The Double Slit Experiment and Quantum Mechanics*

Richard Rolleigh

2010

1 Introduction

When we first studied quantum mechanics in the 1960's, my colleagues and I were astounded by strange and weird concepts like wave particle duality, the uncertainty principle, nonexistence of trajectories, and collapse of the wave function. Today, sixty years later, those same concepts have become part of our culture through television shows like Star Trek, Sliders, Quantum Leap, and the NOVA series. However, I suspect that today's students find it almost as difficult as we did to accept a physical theory that contradicts so strongly the Newtonian mechanics that we learned intuitively as children.

We know that moving objects have trajectories because we have played baseball and soccer. We know that inanimate objects like baseballs have a well defined nature and that their behavior is totally determined by initial conditions and the forces acting on them. All inanimate objects familiar to us obey Newton's laws. Yet the quantum physicists tell us that all these familiar things are made up of microscopic particles that do not obey Newton's laws at all. What rational person would believe this rubbish? In support of their ridiculous claims, the quantum physicists give us convoluted explanations of esoteric experiments and even more convoluted explanations of even more esoteric mathematics.

What is needed is a simple experiment that we can all understand and that unequivocally demonstrates the more disturbing properties of microscopic particles. It would also be nice if the experiment had actually been done and the results corroborated the strange predictions of quantum mechanics. Richard Feynman described just such an experiment in 1963: the double slit interference experiment that you studied in introductory physics.[1,2]

The double slit experiment (DSE) was first reported to the Royal Society of London by Thomas Young in 1803. Young did the experiment with light waves (photons) and measured the interference bands by observing the brightness of the light. Feynman proposed using modern technology to either do the experiment with electrons or do it with photons and detect individual photons. Clinton Davisson and Lester Germer had demonstrated electron diffraction in 1927, but this is one of those esoteric experiments referred to previously. The Feynman double slit experiment with individual electrons or photons is easier to understand and confronts us with inescapable evidence of the weirdness of microscopic particles. The experiment was not done in the form that Feynman described until 1972.[3] The experiment has since been repeated in a multitude of forms that include all the aspects described here.[4]

Feynman's description was designed for non-scientists, so I will modify it in recognition of your advanced understanding of physics and mathematics. These modifications will also help me avoid copyright infringement litigation.

2 Intrinsic properties of particles that motivate the experiment

Electrons and photons (and all other microscopic particles) exhibit two important properties that are crucial to the importance of this experiment. The first is that they all obey interference phenomena just like waves. You have probably observed interference of light waves passing through a double slit apparatus. It is firmly established experimentally that electrons behave the same way. In fact, double slit interference has been demonstrated with

[1] Richard Feynman, **The Feynman Lectures on Physics**, (Addison wesley 1963), Volume III, Chapter I.

[2] Richard Feynman, **The Character of Physical Law**, (MIT 1965), Chapter 6.

[3] Am J of Physics, **41**, p 639 - 644, 1972.

[4] The latest was in 2008. For exact references, see http://physicsworld.com/cws/article/indepth/9745 and http://en.wikipedia.org/wiki/Bell_test_experiments#Loopholes.

electrons,[5] neutrons,[6] atoms, [7] and buckyballs.[8]

The second important property that electrons, photons, and all other microscopic particles share is that they are always detected as individual particles, not as waves. When you did the Milikan oil drop experiment, you observed the motion of oil drops (or perhaps spheres made of teflon, plastic, or glass) containing a small discrete number of electrons. If any of those drops behaved as if it contained a fractional number of electrons, you were probably suffering from eyestrain. It is easy to believe that particles like electrons, protons, and neutrons are always detected as a whole particle and never as a piece of a particle. However, you may have imagined that you see light much as you hear sound, and since sound is clearly a wave, light must be too. You would be wrong: you see light very differently from how you hear sound. Your retina is covered with many tiny rods and cones, and when you see anything, individual photons are absorbed by these rods and cones. Each photon causes a discrete electrochemical excitation that is transmitted along the optical nerve. This is a very different process from that of your eardrum which moves as a unit due to air pressure variations spread over the entire eardrum.

Let me say this again to emphasize it. Your eyeball is covered with a large number of photon detectors. When you see something, each detector counts the number of photons it received and transmits that number to the brain. Some of the detectors (the cones) can detect the energy of the photons, and they transmit that value to the brain also (thus providing color vision). Your eyeball works much like the detector portion of a digital camera. You have never observed a light wave in your life, but you have added up the numbers of photons striking different places on your retina to create a diffraction pattern.

To me, the most convincing evidence that all particles, including photons, are always detected as individual and whole particles was observing the output of a particle detector on an oscilloscope. The output is a series of pulses. Each pulse represents the passage of one particle (a photon, an electron, or whatever) through the detector. You get the same effect with an old fashioned geiger counter: each click represents the passage of a particle through the detector. If you have never had the opportunity to observe this,

---

[5] American Journal of Physics, Volume 42, pages 4-11, 1974

[6] Reviews of modern Physics, Volume 60, pages 1067 -, 1988

[7] Physical Review Letters, volume 66, page 2689 - , 1991

[8] Letters to Nature, Wave Particle Duality of C_{60} molecules, Markus Arndt, 1999

you should at least read Wikipedia's article on particle detectors.

All microscopic particles, including photons, exhibit these two properties: they form interference patterns when passed through a double slit apparatus and they are detected individually as whole units. Never is a piece of one detected. The pictures in the referenced articles clearly demonstrate that individual particles are being detected as whole units, and that they form an interference pattern as more and more of them are detected. These experiments have been done with a great variety of microscopic particles, including photons. The results of the experiments have all been the same for all of the various particles. I will henceforth just use the generic word 'particle' and not specify whether I am speaking of an electron, photon, neutron, proton, buckyball, or whatever. They all behave the same in these experiments.

3 The basic double slit experiment with particles

In the basic experiment, we pass a large number of particles through the double slit apparatus and let them strike detectors attached to the screen as illustrated in Figure 1. Of course we will have to take care that our particles are all going in the same direction and all have the same wavelength. In other words, we need a columnated beam of particles that all have the same momentum because the de Broglia wavelength for all particles (including photons) is just Planck's constant over momentum,

$$\lambda = h/p.$$

For photons, we can generate the particles with a mercury lamp and various filters and lenses just as you did when you performed the photoelectric experiment. For charged particles, we can use an apparatus similar to the electron gun that you used when you performed the Thompson e/m experiment in introductory physics.

The screen on the right side of Figure 1 is covered with many closely spaced particle detectors whose positions are indicated by the variable z. For each experiment, we will pass a few billion particles through the slit apparatus and record the number of particles striking each detector. We will then make a histogram of the number of particles arriving at each detector as a function of detector position.

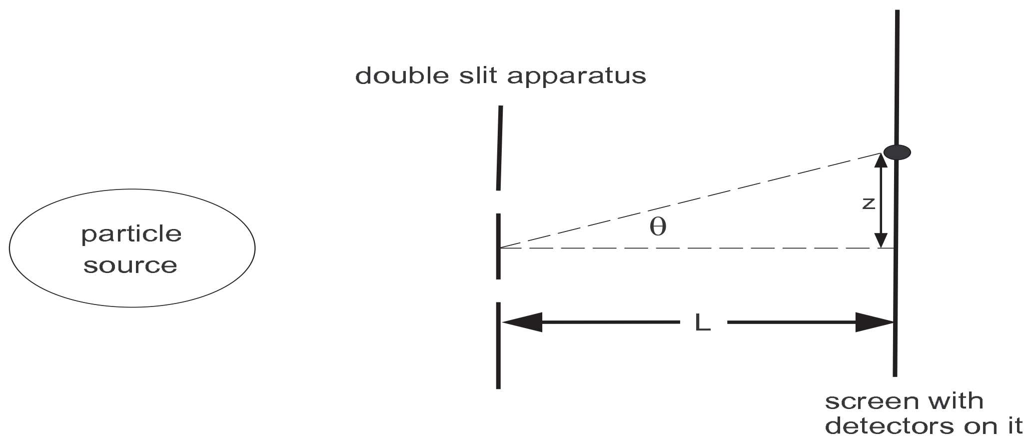

Figure 1: Double Slit Apparatus

First we close the lower slit requiring all the particles to pass through the upper slit. The histogram we observe is illustrated in figure 2. This is the same as the single slit diffraction curve produced by monochromatic waves that pass through a single slit that is 12.72 times as wide as the wavelength and then strike a screen one meter away. We could obtain this same single slit pattern by either using photons with wavelength 550 nm (green light) and a slit width of 7 μm or electrons accelerated through a potential of 200 Kv and a slit width of .035 nm. The precise expression for single slit diffraction is

$$I(\theta) = I_{\max}\left(\frac{\sin\alpha}{\alpha}\right)^2, \tag{1}$$

where $\theta = \tan^{-1}(z/L)$, $I(\theta)$ is the intensity at the angle θ, $I_{\max}$ is the maximum intensity at $\theta = 0$, α is

$$\alpha = \frac{\pi a\,\sin\theta}{\lambda},$$

a is the width of the slit, λ is the wavelength of the monochromatic light or the de Broglie wavelength of the particle (if it has mass), and L is indicated

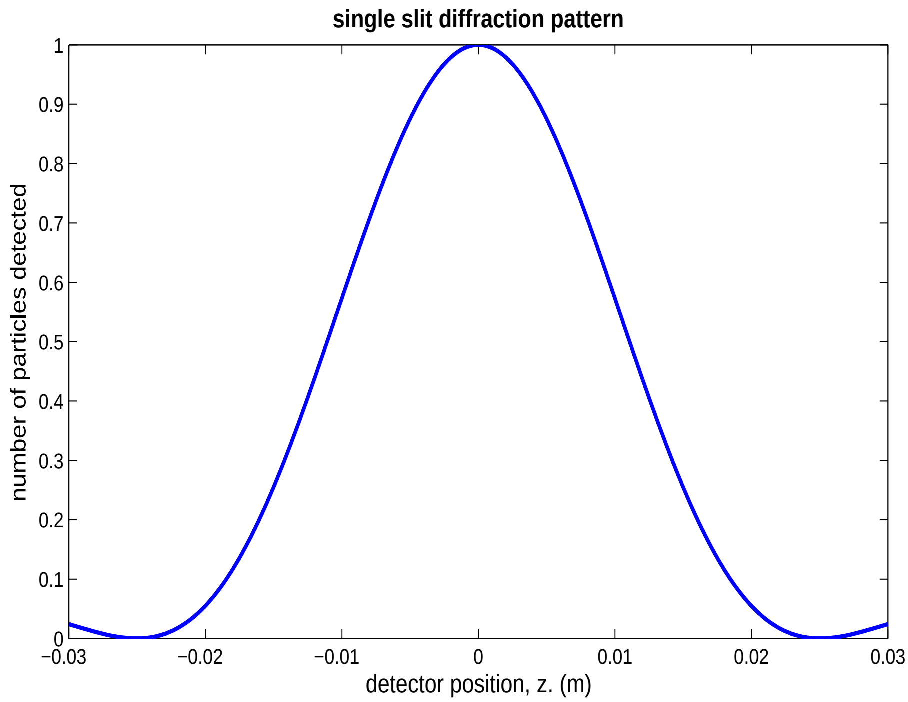

Figure 2: Single slit histogram

in Figure 1. Derivations and explanations of Equation 1 can be found in most introductory physics texts. Another source is the URL
`<http://en.wikipedia.org/wiki/Fraunhofer_diffraction_(mathematics)>`.

Of course we could close the upper instead of the lower slit thereby forcing the particles to go through the lower slit. The result is exactly the same except the pattern is displaced down by the distance between the slits. That distance is less than .1 mm so we can't tell the difference in the curves.

When we open both slits so the particles can go through either slit, we see something entirely new. Figure 3 illustrates the histogram we observe. Monochromatic waves passing through two slits separated by 145 times their wavelength would produce the same pattern on a screen one meter from the two slits. We could obtain this same double slit pattern by either using photons with wavelength 550 nm (green light) and a slit separation of 80 μm or electrons accelerated through a potential of 200 Kv and a slit separation of .4 nm. The precise expression for the double slit interference curve is

$$I = I_{\max} \cos^2 \left(\frac{2d \sin\theta}{\lambda} \right) \left[\frac{\sin\left(\pi a \sin\theta/\lambda\right)}{\pi a \sin\theta/\lambda} \right], \tag{2}$$

where d is the slit separation, a is the slit width, θ is the angle in Figure 1, $\lambda = h/p$, h is Planck's constant, and p is the momentum of the particle. Derivations and explanations of this expression can be found in most introductory physics texts. Perhaps a more convenient reference is `http://en.wikipedia.org/wiki/Double-slit_experiment`.

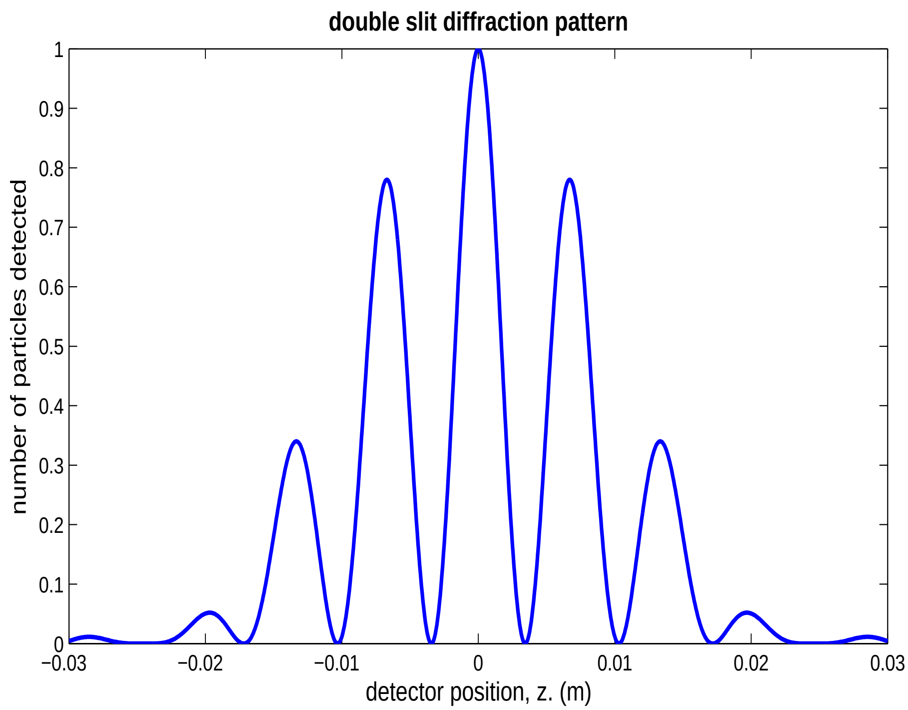

Figure 3: Double Slit histogram

We see that if we force the particles to go through only one slit, we obtain a single slit pattern. If we allow the particles to go through both slits, we obtain a double slit pattern.

It is important to note that the shape of the double slit pattern depends on the distance between the slits. If you increase that distance, the interference maxima get closer together. The only rational interpretation of this is that in order for the particles to form a double slit pattern, either each particle must interact with both slits or some particles pass through the upper slit and some pass through the lower slit, and the particles then interact with each other to form the double slit pattern. The second possibility will be discredited by the next experiment.

4 The double slit experiment with one particle at a time

In order to test the conjecture that some of the particles pass through the top slit and some pass through the bottom slit, and then they interact with each other to form the interface pattern, we do the experiment with only one particle at a time passing through the double slit apparatus. If the particles had to interact with each other to produce a double slit pattern, then passing one particle at a time through the apparatus would destroy the pattern. However, we find that even if we pass only one particle at a time through the apparatus, we still get the two slit interference pattern. This was verified by the experiments reported in references 3 and 4.

Up to this point the particles behave just like classical sound waves except for the way they are detected. If you close one slit, each particle goes through the other slit just as sound waves would. If you open both slits, each particle interacts with both slits just like sound waves. With sufficient time, enough particles will accumulate to form an double slit pattern just like sound waves. The only feature that distinguishes particles from sound waves so far is that only one detector at a time on the screen detects a particle. If we were using sound waves, all the detectors located in bright fringes would fire at the same time. We cannot turn down the amplitude of the sound wave until only one quanta of sound energy passes the slits at a time because sound wave energy is not quantized.

Since each particle interacts with both slits, each particle's energy must get divided so that some goes through each slit. We try to detect that in the next experiment

5 Detect which slit

It is not difficult to build a particle detector that doesn't absorb all of the particle's energy. If you study the design of particle detectors in Wikipedia, you will understand that by adjusting the length of the detector along the direction of the particle's motion, you can adjust the amount of energy absorbed from zero to 100 per cent. Of course as you reduce the amount of energy absorbed, you decrease the probability that the particle will be detected.

In order to detect how much of each particle goes through each slit, we place detectors after each slit. If we make the slit detectors very sensitive so that they detect everything that goes through their respective slit, we observe that each particle goes through one slit or the other. No particles divide their energy between the slits. Clearly, the particles are not interacting with both slits. How can they then make a double slit pattern? Well, **they don't**! When we turned on the slit detectors and formed a histogram from the outputs of the detectors on the screen, we got the superposition of two single slit patterns. These patterns are so much alike that their sum looks just like the single slit pattern in Figure 2. It seems that detecting which slit they go through forces them to go through one slit or the other and also forces them to produce two single slit patterns instead of a double slit pattern. Although this experimental result may be intuitively disturbing, it is nice that it agrees with the predictions of quantum mechanics.

This latest particle behavior is quite distinct from that of sound waves. If we measured how much of sound wave energy went through each slit, we would find that the sound wave split its energy equally between the slits and still formed a double slit pattern. Particles on the other hand, choose one slit or the other (when we measure which slit) and form a single slit pattern.

This experiment has been done with photons[9] and with atoms[10]. The method they used to determine which slit the particle traversed involved an entangled photon and measurements made on it. We may have time to discuss these experiments in more detail after we have studied entangled states. Despite the esoteric nature of these experiments, they fully corroborate the results I have described in this section.

---

[9] Phys Rev letters, **84**, pp 1 - 5, January, 2000,

[10] Phys Rev Letters, **81**, pp 5705 - 5709, December, 1998

6 Weakly detect which slit

An incorrigible sceptic might argue that in the previous experiment we destroyed the double slit pattern because our slit detectors were too sensitive. They interfered with the particles too much. The obvious solution is to make the detectors absorb less of the particles' energy and thus be less sensitive. If we do this, the slit detectors will miss some of the particles that eventually are detected by the screen detectors. Our data will fall into three classes:

- Particles that are detected traversing the upper slit and then striking the screen,
- particles that are detected traversing the lower slit and then striking the screen, and
- particles that are not detected by either slit detector yet we know they were there because they were detected at the screen.

The percentage of particles in the third group will increase if we decrease the sensitivity of the slit detectors. If we form histograms of each class, the first two classes will form single slit patterns while the third class will form a double slit pattern.

There is no way to escape the conclusion that we determine how the particles traverse the double slit apparatus by what we choose to measure or not measure. If we measure which slit, the particles accommodate and go through one slit or the other and then strike the screen at places that form a single slit pattern. If we do not measure which slit, the particles strike the screen at places that form a double slit pattern. Since the double slit pattern depends on the distance between the slits, the particles must interact with both slits if we do not detect which slit they traverse.

I hope you are not uncomfortable with all this because it will get worse in the next section.

7 Delayed Choice Experiment

The previous experiment tells us that turning on the slit detectors forces the particles to traverse only one slit and turning off the slit detectors forces the particles to interact with both slits. What if the decision to turn the slit detectors on or off is made after the particle has already passed through the double slit apparatus? This is not too hard to do with the accurate timing available today and the existence of particle storage devices that can hold a particle isolated from all influence for several ns.

We place a particle storage device between each slit and its corresponding slit detector as illustrated in Figure 4. For photons, the storage device is just

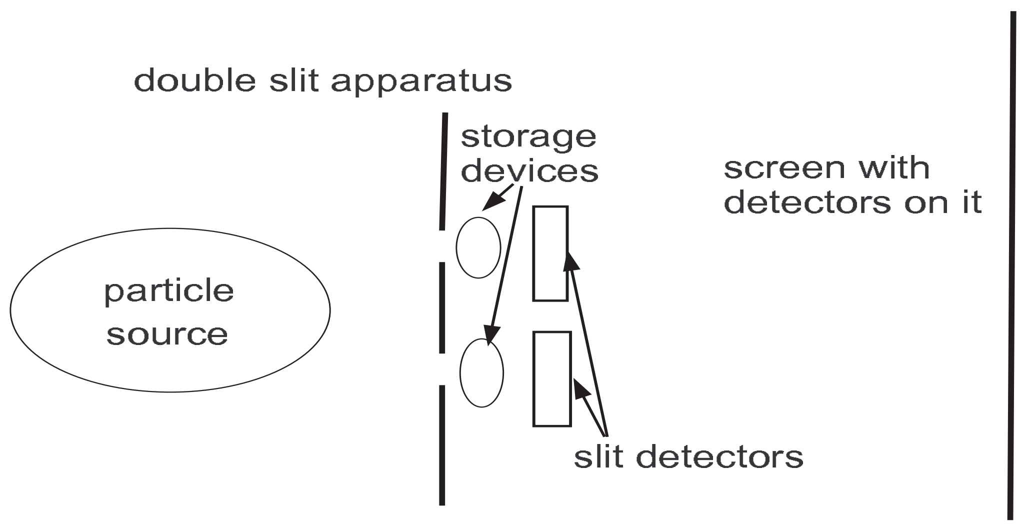

Figure 4: Double Slit with storage devices and slit detectors

an optical fiber loop, and for charged particles it is just a magnetic field that causes the particle to go in circles. Suppose the storage devices will delay the particles for 10 ns. We randomly change the settings on the detectors every 10 ns. This forces the particles to interact with a detector whose setting was determined *after* the particle interacted with the slits. But the setting on the detector will still determine how the particle interacted with the two slits just as it did in the previous experiment. Although this delayed choice experiment has not been done exactly as described here, slight variations have been done a number of times,[11] always with the results described here.

[11] Science, **315**, no, 5814, pp 966 - 968, (2007) and references 9 and 10 here.

8 Interpretations

The only credible interpretation of the experimental results is that the act of measurement seems to influence the behavior of the particle, and that this influence can go backwards in time. There are a plethora of philosophical schemes to explain this strange behavior, but physicists have reduced them to two competing paradigms: realist and orthodox that Griffiths describes on page three of your text. The old school name for these interpretations are the hidden variables and Copenhagen interpretations respectively, and you will see these names in much of the older literature (Copenhagen equals orthodox and hidden variables equals realist). You should be cautious in your usage of the terms 'realist' and 'realism' because they are widely used in philosophy, art, literature, and politics and they mean different things to different people.

How do the realist and orthodox paradigms interpret the experiments we have been discussing? The realist would insist that the path of the particle through the slits (whether it went through only one slit or interacted with both slits) was real and had a precise value before the particle entered the storage device. The realist would also have to conclude that at least for some of the particles, the path through the slits was changed when the particle passed through the slit detector *after passing through the slits*. The action of the slit detectors exerted an action backwards in time that changed the value of a physical property in the past. We physicists have a strong aversion to changing the past. In other words, we have a strong belief in causality. By causality, I mean that if a physical property had a value yesterday, then there is nothing you can do today to change what its value was yesterday.

The orthodox position on the other hand is that the path of the particle through the slits is never real even if the particle is detected by one of the slit detectors. When you detect a particle in the top slit detector, what is real is the localization of that particle in the top slit detector at that time. Although that reality is consistent with the particle having gone through the top slit and not interacting with the bottom slit, to conclude that the particle was really in the top slit at an earlier moment of time is more than most orthodox adherents would claim. They would be more likely to say that until the particle is detected by either of the slit detectors or by a detector on the screen, it has the potential to land anywhere on the screen. If it is detected by a slit detector, then the probabilities of where it will land on the screen are modified by that detection. If it is not detected by a slit detector, then the probabilities of where it will land are modified in a different way.

The orthodox position is that even though the particle was measured by a slit detector, and the only way it could have gotten to the slit detector was through the slit, this does not require that the particle was ever really in the slit at any time. This position may seem to be evasive, but there are well established experimental results that demonstrate this very thing. I am speaking of the tunneling of particles through potential barriers that require more energy than the particle has. This phenomena has been well known for so long that tunneling diodes and tunneling electron microscopes are based on it. The particle clearly moves from one side of the barrier to the other because it is detected first on one side, then on the other. However, it can't ever be in the barrier without violating conservation of energy.

Some people will argue that the orthodox interpretation claims that the detection of a particle in one of the slit detectors makes the path of the particle through the slits real retroactively in time. Therefore, they conclude that the orthodox interpretation requires that reality be created in the past and that this is just as large a violation of causality as the realist position that requires that reality be changed in the past. However, the previous two paragraphs expose the fallacy of this argument. The orthodox position neither creates nor changes anything in the past because it claims there is nothing real in the past except what was measured in the past. Although what I have just stated is true, it will have to be clarified when we consider entangled particles and instantaneous creation at a distance (nonlocal creation).[12]

9 Measurement

I have shown experimentally that the process of measurement changes the state of the system, and I have argued that it either changes the past, projects reality into the past, or ultimately defines what is real. Yet I have not provided a precise definition of what constitutes a measurement. That was rather sloppy of me wasn't it. Let me remedy the situation.

I think Niels Bohr said it best when he said that a measurement is an

[12] It is true that the orthodox interpretation requires nonlocal creation of reality. It follows that different observers will disagree on the order in time in which measurements were made. Consequently, they will disagree on which measurement actually created the reality. But neither observer will observe creation going backwards in time.

'irreversible magnification'. You will understand this better if you study the operation of particle detectors. The basic unit of charge is 1.6 $(10)^{-19}$ C (actually one third of that if you consider quarks). We just can't measure this small a charge. However, if any particle that produces an electromagnetic field (this includes uncharged particles like photons and neutrons) passes through some types of matter (semiconductors and gases), then it will transfer small fractions of its energy to many electrons and raise them to the conduction band in a semiconductor or free them from the gas molecules in a gas. If there is a large accelerating potential present, these electrons gain tremendous energy from the external field, and they will liberate more electrons. This produces an avalanche effect. The result is that the single particle being detected produces a pulse of many electrons whose combined charges can be detected. This is obviously an irreversible amplification just like a snow avalanche is. When a measurement occurs, entropy increases, disorder increases, energy moves from high concentration to low concentration, and the measurement can't be undone.

I like the idea that the only things that are real are those things that can't be undone. If we could go back in time and change reality, it wouldn't be very real would it? I also like the way that my personal interpretation of quantum mechanics integrates so well with the second law of thermodynamics.

10 Impact on the theory

How do we use these experiments to guide us in the construction of a theory of microscopic particles? Any useful theory predicts things, so we must first decide what properties of microscopic particles are predictable. For something to be predictable, it must be a consistent measurement result. The positions at which individual particles land on the screen are not consistent: each particle could land in any bright fringe. Positions are not predictable. What is consistent is the probability of each particle's landing at any position, i. e. the probability density function (pdf) of each particle's position. The pdf of position is just the double slit interference pattern illustrated in Figure 3. It is reproduced any time you repeat the experiment and it is predicted by Equation 2. We will find that all physical variables exhibit this behavior in all experiments with microscopic variables: specific outcomes are not consistent but the probabilities of all possible outcomes are. The only time a specific outcome is predictable is when a measurement is performed, a specific value is obtained, and then the identical measurement is repeated on the same system before it has time to interact with anything.[13] In this case, the same result will be obtained the second time.

Note that the actual value of a physical variable (position, momentum, etc.) is not predictable because identical measurements of the physical variables of identically prepared systems produce different results. The fact that the actual value of a physical variable is not predictable arises from experiment, not theory. How will this make quantum mechanics different from classical mechanics? In classical mechanics, the values of all physical variables are predicted as functions of time. Time is the only independent variable and all the physical variables are dependent variables in classical mechanics. Prediction of the physical variables as functions of time is the program of classical mechanics. What is the program of quantum mechanics? Quantum mechanics predicts the probability distributions of all the physical variables. In addition to time, all the physical variables are independent variables, and the probability distributions are dependent variables.

If we return to the double slit experiment without slit detectors, we see that the probability of any one particle striking at z on the screen is predicted quite accurately by Equation 2. However we design our quantum theory, we want it to reproduce Equation 2 as the probability distribution for the positions of where the particles will strike the screen. Let us review how Equation 2 was obtained so as to get ideas on how to build our quantum theory.

Equation 2 is just the time average of the magnitude squared of the solutions to the classical wave equation that match the boundary conditions imposed by the slit and the screen. The classical wave equation in three dimensions is

$$\boldsymbol{\nabla}^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = 0, \tag{3}$$

where c is the wave's phase speed. There are many (an infinite number) of independent solutions to this equation, but boundary conditions limit the solutions appropriately for each situation. An appropriate solution for our

[13] Note that position can not qualify for this special case of consistency because you can not obtain a specific value from a position measurement device. Since detectors are finite in size, you cannot detect a particle to be at x. Rather, you detect it between x and $x+dx$. The only variables that can qualify are those that are quantized like energy and angular momentum.

problem is

$$
\begin{aligned}
\Psi &= \left(\psi_1(x,z,t)+\psi_2(x,z,t)\right)/\sqrt{2}, \quad \text{where} && (4)\\
\psi_1(x,z,t) &= \int_{d/2-a/2}^{d/2+a/2}\int_{-b/2}^{b/2} f(x,-y',z-z',t)\,dy'\ dz'\,, && (5)\\
\psi_2(x,z,t) &= \int_{-d/2-a/2}^{-d/2+a/2}\int_{-b/2}^{b/2} f(x,-y',z-z',t)\,dy'\ dz'\,, && (6)\\
f(x,y,z,t) &= \frac{\mathrm{e}^{(2\pi i(r-ct)/\lambda)}}{r}, && (7)\\
r &= \sqrt{x^2+y^2+z^2}
\end{aligned}
$$

and where b is the height perpendicular to the page of the slits in Figure 1.

Note that this solution is just the sum of two solutions: ψ_1 represents the waves passing through the top slit, and ψ_2 those passing through the bottom slit. Each solution is the integral over the corresponding slit of radially outgoing waves of wavelength λ and phase speed c. If you perform the integrals, square the absolute magnitude of the result, set $x = L$ (the distance from the slits to the screen), and average over time, you will reproduce Equation 2 in the limit that $b >> \lambda > a$ and $L >> \lambda > a$.[14]

The tremendous success of this approach suggests that we base our theory on the following two ideas:

- for every particle, there exists a wave function that is a solution of equation 3 that also meets the boundary conditions imposed by how the particles were prepared, and
- the probability density function of the particle's position is the magnitude squared of the particle's wave function.

There are three problems with this proposed theory. They are

- The proposed wave equation does not include the potential energy of the particle. We know that the potential energy at a point must influence the probability that the particle will be found at that point. For example, we do not expect to find a particle in a region where the potential energy is larger than the total energy of the particle. Also, it

[14] You will need to multiply by a normalization constant because neither the intensity pattern in Equation 2 nor the wave functions in Equations 4 to 7 are normalized.

should be very likely to find a particle in regions where it moves slowly (because it hangs out there a lot). These would be regions in which the potential was only slightly less than the total energy of the particle.

- Another major problem with Equation 3 is that it does not conserve probability,

$$\frac{d}{dt}\int_{-\infty}^{\infty} (\Psi^*\Psi)\ dx\ \neq 0.$$

  The integral on the left hand side is the probability that the particle is somewhere. We can normalize ψ so the integral is one today, but since its time derivative is non-zero, it may be two tomorrow. What does this mean? The only reasonable interpretation is that there are two particles tomorrow. There is strong experimental evidence that electrons, protons, and all other baryons and leptons are conserved. Any theory that does not conserve probability cannot describe these particles.

- The proposed theory is incomplete. It only predicts the probability distribution of position. What about momentum, angular momentum, energy, and all those other interesting physical properties?[15]

First let us fix the problems with the wave equation. We know that equation 3 is not acceptable because it does not include potential energy and because it does not conserve probability. I am sure that all of you also know that the Schrodinger equation,

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V(x)\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{8}$$

does include potential energy and does conserve probability. Since our double slit experiment involves three dimensions, I will write down the Schrodinger equation in three dimensions,

$$-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\Psi + V(\boldsymbol{r})\Psi = i\hbar\frac{\partial\Psi}{\partial t}. \tag{9}$$

There are many ways to arrive at the Schrodinger Equation, but I will not spoil your discovery of that exciting literature by saying too much here. Rather, I will make three comments, two brief and one not so brief.

[15] I personally would be very interested in the probability distribution of charm when choosing a particle with which to interact.

The solutions of the classical wave equation presented in Equation 5 are also solutions to the three dimensional Schrodinger equation with no potential energy. We can conclude therefore that if we use the Schrodinger equation instead of the classical wave equation, our theory will accurately describe particles that traverse the double slit apparatus without any potential energy.

It is the second time derivative in the classical wave equation that destroys probability conservation. If you have worked problem 1.41 in your text, then you are probably aware that the Schrodinger equation preserves probability because it has a first time derivative instead of a second time derivative. If we set the potential energy $V(\boldsymbol{r})$ equal to zero, then the only difference between the classical wave equation and the Schrodinger equation is that the former has a second time derivative while the latter has only a first time derivative. The change from the classical wave equation to the Schrodinger Equation for zero potential energy is the minimum that must be done to conserve probability.

How can we use the double slit experiment to discover how to add potential energy to our theory? If we use charged particles, we can easily add potential energy to the double slit experiment by placing a potential difference between the particle source and the two slits. We can then measure the effect if any that this will have on the interference pattern. When we do this, we find that increasing the particles' potential energy increases the distance between the interference fringes. Adding potential energy has the same effect as decreasing the particle's wavelength that appears in the solution given in Equation 4. Let us determine experimentally the relationship between the initial energy E, the potential energy V, and the wavelength λ. Doing this, we discover the following relationship:

$$-\frac{\hbar^2}{2m}\frac{4\pi^2}{\lambda^2} + V = E. \tag{10}$$

It is important to remember that Equation 10 is an empirical equation that we determined by varying V and measuring λ (we measure λ by measuring the positions of the interference fringes).

We know that Ψ in Equation 4 is a solution to the Schrodinger equation without potential. Let us substitute Ψ into the Schrodinger Equation without potential and see how we would have to modify it to get our empirical result, Equation 10. Substituting Ψ of Equation 4 into Equation 9 without potential, we obtain

$$-\frac{\hbar^2}{2m}\frac{4\pi^2}{\lambda^2}\Psi = \frac{hc}{\lambda}\Psi. = hf\Psi = E\Psi, \tag{11}$$

Where I used the fact that the wave's phase speed c is the product of frequency and wavelength, $c = \lambda f$, and the Planck result that $E = hf$. Comparing Equations 10 and 11, we see that we need to add a term $V\Psi$ to the left side of Equation 11 to make them agree. Doing this reproduces the Schrodinger equation with potential. We see that the Schrodinger Equation incorporates potential energy in precisely the manner required to explain the results of the double slit experiment with potential energy added.

Now that we have fixed our theory so that it conserves probability and incorporates potential energy correctly, we will consider the limitation that the only physical variable whose probability distribution our theory predicts is position. Actually, this limitation should be no surprise since we built the theory from the double slit experiment and that experiment only measures position. However, we can get an idea of how the theory will handle other variables if we modify the experiment slightly and if you will allow me to be a little sloppy with normalization.

We modify our source so that more particles reach the top slit than the bottom slit. Say a^2 is the fraction that hits the top slit and $b^2 = 1 - a^2$ is the fraction that hits the bottom slit. This changes the interference pattern, and we find experimentally that we can predict the new pattern accurately if we replace the solution of Equation 4 with

$$\Psi = a\psi_1 + b\psi_2. \tag{12}$$

This makes sense both from our proposed quantum theory and from classical wave theory. Since the fraction of particles going through the top slit is a^2, our quantum theory requires the intensity of the wave for the top slit to be multiplied by a^2. Classical wave theory requires that we multiply the amplitude by a in order to multiply the intensity by a^2. Consequently, we must multiply the wave for the top slit ψ_1 by a and the wave for the bottom slit ψ_2 by b.

Now suppose that we place detectors after each slit. Clearly, the top detector will detect $a^*a = a^2$ of the particles and the bottom slit will detect $b^*b = b^2$ of the particles. I choose to use a^*a instead of a^2 because that allows a and b to be complex without changing our results. There will be cases in the future in which a and b might be complex For each particle, the probability of its going through the top slit is a^*a. So the possible outcomes of a 'which slit' measurement are top and bottom with probabilities a^*a and b^*b respectively.

The states ψ_1 and ψ_2 are called pure states for the 'which slit' measurement. If the system is in the state ψ_1, we know that a 'which slit' measurement will result in the top slit. We also know that whatever the initial state, if a 'which slit' measurement results in the top slit, then after the measurement the system is in the state ψ_1. The initial wave function

$$\Psi = a\psi_1 + b\psi_2$$

is a superposition of pure states. It is called a superposition state, a mixed state, or just the state function. The measurement is described by saying that it causes the initial state function Ψ to collapse instantaneously to a pure state of the measurement. And not just any pure state, it is the pure state corresponding to the value that was measured. Philosophers describe this by saying that before the measurement the particle has various mutually exclusive potential attributes. The measurement destroys some of those potentials and actualizes only one.

We now have a recipe for predicting the probabilities of all the possible outcomes of any measurement. The recipe is

- Construct the state function Ψ. It must satisfy Schrodinger's equation and incorporate all the knowledge we have about the initial state of the system.

- Find the pure states of the measurement. This sounds scary, but actually you have already had much of the math, and the first semester of quantum mechanics is devoted to learning how to find the pure states.The pure states are just the eigenvectors of the operator corresponding the the classical variable being measured.

- write the state function as a superposition of the pure states.

- the probability of measuring any particular value α is the magnitude squared of the coefficient in the state function superposition of the pure state that corresponds to α.

Our basic theory is complete. Now we need to learn how to find the pure states.