The Quantum Theory of Measurement

Verily, there are just two states	of this Person,
The state of being	in this World
And the state of being	in the other World.
There is a third state —	and that is of Dreams.
By standing in this state	one sees both of those states,
Of being in this World	and of being in the other World.

Br'hadāranyaka, **IV**.3.9

We have seen how quantum mechanics grew out of a jumble of ideas about matter waves, wave-particle duality, wave-packets and probability amplitudes, together with a new differential equation — that of Schrödinger — leading to a calculus of the microworld. It took a few more years for these ideas to crystallise and produce a set of postulates, on the basis of which quantum mechanics has rested ever since. Most of this work of interpretation arose out of the work of Niels Bohr and his collaborators — chiefly Werner Heisenberg and Wolfgang Pauli — at the Institute for Theoretical Physics[1] in Copenhagen, and hence has been dubbed the **Copenhagen interpretation**. In this lecture, we shall try to describe the main ideas which make up this interpretation.

The fundamental idea in quantum mechanics, as in relativity, is that it is only *measurable* quantities that have scientific validity. The only quantities for which the laws of physics are applicable are such measurable quantities, called **observables**. The sole purpose of empirical science, in this limited view, is to use known observables to predict other observables, which are known or will subsequently be known[2]. In order to do this, scientists frequently have resort to a 'model', which is a collection of ideas (often non-testable) useful to form a link between the already-known observables and the ones to be predicted. Theoretical physics constitutes the building and analysis of such models.

What is the main difference between the laws of the microworld and the laws of the macroworld? It arises from the fact that our knowledge of both arises from the *probes* we use to study it. Typically the probes we use are light waves or photons, which scatter off the object being

[1]now the Niels Bohr Institute.

[2]In doing so, empirical science closely follows the epistemological ideas originally due to George Berkeley in the eighteenth century. According to Berkeley, for example, a tree existed only insofar as it was seen by men to exist. Of course, to the criticism that in a Universe populated by one man and a tree, the tree would cease to exist every time the man blinked, and then be re-created when he opened his eyes, Berkeley countered that even when the man was blinking, the tree was watched by God, and hence it existed. Hence, also, God existed. (Incidentally, Berkeley was a clergyman).

probed and are recorded by the eyes or a camera or some suitable apparatus. We can also use electrons or neutrons or α-particles or, in fact, any small object as a probe. Now when a photon, for example, scatters off a macroscopic object — like a cricket ball — it hardly disturbs the motion of the cricket ball. For a 163 gm ball moving at 50 Kmph, the momentum transfer in head-on elastic scattering by a γ-ray photon of wavelength 0.1 Å is around $\frac{1}{100\ 000\ 000\ 000\ 000\ 000\ 000\ 000\ 000}$ of the ball's momentum. Thus, we can know the position of the electron within an accuracy of the order of 0.1 Å, and its momentum is practically unaffected. However, when the same photon hits an electron, travelling at 1% of the velocity of light, the momentum transfer is nearly four times the electron's momentum! Clearly, any information obtained from this scattering is going to be erroneous. In fact, though we do know the position of the electron within an accuracy of the order of 0.1 Å, the γ photon has given the electron a sharp smack, so that we do not know where it has gone the next instant. Obviously the solution lies in using a smaller — or less energetic — probe. Let us reduce the photon energy by an order of magnitude, which means the wavelength will go up to 1 Å. We now give a much smaller impetus to the electron, about 40% of its momentum. However, we have paid a heavy price: we no longer know where it is within a region of dimension about 1 Å across (comparable to the size of a hydrogen atom). Decrease the momentum transfer still further, and the region where we know the electron to be will become even larger[3]

This argument, known in textbooks as Heisenberg's **gamma-ray microscope**, is thus an epistemological justification for the famous position-momentum **Uncertainty Principle** of Heisenberg, usually written in the compact form

$$\Delta x \ \Delta p > \hbar \ .$$

This tells us that there is a deep-rooted fuzziness in our knowledge of the microworld — accurate positions are accompanied by a hard impetus to the measured object; on the other hand, a soft impetus means a sacrifice of accuracy in position. This is because, in the microworld, all available probes are big enough to *disturb* the system. We must note that similar ideas were already present, to some extent, in classical physics. For example, in electrostatics, we define a *test charge* q as a charge which is introduced into the system, but is too small to disturb it — which we achieve by taking the limit of vanishing charge $q \to 0$. Similarly a gravitational potential will be defined in terms of a test mass $m \to 0$ of infinitesimal proportions. So long as the probed object is large enough to neglect the disturbing action of the probe, it is meaningful to apply this limit. Quantum mechanics simply denies that we can take such probes to the vanishing limit when dealing with the

[3]The wave-particle duality discovered by de Broglie establishes the fact that this fuzziness is a universal phenomenon and not something that has to do with photon probes.

microworld. It is, therefore, a subject which has developed out of the discovery of the fact that elementary masses and charges are small but finite. Accordingly, we have to develop a science of measurement for a bull-in-a-china-shop world, where the observer changes and disrupts what he observes. An analogous situation would be if a zoologist tries to study a jellyfish which lives in deep waters, which dies and bursts when it is trawled and brought to the surface. The reader will agree that the task is a difficult one.

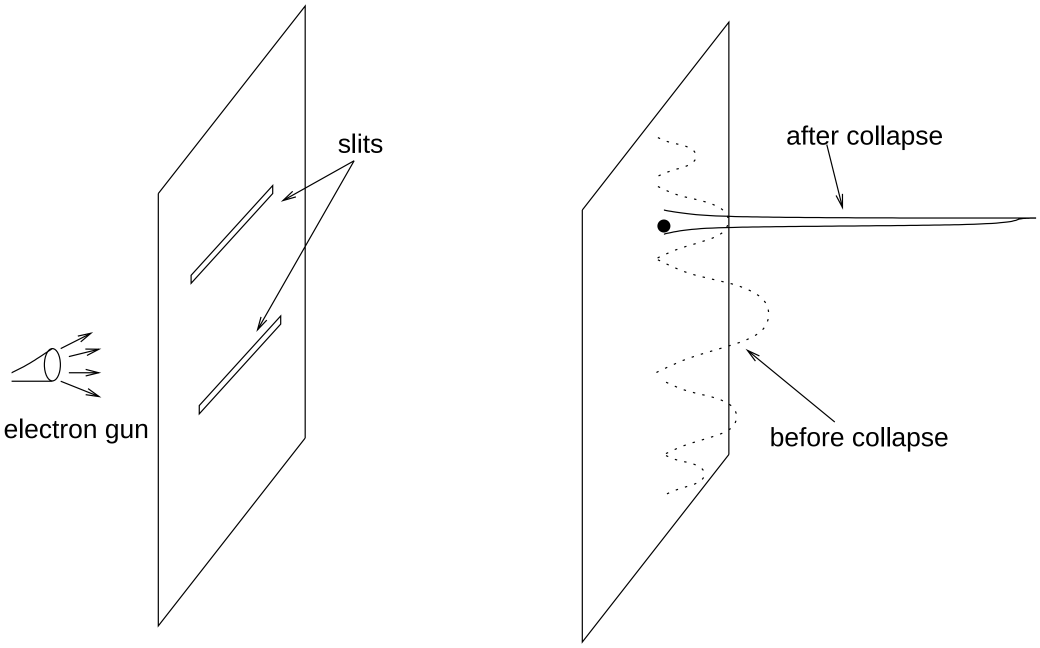

Figure 1. *Illustrating the double slit experiment with electrons.*

It is, in fact, amazing that in such a world, it is at all possible to construct a theory and predict anything. However, that is just what quantum mechanics does. It does this by adopting a pragmatic stance. Let us take an example from daily life. If your friend tells you that that he may or may not go to the cinema with you tomorrow, you will probably not try to convert this uncertainty into a certainty. What you would most likely do is immediately ask what are the chances that he may go with you. He may say 50-50 or 80% or maybe 99.9%. In other words, where there is uncertainty or fuzziness of information, the only question that is meaningful to ask is that of *probability*. It is useless to ask exactly where an electron is, because we can never know a position *and* keep the electron there. But it is relevant to demarcate a fuzzy region of some finite size and ask how probable it is that the electron resides in it somewhere, accepting also that its momentum is somewhere between known limits.

It was first realised by Max Born that the *wavefunction* associated with a particle by de Broglie and by Schrödinger tells us just such a probability. This wavefunction is a complex number, *i.e.* a pair of real numbers. One real number — the modulus — is a measure of the probability and the other real number — the phase — is a determinant of its time evolution. In fact, the wavefunction contains within it all the information about the particle which we will ever need. Even in classical mechanics, if we know where a particle is and where it will be after a given time, we can obtain all other information about the particle. The situation is the same in quantum mechanics, except that the Heisenberg fuzziness forces us to use a wavefunction instead of a point position.

The Schrödinger equation tells us how this wavefunction evolves with time. Since the wavefunction is what has replaced the position of a particle in classical mechanics, it is clear that Schrödinger's equation has replaced Newton's equation of motion in governing the time-evolution of a system. For this reason it is often called the **Schrödinger equation of motion**. Since the microworld has its own mechanics, it is natural that it should have its own equation of motion. However, as Ehrenfest showed, and as we shall presently establish, the Schrödinger equation is compatible with Newton's equation of motion when certain average motions are considered.

It is of paramount importance to note that a wavefunction is not a measurable quantity. One cannot measure a probability — the very act of measurement will force the probability to change. Suppose, to continue the same analogy, your friend does turn up the next day to go to the cinema. His intentions are now 'measurable' and the probability has jumped up to 100%. If he does not come, then it has dropped to 0%. These are the only measurables. Similar is the case when an electron passes through a double slit, falls on a zinc-sulphide screen and produces a scintillation. In the instant before it has fallen on the screen, all we can tell is that there is some probability for the electron to fall in various places on the screen. This will be more where the probability waves interfere constructively (bright fringes) and less where they interfere destructively (dark fringes). However, the moment the electron does fall on the screen and a scintillation is seen by the watching scientist, its probability has become a delta function at that point on the screen — for we know that it is precisely there and cannot be anywhere else. Thus, in a single instant, the spread-out wavefunction has been converted to a delta function. Technically, this is called *collapse* of the wave function, and this idea is known as the **Collapse Hypothesis**. As soon as we know that the electron has caused a scintillation on the screen, we can compute its trajectory and work out, for example, which slit it came through. However, till the collapse happened, we did not know this: the electron came through *both* the slits! To remove possible effects of electrons interacting in a bunch, it is possible to do the experiment with one electron at a tine, i.e. ensuring that only after one electron has hit the screen does the next electron get emitted from the source. The result is the same: each electron gives a point scintillation, but as these grow in number, the fringe pattern begins to show up. Thus, clearly, the electron is a point particle riding a wave of probability. This wave goes through both the slits, forms an interference pattern and guides the individual electron to a point on the screen. We can never predict, á priori, where a lone electron will go, which slit it passed through, and so on. But we *can* predict what a large ensemble of electrons will do.

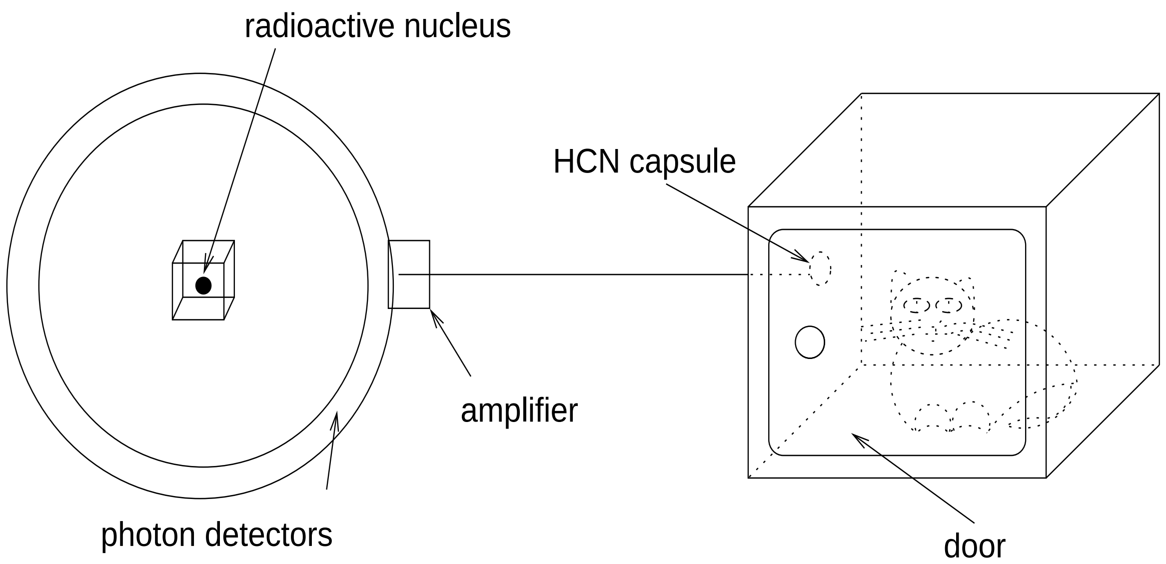

Figure 2. *Illustrating the paradox of Schrödinger's cat.*

Now, it may be argued that the collapse is a result of the action of the observer in placing a screen to intercept the motion of the electron and hence is a cause of disturbance of the system. For if we move the screen back a little, then at the point where the collapse occurred earlier, there will be a probability wavefunctions carrying the typical fringe pattern. We can thus argue that it is the action of the observer in seeing the scintillation that made the electron fall on a particular point and fixed its trajectory[4] The Copenhagen interpretation has no quarrel with this, but Schrödinger was distressed by the idea. In fact he constructed the following conundrum which takes this idea to an absurd limit.

Consider a box with a radioactive nucleus $\mathcal{N}$ inside it. You know that this will undergo a gamma decay $\mathcal{N}^* \to \mathcal{N} + \gamma$ with a half-life $T_{1/2}$. This box is surrounded with detectors, none touching the box. The moment the nucleus decays, the emitted gamma photon penetrates the walls and sets up an impulse in the detector. This is amplified and the cor-

[4] This, of course, is a completely Berkelian point of view.

responding electrical signal is used to break open a capsule containing cyanide gas inside a closed opaque box containing a live cat. Thus, the moment the nucleus decays (up to the time taken for the signal to travel down the wires, which is at the speed of light), the cat dies. Till this happens it is alive. Now wait for a time equal to exactly one half-life $T_{1/2}$. At this time it is equally probably for the decay to have happened or not to have happened, and hence it is equally probable for the cat to be dead or alive. The 'wavefunction' of the cat corresponds to a half-alive—half-dead state! Now open the box (carefully!). Either a live cat will jump out of it, or you will see a dead cat lying inside. Opening the box is a measurement. It has caused the cat to pass from half-alive to alive or from half-dead to dead. If this seems absurd, what is worse is that it has caused the nuclear state to jump to a decayed or intact state, though the box containing the cat and the box containing the nucleus are not physically connected in any way. Somehow, the information has jumped up the causality ladder and reached the nucleus across the intervening space. **Schrödinger's cat** problem, as it is called, has given rise to many a lively debate among proponents of the Copenhagen interpretation, who maintain that this is perfectly possible, and others (notably Einstein) that this sort of 'spooky' action-at-a-distance is not possible.

Is such a collapse consistent with special relativity? Consider the double slit experiment. If the fringe pattern has shrunk in an instant to a single point, the information must have propagated along the pattern with infinite velocity. One can get around this by saying that wavefunctions and trajectories are not measurable, but a *gedanken* experiment[5] concocted by Einstein, Podolsky and Rosen (EPR) translates this in terms of measurable quantities. The idea is extremely simple. Suppose a particle with zero angular momentum decays into two photons of opposite polarisation, moving in opposite directions. Consider the right-moving photon and let it hit a detector after time t. Assume the detector can measure its polarisation. Just before it hit the detector, the probability of the right-moving photon having either polarisation was 50-50. However, the moment it hit the detector, one polarisation jumped to 100% while the other dropped to zero. That this collapse is a consequence of the measurement is something which we immediately realise. However, the crucial point is that the moment we know the polarisation of the right-moving photon, we also know the (opposite) polarisation of the other photon, which, by now, has moved far away. The information about polarisation of the far electron has jumped instantly across the intervening distance $2ct$ and caused the wavefunction of the far electron to collapse. Since polarisation

[5]Literally 'thought' experiment, which means that it is a hypothetical experiment conducted in ideal circumstances. Such 'thought' experiments are liberally used in classical physics too, notably in thermodynamics, *e.g.* the Carnot engine. The first 'thought' experiment was probably introduced by Galileo who wanted to illustrate what is known as the Galilean principle of relativity.

is certainly a measurable quantity, information has propagated *instantly*.

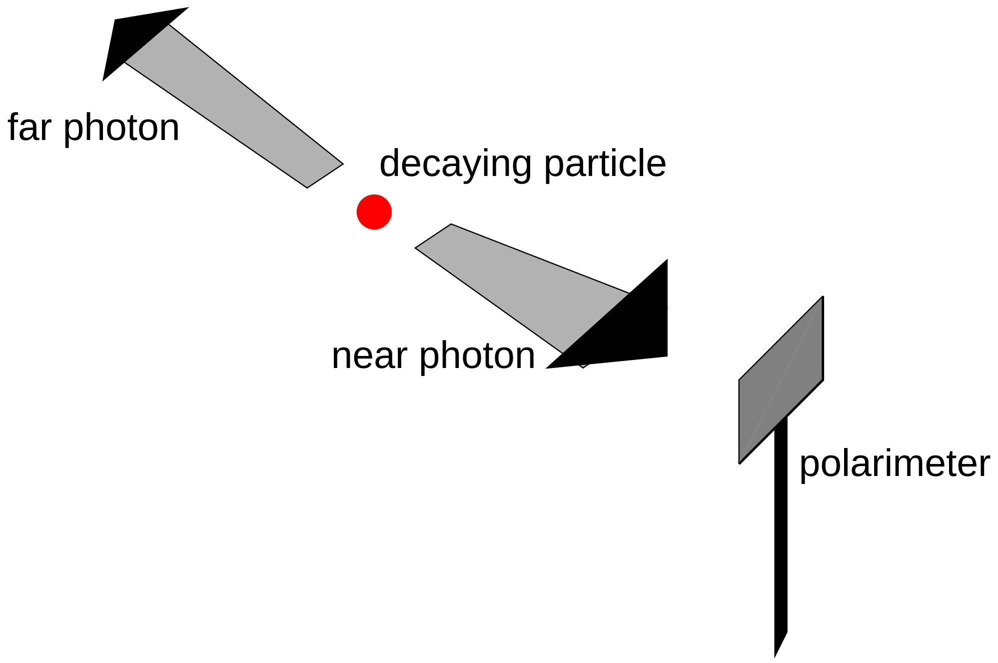

Figure 3. *Illustrating the EPR paradox.*

The **EPR paradox**, as it is usually called, is a paradox only when a single pair of photons is considered. If it is an ensemble of such pairs, then there will be equal numbers of polarisations on either side, and relativity is saved. For no more information comes out after the measurement(s) than was available before. This led Einstein to propose the so-called **statistical interpretation** of quantum mechanics. According to this, quantum mechanics is only valid for large numbers of microsystems, not for individual ones. The laws governing these individual systems are unknown and waiting to be discovered. Just as probabilities in statistical mechanics represent our ignorance of the actual particle trajectories, so Einstein supposed microsystems to be governed by a new set of variables — *hidden variables* — which average out to give the same quantities calculated using the probabilities. However, the 'true' laws of the microsystems should be deterministic. According to Einstein, he could not believe that God participated in 'that game of dice', proposed by the Copenhagen physicists.

For a majority of systems, such as atoms and solids, it was not possible to isolate microsystems in the days of Bohr, Born and Einstein. Thus, the statistical interpretation and the Copenhagen interpretation worked equally well and soon the majority of scientists lost interest in the philosophical argument and concentrated on applications instead. Only in the 1990's has it been possible to test the EPR paradox experimentally for a single pair of photons. This follows from the development of delicate instruments which can measure effects induced by single photons. The victory is for the Copenhagen interpretation. It has also been possible to do Schrödinger cat-like experiments (without involving real cats!), and once again, it is found that the 'spooky' action-at-a-distance does indeed occur. This is called *violation of microcausality*. We recall that the restriction on information travelling faster than light arose from the weak principle of causality[6]. In the quantum world, then, it is possible for the effect to precede the cause, which is another way of saying that time can run backwards.

This may seem crazy at first, but a little reflection will show that there is no real cause for distress. Causality, or the 'arrow of time' as it is often called, is not a microscopic principle. The microscopic equation of motion, be it Newton's equation of motion, or Schrödinger's equation or the relativistic Klein-Gordon and Dirac equations, is always invariant under time reversal $t \to -t$. The irreversibility of time evolution is related to uni-directional increase in entropy, which is only true of ensembles[7] and is enshrined in the Second Law of Thermodynamics. It follows that causality is a statistical principle, valid only for an ensemble of microsystems. Now we note that all violations of microcausality are washed out when an ensemble is taken. Thus, paradoxes associated with collapse are not really paradoxes, though they may seem bizarre.

Bizarre effects like the EPR paradox and Schrödinger's cat are fascinating, but they do not form the real meat of quantum mechanics. Let us come back to our original objective, which was to obtain meaningful information from a system in the case when the only probes we can use will end up in changing the state of the system. We recognise that a measurement on a system changes the original wavefunction into a new wavefunction. Ergo, the action of a measurement on a system must be mirrored by that of an *operator* on the wavefunction. For an operator is just that which takes one wavefunction to another wavefunction. Thanks to Schrödinger, we have already stumbled on the differential operators for position and momentum. Now if an operator/measurement is going to change the wavefunction, what will be the result of the measurement? We do not know the answer in the general case. Fortunately, there exist special wavefunctions which remain unchanged by differential operators. These are called *eigenfunctions*. If $\hat{\mathcal{O}}$ be a differential operator (we use the $\hat{}$ symbol to denote an operator as opposed to a simple number), then an eigenfunction will satisfy the equation

$$\hat{\mathcal{O}}\ \phi_\lambda(\vec{x}) = \lambda\ \phi_\lambda(\vec{x})$$

[6]The *weak* principle of causality states that the cause must precede the effect; the *strong* principle of causality states that the same cause must always produce the same effect. Both are violated in quantum mechanics, at the single-particle level.

[7]It can be ultimately traced to the expansion of the Universe.

where λ is, in general, a complex constant. In other words, the action of the operator has been to *scale* the function, rather than to change it. Now, if a system is in a state described by a wavefunction which is described by an eigenfunction — we call it an *eigenstate* — then the measurement can be repeated many times without ever changing the wavefunction, *i.e.* the state. Mathematically,

$$\hat{\mathcal{O}}^n \ \phi_\lambda(\vec{x}) = \lambda^n \ \phi_\lambda(\vec{x})$$

for any power n.

What result does the measurement on an eigenstate give? We note that the scale factor λ — known as the *eigenvalue* — has the same dimension as the operator, and hence it is reasonable to assume that it must be the outcome of the measurement. This is what was meant by the 'operator equivalence' used by Schrödinger. We immediately run into a problem: if λ is complex, we have *two* real numbers corresponding to the result of a measurement. We are thus forced to take only a certain class of operators, namely, those which have real eigenvalues. Such operators do exist, and they are called *Hermitian operators*. The defining property of a Hermitian operator $\hat{\mathcal{O}}$ is

$$\int d^3\vec{x} \ \psi^*(\vec{x})[\mathcal{O}\psi(\vec{x})] = \int d^3\vec{x} \ [\mathcal{O}\psi^*(\vec{x})]\psi(\vec{x}) \ ,$$

and it is straightforward to verify that the Schrödinger operators for position and momentum are Hermitian. An interesting and useful property of Hermitian operators is that any two different eigenfunctions can be made to satisfy the condition

$$\int d^3\vec{x} \ \phi_i^*(\vec{x}) \ \phi_j(\vec{x}) = \delta_{ij}$$

which is known as the *orthonormality condition.*

Let us go back for a moment to the double slit experiment. Once the measurement has been done, the electron has stuck to the screen somewhere. Every time a measurement is now made, no further collapse will occur. Each measurement will simply show the electron to be at that spot. This means that every time we try to measure position, we will leave the collapsed wavefunction $\delta(\vec{x} - \vec{x}_0)$ unchanged, and pull out the quantity $\vec{x}_0$ as the result of the measurement. Clearly, this has to be an eigenstate! We thus realise that when the particle is not initially in an eigenstate, the process of measurement *forces* the wavefunction to collapse into an eigenfunction and then the result of the measurement is the corresponding eigenvalue. To get back to the analogy of going to the cinema, it is as if you press your friend to give you a definite answer immediately, and he finally says either yes or no. Once his word is given, he repeats yes or no every time you ask.

Very well, we have argued that a measurement causes the wavefunction to collapse into an eigenfunction. But to which eigenfunction, for there is usually more than one? We cannot tell this á priori, just as you cannot tell á priori if your friend is going to say yes or no. But if he was going to go out with you with 80% probability, it is surely 80% probable that he will say yes and 20% probable that he will say no. The situation in quantum mechanics is rather similar. There is no way to tell á priori which eigenstate the system will collapse to, but there is a fixed probability for each particular eigenstate to be chosen. If the measurement is performed on each of an ensemble of identically-prepared systems, the number of times the collapse will occur to a particular eigenstate should be proportional to the corresponding probability. To fix our ideas, let us concentrate on a dynamical variable $\mathcal{O}$, to which corresponds the Hermitian operator $\hat{\mathcal{O}}$, whose eigenfunctions satisfy the equation

$$\hat{\mathcal{O}}\phi_i(\vec{x}) = \lambda_i \ \phi_i(\vec{x}) \quad i = 1, 2, \ldots$$

Now, if the (fractional) probability that a wavefunction $\psi(\vec{x})$ will collapse on measurement of $\mathcal{O}$ to the eigenfunction $\phi_i(\vec{x})$ is given by Π_i, then the number of times the corresponding eigenvalue appears as the result of the measurement is given by $N_i = N\Pi_i$, where N is the total number[8] of measurements made on an ensemble of identically-prepared systems. The *average value* of the dynamical variable computed over the ensemble is simply

$$\langle\lambda\rangle = \frac{\sum_i N_i \ \lambda_i}{\sum_i N_i} = \frac{\sum_i \Pi_i \ \lambda_i}{\sum_i \Pi_i} = \sum_i \Pi_i \ \lambda_i \ \ \text{since} \ \ \sum_i \Pi_i = 1 \ ,$$

and this is a reasonable indicator of the value of $\mathcal{O}$ when averaged over an ensemble. We call it the *expectation value.*

How do we calculate the probabilities Π_i? First we note that wavefunctions should satisfy the usual **Principle of Linear Superposition** on which all analyses of interference and diffraction are based. In mathematical term this means that if $\psi_1(\vec{x})$ and $\psi_2(\vec{x})$ are acceptable wavefunctions, then so is a linear combination of the form

$$\psi(\vec{x}) = c_1 \ \psi_1(\vec{x}) + c_2 \ \psi_2(\vec{x})$$

where the c_1 and c_2 are (complex) constants. Further, we need to assume that any arbitrary wavefunction can be expanded as a linear combination of eigenfunctions of a Hermitian operator, *i.e.*

$$\psi(\vec{x}) = \sum_i c_i \ \phi_i(\vec{x}) \ ,$$

where the coefficients c_n are, in general, complex. This result is known as the **Completeness Hypothesis**, and it is easily seen to be identical with Fourier's Theorem in the case of plane

[8]We assume that $N \gg 1$, since it is only in this limit that actual numbers will correspond to probabilities.

waves, or eigenstates of the momentum operator. For other operators it can be established on a case-to-case basis. We simply assume this as a property of the set of all acceptable wavefunctions, or, in other words, only such functions are acceptable as wavefunctions which can be thus expanded[9]. The normalisation condition for the wavefunction then gives

$$\int d^3\vec{x}\ |\psi(\vec{x})|^2 \ = \ \sum_i |c_i|^2 \ = \ 1 \ ,$$

where we have used the orthonormal property of the eigenfunctions. Recalling that $\sum_i \Pi_i = 1$, we have good reason to guess that $\Pi_i = |c_i|^2$. However, we can strengthen this by the following argument.

Using the same orthogonality property of the eigenfunctions, we can solve for the coefficients

$$c_i = \int d^3\vec{x}\ \phi_i^*(\vec{x})\ \psi(\vec{x}) \ .$$

The integral on the right is called the *overlap integral* and it indicates, how close, in some sense, the wavefunction $\psi(\vec{x})$ is to the eigenfunction $\phi_i(\vec{x})$. For example, if the wavefunction is one of the eigenfunctions, say $\phi_n(\vec{x})$ for some $i = n$, then $c_n = 1$ and the rest of the c_i vanish. In this case, we realise that every measurement of the variable $\mathcal{O}$ will yield the eigenvalue λ_n, *i.e.* the probability $\Pi_n = 1$ and all the other Π_i vanish. Clearly, there is a fundamental correspondence between the overlap c_i and the probability Π_i, such that when $c_i = 1, \Pi_i = 1$ and when $c_i = 0, \Pi_i = 0$. Since c_i is complex and Π_i is real, the only possibility is that $\Pi_i = |c_i|^\alpha$, where α is an unknown positive power. Thus, we conclude that

$$\langle\lambda\rangle = \sum_i |c_i|^\alpha\ \lambda_i \ ,$$

and, moreover, that this is true for every variable $\mathcal{O}$ and every operator $\hat{\mathcal{O}}$, which means that α is a universal constant[10]. It only remains to determine α. This is done by considering the position operator $\vec{x}$ and is performed in the following way.

If the wavefunction $\psi(\vec{x})$ is to be interpreted as the probability density of finding the system in a volume element $d^3\vec{x}$ around the point $\vec{x}$, then clearly the average value of position is given by

$$\langle\vec{x}\rangle = \int d^3\vec{x}\ |\psi|^2\ \vec{x}$$

[9]Those well-versed in linear algebra will recognise this space of wavefunctions to be a **Hilbert space**.

[10]There is an assumption here, for the above argument would work perfectly well if there were different α's for different variables $\mathcal{O}$. However, we recognise that α builds a bridge between overlap integrals and probabilities, and it is a physical assumption that they should always be related in the same way.

which is the normal way to define a statistical average in terms of a frequency distribution, where probability is taken to be identical with the frequency. Substituting the expansion $\psi(\vec{x}) = \sum_i c_i \phi_i(\vec{x})$ in this equation leads to the identification

$$\langle \vec{x} \rangle = \sum_i |c_i|^2 \; \vec{x} \; .$$

Recognising that the eigenvalue equation is just

$$\hat{\vec{x}} \; \psi(\vec{x}) = \vec{x} \; \psi(\vec{x}) \; ,$$

it follows that $\lambda_i \to \vec{x}$ and therefore, that $\alpha = 2$. Hence $\Pi_i = |c_i|^2$.

The probability of the collapse of wavefunction $\psi(\vec{x})$ occurring to the eigenfunction $\phi_i(\vec{x})$ is thus

$$\Pi_i = |c_i|^2 = \left| \int d^3\vec{x} \; \phi_i^*(\vec{x}) \psi(\vec{x}) \right|^2 \; .$$

The overlap integral is thus a probability amplitude. We can rewrite the expectation value as

$$\begin{aligned}
\langle \lambda \rangle &= \sum_i \Pi_i \; \lambda_i \\
&= \sum_i |c_i|^2 \; \lambda_i \\
&= \sum_i c_i^* \; \lambda_i \; c_i \\
&= \int d^3\vec{x} \; \psi^*(\vec{x}) \left[\hat{\mathcal{O}} \; \psi(\vec{x}) \right]
\end{aligned}$$

which is a convenient expression to calculate. The last line of the above follows from the orthonormalisation relation. If the wavefunction $\psi(\vec{x})$ is not normalised, as it usually is, the expression reduces to

$$\langle \lambda \rangle = \frac{\sum_i |c_i|^2 \; \lambda_i}{\sum_i |c_i|^2} = \frac{\int d^3\vec{x} \quad \psi^*(\vec{x}) \left[\hat{\mathcal{O}} \; \psi(\vec{x}) \right]}{\int d^3\vec{x} \quad \psi^*(\vec{x}) \; \psi(\vec{x})}$$

which is the most common form.

Taking the expectation value of a variable $\mathcal{O}$ to be an average, we can now formally *define* the uncertainty as the square root of its variance,

$$\Delta\mathcal{O} = \sqrt{\langle (\mathcal{O} - \langle \mathcal{O} \rangle)^2 \rangle} = \sqrt{\langle \mathcal{O}^2 \rangle - \langle \mathcal{O} \rangle^2} \; ,$$

which means that $\Delta\mathcal{O}$ is just like a standard deviation. Recall that $\langle \mathcal{O} \rangle$ is an average calculated over an ensemble of identically-prepared systems. It is now possible to calculate the uncertainty for various operators and wavefunctions, and verify that the uncertainty product, in certain cases, cannot be made smaller than a number of the order of $\hbar$.

Now that we have a formal definition of uncertainty, we can state the **Heisenberg Uncertainty Principle** explicitly. It assumes the form

$$\Delta x\ \Delta p \geq \frac{\hbar}{2}$$

where $\Delta x = \sqrt{\langle (x - \langle x \rangle)^2}$ and $\Delta p = \sqrt{\langle (p - \langle p \rangle)^2}$. However, though this is enough to describe a particle, it is not adequate to describe an arbitrary system. For that, we consider a system with N degrees of freedom, which is described by N generalised coordinates q_k ($k = 1, 2, \ldots, N$), each corresponding to a generalised momentum π_k[11]. The pair of variables (q_k, π_k) are said to be *canonically conjugate*. In quantum mechanics, to each variable q_k, there will correspond an operator $\hat{q_k}$ and to each π_k there will correspond an operator $\hat{\pi_k}$. Given a wavefunction, then, we can calculate expectation values of these operators and also their uncertainties Δq_k and $\Delta \pi_k$. The formal statement of the Uncertainty Principle is, then,

$$\Delta q_k\ \Delta \pi_\ell \geq \frac{\hbar}{2}\ \delta_{k\ell}\ ,$$

which means that it is possible to measure canonically non-conjugate variables exactly, but when a pair of canonically conjugate variables is measured, we have to reckon with a finite uncertainty product.

The uncertainty principle as stated above has many manifestations. To begin with, let us calculate the uncertainty in the variable $\mathcal{O}$ when the system is in an eigenstate $\phi_i(\vec{x})$ of the $\hat{\mathcal{O}}$ operator, belonging to the eigenvalue λ_i. We have

$$(\Delta \mathcal{O}) = \sqrt{\langle \mathcal{O}^2 \rangle - \langle \mathcal{O} \rangle^2} = \sqrt{\langle \lambda_i^2 \rangle - \langle \lambda_i \rangle^2} = \sqrt{(\lambda_i^2) - (\lambda_i)^2} = 0\ ,$$

which means no uncertainty[12]. This is consistent with the fact that we expect a measurement of $\mathcal{O}$ on the state $\phi_i(\vec{x})$ to yield the definite value λ_i over an entire ensemble of identically-prepared systems. Now, if the uncertainty vanishes over an eigenstate, then an alternative statement of the uncertainty principle would be that no pair of canonically conjugate variables have common (simultaneous) eigenstates.

[11] The definition of a generalised momentum, in classical mechanics is given by $\pi_k = \partial L / \partial \dot{q}_k$ where $L = T - V$ is the Lagrangian for the system.

[12] We have to add the rider that these considerations do not apply to the position operator $\vec{x}$, since all wavefunctions are eigenfunctions of this operator. The apparent paradox can be resolved if we realise that the wavefunction does not full represent a state but is the projection of a state vector onto the (continuous) position basis.

We now prove a simple theorem:

> If two operators $\mathcal{O}$ and $\mathcal{O}'$ have a common set of eigenstates, then it is necessary and sufficient that they form a commuting pair.

To establish that the condition is necessary, let $\phi_i(\lambda_i, \lambda_i')$ be a common eigenstate of the two operators, such that

$$\begin{aligned} \hat{\mathcal{O}}\ \phi_i(\lambda_i, \lambda_i') &= \lambda_i\ \phi_i(\lambda_i, \lambda_i')\ , \\ \hat{\mathcal{O}}'\ \phi_i(\lambda_i, \lambda_i') &= \lambda_i'\ \phi_i(\lambda_i, \lambda_i')\ . \end{aligned} \tag{1}$$

Then it follows that

$$\hat{\mathcal{O}}\hat{\mathcal{O}}'\ \phi_i(\lambda_i, \lambda_i') = \lambda_i\lambda_i'\ \phi_i(\lambda_i, \lambda_i') = \hat{\mathcal{O}}'\hat{\mathcal{O}}\ \phi_i(\lambda_i, \lambda_i')\ ,$$

or

$$[\hat{\mathcal{O}}, \hat{\mathcal{O}}']\phi_i = 0\ ,\ \text{where}\ [\hat{\mathcal{O}}, \hat{\mathcal{O}}'] = \hat{\mathcal{O}}\hat{\mathcal{O}}' - \hat{\mathcal{O}}'\hat{\mathcal{O}}$$

is called the *commutator* of the two operators and is a linear operator on the space of eigenfunctions. If the operators have a common *set* of eigenfunctions, then this relation holds for all values of i. Now, using the completeness hypothesis to expand an arbitrary wavefunction in terms of the eigenfunctions, we get

$$\psi = \sum_i\ c_i\phi_i$$

and hence

$$[\hat{\mathcal{O}}, \hat{\mathcal{O}}']\psi = \sum_i\ c_i[\hat{\mathcal{O}}, \hat{\mathcal{O}}']\phi_i = 0\ .$$

Since ψ is an arbitrary function, it follows the commutator must vanish, *i.e.*

$$[\hat{\mathcal{O}}, \hat{\mathcal{O}}'] = 0\ , \Rightarrow \hat{\mathcal{O}}\hat{\mathcal{O}}' = \hat{\mathcal{O}}'\hat{\mathcal{O}}\ ,$$

which means that the operators commute.

It is trivial to establish that the condition is sufficient. Let us assume that $\hat{\mathcal{O}}\hat{\mathcal{O}}' = \hat{\mathcal{O}}'\hat{\mathcal{O}}$. Now let

$$\hat{\mathcal{O}}\ \phi_i(\lambda_i) = \lambda_i\ \phi_i(\lambda_i)\ ,$$

for some ϕ_i. Then

$$\mathcal{O}(\mathcal{O}'\phi_i) = (\mathcal{O}\mathcal{O}')\phi_i = (\mathcal{O}'\mathcal{O})\phi_i = \mathcal{O}'(\mathcal{O}\phi_i) = \mathcal{O}'(\lambda_i\phi_i) = \lambda_i(\mathcal{O}'\phi_i)\ .$$

This means that $\mathcal{O}'\phi_i$ is an eigenfunction of $\mathcal{O}$ belonging to the same eigenvalue λ_i. It follows that $\mathcal{O}'\phi_i$ and ϕ_i must be the same[13] function, up to a scale factor, which we call λ'. Thus

$$\mathcal{O}'\phi_i = \lambda'\phi_i \ ,$$

which tells us that ϕ_i is also an eigenfunction of $\mathcal{O}'$ belonging to the eigenvalue λ'.

What is the relevance of this theorem in the context of the uncertainty principle? We saw that canonically non-conjugate variable can have simultaneously vanishing uncertainties, and hence can have simultaneous eigenstates. This, by the above theorem, means that the corresponding operators must commute. Conversely, canonically conjugate variables will correspond to operators which do not commute, *i.e.* the commutator $[q_k, \pi_k] \neq 0$. The value of this commutator determines the uncertainty product and hence fixes the difference between classical and quantum mechanics. Clearly it must vanish in the classical limit, which we defined as $\hbar \to 0$. Moreover, in this limit, when the disturbance is small, any wavefunction should be an eigenfunction of the commutator, for otherwise one of the operators would induce a large disturbance in the system. In other words the commutator must be a constant operator, and thus, we can write

$$[q_k, \pi_k] \propto \hbar \Longrightarrow [q_k, \pi_k] = \beta\hbar$$

and it only remains to fix the constant of proportionality β. We do this by considering the familiar variables x and p, for which the operators are known:

$$\hat{x}\psi = x\psi \ , \qquad \hat{p}\psi = -i\hbar\frac{\partial}{\partial x}\psi \ .$$

Now, for an arbitrary wavefunction ψ

$$\begin{aligned}
[\hat{x}, \hat{p}]\ \psi &= (\hat{x}\hat{p} - \hat{p}\hat{x})\ \psi \\
&= -i\hbar\left[x\frac{\partial\psi}{\partial x} - \frac{\partial}{\partial x}(x\psi)\right] \\
&= -i\hbar\left[x\frac{\partial\psi}{\partial x} - x\frac{\partial\psi}{\partial x} - \psi)\right] \quad = \quad i\hbar\ \psi
\end{aligned}$$

so that we immediately identify $\beta = i$, which means $[\hat{x}, \hat{p}] = i\hbar$. We can now write down the **canonical commutation relations**

$$\begin{aligned}
[\hat{q}_k, \hat{q}_\ell] &= 0 \\
[\hat{q}_k, \hat{\pi}_\ell] &= i\hbar\ \delta_{k\ell} \\
[\hat{\pi}_k, \hat{\pi}_\ell] &= 0
\end{aligned}$$

[13]This argument holds true for *non-degenerate* eigenfunctions, *i.e.* the case when every eigenfunction belongs to a different eigenvalue. If two or more eigenfunctions have the same eigenvalue, the proof is more difficult, but the theorem still holds. Any textbook on linear algebra will have the proof.

which holds for all variables q_k and π_k. It can be shown quite generally that these canonical commutation relations are completely equivalent to the uncertainty relations

$$\begin{aligned}
\Delta q_k \ \Delta q_\ell &= 0 \\
\Delta q_k \ \Delta \pi_\ell &= \frac{\hbar}{2} \ \delta_{k\ell} \\
\Delta \pi_k \ \Delta \pi_\ell &= 0 \ ,
\end{aligned}$$

but the proof of these lies beyond the scope of these lectures.

Do we have a physical understanding of these commutation relations? Indeed we do. We argued that a measurement of a variable $\mathcal{O}$ on a system causes it to collapse into an eigenstate of $\hat{\mathcal{O}}$. We do not know á priori which eigenstate this is going to be, and this is reflected in a finite uncertainty $\Delta\mathcal{O} \neq 0$. If we now measure a variable $\mathcal{O}'$ on this eigenstate, and this operator commutes with $\mathcal{O}$, we realise that the collapsed state is an eigenstate of the new variable $\mathcal{O}'$ as well, so the measurement will yield the appropriate eigenvalue with zero uncertainty, and moreover, the system will remain in the same state. No further collapse will occur. Since $\Delta\mathcal{O}' = 0$, the uncertainty product $\Delta\mathcal{O} \ \Delta\mathcal{O}'$ vanishes. If the measurements are done in reversed order, then $\Delta\mathcal{O}' \neq 0$ and $\Delta\mathcal{O} = 0$, so that, as before, the uncertainty product $\Delta\mathcal{O} \ \Delta\mathcal{O}'$ vanishes. However, if the operators do not commute, then, after the first measurement, the collapsed state will not be an eigenstate of the new operator $\mathcal{O}'$ and a second collapse will occur, this time to an eigenstate of $\mathcal{O}'$. Thus, both the measurements will yield finite uncertainties, and the uncertainty product will be non-vanishing. The uncertainty principle is, therefore, a quantitative measure of the disturbance (collapse) induced in a system by an external measurement, and its magnitude is measured by the reduced Planck constant $\hbar$.

In the classical limit $\hbar \to 0$, these effects can be neglected, and we are back to deterministic physics, *i.e.* no uncertainties. In fact, if quantum mechanics is applied to a macroscopic system, all possible dynamical operators will commute and any state will be a simultaneous eigenstate of all of these. In this case, we might as well work with the eigenvalues. To express this mathematically, note that in the classical limit

$$f(\hat{q}_1, \hat{q}_2, \ldots, \hat{\pi}_1, \hat{\pi}_2, \ldots)\Psi = f(q_1, q_2, \ldots, \pi_1, \pi_2, \ldots)\Psi$$

where Ψ is any wavefunction and $f(q_1, q_2, \ldots, \pi_1, \pi_2, \ldots)$ denotes any relation between the eigenvalues $q_1, q_2, \ldots, \pi_1, \pi_2, \ldots$). Since classical mechanics is concerned with such relations only, we can dispense with the operators and the null (or negligible) disturbance they cause.

The converse of this result[14] provides us with an extremely powerful tool to construct the operators corresponding to dynamical variables in quantum mechanics. If a relation between eigenvalues (classical dynamical variables) is true, then a similar relation must hold between the corresponding operators in quantum mechanics. This is the famous **Correspondence Principle** of Bohr. It does not help to construct the basic non-commuting pair(s) of variables; for example, we cannot write $\hat{p} = m\frac{\partial}{\partial t}\hat{x}$ because this does not have the non-commutativity built in. However, the other variables, which commute with $\hat{x}$ and $\hat{p}$ can be written down easily. Thus we have

Dynamical Variable	Classical formula	Quantum Mechanical formula
angular momentum	$\vec{L} = \vec{x} \times \vec{p}$	$\hat{\vec{L}} = \hat{\vec{x}} \times \hat{\vec{p}}$
kinetic energy	$T = \vec{p}^2/2m$	$\hat{T} = \hat{\vec{p}}^2/2m$
potential energy	$V = V(x,t)$	$\hat{V} = V(\vec{x},t)$
Hamiltonian	$H = \vec{p}^2/2m + V(x,t)$	$\hat{H} = \hat{\vec{p}}^2/2m + V(\vec{x},t)$

and so on.

What is the status of *wave-particle duality* in the quantum theory of measurement? We see that the moment we have introduced a fuzziness in our information about the position and momentum of a particle, we have given it a wave aspect. This is why we have to work with a wave function — a probability amplitude — rather than with the exact values. The wave aspect, therefore, represents our inability to find the exact values of canonically conjugate variables and thus, ultimately, is a mirror of the fact that we cannot observe a microscopic system without disturbing it. The above discussion of the Uncertainty Principle makes it clear that these will be manifested when the experiment, in some way, requires a simultaneous measurement of at least one pair of canonically conjugate variables. If the measurement involves only commuting variables, then a particle will behave as a classical particle, such as a photon in Compton scattering. We thus have a firm basis for Bohr's **Complementarity Principle**, which says that in a given experiment, a microscopic system will exhibit either wave or particle properties, but not both. We now see that it is simply related to whether the experiment requires observation of canonically conjugate or non-conjugate variables.

Finally, let us note that all of the above discussion is just a series of plausibility arguments, conjectures and physical insights and one can find lots of flaws in them if one wishes. Thus, it is traditional to embody the quantum theory of measurement and its consequences in a set of ten postulates. These embody everything we have discussed in a compact form and, for those who like an axiomatic approach, we could well have started by postulating

[14] conjectured, not proved

them as the basic laws of dynamics. The ten postulates are listed below.

§ Every physical system is described by a wavefunction ψ, which contains all knowable information about the system. The wavefunction is normalised in the sense that $\int d^3\vec{x}\ |\psi|^2 = 1$. The wavefunction description is necessitated by the essential fuzziness caused by the fact that measurements disturb a system.

§ Dynamical variables correspond to linear Hermitian operators on the space of wavefunctions. These have real eigenvalues and orthonormal eigenfunctions.

§ The linear superposition of two wavefunctions is an acceptable wavefunction and every wavefunction can be written as a linear superposition of eigenfunctions of any Hermitian operator corresponding to a dynamical operator (*Completeness Hypothesis*). This restricts the set of acceptable wavefunctions.

§ A measurement of a dynamical variable on an eigenstate of the corresponding operator produces the corresponding eigenvalue as a result.

§ A measurement of a dynamical variable on an arbitrary state of the corresponding operator causes the wavefunction to collapse to an eigenfunction (*Collapse Hypothesis*).

§ The probability that the arbitrary state ψ will collapse to a particular eigenfunction ϕ_i is given by the squared modulus of the overlap integral, *i.e.* by $|\int d^3\vec{x}\ \phi_i^*\ \psi|^2$, and hence a set of measurements on a set of identically-prepared system will yield the expectation value $\langle\mathcal{O}\rangle = \int d^3\vec{x}\ \psi^*[\hat{\mathcal{O}}\psi]$ as an average. This is an explicit statement of the *Born interpretation* of the wavefunction in terms of probability.

§ The operators corresponding to dynamical variables are such that the canonical commutation relations hold, *i.e.*

$$[\hat{q}_k, \hat{q}_\ell] = 0\ , \qquad [\hat{q}_k, \hat{\pi}_\ell] = i\hbar\ \delta_{k\ell}\ , \qquad [\hat{\pi}_k, \hat{\pi}_\ell] = 0\ .$$

This, as we have seen, is another way of writing the *Heisenberg Uncertainty Principle.*

§ Except for the basic non-commuting pair(s), all other dynamical operators obey the same algebraic relations as they do in classical mechanics and, in the limit $\hbar \to 0$, this is true of the basic pair(s) as well (*Correspondence Principle*).

§ In a given experiment, the wave or particle properties of a microsystem will show up, but not both (*Complementarity Principle*) and this depends on the requirement of measuring commuting or non-commuting variables during the experiment.

§ The time evolution of a system is governed by the *Schrödinger equation*

$$\hat{H}\psi = i\hbar \frac{\partial \psi}{\partial t} \ .$$

This is true in the non-relativistic limit. In the relativistic case, there are other equations of motion, which reduce to the Schrödinger equation for low energies. We note that in quantum mechanics, time is not a dynamical variable, but just an evolution parameter. There is thus no time operator, no time expectation values and no time uncertainty.

These postulates form the essential structure of quantum mechanics and we shall, in the succeeding lectures, try to build up the structure by considering various examples and special cases. In doing so, we shall see how quantum mechanics can successfully explain all the lacunae and ad hoc assumptions made in the old quantum theory. In fact, till the present day, we have no better theory to describe the microscopic world than quantum mechanics.

Finally, we note that the Copenhagen interpretation, built out of conjectures by Born, Bohr, Heisenberg, Pauli and other lesser lights in the 1920's and 1930's, has successfully withstood the test of time and every attempt to test it or to prove it wrong has ended up by strengthening it. Today we have a deeper and more complete understanding of quantum mechanics than these pioneers had, but the essential structure remains quite unchanged. The best proofs, however, are those which came in the wake of the new ideas — in every field, application of quantum mechanics yielded new results, which always proved a better explanation of experimental results than the classical one. The justification of the postulates of quantum mechanics is, thus, ä posteriori. In the final resort, this is all that science requires or will ever require.