Experiment and Probability

Dylan Casey

Lecture given at St. John's College 18 March 2005

1 Introduction

Why do we do expermients? Because they are convincing. Because they are authoritative. Because we learn something about the world. We *know* because of having done experiments. But *why* are experiments convincing? What are we convinced of when we do an experiment?

Some experiments are convincing because there is something manifestly obvious or clear about them. You see something happen with your own eyes right in front of you. Laying your hands on it, putting it together, you are convinced.

When we stand in the pendulum pit during the winter of Freshman Lab, holding a corked tube of water high above the ground, and see right there in front of our eyes that the water does not keep going up with the tube if we take it past 32 feet above the ground, we are convinced. Pascal must be right. I see it with my own eyes.

Other experiments are much more difficult in this regard, even ones we do here at St. John's. Consider the experiments we do examining the atomic structure of matter in the spring of Freshman lab. We know that we are supposed to get the right number of whole units for each of those substances, as we heat and burn pairs of them. That's what Guy-Lussac, Lavoisier, and those guys tell us, right? Most of the time, though, for a number of reasons, the result just isn't manifest before us. Often we just " get it wrong". We know that we got it wrong because we defer a bit to the authority of Lavoisier and 250 years of modern chemistry. Still, here of all places, where we try to look behind authority and understand what it rests on, to tie it down like the statues of Daedelus, it doesn't quite convince us. Not the way 32 feet of water in a corked tube does.

I think that the reasons why our reproduction of Pascal's experiment on pressure is so convincing and those of Guy-Lussac, Lavoisier, and the early chemists on the atomic nature of matter less so, are related simply to our relative confidence about the circumstances of each experiment. In the end, in doing those experiements, we just believe there is less wiggle-room in the explanation given by pressure for the 32 feet of water in a corked tube than in the atomic theory of the elements. I think the way to talk about this difference in confidence with precision and clarity is by talking about probability.

In my lecture tonight, to try to get at this question of the confidence, and authority of experiment, I'm going to tell you the story of a pretty complicated experiment, where we become convinced about the nature of things we can't see or even hope to see. Ever.

At this point, I ought to make a small apology. This lecture, particularly the long middle part that is to follow, is quite different from what we are used to listening to here in this auditorium. Not unheard of, but atypical. I won't lie to you. There are going illustrations, pictures, plots, equations, and numbers. For myself, I don't think that science can be done without them, so after some hand-wringing, I decided not to try. There is undeniably a bit of show-and-tell here, but I hope, at the service of reflecting even more fully on the questions at hand.

2 Narrative of a Number: Smash, Detect, Induct

Once upon a time there was a measurement made by some high-energy particle physicists with the following result:

$$\sigma(Z \rightarrow ee) = 224 \pm 25pb^{-1} \quad (1)$$

or *The cross section for the production of the Z boson decaying into an electron and positron in proton/antiproton collisions at 1.8 TeV is* $224 \pm 25pb^{-1}$.

Please don't walk out now. I promise to make some sense of that.

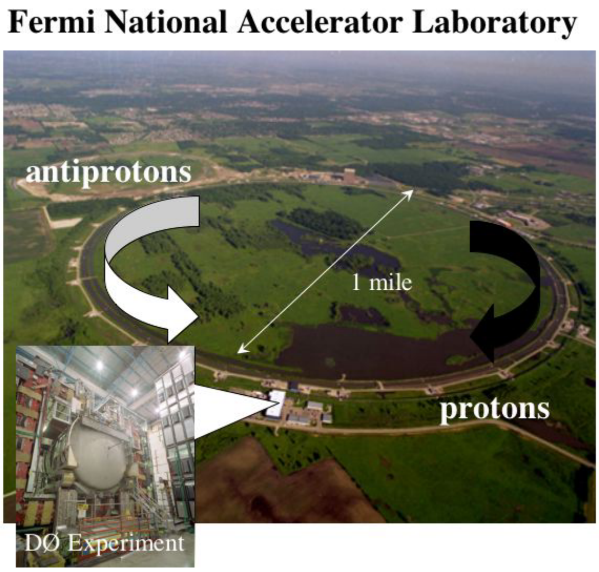

Figure 1: Aerial photo of Fermi National Accelerator Laboratory, located about 40 miles west of downtown Chicago.

2.1 THE WORLD'S SHORTEST DESCRIPTION OF A PARTICLE PHYSICS EXPERIMENT

The phrase *cross section for Z boson production in proton-antiproton collisions* is not jargon so much as a bunch of nouns that you might have never seen strung together. Let me unpack it a bit.

You've likely heard of protons. They are the positively charged constituents of the nucleus of an atom, the other part of the nucleus being the neutrally charged neutron. Antiprotons are just like protons, but all the charges, in particular the electric charge, are opposite that of the proton. So, since a proton has positive electric charge, the antiproton has negative electric charge. For this discussion everything else can be considered to be the same between protons and antiprotons, in particular, the mass of a proton is the same as the mass of an antiproton. "Proton-antiproton collisions" means just what it seems to mean – protons and antiprotons hitting one another.

They don't do this by themselves very often, mainly because antiprotons aren't very common. However, it is made to happen at (among other places) Fermi National Accelerator Laboratory (Fermilab) in Illinois, just west of Chicago.

Exploiting the fact that charged particles that are moving in a magnetic field bend in a direction according to their charge, that is, positively charged particles moving in a magnetic field bend in the opposite direction of negatively charged particles, the protons and antiprotons circulate in the accelerator at the same time and the beams are bumped into one another at specified locations in order to collide them with one another. The illustration in Figure 1 shows an aerial view of the laboratory. The particles travel inside a pipe that is housed in a tunnel underground. The tunnel is about three miles long, having a diameter of about a mile. The curved arrows show the directions for the protons and antiprotons in the accelerator. One of the locations where collisions are measured is at the DØ experiment and is the experiment where the measurement I am talking about was made.

As a matter of practice, closely packed bunches of protons and antiprotons are hurled at one another rather than individual particles in order to increase the likelihood that a collision will actually occur. To get a sense of it, try this at home. Get a friend and two buckets of rocks. Any friend will do, but make the rocks no bigger than a nickel for safety's sake. The two of you now attempt to collide two of the rocks. Likely, you start by picking up one rock each and hurling them at one another, trying to get them to collide. After failing miserably, you may, in fact, try to be clever and stand far apart, allowing one of you to throw without aiming while the other attempts to hit the flying rock with her own throw. After failing miserably at that, before giving up, you might each grab a handful of the rocks and throw them at each other at the same time. You'll likely meet with success on this try, and at least one pair of rocks will collide. Also, you will have witnessed one of the basic principles of accelerator physics and had a bunch of fun at the same time. Incidentally, face shields, helmets, and some light body armor are recommended as protection when doing that demonstration.

Protons and antiprotons themselves have hard bits inside them analgously to the nucleus in an atom. We study the experiment that showed atoms have hard bits inside of them in Senior Laboratory. It's called the Rutherford Scattering Experiment. Before Rutherford's experiment, it was understood that everyday stuff like metal and rocks is made up of small pieces of stuff (atoms) and that those atoms have at least one discernable part, an electron, which has all the peculiar properties of something possessing electric charge. Furthermore, we knew that, as a general rule, ordinary matter does not exhibit peculiar properties associated with being charged. As a general rule, matter is electrically neutral. So, since atoms make up matter that is electrically neutral and they have a negatively charged part, the electron, there must also be some postitively charged part of the atom as well. The question facing Rutherford was how that positive charge was distributed. Was it smeared out throughout the atom, or concentrated in some way? To sort this out, he did what anyone would do, and threw alpha particles at pieces of metal and watched what happened.

Figure 2) shows an illustration of the experiment. The alpha particles are bits of positively charged matter radiated from some kinds of atoms, shown as the group of green and blue balls. They are thrown (shown as the incoming red arrow) at some stuff, generally a piece of metal (shown by the black bar).

If the positive charge of the atom necessary to balance the negative charge of the electrons in the matter was smeared out, then the postively charged alpha particles would pass through the metal (alpha particles being radiation, that is very small bits of matter), getting pushed a little by the charge inside the metal as they passed by, but generally going straight through, as denoted by the outgoing red arrow on the righthand side of the black bar. While that is exactly what happened some of the time, Rutherford found that some of the time, the incoming alpha particles bounced backwards from the metal. The explanation for this is that the positive charge of the atom is concentrated in a relatively small bundle inside the atom. In the right corner of the picture I've illustrated a close-up view of the metal. The orange circles denoting the general size of the atom. The large black circle denoting the nucleus, the concentrated positive charge inside the atom. Most of the time, something passing through the metal doesn't get close enough to the nucleus to be deflected much at all. (Whoosh) However, sometimes it does, resulting in a what we call a hard scattering event. (Whoosh, bang).

Just as the protons in the nucleus of an atom are constituents of that atom, a quark is a constituent of a proton. (In fact, the experiment done in the late 1960s that finally convinced us that protons have hard bits inside them was just a repeat of the original Rutherford scattering experiment at much smaller distances). So, just like in the Rutherford experiment where we can cause one particle (the alpha particle) to hit the hard bits inside the atom and thereby examing the stucture of the atom, we can hit things against the hard bits insides protons, the quarks, and by examining what comes out, learn something about how matter is put together. So, now we know what "proton/antiproton collisions" means.

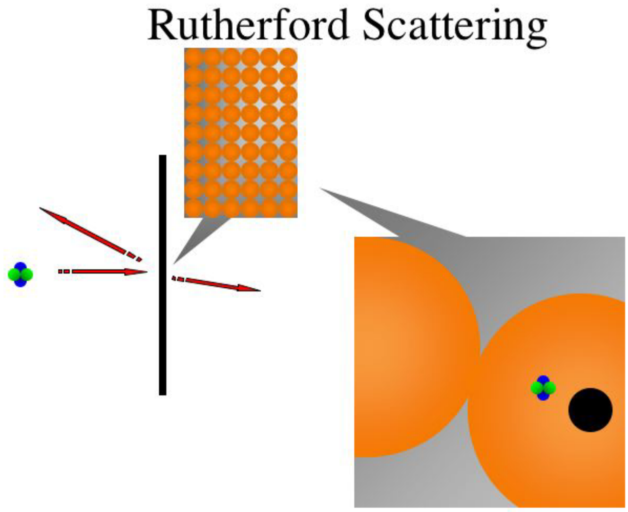

Figure 2: Illustration of Rutherford Experiment

What about the Z boson? While there is in principle much more to say, for this discussion we will understand the Z boson as a particle that can decay into a quark and antiquark (among other things) and because of that can also be created by colliding quarks and antiquarks of sufficient energy together. The details here are complicated. This is really at the heart of understanding radiation, which is really about understanding quantum mechanics and quantum field theory. Let's not go there. The upshot is that I can smash two things together and get a third, new, thing, if that new thing can decay into the first two. This process is just the reverse of radioactive decay in which one thing (e.g.an uranium nucleus) decays into two (or more) things (e.g.an isotope of uranium and an alpha-particle). So, the $Z \to ee$ part just tells us that we're talking about Z boson particles produced in proton/antiproton collisions (colliding quarks and antiquarks) which decay into an electron and a positron.

The term "cross section" is just a precise way of talking about the likelihood of producing something in a collision, saying that if you bang a proton against an antiproton, how often you will get a Z boson. The likelihood depends upon what is being collided (in this case, the quarks that are inside the proton and antiproton), what energy the collision occurs at (that is, how hard the proton and antiproton are smashed together), and what is being produced (in this case, a Z boson). In an everyday collision, hitting a billiard ball for instance, the likelihood of hitting the ball corresponds directly to its size in silhouette – its cross section. In a quantum mechanical collision, like that between two protons, the production cross section refers to the likelihood that a given particle will be produced.

The cross section doesn't tell us by itself how many scatterings will occur. The number of scatterings – the number of rocks that collide – will depend upon how many are thrown. The "how many are thrown" part is called the luminosity. Thinking back on our earlier rock-throwing demonstration, if the handfuls of rocks we throw at each other are bigger, the number of rocks that collide will be bigger.

So, I've now smashed two things together. Now, we need to look at what happens. We need to detect. Then we need to induct. But first, DETECT.

2.2 DETECT: Seeing = Detecting + Reconstructing

However, we don't see Z bosons in exactly the same way we see rocks, just as we don't see the nucleus of an atom in exactly the same way. We can't do it with just our eyes.

In the Rutherford scattering experiment, the nucleus of an atom is seen by throwing small particles at atoms and inferring the properties of the atom based upon what happens. In the experiment, an alpha-particle is thrown and an alpha-particle is observed, but with a new momentum and direction. The Z boson has the added problem that it itself is not a stable particle and decays very quickly into pairs of other particles – electrons and positrons, muons and antimuons, quarks and antiquarks. It is these more stable particles that are observed in the end.

Figure 3 is an illustration of what I'm talking about. On the left is the proton and the right the antiproton. They contain quarks which are shown as the solid blue circles. At sufficient energies, the colliding quark and antiquark will form a new particle, the Z boson, shown as the black circle. This new particle has the combined momentum and energy of the quark and antiquark that combined to create it, shown by the black arrow. This is analogous to two sticky gumballs hitting one another and sticking together. The fused pair of gumballs will travel off in a direction that is the combination of the momentum and energy of the separate incoming gumballs.

The Z boson itself is unstable and decays well before it has moved the distance of a single atom. As mentioned, the Z always decays into pairs of particles and antiparticles, e.g., electron and positron, quark

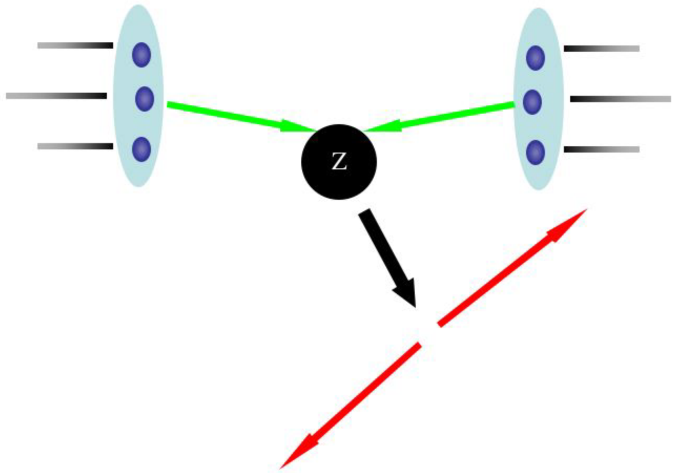

Figure 3: Illustration of the production of Z bosons in proton/antiproton collisions.

and antiquark. We are most interested in the electron/positron case because, as it turns out, electrons are pretty easy to see and measure precisely. Just as the Z boson has the combined energy and momentum of the original colliding quarks, the electron and positron produced in the decay have the combined momentum and energy of the original Z boson. They are denoted by the red lines.

So here we are. We've collided a proton and antiproton together, created a Z boson and it has decayed into an electron and positron. Electrons and positrons are stable particles – they do not decay. This means that they are things that we can "see", that is we can detect them. We call the device used to "see" the electrons *detectors*. Seeing an electron in a particle physics experiment is in most ways the same as seeing a cell through a microscope. A particle detector is, in most respects, a less fancy version of your eye, and in that manner, an extension of the microscope. In a microscope, light that has scattered off something small, like a cell, is focused into your eye by the lens of the microscope. Each piece of light that is scattered activates a part of your retina and this signal is transmitted to your brain via your optic nerve, your brain puts all the signals together, and you see something. As a mechanical account of seeing, this is exactly how a particle detector works. As the electrons from the Z boson travel through the material of the detector, they bump into the electrons bound to the atoms, sometimes knocking them away (called ionization) and sometimes just adding energy to them, after which they give up that energy as pieces of light (photons).

Each of these little collisions deposits a little bit of energy in the detector, which we can observe by turning it into an electrical voltage or current. As a mechanical analogy, you might imagine a detector as a tank of water and an electron as a rock thrown into the water. We can detect the rock and its path by looking at the trail of bubbles left behind. Figure 4 shows a schematic illustration of this analogy. Such a system is actually closer to fact than might be thought at first blush. Figure 5 shows a famous picture from a particle detector called a bubble chamber. Each of the lines in the picture is actually a trail of very small bubbles cause by energetic charged particles traveling through a special liquid. It is actually a photograph – the characteristics of the particles are determined by measuring the lines on the photograph much the same as one would a picture of a cell from a microscope.

As I mentioned earlier, the place where this measurement was done is Fermilab just outside of Chicago. The picture I showed earlier was of the accelerator, the machine used to collide the proton and antiprotons together. The detector used to examine the collisions in this case is called the DØ detector. Figure 6 is a relatively boring picture of the detector itself, mainly showing how big it is.

The detector is about three stories tall and weighs about 5000 tons. While we call DØ *a* detector, it's really millions of little detectors covering as much of the volume around the collision point as possible, just like one might say your eye is a detector, really made up of many, many smaller detectors, the rods and cones in your retina.

Each of the detectors produces a change in current or voltage when a charged particle, like an electron, passes through them. All these little currents and voltages are like the individual signals from the rods and cones in your retina. In order to make a picture, they need to be put together. In the case of your eye, your brain does the work directly. In the case of a particle detector, we use computers and algorithms to recognize the patterns of the energy deposits, along the way making decisions about which things are electrons or quarks or whatever.

Figure 7 is a picture of the output of one of the detector systems. Here, we're looking at the detector with the beam going down the center into the picture. The collisions occur in the center of the picture. The "donut" shows the energy measurement in the primary energy detector in DØ the calorimeter. Each segment of the donut is a section of angle going around the beamline. The two narrow red sections show

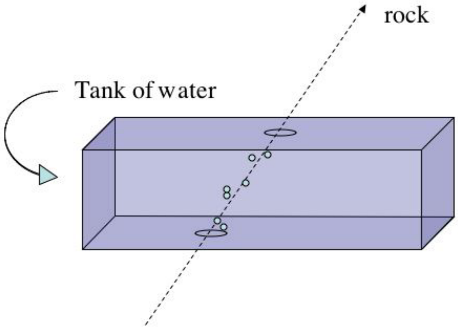

Figure 4: Illustration of a water tank used to detect rocks.

the electrons produced from the Z boson. The wider sections of red and blue show where quark-jets have struck the detector. Notice that the energy is more dispersed in angle than the electrons. That fact is used as one of the ways of separating the electrons and positrons from everything else.

One might ask at this point, if you can't see the Z boson itself, how do you know it is a Z boson? As I mentioned before, the electron and positron that are produced by the Z boson together carry all the energy and momentum of the original Z boson – the decay conserves energy and momentum. Each particle is described by its momentum and energy. The set of particular momentum and energy values are together called the *momentum four-vector* – three values for the momentum in each spatial direction (e.g. the Cartesian coordinate directions x, y, and z) and one value for the energy. There are rules for manipulating these vectors together when describing collisions and decays of particles just like there are for adding together vectors describing the speeds and directions of colliding cars. In general, however, the particular values of the energy and momentum change depending upon the relative velocity of the particles – the closer the speeds of the particles are to the speed of light, the more change there is in the relative momentum and energy. In high energy particle physics experiment, the speeds of the particles are very close to the speed of light. We call such systems of particles "highly relativistic." Fortunately, some quantities describing the particle are relativistically invariant, that is, unlike the momentum and energy of the particle, they do *not* change

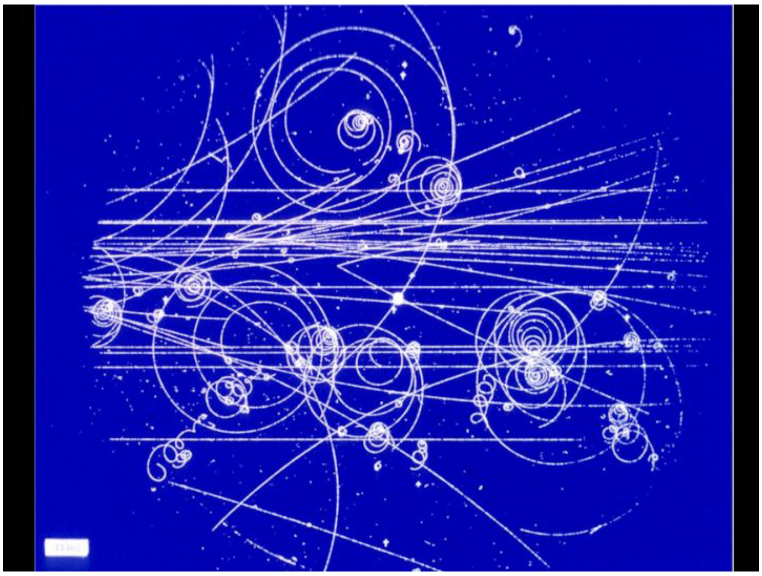

Figure 5: Picture taken from a bubble chamber detector at CERN (a particle physics laboratory in Europe) years ago.

value depending upon the relative speed of the particle. One of these is the "invariant mass." The invariant mass is the magnitude of the momentum four-vector. Those of you who have studied a bit of Einstein might remember the relation:

$$E^2 = p^2 + m^2 \tag{2}$$

where E is the energy, p is the momentum, and m is the mass of the particle. The energy and momentum in this relation change relativistically. The mass does not and is the magnitude of the vector four-momentum – the invariant mass.

Thus, measuring the rate of production of Z bosons means measuring the energy and momentum of electrons and adding the momentum four-vectors together to form the momentum four-vector for one particle and then looking at the invariant mass of the particle. Since we're interested in how many Z bosons are made in proton/antiproton collisions, we just count the number of Z bosons we see.

Figure 2.2 shows a graph of the number of events observed as a function of the invariant mass made made from an electron/positron pair. Here we see the invariant mass of most of the pairs is about 92 GeV. (GeV is the standard unit of mass in particle-physics speak.)

Figure 6: Pictures of DØ detector. The left picture shows the (relatively) small pipe that the colliding beams travel through.

Let me summarize where we are right now. We're colliding protons and antiprotons by throwing bunches of them at one another. When they collide, sometimes they form a Z boson, which sometimes decays into an electron and positron, which are themselves detected by tracing the little bits of energy they leave behind as they travel through the material of the detector. These little bits of energy have locations and magnitudes. From them, we generate energy-momentum vectors – quantities describing the direction and magnitude of the energy and momentum of the particles. We are able to talk about Z bosons by combining the momentum-energy vectors of the electrons into a single momentum-energy vector for the Z boson.

When we ran the experiment in the mid-90s, we had hundreds of billions of collisions between protons and antiprotons inside the detector, each of which we call an event. Out of those collisions, we identified a few hundred thousand events as likely having at least two electrons and we saved output of the detectors for those events as data files so that we could analyze them later in a more refined manner. After looking at that pool of events more closely and selecting some of them based upon a more refined examination of the characteristics of events, we selected 6407 events as Z boson.

This is not the end of the story however. What we want to know is how often Z bosons are created when we collide protons and antiprotons at a certain energy. We'd get very far along that path by knowing how

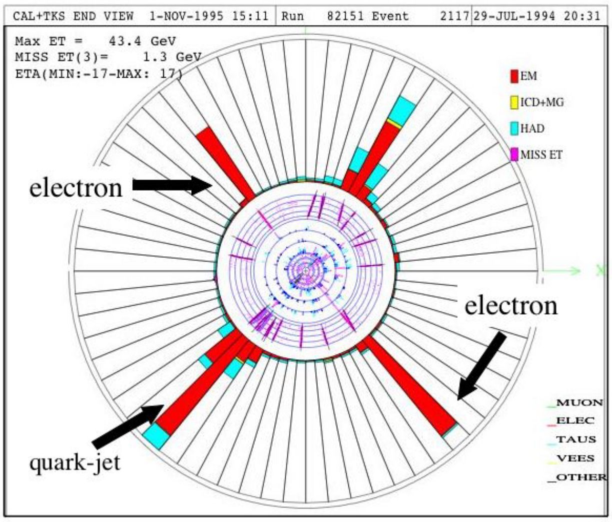

Figure 7: One version of the graphical output from the DØ detector showing the energy observed in a Z boson event.

many Z bosons were created in all the collisions that happened in our detector. What we have are 6407 events that we are calling Z bosons. But are they Z bosons? All of them? Did we miss any? We counted Z bosons by looking for electrons and positrons. Did we make any mistakes in that? Sorting these things out and getting from the counting of things we see to what lies behind it is what doing a measurement is all about. It's about reasoning in the face of uncertainty. It brings us to the last chapter of our story – Induct.

2.3 INDUCT: The Probability Calculus or Getting the Number Out

All in all, that's not so bad, right? Knowing how many protons and antiprotons we throw at each another, we expect that the total number of observed events would correspond to the cross section (σ_Z) times the luminosity ($\mathcal{L}$):

$$N_Z = \mathcal{L}\sigma \tag{3}$$

The problem, as with so many things, is what you see is not necessarily what you get. For instance, what we identified as an electron in the detector may or may not really be an electron. It could be a quark masquerading as an electron, for instance. Remember, we used pattern-recognition to determine whether it

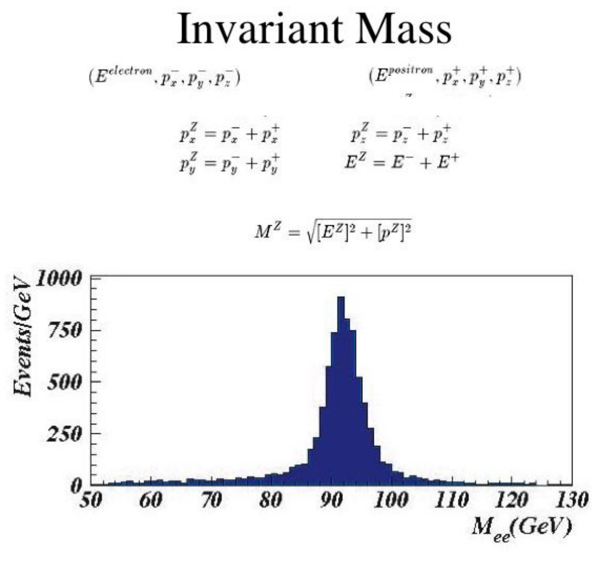

Figure 8: Graph of the invariant mass of Z boson candidates at DØ.

was an electron, and if certain parameters in shape, size, and energy had certain values, then we called it an electron. In the end, though, for each splash of energy in the detector that we call an electron, we are only so confident that it really is one. Importantly, we're not 100% confident. Also, we may have identified the electrons correctly, but they may have been produced in a different physics process. They may not have been the result of the decay of a Z. When we identify something as a Z boson that wasn't a Z we call it *background*. This can happen from mistakes in identifying the things we are looking at (the electrons and positrons) or because other physical processes, while different from the Z production, provide the same sort of signature – an electron and positron. Still, we want to know how many Z bosons were actually produced based on how many Z bosons we saw and what we know about Z boson production in proton/antiproton collisions. In our count of the number of Z bosons observed, there is some percentage that are not actually Z bosons because we know we aren't perfect. So, in trying to sort out what the cross section for Z boson production really is, we need to include the fact that our detected number of Z bosons includes background.

So, at the very least, the number of Z bosons that we have counted in our data are a combination of Z bosons that were actually produced and ones that we mistakenly called Z bosons. The previous relation for the number of Zs produced becomes:

$$N_Z = \mathcal{L}\sigma + b \tag{4}$$

where b denotes the number of background events observed.

Look at the plot of invariant mass I showed a minute ago. The Z has a mass of about 92 GeV. The very low number of events in the low mass region on the left and the high-mass region on the right are almost exclusively background events – they are almost certainly not Z bosons. If the Z didn't exist, as the mass of the electron/positron pair that I detected increased, the number of events I observed would decrease smoothly like the sound of a car as it drives away from you. The "bump" in the middle is the smoking gun signature of a decaying particle. If the two particles found (*i.e.* the positron and the electron) are the products of the decay of another particle, then the invariant mass calculated by assuming that they decayed from some other particle will show this sort of bump (often called a resonance). If you were to search for a new particle, something never seen before, you would examine this sort of plot, the number of events observed as a function of the invariant mass of the two particles, looking for a bump in the number of events. For a particle physicist seeing a bump in the mass distribution is like seeing the water in a corked tube stop at 32 feet above the ground.

So much for the background. We still have other issues to contend with. In addition to counting some things as Z bosons that are not Z bosons, we also don't even count all the Z bosons themselves. Part of the problem is that the probability of identifying a real electron as an electron based upon the shape, size, and energy as seen in the detector isn't 100%. We sometimes make mistakes. Also, we don't see all the electrons that are there just because we aren't looking everywhere. The detector just isn't hermetic, it is full of cracks and non-instrumented sections. Together, we call these sorts of mistakes and misses the efficiency. In this case, we make mistakes in identifying events with two electrons about 30% of the time – our efficiency is about 70%. Also, the electrons from the Z boson end up in places we don't look about 25% of the time. So in the end, the number of Z bosons we observe comes from the number of Z bosons that were actually produced, times the efficiency of identifying the two electrons in that Z boson, plus the number of background events. I need to make one more modification of my relation:

$$N_Z = \mathcal{L}\sigma\epsilon + b \tag{5}$$

Much of the work in making a measurement lies in determining the values of the background and the efficiency and how well we know each of these things. Often, as much time, thought, and energy goes into such determinations as went into taking the data in the first place. For tonight, I'm not going to provide any more details on this point. If there are questions about the details of how these things are done, we can discuss them in the conversation period.

Now, determining the number of events in my data, the value of d, has really just been counting. I've counted 6407 events that satisfy my selection criteria for Z bosons. My model for what composes that counted number is embodied in the above equation. However, N_Z isn't the same and the number of data events I've measured. It's not the same as d_Z. The counting I'm doing is special, reflecting and underlying likelihood with an average. Here, we are trying to understand what is underlying by taking a sample.

If you flip a coin many times, and record the number of times each of the sides turns up, you'll typically find that each comes up about half of the time. However, it's not that unusual to make the same flip 2 or 3 times in a row. In fact, over 100 throws of the coin, you will only sometimes get 50 heads and 50 tails. Easily, one will get 45 heads and 55 tails. Very unlikely, one may get 95 heads and 5 tails. Counting heads and tails is also a special kind of counting. In order to talk about this sort of thing precisely and, for instance quantify a question like "what is the likelihood that the next flip will be a head" or "what is the likelihood I will see N_Z Z bosons for a given luminosity and production cross section" we need to use probability the probability calculus.

2.4 Probability Calculus

The basic rules of the probability calculus are well understood. There are some variations in details, but the basics of how to add, subtract, and multiply probabilities is well-defined. As I outlined at the beginning, the problem is one of induction, of talking about what has happenned based upon what we see, especially when we are uncertain about many parts of what we are looking at.

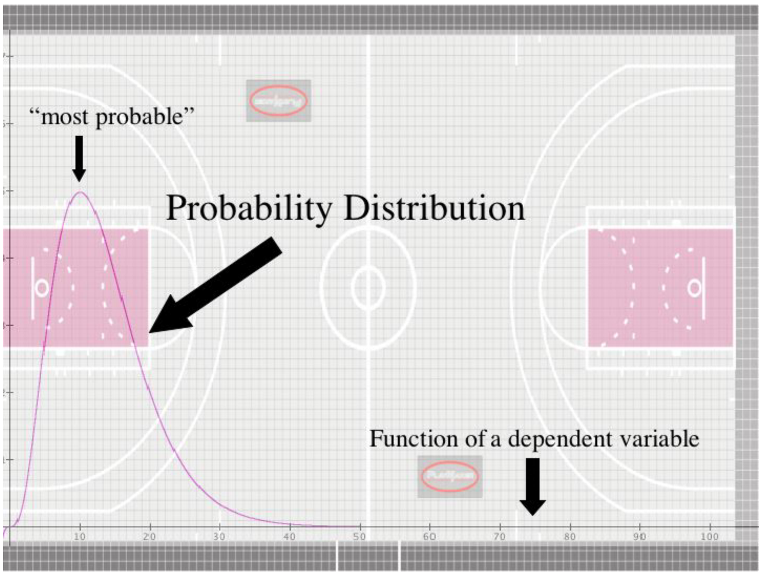

Figure 9: Basketball

2.5 What is probability?

Probability is quantification of how well we know something. As a matter of everyday practice, we do this all the time in betting analogies. There's an 80% chance he'll make the free-throw. She probably loves you. The defendant is guilty beyond a reasonable doubt. There is a 30% chance of snow-showers tomorrow. I'm going to skip the mathematical details here. For now, I'll just say that probability is a number betwen 0 and 1 quantifying our degree of certitude about the truth of a proposition, where a probability of 0 means certain falsehood and a probability of 1 means certain truth. In everyday speech we often change the scale to a percentage scale of 0% to 100%, but it's really just the same idea.

Probabilities can denote the certitude of individual propositions, but can also denote the certitude over a range of parameters. A good example would be the probability of making a jumpshot in basketball as you go from close to the basket to very far from the basket. The probability will be very small if you're standing on the baseline, get higher as you get out from underneath the basket, maybe peak at about three or four feet and decrease as you get further away. Figure 2.5 is an illustration of this. The red curve is a probability density or probability function. Here the probability of sinking a shot depends upon the distance from the basket.

As I mentioned a moment ago that the kind of counting done in this sort of experiment is special. The counting of outcomes when flipping a coin is a counting of outcomes based upon each event having one of two possible outcomes. It is described by a binomial probability distribution. The counting of Z bosons is a counting in time interval in which each success is considered to be unrelated to the previous one; each event is independent of the other. Such a probability distribution describes the probability of obtaining a given number of successes in a single trial for a given average number of successes produced. (get a better statement of a poisson distribution)

In our case of counting, this probability distribution gives us the probability that we obtain a particular count, d_Z given that some average number N_Z was produced. Of course, for us, N_Z is more complicated. Our probability here is really the probability of obtaining d_Z observed events given a particular combination of cross section, efficiency, and background. Mathematically, I would write it like this:

$$P(d_Z|\sigma, \epsilon, b, I) \tag{6}$$

I've included the I to denote the other assumed information that I am not enumerating. For instance, here I am assuming that I know the luminosity perfectly well, so I am not including it as a condition. This is just a mathematical statement of "the probability of observing d_Z Z bosons given the cross section (σ), detection efficiency (ϵ), and background (b)". It is a probability function with the dependent variable d_Z. Figures 2.5 and 2.5 shows a few examples for different values of N_Z. Most probability distributions have this sort of shape. The probability is very small for all values of the parameter except in some restricted region. We talk about how restricted the region is and therefore how well we know the parameter, using probability.

Each of thee plots is a distribution for the value of the probability of d_Z given a value of the efficiency, background, and production cross section, we have the probability distribution for our data. Let me say that again: for a given value of the efficency, background, and production cross section, we know the likelihood of the data. If that sounds a bit backwards to you, you would be right. It amounts to having the probability of being 10 feet from the basket, given that I made the shot. What I really want to know is the probability that I'll make the shot given that I'm 10 feet away. When we make a measurement, we have data and want

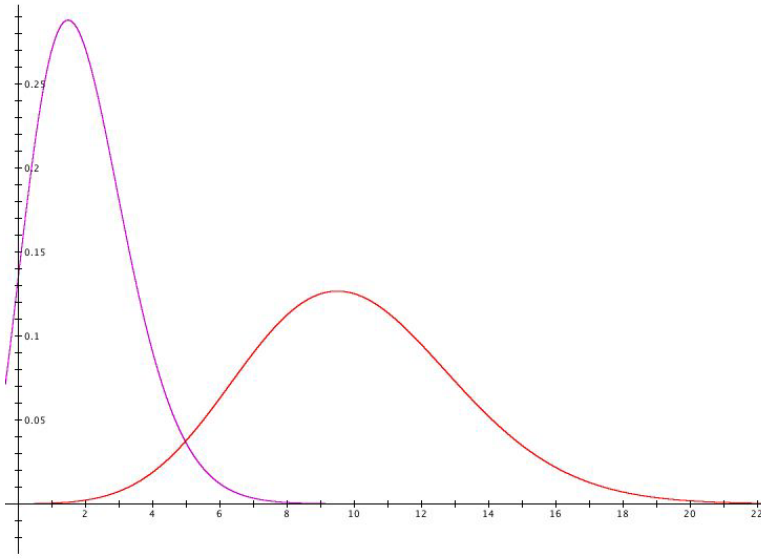

Figure 10: Poisson Distributions

to find out about things like the productions cross section for Z bosons, not the other way around. What we want is $P(\sigma_Z|d_Z)$ – the probability distribution for the production cross section given our data.

This is the heart of the problem of science. This is the problem of induction. This is the problem of reasoning in the face of uncertainty. This problem is addressed with probability.

We need to invert our probability distribution. Mathematically, we need to make the cross section the dependent variable, obtaining the probability as a function of cross section for a given number of observed events. To invert this probability, we use a special probability relation called Bayes' Theorem. Applied to this case, assuming all the probabilities are properly normalized, Bayes' Theorem looks like this:

$$P(\sigma, \epsilon, b, |d, I) = P(d|\sigma, \epsilon, b, I)P(\sigma, \epsilon, b, |I) \tag{7}$$

Now this is much closer to what I want to get. On the left is the *joint posterior probability*, in this case, the probability distribution for the cross section, , the efficiency, and the background given my observed data d_Z. On the right there is the *likelihood*, the probability distribution for my data given what I know about how my data is produced. This is the Poisson distribution that I started out with. Also, there is the *joint prior probability*, what I know about everything that goes into generating the data prior to the measurement.

There is, unfortunately, a problem here. I don't really want a *joint* posterior probability. I'm not really

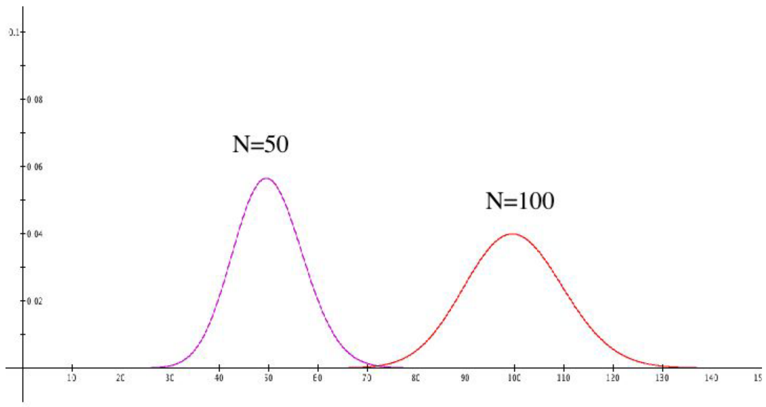

Figure 11: Poisson Distributions

interested in the combined probability of the cross section, efficiency, and background and more than I'm interested in the combined probability of planes landing at BWI and St. John's college having a Friday Night Lecture this week. I really just want $P(\sigma|d_Z)$, the probability distribution of the cross section given the number of Z bosons I've seen. When I invert the probability, I'm left with a function with several dependent variables. I need to get rid of the dependence on the efficiency, background, and luminosity, but I need to do so by considering all the possible ways in which they are combined. In general, I don't know the efficency and background perfectly well. I know something about them, maybe an average or a likely range of values. Here, I also know that the detection efficiency isn't dependent upon the background or the production cross section. Fortunately, they are all independent from one another. This means that the complicated joint prior probability, which is a single function depending upon several variables can be written as a product of probabilities each depending upon one variable. This is like saying that the joint probability of it raining today and the probability of my giving a lecture is merely the product of those probabilities.

To get the probability distribution for the cross section by itself, I need to consider the likelihood of my data given all the possible values for the all the things that go into generating my data. I need to evaluate the likelihood distributions accounting for all possible combinations of efficiency and background. In calculus we add up a bunch of numbers by doing an integral. So, Bayes' Theorem then leads to this:

$$P(\sigma|d_Z) = \int P(d_Z|\sigma, \epsilon, b, I)P(\epsilon|I)d\epsilon P(b|I)db P(\sigma|I) \tag{8}$$

This equation is saying the following. The probability for a given cross section value, σ given the number of Z boson events I've observed is given by the likelihood for that number of Z boson events for that given cross section and what I know about the background adn efficiency – the others factors that go into creating the data I observe. On the left side, $P(\sigma|d_Z)$ is a function that depends upon the cross section, σ. For each value of the cross section, there is a correspond value for the probability. The individual probabilities, $P(b|I)$ for instance, are generally well-described by Gaussian distributions – the traditional, bell-shaped curve, which has its most probable value at the average and is described by a characteristic width. This characteristic width is what quoted as the uncertainty on the parameter. Technically, it means that we are 68% confident that the value of the parameter lies between the average minus half the width and the average plus half the width. This is true for both the background and the efficiency; they are both given by Gaussian probability distributions in which the averages and uncertainties were determined by other measurements.

The last quantity on the right is, in some ways, the most troubling. $P(\sigma|I)$ is the prior probability for the cross section. It is what I know about the cross section before I measure it. Though I am trying to measure the cross section σ, I cannot avoid the fact that what I know about the cross section goes into determining its value. I have two sets of choices here. One is to make the measurement of the cross section as independent as possible from the any predispositions. I need a *least informative prior* – a probability function that provides as little information as possible about what the value of the cross section is so that the selection can be driven by my data as much as possible. There is a large industry of discussion about what makes for a satisfactory least informative prior. Often it is merely a flat distribution, in which every value of the parameter is assigned equal likelihood. It's the kind of thinking that makes us assign a probability of 1/2 to each side of a coin flip. Barring other information, we consider that there are only two possibilities for the flip (ignoring landing on the edge, for instance) and assign each case an equal value. There are other prior probability distributions, but I'll leave such a discussion for another time.

Going back to the integral, what is going on with each term? The first term on the right, the likelihood, depends upon the number of data events. The more data I have, the more I restrict the possible number of events. This dependence upon the amount of data is often called the *statistical uncertainty*.

The other distributions for the background and the efficiency do two things. First, by determining at least one value of the background as possible, the value of the cross section is being determined. Second, as the range of possible values increases, the range of possible values of the cross section increases. If the background and the efficiency are single-valued, the integrals are very easy to do:

$$P(\sigma|d_Z) = P(d_Z|\sigma, \bar{\epsilon}, \bar{b}, I)P(\sigma|I) \tag{9}$$

The only remaining dependence of the probability of the cross section is on the amount of data we have. There is only statistical uncertainty and no systematic uncertainty.

What happens if I don't actually know the background very well? If the background is very small, the addition it makes to the number of data events is also very small, so even a large uncertainty in the value of the background won't matter much to the determination of the cross section. However, if the background is significant, and the uncertainty is large, then the number of observed data events has a large contribution from the background and a large uncertainty due to our not knowing the background very well. The effects of not knowing the values of the nuisance parameters well is often called the *systematic uncertainty*.

Imagine it this way. You are a painter painting a room. The walls are white and freshly painted. There are a window and a door that need to have the trim freshly painted. Red trim on the door and blue trim on the window. (It's patriotic.) For the window trim, you're given a small, 1/8" brush usually used to paint landscapes on canvases. Painting the window trim takes a very long time, but when you are done, the edges where the trim meets the wall are perfect. Not a drop or smudge of blue paint is on the bright white wall. Now, for the red trim on the door, you're given an 18" wide, deep knapp wall roller. You finish the painting job in about 45 seconds, but the doorway now looks like the entrance to a slaughter house. Red paint is smeared all over the wall where it meets the moulding. Whereas the window trim kept it's three inch wide appearance even though you painted it blue, the trim around the door now looks 18" wide. You can't really see the original trim that you were supposed to have painted. The red paint has completely overwhelmed it. The statistical uncertainty quantified by the likelihood is like moulding and the effect of the systematic

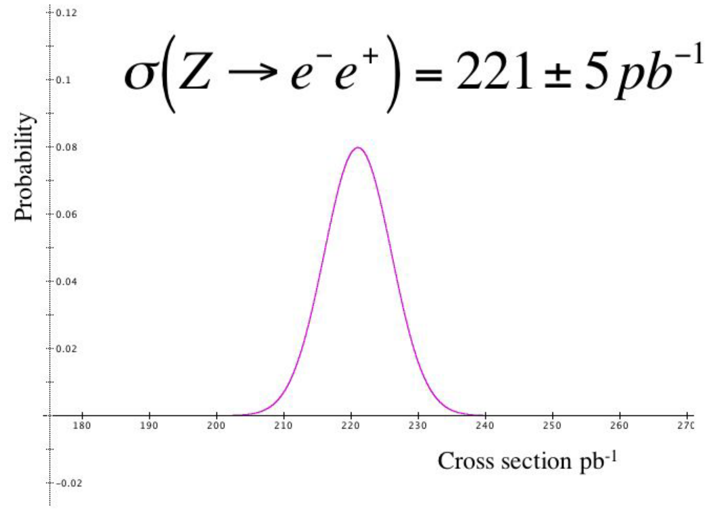

Figure 12: Probability distribution for the cross section

uncertainty is like the paint. The size and shape of the window moulding is undisturbed by the blue paint. The door moulding disappears behind the red paint.

The integral here is a formal way for me to fold my understanding of how I got me number of events into what I've learned from the counting itself. So, here's the final value for the measurement. After all this talk about probability distributions, what is $\sigma(Z \rightarrow ee) = 221 \pm 5pb^{-1}$? Figure 2.5 shows the final probabiltiy distribution, taking everything into account. As you can see, the value 221 is the average value of the cross section, and in our case, the most probable value. The ± 5 is the standard deviation of the probability distribution. On the plot, the range between 216 and 226 pb^{-1} comprises a region of 68% confidence that the cross section lies in that range. They are both statistics used to summarize an underlying probability distribution. They are summary statistics for quantifying how well we know the thing measured, the cross section, given the data we've obtained and everything we know about our detector and how it affects how the data is obtained. The numbers quoted don't tell us what the cross section is, they tell us how well we know it, given the things we have assumed in making the measurement.

Before concluding the story, I should note two things. First, as a matter of completeness, the actual uncertainty on the cross section is a little more than twice the 5 pb-1 that I quote to you because the luminosity (remember that? the number of rocks being thrown?) is acutally known much more poorly than the other systematic uncertainties in the measurement. Secondly, I should note that probability as I used it here is not the only version or account of probability there is. In particular, my claim that probability is a quantification of certitude or an amount of knowledge is not without controversy. The other primary version of probability theory considers probabilities long-run frequency distributions for the occurance of events, not quantifications of what I know or want to know. The results of this measurement and most others in fact, don't change in any signifcant manner whether you use one or the other version of probability theory. Some important results are different, but I will have to save that for another lecture.

So, that is the end of my little narrative. The number $\sigma(Z \to ee) = 221 \pm 5pb^{-1}$ lived happily until another more precise measurement took its place.

3 Reflections

I said at the beginning that we do experiments because they are convincing, because they are authoritative, because we learn something about the world. That we *know* because of experiments.

Experiments are authoritative because experiments tell us about the world as we examine it. Whether, counting the number of electron/positron pairs in proton/antiproton collisions or lifting a corked tube filled with water high above the pendulum pit, we are examining something happening in the world. How is this convincing us of anything? We come to the examination with a notion of how it works, how the data comes to us. In the case of the corked tube, we know that when we corked it, there wasn't any air in the tube. Actually, if there is air in the tube, it is negligible – so little, we don't need to account for it. We also know that we corked it well – it won't leak. At least, the probability of it leaking very much is very small. We see the water level with our own detectors, our own eyes. As we raise the tube past 32 feet above the ground, the water level in the tube stops rising with the tube. It could be explained by air getting into the tube, but that's unlikely. We corked it well. We can perform some leak tests, and quantify the maximum leak rate, if we like. Still, we find probability of a high rate of leaking air is small. Other things have to be true about our seeing the water level. The lights have to be on – there needs to be enough data (light scattering from the water and the tube) for us to detect the level of the water moving up with the tube and then stopping.

We have to be close enough to the tube in order to resolve the difference between the tube and the water with our detectors, our eyes.

We also have hypothesis that doesn't require air to leak into the tube – Pascal's. Pascal's hypothesis is that the water is held up in the tube by pressure on the water from outside and that the amount of that pressure determines the amount of water that can be held up in the tube. That hypothesis becomes more and more likely as all the pieces fit together – that is, we know enough about the range of possibilities of each piece of our observation to add up to the hypotheses being probable. If we are still skeptical, we might perform other sorts of measurements to see if Pascal's theory holds for other situations. Still, as far as the tube of water is concerned, Pascal's hypothesis works. In the end, we consider experimental complications, a leaky cork for instance, to be very unlikely in accounting for what we've seen. We also have a hypothesis with very little wiggleroom. We can measure the volume of water well. We can measure the height of the tube fairly well. We can measure the air pressure in the room. All of which constrain how much flexibility there is in the explanation; all of which constrain how well I know what I've measured. All of which limit my wiggling.

This experiment is so convincing that if we've never seen the demonstration, we ooo and ahh upon witnessing it. Shooting the monkey is like this too. You may never have considered that a ball dropped from you hand falls to the ground just as fast as that same ball fired from a cannon parallel to the ground does. But, when you see someone stand on a table, aim a dart gun straight at an unsuspecting stuffed monkey dangling above the floor, shoot the gun, releasing the monkey at the same moment and you see the dart hit that monkey right between the eyes at it falls to the floor, you will be convinced.

On the other hand, what's wrong with the chemsitry experiments? Why are they less convincing of anything? Part of the problem is that they rely on making precise measurements of amounts of stuff. As we do them here at St. John's, we pay fairly little attention to quantifying how precise we're being, but we have a gut feeling (a correct gut feeling) that we just aren't being very precise as a general rule. There are large uncertainties in making any of the measurements; we just don't know as much as well. As originally performed, the measurements were notoriously difficult. Systematics uncertainties constantly plagued them. Furthermore, we usually just don't pay close attention to these uncertainties and don't keep track of them. So, often we're left a little confused and maybe a great deal uncertain for good reason – we haven't constrained our certainty very much at all. Furthermore, regarding the atomic nature of matter, we may be just a little more skeptical because we can't see the bits of matter like we can see the water level in Pascal's tube. The analog might be if we were to break apart some matter into smaller and smaller pieces until we couldn't break them apart any more, we might be more convinced. We can't do that, so we're left testing some hypotheses and checking and rechecking how well we know each of the pieces. Once we are convinced that our mass measurements are precise, we can look at the ratios of numbers themselves and see if our hypothesis is correct. Being convinced by the conclusion means being convinced by the parts.

In a nutshell, what makes an experiment convincing is having very little wiggle-room, whether that wiggleroom be in the technical components of the measurement – the detectors for instance – or in the amount of data we have, or in the character of our expectations. Making a measurement means quantifying that wiggleroom. We quantify wiggleroom with probability. We make the inductions about what is behind out hypothises using probability, particularly Bayes' Theorem. We do this in generating the results of expensive, complicated, difficult particle physics experiments when we measure a physical quantity. We also do this in everyday speech when we quantify our expectations as in "I'll probably enjoy the lecture". Saying so really means that there's more than a 50% likelihood that you'll enjoy the lecture. Hopefully, the experiment you have made tonight has lined up with that probability. Thank you and good night.