotates under measurement, but it certainly does not make much sense when asking about lunch bags.

The above is a representation in as simple a language as possible of some puzzling properties of quantum mechanics. The essence is that according to quantum mechanics (sometimes) we cannot create an ensemble that for sure will yield definite values for two non-commuting observables. This is the content of the Heisenberg uncertainty principle. The precise formulation would be

$$\Delta A \, \Delta B \geq \frac{1}{2} \left| \left\langle \left[\hat{A}, \hat{B} \right] \right\rangle \right|.$$

For a proof and discussion see Cohen-Tannoudhji pages 286-289.

It is often stated as "one cannot measure the precise value of $\hat{A}$ and $\hat{B}$ simultaneously". This is a very incomplete statement of the principle and it has led to all kinds of ingenious constructions to violate the principle. It is much easier and complete to interpret the principle in a different way. There is no problem to measure $\hat{A}$ or $\hat{B}$, and for each measurement (either $\hat{A}$ or $\hat{B}$, but not both) on an individual system you get definite results. However, for certain pairs of eigenvalues of $\hat{A}$ and $\hat{B}$, (a_i, b_j) say, it is in principle impossible (according to QM) to prepare an ensemble such that all of the measurements on this ensemble yield precisely the result a_i if you measure $\hat{A}$ and b_j if you would measure $\hat{B}$. In contrast there is no problem in preparing an ensemble such that every member would yield a_i if you measure $\hat{A}$. You might put in some effort to appreciate the precise translation of the mathematical formulation of the uncertainty principle into words. It is a little easier if the commutator $\left[\hat{A}, \hat{B}\right]$ is a constant, since then *no* ensemble will yield the same value for A and B for all elements in the ensemble. So necessarily there is a spread, and the minimum spread depends on the commutator. In the general formulation the mimimum spread depends on an expectation value and hence on the state under consideration. Note that quantum mechanics actually does not preclude that individual systems have definite values for all observables. It does say that within the realm of quantum mechanics you cannot create an ensemble to prove it. Also note that it is impossible to discuss the uncertainty principle using a single system. It is perfectly possible to have an experiment where you measure $\hat{A}$ then $\hat{B}$ then $\hat{A}$ then $\hat{B}$ and find nothing weird: measurement of $\hat{A}$ yields a_i twice, while the measurement of $\hat{B}$ yields b_j in both cases. This is quite a possible outcome of this experiment. But beforehand you cannot be certain that it will happen that way. It is impossible to create an ensemble where all elements necessarily behave in this fashion. Of course you might be lucky and by chance, using small enough ensembles one can easily violate the Heisenberg uncertainty principle. That is all part of statistics.

Let us discuss another hair raising situation. A long standing controversy is the so-called Einstein-Podolski-Rosen Paradox (EPR). EPR sought for the properties of *individual*

systems obeying the laws of quantum mechanics. In essence all parties can agree on the fact that a measurement can change the system. So in the example above if I ask Mary if she has a coke, afterwards she might no longer have the roast beef sandwich that she started out with. However, the issue at hand is different. EPR thought it would be possible that each lunch bag has a definite content before measurement, and we are simply looking what is in it. By looking at one piece of information we might, in the convoluted act of measurement, change another piece of information in ways that are hard to predict. This would then be the reason that one cannot prepare well specified ensembles, which are themselves prepared by measurements. It might be that we simply have too little control over the act of measurement (at present ?). Quantum aficionados tend to think differently about what happens during a measurement on an individual system. Their idea is that by measuring you force the microsystem to take a position. It is like flipping a coin at the moment of measurement. "Choose my dear electron! Up or Down?" By the act of measurement you force the system into an eigenstate of the corresponding observable, and it does so with probabilities predicted by quantum mechanics. The precise outcome of an individual experiment is unpredictable in principle. If one reads initial accounts of the Heisenberg uncertainty principle however, they very much reflect the viewpoint of EPR. Heisenberg himself for example discusses how measuring position necessarily changes the momentum of an electron. The later accepted viewpoint according to the so-called Kopenhagen interpretation is rather convoluted in that they use classical mechanics to describe the measuring apparatus and so there is a mysterious connection between the quantum and classical system. However, I think that the above stated position of the quantum aficionado reflects the attitude of many scientists in the field. It was my position until I wrote these notes.

Let us adjust our lunch bag parabel a little so that we can describe the EPR line of thought in trivial terms. What if we could gain information about what is in a lunch bag without asking a question? Let us set up the experiment in a tricky way. Say we know that the lunch bags are handed out in complementary pairs. Each pair contains both turkey and roast beef, an apple and an orange, a coke and a sprite. So a pair of lunch bags might consist of *tca*&*rso* or *rso*&*tca* and so forth. We look when the lunch bags are

handed out and keep track of the corresponding pairing of the kids. The actual quantum experiment consists for example of two spin 1/2 atoms in an overall S=0 state. Back to the kids. Let's say, Lois and Clark form a pair. Now we take Clark out on the playground and ask him about his sandwich. "Turkey he says. I would like salami!" Lois doesn't even know we asked, but we now know that she has roast beef without asking her (*or perturbing her lunch bag*). If we would ask her she would say roast beef 100% of the time. However, we don't need to ask her about her sandwich as we know already. Instead we ask Lois about her drink. "I have a coke she says". After we ask the coke question she might no longer have roast beef, but if we assume she has something definite in her lunch bag, *before* the coke question it was most definitely roast beef and a coke. So this is a smart measurement that shows it makes perfect sense that every lunch bag has something definite in it and by measuring we simply find out what it is. Only, by asking one specific question we might change the content of the lunch bag in other respects, and in unpredictable ways. At the time EPR wrote their paper this interpretation was in no conflict with any piece of data whatsoever. It was just an interpretation that should have appealed as something far more rational than flipping a coin at the time of measurement. If we take the alternative quantum interpretation about what actually happens, the EPR experiment is seen to take on all of its weirdness. Asking Clark what is in his lunch packet forces him to take a position. Clark flips a coin to make a decision. "Turkey". If we now would ask Lois about her sandwich she will say roast beef for sure. So she flips her coin too, but it always yields the same result. If we wouldn't have asked Clark it would give a fifty-fifty result, but now it yields a 100%. Now Lois nor her coin knows anything about our asking Clark. To put it in the extreme: flipping a coin in Tokyo determines the outcome of the flipping of the coin in New York. That doesn't make sense. The EPR interpretation is far more reasonable: if we assume there is something definite in the lunch bag, there is nothing strange about us knowing what is in Lois's lunch bag if we know what Clark has, given they form a perfect pair.

However, EPR did something more. They claimed that physical theories should describe 'reality', which means that quantum mechanics should allow for ensembles of completely specified lunch bags. This it did not, and therefore the theory was not quite up to par.

Quantum theory was incomplete. In order to describe ensembles of well specified $\hat{S}_x, \hat{S}_y, \hat{S}_z$ the structure of the theory needs to be changed completely. If we use the concept of Hilbert space, operators and eigenvalues it can not accommodate EPR's reality. Quantum theory was too successful to discard it, just because of a difference in interpretation that had apparently no measurable consequences: Is there something definite in the lunch bag before you look at it?

This was the situation until John Bell came around. He showed that the EPR interpretation might lead to different results from the usual quantum theory for some experiments. And he used the EPR experiment to show it. This is how it works in terms of lunch bags. If EPR's postion is right then in fact I can construct what was in Lois's lunchpacket from the pairing experiment. From Clark's answer I know she had roast beef, and by our question we also know she has a coke. We are simply assuming that the question to Clark could not possibly have affected Lois's lunch box. There is no unpredictable act of measurement that has a range from New York to Tokyo. Let us assume therefore for the sake of argument that EPR are right. Every lunchbox has a definite content and by doing the pairing experiment I can determine two items in a luchbox. Now we take our whole class and do three types of experiment starting from identical ensembles in each experiment. In the first experiment we use the pairing experiment to determine if somebody has a turkey sandwich and a coke. By assumption she would then have either an orange or an apple as the third item. If we do this for the whole first ensemble we can write

$$n[t,c] = n[t,c,o] + n[t,c,a]$$

where $n[t,c]$ denotes the number of kids in the enesemble that have both a turkey and a coke, and so forth. In the next group we determine the number that has a sprite and an orange, in the third group turkey and orange. In total we would then have the following relations, assuming the minimal EPR conditions

$$n[t,c] = n[t,c,o] + n[t,c,a]$$
$$n[s,o] = n[t,s,o] + n[r,s,o]$$
$$n[t,o] = n[t,c,o] + n[t,s,o]$$

From this we can derive the so-called Bell inequality:

$$n[t,c] + n[s,o] \geq n[t,c,o] + n[t,s,o] = n[t,o]$$

This is an inequality that one can test in an actual experiment, as one can make a spin zero pair, let it fly apart and measure the spin in different directions for the particle in Tokyo and the particle in New York. We will discuss the full details of the precise quantum treatment later on. But the outcome is that the usual treatment of quantum mechanics is in conflict with the above analysis based on the assumptions of EPR. Quantum mechanics violates Bell's inequalities. At the time of the EPR paper (1935) people couln't really say if EPR was right or the standard Kopenhagen interpretation was right. Neither did contradict any experiment. It appeared simply a matter of interpretation. With Bell however, there was a testable hypothesis. Experiments were done in the seventies, and the experiments by Alain Aspect are perhaps best known (though not the first). The technical details and fine print are rather involved, but the conclusion was that the traditional laws of quantum mechanics are correct. So one cannot assume that individual particles actually have definite values for $\hat{S}_x, S_y, S_z$ and we are simply determining what they are, although perturbing these values in the process.

Does this mean that we have to accept the alternative interpretation? Flipping a coin in Tokyo determines the outcome of the flipping of a coin in New York? Not in my opinion. This 'making a choice during the measurement' aspect appears to be an act of human imagination. Us trying to understand what we cannot grasp. I think it is better to take a very mundane position. Quantum mechanics describes the statistical outcomes of complete experiments. In doing the measurement in Tokyo I am preparing a specific ensemble. The subsequent measurement in New York is described using this new ensemble. From the perspective of quantum mechanics the pairing experiment is no different from first asking Lois if she has roast beef and then asking if she has a coke. A measurement is a measurement, and if a measurement in Tokyo tells you something about the situation in New York, you have to adjust the ensemble accordingly. Of course this is nothing more than using the laws of quantum mechanics which are very definite for this type of experiment. What is hard to understand is how there can be such a strong correlation between two distant particles, which cannot be assumed to individually have definite properties, while as a pair they do. Quantum mechanics gives us the

mathematical prescription but it does go against common sense, and this is illustrated very vividly by the flipping the coin at the time of measurement picture. This phenomenon is called entanglement in the literature. I have a set of excersises that has you work out the quantum mechanics of Bell's experiment, if you are interested. It is a fair amount of work, but you have all of the back ground to do this.

Further reading and references:

MS: McQuarry and Simon: Physical Chemistry, a molecular approach, University Science books, ISBN 0935702-99-7. This contains an introduction to quantum mechanics, and it has a number of math chapters that give a quick overview of some needed material.

Levine: Ira N. Levine, Quantum Chemistry, fifth edition, Prentice Hall, ISBN0-13-685512-1, Chapters 3 and 7.

C-T: Claude Cohen-Tannoudji, Bernard Liu and Franck Laloë, Quantum Mechanics, volumes I and II. Wiley, This is a bible of quantum mechanics. It contains a wealth of material. It is an advanced textbook. ISBN 0-471-16433-X