Lecture 1

INTRODUCTION AND BACKGROUND

Let's start out by putting our motto on the board:

"PROBABILITY THEORY IS NOTHING BUT COMMON SENSE
REDUCED TO CALCULATION" (Laplace).

This is the motto and this is the exact summary of everything I'm going to tell you in all these talks.

Our main concern is with applications of probability theory, but we're going to have to spend some time on foundations of probability theory for a very simple reason. Before you can apply any theory to any problem, you first have to make the decision that the theory applies to the problem. It turns out that this is not always an easy decision to make. In most of the problems in science and engineering where you might think of using probability theory, your decision as to whether its use is really justified can depend entirely on how you approach the fundamentals of probability theory itself. In other words, what do we mean by probability? Before we can discuss any applications, we'll have to make up our minds about that.

My purpose in these talks is to show that, with a little different approach to fundamentals than the one usually given nowadays, we can extend the range of practical problems where probability theory can be used, and in some known applications we can simplify the calculations.

1.1 Historical Remarks*

Before going into details a few historical remarks might be of interest, to show how it could happen that a person who is a rather strange mixture of theoretical physicist and electrical engineer could get really worried about the foundations of probability theory. The things that I'm going to talk about here arose from my attempts, over a period of ten years, to understand what statistical mechanics is all about and how it is related to communication theory. In 1948 I was very fortunate in being a graduate student in Princeton, and I took a course in statistical mechanics from Professor Eugene Wigner, who went very carefully into the various approaches to statistical mechanics and in particular, pointed out the unsolved problems that still existed. I was impressed by the fact that everyone who has written about the fundamentals has a very ready way of resolving all the famous paradoxes; but that no two people have done this in the same way.

It was just during this year that Shannon's papers (Shannon, 1948)**, announcing the birth of information theory, appeared. I discovered them accidentally in the Princeton library, took them back to my room, and disappeared from the face of the earth for about a week. When I finally came out, I ran through the halls of Princeton explaining to anybody who would listen to me (and a few who wouldn't) that this was the most important piece of work done by any scientist since the discovery of the Dirac equation. It's almost impossible to describe the psychological effect of seeing our old familiar expression for entropy derived in a completely new way, and

---

*This and the following Section describe the history and motivation of the work reported. The reader who does not care about this and wants to get on with the constructive development can turn immediately to Lecture 2.

**Insertions of this type refer to the General Bibliography in Appendix A.

then applied to problems of engineering which apparently have no relation to thermodynamics. But all of the inequalities, which are often associated with the second law of thermodynamics, turn out also to be statements of the greatest significance in an entirely different context. It seemed to me that there must be something pretty important that we could learn from this situation.

This feeling was shared by a number of physicists and there was quite a rush to exploit all these wonderful new things. But then something went wrong. Quite a few papers appeared in the physics journals inspired by Shannon's work, but there was a scarcity of new results useful to physics. This caused a psychological reaction, and by 1956 Information Theory had acquired a bad reputation among physicists.

I think the time has come now when physicists might find it worthwhile to take a sober second look at Information Theory and what it can do for them. And with the benefit of hindsight, we can see what went wrong in those first few years. The first efforts were based only on a mathematical analogy between statistical mechanics and communication theory, in which the appearance of the same mathematical expression was the dramatic thing. The essential link between them--the thing I want to try to show here--is not one of mathematics, but something more subtle. Until you see what the link is, you can't expect to get results out of this situation. Now let's see why this is so.

The mere fact that a mathematical expression like

$$\Sigma p_i \log p_i$$

shows up in two different fields, and that the same inequalities are used, doesn't in itself establish any connection between the fields. Because after all,

$$e^x, \qquad \cos\theta, \qquad J_o(z)$$

are expressions that show up in every part of physics and engineering. Every place they show up, the same equalities and the same inequalities turn out to be useful. Nobody interprets this as showing that there is some deep profound connection between, say, bridge building and meson theory. The reason for that is the underlying ideas are entirely different.

Now the essential content of both statistical mechanics and communication theory, of course, does not lie in the equations; it lies in the ideas that lead to those equations. And at first glance there doesn't seem to be any relation at all between the kind of reasoning that the physicists go through in statistical mechanics and the kind of reasoning that Shannon went through. We might describe this by paraphrasing a statement of Albert Einstein (Einstein, 1946) that I like very much: <u>Science is fully justified in identifying these fields only after the equality of mathematical methods has been reduced to an equality of the real nature of the concepts</u>. You recall that Einstein insisted on exactly this point in connection with gravitational and inertial mass. It had been known, for 200 years before Einstein was born, that gravitational mass and inertial mass were experimentally proportional to each other; by proper choice of units you can make them numerically equal. Einstein refused to identify them; i.e. to accept this empirical equality as a general principle of physics, until he could reduce inertial mass and gravitational mass to the same <u>concept</u>. He had to pay a rather high price to do this. Before he could find a viewpoint from which he saw them as special cases of the same idea, he had to invent General Relativity.

It is interesting to note that this principle was appreciated equally well by J. Willard Gibbs, many years earlier. In his response to the American Academy of Arts and Sciences of Boston, on the occasion of his being awarded the Rumford Medal (January 12, 1881), Gibbs remarked: "One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity." Gibbs had shown in his famous work of 1878 that classical thermodynamics appears particularly simple if we regard entropy as the fundamental quantity; from its dependence on energy, volume, and mole numbers all thermodynamic properties of a system are determined.

These examples could be used with profit in all parts of science. We won't commit any serious error of methodology if we try to follow the examples of Gibbs and Einstein in our problem, because it's really a very similar sort of thing. So the job as I saw it was not to try to invent any new fancy mathematics. That would presumably come later if we were successful; but the immediate job was to try to find a viewpoint from which we could see that the reasoning behind communication theory and statistical mechanics was really the same. As it turns out, to do this requires a rather drastic reinterpretation of both fields; and this reinterpretation clears up several outstanding difficulties in each field.

1.2 The Gibbs Model.

Now to state the problem a little more specifically, I'd like to go very briefly into the version of statistical mechanics that Gibbs gave us (Gibbs, 1902), and try to show the sense in which my work is not only an attempt to generalize his theory, but also an attempt to make use of another lesson in methodology which he gave to science.

Most of the discussions about the foundations of statistical mechanics consist of Mr. A criticizing the basic assumptions of Mr. B and this process is always fruitless and inconclusive. It never leads to any useful results.

However, there is one person who has kept free of that, and his name is J. Willard Gibbs. I think of all people who have written on statistical mechanics, he is the only person who has stayed above this kind of criticism. He did this by a very clever trick. He avoided criticism of his assumptions by not making any assumptions, and by pointing this out to the reader in the preface to his book.

Gibbs simply constructed models in which he assigned certain probabilities for certain situations, and in introducing them he did not say a word about why he chose those particular probabilities. In the preface he tells us that the reason for this has something to do with difficulties which the theory faced in his day, and in particular he mentioned the fact that the experimental specific heat of diatomic gases comes out only 5/7 of what he expected it to be on the basis of his theory. There are a few other difficulties. The paradox about entropy of mixing, for example, and the fact that his theory failed to predict the actual values of equilibrium constants and vapor pressures until you added still more assumptions.

I like to think that there is another reason why Gibbs operated this way. It was maybe even more compelling that the temporary difficulties. Of course, all those difficulties we recognize today as signaling the first clues to the quantum theory. We all know that Gibbs was a very shrewd old gentleman who was a master of science as it existed in his day. I think he was equally well a master of psychology. He realized that the physics of his day and the probability theory in his day didn't provide any really convincing arguments to justify the probability assignment of his canonical ensemble in terms of more fundamental things. And yet, his work had shown that it had all the formal properties which convinced him that it must be right. It clearly was the neatest, most elegant, and simplest way of describing thermodynamics.

Suppose you were in a situation like that. Which is the best way to proceed? I think Gibbs said to himself, "If I try to say a single word to justify this canonical distribution, if I try to invent any argument to back it up, then almost everybody who reads this work will conclude, quite irrationally, that the validity of my equations depends on the validity of those arguments. But I know in my bones that this theory is right independently of any arguments I am now able to give, because it has formal properties which make it superior to any other. So I will say as much as possible about what I know, and as little as possible about what I don't know. The real justification will have to come later." So he simply introduced his canonical ensemble by entitling a chapter "On the Distribution in Phase called Canonical, in which the Index of Probability is a Linear Function of Energy," and that was it. He goes right on into the discussion.

So you can't say to Gibbs, "How do you know that this is the right probability distribution?" He'd be perfectly justified by answering something like this: "I didn't say it was the right probability distribution, and I'm not sure the question has any meaning. I'm simply constructing a model for my own amusement. My canonical probability assignment is not derived from anything, it's not an assumption about anything. It's a <u>definition</u> of which model I propose to study. After this model is set up, we can compare its predictions with experimental facts and see how far this model is able to reproduce thermodynamic properties of systems. If the model turns out to be successful, then it will be worthwhile to consider whether, and in what sense, we might consider it to be correct."

I think that's a very clever attitude to take - it avoids so much useless argumentation. It's a good example also of the methodology we really have to use in all theoretical physics. If we had to be sure we were right before starting a study, we would just never be able to do anything at all. We have to start out by arbitrarily inventing something, some model, which we don't attempt to justify in terms of anything deeper at the time, and see where it leads us. Every once in a while we find that we can invent a model which has very great success in reproducing observed phenomena, and whenever this happens we get convinced that there must be some deeper reason why this model is correct. Then we repeat the process. We try to invent another model operating at some deeper level, from which we can deduce the features of our old model. The exciting thing about this is that when we finally succeed, we always find that the new model is much simpler than the old model, but at the same time is much more general.

There are all sorts of examples of this in the history of science which you all know about; for example, in electromagnetic theory, the experimentalists had produced a large number of separate equations and rules of thumb--the work of Coulomb, Ampere, Faraday, Henry, and so on. And then these were all summed up in Maxwell's equations. Maxwell's equations are much simpler than this series of models which they replaced; but still they are more general, and predicted new phenomena which the experimentalists hadn't found. In fact, Maxwell's equations proved to be so general that to this day, a century later, they still provide the theoretical basis for all of electrical engineering.

Perhaps the best example of all is the tremendous complication which spectroscopy got into by the early 1920's. All the rules of thumb that were developed in predicting what spectral lines would occur and which ones would not, estimating where they would be, and so on. These rules of thumb were quite successful, of course. You could use them for practical prediction. But then we have the Schrödinger equation, which suddenly in a single differential equation says everything that all these rules ever said, and much more; so much more that we are still finding new things from it.

How has the Gibbs model fared? We've had it for 70 years now. It has fared very well, except for these minor changes which have something to do with quantum theory. We find that in every case where you can work out the mathematics, the model has been successful in reproducing observed properties of matter in the limiting case of thermal equilibrium. There are some equilibrium cases where the mathematics is rather resistant to calculation, particularly the phenomenon of condensation; and we don't really know whether the Gibbs model exhibits condensation for general attractive forces, in the sense of being able to prove it rigorously. But I don't think anyone doubts that the Gibbs model would be successful here if we were just better mathematicians than we are. So for the sake of the argument, let's just grant that the Gibbs model has turned out to be completely successful in reproducing all features of equilibrium thermodynamics.

Because of its success, naturally, attempts would be made to justify the Gibbs model in terms of something deeper. Unfortunately, these attempts do not seem to have been successful; at least I don't think there is a single one of them which is so considered by any clear majority of the physicists who worry about these things.

It hasn't been easy to get rid of the idea that the ultimate justification of the Gibbs model must be found somehow in the laws of physics. By this we mean particularly, say, the Schrödinger equation or the Hamiltonian equations of motion on a microscopic level. For this reason you have this enormous amount of work that has been expended on "ergodic" approaches to statistical mechanics, in which we tried to prove that the time average of some quantity for a single system would, in consequence of the equations of motion, be equal to an average over the Gibbs ensemble. But the results of this approach have remained inconclusive, and it has done nothing to extend the Gibbs model to more general situations, as real advances in understanding always do.

More specifically, while the ergodic arguments have led to a number of important theorems (such as reduction of the original problem to that of metric transitivity), they have led to no definite conclusions proved applicable to real physical systems even in the equilibrium case; and they have provided no clues as to how a general theory of irreversible processes might be set up.

I don't want to go at this point into any detailed criticism of past attempts to justify the Gibbs model, because that would take a lot of time and would again be one of those fruitless and inconclusive kinds of criticism which leads nowhere. But I'd like to indicate why it seems to me that any appeal to the laws of physics may miss the point. It is simply that the problem is not to justify any statement about physics. The problem is to justify a probability assignment, and you can't deduce probability from certainty. No matter how profound your mathematics is, if you hope to come out eventually with a probability distribution, then some place you have to <u>put in</u> a probability distribution; and nothing in the equations of motion tells you what distribution to put in. They can give you only relations between probabilities, at different times.

You might note that this argument has nothing to do with whether we're considering classical or quantum statistical mechanics. In classical theory we have our precisely defined states where we've specified the value of every coordinate and every momentum to arbitrary accuracy, and the equations of motion then determine uniquely what every coordinate and momentum must be at some other time. In quantum theory we don't use that method of description, but we still have our precisely defined states. They now are points in a linear vector space, or Hilbert space, whose motion is uniquely determined by the Schrödinger equation.

The analogy goes a good deal deeper; Liouville's theorem in the classical case finds its analog in the fact that in quantum theory the equations of motion induce a unitary transformation, which is therefore a measure-preserving transformation, in the Hilbert space. The fact that the total phase volume below a certain energy is finite in the classical case, has its analog in the fact that the linear manifold spanned by all eigenfunctions of the Hamiltonian with energies below a certain value, is a finite-dimensional vector space. These are about the only properties which are actually used in the ergodic arguments. Therefore practically everything that has been said about these problems in classical statistical mechanics carries over immediately to quantum theory.

One of our major objectives is to justify the Gibbs canonical probability distribution in terms of something more fundamental. The only thing we could accomplish by applying the laws of physics is that we could carry out transformations and express the <u>same</u> distribution in terms of some other parameters. But the distribution of Gibbs is already as simple as any we could hope to get in this way, and afterwards we would still be faced with exactly the same problem; to justify some probability assignment.

It seems to me that if we're ever going to justify the Gibbs model in any meaningful way, we'll have to justify it directly on its own merits, without considering the laws of physics at all. In other words, the problem is to find a viewpoint from which we can see that the Gibbs model, and Shannon's model of a communication process, are special cases of a general <u>method of reasoning</u>.

In the next two lectures, we're going to take what may seem like a rather long detour, and study the general problem of plausible reasoning--also known by the more highbrow, and more restrictive, name of inductive reasoning (I'm not going to bother to distinguish between these terms).

But if you'll bear with me, I think you'll find that we can give, not quite rigorous theorems, but very powerful heuristic arguments, which indicate what this more general viewpoint is.