First a quick summary of Classical Mechanics:

There is a multi-dimensional phase space made up of all the positions q_i and all the momenta p_i of a particle or system. In general the dimension is 6N (where N is the number of particles). If I want to know everything there is to know about a system, all I need to know is where the particle is in phase space at $t = 0$. Then I can use Hamilton's equations of motion to describe everything before and after ($t < 0, t > 0$).

Once consequence is that if I propose to make a measurement ($x, p, \ldots$), I can do so with absolute certainty. It makes sense to talk about the trajectory of the particle even if it is not continuously observed (e.g. a car moves behind a red truck, it still exists). There is *one and only one* path in phase space that the physical system can follow from an initial point - governed by Hamilton's Equations of Motion.

This picture gets modified if I don't know the initial state perfectly well (because of insufficiently precise measurement instruments). In that case we can only describe the probability of a certain final state. Thus, probability exists already in classical mechanics (e.g. thermal processes). In this case, we can define a probability distribution (as opposed to a single point) in phase space:

$$P(q_i, p_i) \rightarrow probability\ density$$

$$Prob(system \in V) = \int_{V\ in\ phase\ space} P(q_i, p_i)\Pi_i dq_i dp_i$$

One classical example is the Gaussian Probability Distribution:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

If we integrate $g(x)$ over an area, then we will get the probability of finding x within the interval Δx. You must have a finite interval to get a finite probability.

$$Prob(x \in [\,]) = \int_{[\,]} g(x)dx$$

Probability Theory Highlights

Probabilities should never be < 0. The probability of anything should be between 0 and 1.

$$\int_{-\infty}^{\infty} g(x)dx = 1$$

$$\langle x \rangle = \int_{-\infty}^{\infty} xg(x)dx = \mu$$

$$where\ \mu \rightarrow\ \textbf{mean}\ and\ \langle x \rangle \rightarrow\ \textbf{expectation value}$$

The average distance from $\langle x \rangle$ is the Root Mean Square (RMS) deviation:

$$\Delta x = \sqrt{\langle x - \mu \rangle^2} = \sigma$$

or

$$\sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

or

$$\sigma = \sqrt{\int_{-\infty}^{\infty} x^2 g(x)dx - \mu^2}$$

Liouville's theorem tells us that if we know that the system was initially in a volume V of phase space, the size of the volume it could be in after some elapsed time is still V (although it could be a different shape or position). This doesn't contradict the 2nd law of thermodynamics, but is negated by quantum theory (as we will see): here, the uncertainty (possible volume in phase space) increases with time for most unbound systems.

Waves and Particles

The ideal system in classical mechanics is a point particle. The electron, e^-, is a perfect point particle. It should be completely specified by $m_e, \mathbf{r}, \mathbf{p}$. However, because of their small mass and size, it turns out that electrons are governed by quantum mechanics, not classical mechanics.

We could set up an experiment where an electron source spits out e^-s at a shield which contains a slit. On the other side of the shield we place a detector and amplifier which is free to move (say, along the x direction perpendicular to the propagation direction) and detect e^-s. In principle you can do this one e^- at a time. After an amount of time you will get a distribution for the number of electrons near any given point x_0: $n_e(x_0...x_0 + \Delta x) = N_{total} Prob(x \in [x_0...x_0 + \Delta x]) = P(x_0)\Delta x$. Here, the probability distribution $P(x)$ encodes the consequence of our insufficient knowledge of where in phase space the electron was initially. However, classically we "know" that each individual electron started out at just one point in phase space (even if we don't know which) and follows a single, well-defined path through the slit to the final position x_0.

Now let's repeat the experiment with the slit in a different location. Now we will get a slightly different (likely shifted) distribution n'_e corresponding to a different probability distribution $P'(x)$.

Now what about 2 slits? Classically you get a sum of the two distributions. Since each electron has to go through one slit or the other (and not both), we can simply add the probabilities for the two possibilities (going through the first slit or the second slit):

$$Prob_{tot} = Prob_1 + Prob_2$$

In actuality the probability distribution looks nothing like what is expected by classical mechanics. Instead we get the typical interference pattern (well-known from Young's double slit experiment with light). Therefore one of our premises must be wrong. The only one that is open to debate is *that every particle follows a defined trajectory.* Now we know of a physical system that doesn't have a trajectory, *a wave.* Waves go through both slits and interfere. The decisive point here is that *whatever encodes our knowledge of the e^- is non-local and it interferes with itself.* Furthermore, we can conclude that our knowledge must be encoded in something like an *amplitude and phase*, and the probability density is then the *square* of that amplitude.

One could argue that in principle one can find out which slit the electron "actually" went through by placing a detector by one of the slits. However, this completely destroys the interference. The conclusion is that by gathering information, we change the amplitude of whatever it is we are observing, and therefore all future measurements will be disturbed.

A typical explanation (first proposed by Heisenberg) goes along the following lines: "The act of measurement disturbs the system itself so it will behave differently". For instance if we were to measure x within a precision Δx (say, by using photons with a wave length shorter than the separation of the two slits), then we would disturb p by Δp, the momentum transferred by the absorbed photon. According to Planck and Einstein, the momentum carried by each photon would be

$$p = \frac{2\pi\hbar}{\lambda}$$

Note that $\hbar c = 197.33$ eV nm. Therefore, this uncertainty only kicks in when dealing with incredibly tiny amounts of energy and distance. We would be giving the particle a kick $\Delta p = \hbar/\Delta x$ which in turn would destroy the interference pattern.

In fact, the correct interpretation of Heisenberg's uncertainty relation $\Delta x \Delta p \geq \hbar/2$ refers not (only) to the disturbance by an actual measurement, but to the *intrinsic* and unsurmountable limits of our knowledge. It says that no matter how well we prepare a system (e.g., an electron in flight), we cannot *predict* the outcome of a measurement of x without some uncertainty Δx and the outcome of a measurement of p without some uncertainty Δp, and the product of these two uncertainties is always larger than the limit given above.