PREDICTIVE STATISTICAL MECHANICS

E. T. Jaynes*

St. John's College and Cavendish Laboratory
Cambridge CB2 1TP, United Kingdom

INTRODUCTION

This workshop is concerned with two topics, foundations of quantum theory and of irreversible statistical mechanics, which might appear quite different. Yet the current problems in both fields are basically the same, two different aspects of a deep conceptual hangup that permeates not only physics, but all fields that use probability theory.

A different way of thinking about these problems is expounded, which has had useful results recently in statistical mechanics and more general problems of inference, and which we hope may prove useful in quantum theory. An adequate account of all the technical details alluded to in the writer's five talks would require a volume in itself, but much of this is now in print or in the publication pipeline. Therefore we try to explain here the original motivation in quantum theory, the formalism that evolved from it, and some recent applications, with references to further details.

QUANTUM THEORY

We think it unlikely that the role of probability in quantum theory will be understood until it is generally understood in classical theory and in applications outside of physics. Indeed, our fifty-year-old bemusement over the notion of state reduction in the quantum-mechanical theory of measurement need not surprise us when we note that today, in all applications of probability theory, basically the same controversy rages over whether our probabilities represent real situations, or only incomplete human knowledge.

If the wave function of an electron is an "objective" thing, representing a real physical situation, then it would be mystical--indeed, it would require a belief in psychokinesis--to suppose that the wave function can change, in violation of the equations of motion, merely because information has been perceived by a human mind.

---

*Visiting Fellow, 1983-84. Permanent address: Department of Physics, Washington University, St. Louis, Missouri 63130, U.S.A.

If the wave function is only "subjective", representing a state of knowledge about the electron, then this difficulty disappears; of course, by definition, it will change with every change in our state of knowledge, whether derived from the equations of motion or from any other source of information. But then a new difficulty appears; the relative phases of the wave function at different points have not been determined by our information; yet they determine how the electron moves.

There is no way quantum theory could have escaped this dilemma, short of avoiding the use of probability altogether. Not only in Physics, but also in Statistics, Engineering, Chemistry, Biology, Psychology, and Economics, the nature of the calculations you make, the information you allow yourself to use, and the results you get, depend on what stand you choose to take on this surprisingly divisive issue: are probabilities "real"?

But in quantum theory the dilemma is more acute because it does not seen to be merely a choice between two alternatives. The "subjective" and "objective" aspects are scrambled together in the wave function of an electron, in such a way that we are faced with a paradox like the classical paradoxes of logic; whatever stand you take about the meaning of the wave function, it will lead to unacceptable consequences.

To achieve a rational interpretation we need to disentangle these aspects of quantum theory so the "subjective" things can change with our state of knowledge while the "objective" ones remain determined by the equations of motion. But to date nobody has seen how to do this; it is more subtle than merely separating into amplitudes and phases.

WHAT IS "REALITY"?

As many have pointed out, starting with Einstein and Schrödinger fifty years ago and continuing into several talks at this Workshop, the Copenhagen interpretation of quantum theory not only denies the existence of causal mechanisms for physical phenomena; it denies the existence of an "objectively real" world.

But surely, the existence of that world is the primary experimental fact of all, without which there would be no point to physics or any other science; and for which we all receive new evidence every waking minute of our lives. This direct evidence of our senses is vastly more cogent than are any of the deviously indirect experiments that are cited as evidence for the Copenhagen interpretation.

Perhaps our concern should be not with hidden variables, but hidden assumptions; not only about the theory, but about what we are measuring in those experiments. Consider a cascade decay experiment. As soon as we say something like "In this experiment we observe two photons emitted from the same atom", we have already assumed the correctness of a great deal of the theory that the experiment was supposed to test. This initial stacking of the cards then affects how we analyze the data.

Bell's theorem was a great, and startling, advance, because it showed us that von Neumann's analysis contained hidden assumptions that had escaped the notice of physicists for over 30 years. But it is not so startling when we note that relativity theory owes its existence to Einstein's perception of a hidden assumption that nobody had noticed in the 300 years since Kepler. In 1949 he wrote, concerning phenomena in two reference frames: "Today everyone knows, of course, that all attempts to resolve this paradox were doomed to failure as long as the axiom of the

absolute character of time, viz., of simultaneity, was anchored unrecognized in the subconscious. Clearly to recognize this axiom and its arbitrary character really implies already the solution of the problem."

We hope it will not take another 300 years to locate the hidden assumptions in Bell's analysis. This is unlikely to be accomplished within the confines of present orthodox thinking; we can hardly expect that a viewpoint which denies the very existence of causal mechanisms will suggest the proper experiment to reveal them.

But to escape from the irrationality of present quantum theory, it is idle merely to complain about its philosophy. The onus is on the dissenter to move outside its area of thought and offer a constructive, experimentally testable, alternative.

The writer's neoclassical theory of electrodynamics is one such effort; but it has not yet been applied to experiments of the type reported here by Aspect, in a sufficiently careful way to draw any conclusions. In any event, many more such efforts may be needed before we can ferret out the crucial hidden assumption, now anchored unrecognized in our subconscious, which is preventing us from tying the mathematics of quantum theory to a rational view of the world.

Do we need more hidden variables? Perhaps eventually, but maybe our immediate problem is the opposite; first we need to get rid of some. To define the state of a classical particle we must specify three coordinates and three velocities. Quantum theory, while denying that even these degrees of freedom are meaningful and claiming to "remove unobservables", replaces them with an infinite number of degrees of freedom defining a continuous wave field. To specify a classical wave field, we need one complex amplitude for each mode; for a quantized field we need an infinity of complex amplitudes for each mode.

So perhaps quantum theory, far from removing unobservables, has introduced infinitely more mathematical degrees of freedom than are actually needed to represent physical phenomena. If so, it would not be surprising if a few infinities leak out into our calculations.

Neoclassical theory was a preliminary attempt to remove the hidden variables that are not actually used (at least, not up to second order) in calculation of the Lamb shift, the anomalous moment, and vacuum polarization. After removing an infinity of irrelevant degrees of freedom, it might be clearer how to add a few new relevant ones. It is not yet completely formulated, and can still be modified in many ways.

Distant correlations, so prominent in all these discussions from EPR to Aspect, are not in themselves difficult conceptually, since a state of knowledge may well be of the form; "if A, then B; but if C, then D". Experiments which merely confirm that these correlations exist hardly surprise us (they would be sensational if they showed that the correlations do not exist). As EPR emphasized, it is not the experimental facts, but the claim of the Copenhagen theory that the wave function containing such correlations is <u>at the same time</u> a complete description of reality, on which one chokes.

Part of Bohr's defense of his position was on the grounds that EPR did not show that his theory contradicts any experimental fact; but that was not the point. Indeed, the difficulty with the Copenhagen theory is most acute just when the experimental facts agree exactly with its predictions; for then we are in the following situation. Since the experimenter can decide, after the interaction with a system S has ceased, which of two

non-commuting quantities (x,p) in S he will be able to predict with certainty, we are again in the realm of psychokinesis if we deny reality to either of them before his decision. Then, since the wave function does not determine either x or p we do not see, any more than did EPR, how it can be called a complete description of reality.

The word "reality" must have had a different meaning to Niels Bohr than it has to others. On this topic, we note a quotation from Heisenberg (1958) which will surprise many: "The conception of objective reality--has thus evaporated--into the transparent clarity of a mathematics that represents no longer the behavior of elementary particles, but rather our knowledge of this behavior."

OUR MIND-BOGGLING PROBLEM

The difficulty in finding a rational interpretation of quantum theory can be illustrated a little more specifically as follows. If one wishes to do so, the amplitudes $|a(t)|$ of stationary states in a wave function

$$\psi(x,t) = \sum a_n(t)u_n(x)$$

can be interpreted, with at least some success, as "subjective" probability amplitudes expressing human information; but they are combined with phases which seem to have no such interpretation. Yet the relative phases strongly affect our predictions in ways that are necessary for agreement with experiment (for example, they determine the polarization of resonance radiation). We seem obliged to consider the relative phases as "objectively real" things.

But then those same phases cause conceptual difficulties; in Schrödinger's paradox, if we observe whether the cat is alive or dead, quantum theory gives reasonable probability statements. But if we choose to measure something that does not commute with "liveness", our predictions will depend on the relative phase with which the live cat and dead cat are superposed in the wave function.

There must be some limit on how far the superposition principle can be applied; at the atomic level we must have it, but at the macroscopic level we would like to get rid of it. Even at the molecular level, it is puzzling; if we take it literally, what prevents us from having a new kind of air in which each molecule is neither nitrogen nor oxygen, but a linear combination of them?

As noted, a mere separation into amplitudes and phases is not enough; further thought convinces us that the amplitudes must also be in part "objective", reviving the difficulty with state reduction. What seems mind-boggling at present is: how can we find a separation into "subjective" and "objective" parts that is invariant under changes in the representation--or at least under a sufficiently wide group of transformations to make physical sense? We do not have a clue about how to do this.

Yet we are optimistic, think that it will surely be solved eventually, and the answer will be simple and obvious (although it may require us to renounce our linear Hilbert space). As Seneca wrote long ago, "Posterity will be astonished that truths so clear had escaped us."

But for the present, to paraphrase Gibbs: difficulties of this kind have deterred the writer from trying to explain the mysteries of quantum theory, and forced him to be content with a more modest goal.

Now let's look at that mind-boggling problem from a different side. A single mathematical quantity ψ cannot, in our view, represent incomplete human knowledge and be at the same time a complete description of reality. But it might be possible to accomplish Bohr's objective in a different way. What he really wanted to do, we conjecture, is only to develop a theory which takes into account the fact that the necessary disturbance of a system by the act of measurement limits the information we can acquire, and therefore the predictions we can make. This was the point always stressed in his semipopular expositions. Also, in his reply to EPR he noted that, while there is no physical influence on S, there is still an influence on the kinds of predictions we can make about S.

With all of this, we can agree entirely. The fact of disturbance of one measurement by another was equally true in classical physics (for example, one cannot use a voltmeter and ammeter to measure the current and voltage of a resistor simultaneously, because of this "complementarity": however you connect them, either the voltmeter reads the potential drop across the ammeter or the ammeter reads the current through the voltmeter).

But in classical physics such limitations on our knowledge could be recognized and taken into account in our predictions without losing our hold on reality; for the separation into what was "objective" and what was "subjective" was never in doubt. The coordinates and velocities remained "objective", while the "subjective" human information resided entirely in the probability distributions over them. The probabilities could vary in any way as our state of knowledge changed for whatever reason; while the coordinates and velocities continued to obey the equations of motion.

Furthermore, the limitations on our ability to make measurements at the microscopic level did not prevent us from discovering the microscopic equations of motion, or from checking them accurately enough to discover their failure in quantum effects. There are lessons in this for the present.

Could we make a theory of microscopic phenomena more like this which, while keeping a firm hold on what is "objective", also recognizes and represents explicitly the role of limited human information in the predictions it can make? Such a theory need not, we think, contradict the successful parts of Bohr's theory; rather it would remove the contradictions that still mar it, thus fully realizing Bohr's goal (indeed, Bohr himself may have had such thoughts; it is known that toward the end of his life he showed an interest in information theory).

Not knowing how to make this separation in quantum theory, we sought first a simpler model in the hope that it might be useful in its own right, and also give some clues for the big problem. The first hope is now fully realized; the second is still in too delicate a condition to be put on public view.

We want to reformulate statistical mechanics in a way that reflects Bohr's thinking by introducing explicitly the role of human information in determining the kind of predictions we can make. This is also closely related to Einstein's thinking; for he regarded present quantum theory as incomplete, related to a complete microscopic theory in much the same way that classical statistical mechanics is related to classical mechanics.

Thus, to enter into pure conjecture (whistling in the dark, some will say); if we could eventually see the predictions of present quantum theory as resulting from the "statistical mechanics" of a deeper theory, this might

realize both Bohr's goal and Einstein's. Of course, in view of Bell's theorem we do not expect this to be possible exactly, in terms of the same variables used in present quantum theory.

NATURE OF PROBABILITY THEORY

In trying to use probability theory for this purpose, we note that one can find very different views as to what probability theory is. It has been termed:

"the theory of additive measure" (Kolmogorov)
"the theory of rational belief" (Jeffreys, Good, Savage)
"the theory of frequencies in random experiments" (Fisher)
"the calculus of inductive reasoning" (de Morgan)
"common sense reduced to calculation" (Laplace)
"the art of conjecture" (Jacob Bernoulli)
"the exact science of mass phenomena and repetitive events" (von Mises)

For over a Century, controversy has raged between those who want us to adopt one of these views and reject others. Clearly, however, each of the above definitions merely reflects the particular problems that the author was concerned with; to insist that we reject all views that are not helpful in his problems is to insist that we work only on his problems.

Here we take the view that all of the above are valid and useful in different contexts, and we are free to choose whichever seems appropriate in ours. But some views are more general than others. Consider, for example, that of von Mises. It is true that probability theory is used successfully in dealing with mass phenomena and repetitive events; but if we insist that it may be used only for that purpose we shall be prohibited from using it for our purpose of representing human information.

On the other hand, the probability theory of Jeffreys (1961) in which probability expresses basically a state of knowledge, applies to a much wider range of problems, automatically including those of von Mises and Fisher; for a state of knowledge may refer to any context. If our problem of interest happens to involve random experiments or mass phenomena, then Jeffreys' theory applies in a natural way, yielding the same or domonstrably better results than those of Fisher and von Mises. For Jeffreys' approach deals easily with technical problems (nuisance parameters, nonexistence of sufficient or ancillary statistics, rectangular likelihood functions, cogent prior information) on which narrower views fail (Jaynes, 1976).

Therefore, while we should not consider it "wrong" to adopt a narrow view if it happens to be adequate for our problem, we have nothing to lose, and may avoid unnecessary technical difficulties, if we always adopt the broadest view of Jeffreys. In our present problem it is necessary to do this, for our explicit purpose is to represent human information of any kind, whether or not it happens to refer to random experiments or mass phenomena.

This choice will avoid not only technical difficulties, but even more disturbing conceptual ones, if our long-range goal is to clarify quantum theory. In our discussions here we have heard such questions as:

"Does the measurement create the state?"
"Does the act of human perception do it?"
"What is the role of consciousness and free will?"
"Are objects real?"

But astonishingly, we have heard also:

"What is the true nonequilibrium ensemble?"

At this point, it is clear that theoretical physics has gone berserk. In quantum theory we have got ourselves into a situation where the objects have become unreal, but the probabilities have become real!

But this too is not peculiar to quantum theory. On deep thought, it will be seen that whenever we allow probabilities to become "physically real" things, logical consistency will force us, eventually, to regard the objects as "unreal". If we are to reach Bohr's goal while at the same time keeping our objects real we must recognize, with Laplace, Maxwell, and Jeffreys, that whenever we use probability it must be as a description of incomplete human knowledge, as it was in classical statistical mechanics.

Therefore, we must go back much further in first principles than merely re-hashing EPR and the QM theory of measurement. We need a basically different way of thinking than scientists are now taught. The conventional attitude asks the question: "How does Nature behave, as determined by the laws of mechanics?" Or, as we heard it put more dramatically here: "I don't like thermo--let's solve the n-body problem and see if all this stuff is really true."

Here we want to ask a different question: "How shall we best think about Nature and most efficiently predict her behavior, given only our incomplete knowledge?" Of course, although it would be impossibly difficult, it would not be "wrong" to solve the n-body problem; the point is that it is unnecessary and would contribute almost nothing to understanding thermo.

To understand and like thermo we need to see it, not as an example of the n-body equations of motion, but as an example of the logic of scientific inference, which by-passes all the detail by going directly from our macroscopic information to the best macroscopic predictions that can be made from that information; a model of what we would like to do in quantum theory.

That this must be possible, at least for thermo, is seen as follows. The fact that thermodynamic and other macroscopic experiments are reproducible shows that most details of those initial microscopic conditions are irrelevant. If control of a small number of macroscopic quantities is sufficient, in the laboratory, to yield a reproducible result, then information about those quantities must suffice for theoretical prediction of that result; there must be an algorithm that goes directly from one to the other.

Predictive Statistical Mechanics is not a physical theory, but a method of reasoning that accomplishes this by finding, not the particular things that the equations of motion say in any particular case, but the general things that they say, in "almost all" cases consistent with our information; for those are the reproducible things.

We are not, however, throwing away the conventional methods or results. Quite the contrary, the statistical mechanics of Gibbs and conventional quantum statistics are contained in it as special cases for certain particular kinds of information, just as the probability theory of von Mises is contained in that of Jeffreys for particular kinds of information.

This concludes our lengthy sermon; the technical problem now before us is: how shall we use probability theory to help us do plausible reasoning in situations where, because of incomplete information, we cannot use deductive reasoning?

A fairly complete treatment of these questions, with extensive physical applications, is given by Jeffreys (1961); we sketch a few of the details. The propositions about which we reason are denoted by letters A, B, C, etc. For example, we might choose

A = "The pressure of the gas is in the range (P, P + dP)"

B = "Its kinetic energy is E"

I = "It has N molecules in a volume V"

As in the usual Boolean algebra, we may construct new propositions by conjunction, disjunction, and negation:

AB = "Both A and B are true"

A + B = "At least one of the propositions A, B is true"

$\overline{A}$ = "A is false"

The symbol $p(A|B)$ stands for "the probability that A is true, given that B is true", and is a real number in $0 \leq p \leq 1$. Thus $p(A+B|CD)$ is the probability that at least one of the propositions A, B, is true, given that both C, D are true, and so on. Since in this system probabilities are a means of representing incomplete information, all probabilities are of necessity conditional on some information; there is no such thing as an "absolute" probability.

The rules for plausible reasoning are simply the familiar product and sum rules:

$$p(AB|I) = p(A|I)p(B|AI) = p(B|I)p(A|BI) \quad , \tag{1}$$

$$p(A|B) + p(\overline{A}|B) = 1 \quad . \tag{2}$$

All other relations can be deduced from these. In particular, if $p(B|I) \neq 0$, we have from (1)

$$p(A|BI) = p(A|I)p(B|AI)/p(B|I) \tag{3}$$

where, since "I" occurs as a condition in all terms, we shall call it the "prior information". Equation (3), usually called Bayes' theorem, represents the process of learning from experience. We start with the prior probability of A, $p(A|I)$ when we know only the prior information, and (3) shows how this is converted into the posterior probability $p(A|BC)$ as a result of acquiring new information B.

Bayes' theorem is undoubtedly the most fundamental principle of scientific inference; by its repeated use we can incorporate long chains of evidence into our reasoning, the posterior probability for one application becoming the prior probability for the next. We readily verify its consistency for this; the final result is independent of the order in which different pieces of information were taken into account. But there has been long controversy about it also, so it is important to dwell a moment on the status of these rules and others derived from them.

Historically, they were given by Laplace in the 18'th Century, on intuitive grounds, and applied by him in many problems of data analysis in astronomy, geodesy, meteorology, population statistics, etc. He had great success with them, using Bayes' theorem to help him decide which astronomical problems were worth working on. That is, are the discrepancies

between calculation and telescopic observation so small that they might well be accounted for by measurement errors; or are they so large that they indicate, with high probability, the existence of some unknown systematic cause not included in the calculations? If so, Laplace would undertake to find that cause. This process (what would be called today a "significance test" by statisticians) led him to some of the most important discoveries in celestial mechanics.

In spite of this success, a reaction set in after Laplace's death as others questioned the validity and uniqueness of the above rules. However, they would be true trivially if we interpret p as an "objective" frequency in a random experiment instead of a mere "subjective" measure of plausibility. So for over a Century probability theory went off into the Fisher/von Mises views that only the frequency interpretation was respectable. In fact, this view succeeded rather well for many years because there were actually many scientific problems where it was adequate; the information available and the questions of interest could be expressed solely in terms of frequencies, and there was no other cogent information.

But as scientific problems became more sophisticated, it became increasingly difficult to adapt frequency interpretations to them. Eventually one had to resort to inventing imaginary universes of repetitions of experiments that could in fact be performed only once, in order to maintain the illusion of a frequency interpretation. More serious, anomalous results began to appear, which could be traced to the failure to take into account cogent information that common sense could see was relevant to the inference, but which the frequency theory could not use because it did not consist of frequency data.

In 1946, R. T. Cox cut through the confusion by a marvelous argument. He had the good sense to ask a constructive question; whether or not Laplace's methods were sound, would it be possible today to make a consistent "calculus of plausible reasoning" along those lines? Cox found that the conditions for consistency of such a theory could be stated in the form of functional equations, whose general solutions could be found. The result was: any method of plausible reasoning in which we represent degrees of plausibility by real numbers is necessarily either equivalent to Laplace's, or inconsistent (in the sense that two methods of calculation, each permitted by the rules, would yield different results). Since 1946, there has been no excuse for anyone to reject the use of equations (1)-(3) as valid scientific inferences. For the details of beautiful applications, see Jeffreys (1961).

THE MASS OF SATURN

Many of these points are illustrated by the famous example of Laplace's analysis of the accuracy with which the mass of Saturn was known at the end of the 18'th Century. Let A stand for the proposition: "the mass of Saturn is in M, M+dM", and denote by D a set of observational data to be taken into account, while I stands for whatever prior information Laplace had about M before the data D were known. Then we may define prior and posterior probability density functions:

$$p(A|I) = f(M)dM \quad ; \qquad p(A|DI) = F(M)dM \quad . \tag{4}$$

Before the data, one did not know much about M, so f(M) was a very broadly spread out function. But Laplace knew at least that M was not zero, else Saturn would not hold its rings and moons or perturb Jupiter; and there would be no data to analyze. Also, he knew that M could not be an appreciable fraction of the solar mass, else Saturn would totally disrupt the

solar system. So f(M) must go to zero at extreme values; but this left an intermediate range of several orders of magnitude, within which f(M) could be taken to be essentially constant.

Of course, f(M) represents only Laplace's state of prior knowledge about M; it can be regarded as a frequency only in a grotesque collection of imaginary universes in which the mass of Saturn takes on all conceivable values. So frequency views of probability would not allow us to do this calculation at all.

In its dependence of M the term $L(M) = p(D|MI)$ in Bayes' theorem is called the likelihood function, proportional to the probability that the data D would be observed if M was the true mass of Saturn. All the evidence from the data resides in this factor, which represents our knowledge of the likely errors in the data due to the imperfection of telescopes and clocks.

This term is worth dwelling on a bit, because it shows both the usefulness and the shortcomings of the (probability = frequency) view. On that view, $p(D|MI)$ means the limiting frequency with which the data set D would be obtained in infinitely many repetitions of the measurement, if the mass of Saturn were held constant at the particular value M. But obviously, if one could actually measure that frequency directly, we would be far beyond any need to estimate M; so how do we, in practice, choose the function L(M)?

For this it is necessary to do a little sub-problem of plausible inference, within the original problem. We suppose that the errors in angles depend only on properties of the telescope, and would be the same whatever the true mass of Saturn and whatever object we are observing. So we shall find the telescope errors by making repeated measurements on some more fixed object like Sirius. The frequency distribution that we find for Sirius is inferred to hold also for Saturn.

This is an inference that common sense leads us to accept at once (although logically it is just the kind of inference that the frequency theory holds to be invalid; frequentists are masters of the art of concealing this). So, in practice we do indeed usually choose the likelihood factor on the basis of frequencies of errors; and the frequentist's procedure does indeed serve our purpose.

But this is so only because the feasible source of information about errors usually happens to come from repeated measurements. What is fundamentally required in (3) is the probability of various errors in the specific case of Saturn now before us. This may or may not be the same as the frequency of various errors in other cases that we are not reasoning about.

Indeed, strictly speaking, the specific case at hand is always in some ways unique and not comparable to others. There may be other cogent information pertaining to the errors in the specific case before us, in addition to frequencies. Then the frequentist's procedure is incomplete, allowing us to take into account only part of the information relevant to our problem. So attempts to uphold frequentist views in all cases can lead to anomalous results, which have become increasingly troublesome in recent applications. The only way to deal with the full problem is to use the Laplace-Jeffreys method, which can in principle take any kind of information into account, because it interprets the concept of probability more broadly.

Likewise, if there is cogent prior information in addition to the data, the Laplace-Jeffreys methods have the means of taking it into account by use of a "informative" prior density f(M) that is not flat, but specifies what values of M are already indicated or contra-indicated by the prior information. Again, frequentist methods which do not admit the existence of a prior probability are helpless to take such information into account.

In recent applications these shortcomings of frequentist methods of inference have become more and more serious. With the advent of the "Generalized Inverse" problem frequentist methods have become totally unusable; yet the Laplace-Jeffreys methods continue to deal with them without any difficulty of principle but with a new technical problem whose solution leads us into predictive statistical mechanics.

GENERALIZED INVERSE PROBLEM

If the likelihood function $L(M) = p(D|MI)$ does not peak sharply at any definite value of M but has a very broad maximum, the frequentist maximum likelihood principle becomes unstable, a small change in the data leading to a large change in the estimate, which common sense rejects as unreasonable. This makes the determination of the prior probabilities a more exacting task; the prior information is no longer overwhelmed by the data, and so it must be considered more carefully than Laplace needed to.

In the limit when the likelihood function develops a flat top, frequentist methods break down entirely; only an informative prior probability can locate a definite estimate somewhere within that flat region. This means that we are faced with a new technical problem: how do we translate verbally stated prior information into a quantitative prior probability assignment?

A typical kind of problem where this occurs is that where we are trying to invert a singular matrix; our data are

$$d_k = \sum_{i=1}^{n} A_{ki}\, x_i \quad , \qquad 1 \leq k \leq m < n \tag{5}$$

where the $\{x_i\}$ represent the "state of Nature" that we are trying to estimate, and A is a known matrix with rank less than n, so there is no inversion of the form $x = A^{-1}d$. Then the likelihood function $L(x) = p(d|xI)$ is rectangular; the data merely partition the set of all x into subsets of possible and impossible values, with nothing to choose within the possible subset. L(x) is only the indicator function of the set of possible states of Nature. So we shall call any problem with a flat-topped likelihood a "generalized inverse problem".

Such problems have proved to be very common in recent applications; after a talk in 1983 the writer was approached by a statistician in Government who said "I suddenly realized that every problem my agency is trying to solve is a generalized inverse problem."

On further reflection it is seen that, from the standpoint of principle, this technical problem is present in every application of probability theory. For, as Cox's derivation showed clearly, the rules (1), (2) express only the consistency of our reasoning, telling us how probabilities of different propositions are related to each other. That is, given some probabilities, they tell us how to calculate others consistent with them. They do not tell us what initial probabilities should be assigned so the calculation can get started.

But for decades probability theory has concerned itself only with building upward from (1), (2), deducing their consequences in the large body of mathematics that fills our libraries. The problem of assigning the initial probabilities represents fully half of probability theory as it is needed for applications; yet its very existence remains unrecognized in most of those books. So we have a great deal of catching up to do; it will require decades to bring this neglected bottom half of probability theory up to the level of power and generality of the top half. But we have made enough progress to date so that the range of useful applications of probability theory is already extended far beyond the confines of the frequentist viewpoint.

ASSIGNING PRIOR PROBABILITIES

Today a number of principles are known by which prior information can be translated, or encoded, into prior probabilities. The simplest and most obvious is symmetry; from the earliest pre-mathematical gropings toward a set of rules for plausible reasoning it was clear to gamblers that if coins, dice, playing cards, roulette wheels, etc. were made with perfect symmetry there was no known cause tending to produce one outcome more than any other; so we had no reason to consider one more likely than another. The only honest way to express this state of knowledge is to assign equal probabilities to all of them.

We hasten to add that this is not to assert that all outcomes must occur equally often, as frequentists invariably accuse us of doing; for of course there may be unknown symmetry-breaking causes (very skillful tossing, shuffling, spinning, etc.). It does, however, mean according to the rules (1), (2) that if we are obliged to predict the frequencies from the incomplete information we have, our "best" estimates, by almost any criterion of "best", will be uniform.

If we have no information about the specific kind of symmetry-breaking influence at work, neither we nor the frequentist can make use of it to alter our predictions. Even if we know that some symmetry breaking is present but do not know which outcome it favors, this information cannot change our estimates, but only increases their probable error.

This again shows the two basically different attitudes noted above; we are not asking how Nature must behave, but only what are the best predictions we can make on our incomplete information. But this is always the real problem before the scientist, as Niels Bohr saw.

The idea of symmetry became more abstract, and more general, in the work of Jacob Bernoulli (1713). He envisaged an underlying population of N inherent possibilities from which we draw (the famous "urn" of elementary probability theory, with its N labelled but otherwise identical balls). If M of the balls are labelled "A" and we draw from the urn blindfolded, the probability that we shall draw an "A" is $p(A) = M/N$, the basic definition of probability used by Laplace. Note that Bernoulli's definition is just the frequency with which we would find "A" if we sampled the entire population without replacement.

But it was recognized already by Bernoulli that in many real problems we do not see how to analyze the situation into ultimate "equally likely" cases. As he put it, "What mortal will ever determine the number of diseases?" A principle was still needed for dealing with cases where we have prior information that makes the known possibilities not equally likely. It is surprising that Laplace did not find this principle, since he saw the problem so clearly.

Now the scene shifts to Boltzman (1877). To determine how gas molecules distribute themselves in a conservative force field such as gravitation, he divided the accessible 6-dimensional phase space of a single molecule into equal cells, with N molecules in the i'th cell. Noting that the number of ways this distribution can be realized is the multinomial coefficient

$$W = N!/N_1!N_2! \ldots N_n! \tag{6}$$

he concluded that the "most probable" distribution is the one that maximizes W subject to the known constraints of his prior knowledge; in this case the total number of particles and total energy:

$$N = \sum N_i = \text{const.} \quad , \qquad E = \sum N_i E_i = \text{const.} \tag{7}$$

where E is the energy of a molecule in the i'th cell. If the number are large, Stirling's approximation gives asymptotically

$$N^{-1} \log W \longrightarrow -\sum (N_i/N) \log (N_i/N) \tag{8}$$

today usually called the "Shannon entropy", although it was in use by von Neumann before Shannon entered the field and by Boltzmann and Gibbs before Shannon was born. This argument, repeated in every statistical mechanics textbook, led to the famous Boltzmann distribution law: our best estimate of N_i is

$$\hat{N}_i = NZ^{-1} \exp(-\beta E_i) \tag{9}$$

where Z is a normalizing factor, and the Lagrange multiplier is found to have the meaning $\beta = (kT)^{-1}$. This is the distribution that can be realized in more ways than can any other that agrees with the information (7). Gibbs generalized this result to the continuous phase space of N interacting molecules, leading to his canonical and grand canonical ensembles, later extended easily to quantum statistics.

At this point, then, (1902) the Principle of Maximum Entropy (PME) was fully in hand and operational, and it has been the de facto foundation tool of statistical mechanics ever since. So why did it take another fifty years to recognize it, ten more to generalize it to nonequilibrium cases, and yet another five to apply it to a problem of inference outside thermodynamics?

The barrier that has held up progress throughout this Century is just that philosophical position that tries to make probabilities "physically real" things. Not until the work of Shannon (1948), which showed that the quantity Boltzmann and Gibbs had maximized is at the same time a unique measure of "amount of uncertainty" in a probability distribution, could the general rationale underlying their procedure be seen. As should be obvious by now, Boltzmann and Gibbs had been, unwittingly, solving the prior probability problem of Bernoulli and Laplace, the principle of symmetry generalized to the principle of maximum entropy.

THE MAXIMUM ENTROPY PRINCIPLE

But for all this time writers on statistical mechanics had interpreted the work of Boltzmann and Gibbs in an entirely different frequentist context, which sought to justify the canonical ensemble as a physical fact, a provable consequence of the equations of motion--that n-body problem--via ergodic theorems.

In the case of a body in equilibrium with a heat bath or isolated, such theorems might conceivably be true in many cases, although one can easily invent counter-examples in which the Hamiltonian is too "simple" for ergodic behavior to be possible in an isolated body. But even when ergodicity can be proved, it cannot ensure that ensemble averages are equal to experimental values, for reasons that were pointed out already by Boltzmann, and discussed in some detail in Jaynes (1967).

When we try to extend the theory to irreversible processes, ergodic theorems are worse than useless, for if the theory is successful the ensemble averages must still be equal to experimental values; but the very phenomena to be predicted express the fact that these are not equal to time averages. The probabilities we use to predict irreversible processes are necessarily only descriptions of human information; i.e., the probability is spread smoothly over the range of possible microstates, compatible with our information.

If a process is reproducible, it cannot be because the probabilities are "real", but only because most of those possible microstates would all lead to the same macroscopic behavior. The maximum-entropy distribution is the safest, most "conservative" distribution to use for prediction, because it spreads the probability out over all the states consistent with our information; and not some arbitrary subset of them. Thus it takes the fullest possible "majority vote", and prevents us from making arbitrary assumptions not justified by our information.

In this sense the maximum-entropy principle expresses nothing more than intellectual honesty; it frankly acknowledges the full extent of our ignorance. Indeed, it measures that ignorance, since the "physical entropy" S of Clausius, which is the "Shannon information entropy" (8) after maximization, is essentially the logarithm of the number of microstates consistent with our information (as Boltzmann, Einstein, and Planck had all noted before 1907).

But in fact the microstates leading to the reproducible behavior are such an overwhelming majority of the possible set that in practice "almost any" probability distribution that concentrates its probability on some subset of them, will lead to substantially the same macroscopic predictions. And any probability distribution--whatever its high-probability domain of microstates--that gives the same covariance functions for the macroscopic quantities of interest, will lead to exactly the same predictions. The maximum-entropy ensemble is sufficient, but very far from necessary, to make correct predictions. Yet some still ask: "What is the true non-equilibrium ensemble?"

But the question can be construed as meaningful in a different sense. If PME predictions are successful, that does not prove its ensemble is "correct" since many other algorithms might also be successful. But if PME fails, the ensemble must have been, in some sense, "wrong". Supposing the data used as constraints were not in error, there seem to be two possibilities: (A) There may be further constraints, essential to determine the phenomena, which we failed to take into account; or (B) Our enumeration of the physically possible microstates (the place where our knowledge of the laws of physics comes in) was wrong; Nature actually uses more states, or fewer, than we supposed.

On possibility (A), if we have included as constraints all the macroscopic conditions that are found, in the laboratory, to be sufficient to determine a reproducible outcome, then there seems to remain only the possibility that the equations of motion have new constants of the motion that we didn't know about. The nature of the error gives us a clue as to

what these new constants might be. This happened in the case of ortho- and para-hydrogen, where an approximate constant of the motion prevented the system from reaching the equilibrium predicted by the canonical ensemble in the time of experimental measurements; and the in existence of allotropic forms like red and white phosphorus, which are stable over many years.

On possibility (B), the failure of PME based on a particular state enumeration would give us a clue to new laws of physics. This, too, has happened more than once. First, the failure ofGibbs' classical statistical mechanics to predict specific heats correctly was the first clue pointing to the discrete energy levels of quantum theory; Nature did not use all the energy values permitted by classical mechanics. There remained a failure to predict vapor pressures and equilibrium constants correctly; this clue was found to indicate that Nature does not use all the mathematical solutions of the Schrödinger equation, only those that are symmetric or antisymmetric under permuations of identical particles.

We point this out to emphasize the completely different way of thinking about the relation of statistical theory to the real world, that is being expounded here. Repeatedly, conventional thinking has led to attacks on the PME viewpoint on the grounds that we make no appeal to ergodic theorems; ergo, there might be unknown constants of the motion which prevent a system from getting to the macroscopic state that PME predicts. Therefore, our prediction might be wrong; and this is seen as a terrible calamity that invalidates our approach.

This point of logic is discussed in more detail in Jaynes (1985), where we note that this conventional thinking is like that of a chess player who thinks ahead only one move. If we think ahead two moves we see that, while the usual success of PME predictions makes them useful in an "engineering" sense, an occasional failure is far from being a calamity.

If we see the formalism of statistical mechanics as a physical theory, then we worry about whether it is "right" or "wrong". If we recognize it instead as a method of reasoning which makes the best predictions possible from the information we put into it, its range of application is seen as vastly greater, not limited to thermodynamics or to physics; and its "failures" are seen to be even more valuable than its "successes", because they point the way to basic advances in science. My colleague Steve Gull has termed this change in attitude, "The Leap" in our understanding of the role of statistical inference in science. Instead of fearing "failure", we look eagerly for it.

This change in attitude is so great that it seemed misleading to use the same term "statistical mechanics" for both viewpoints. To avoid confusion we coined the name "Predictive Statistical Mechanics" to distinguish our line of thought from others.

For some time, this issue could not progress beyond being a mere philosophical difference, because to find a difference in pragmatic results one had to go beyond equilibrium thermodynamics; but the nonequilibrium PME calculations proved to be just as difficult, differing only in details, as the equilibrium ones. Indeed, most current work in statistical mechanics is still struggling with equilibrium calculations, which are hard enough.

It required the advent of the computer before the merit of PME as a predictive tool outside equilibrium theory could be demonstrated; and the impressive recent successes have been in applications outside of thermo-

dynamics. These are quite different from the thermodynamic applications, where it would be impractical to do pencil-and-paper calculations with more than a few simultaneous constraints.

With computers it is possible to locate maximum entropy points with almost any number of simultaneous constraints; spectrum analyses with dozens of constraints have been produced routinely for some fifteen years, and image reconstructions with over a million constraints (i.e., generalized canonical distributions with over a million Lagrange multipliers) have been produced routinely for three years now.

Finally, thanks to these developments, the real powers of the PME method could be seen in a way that transcended all philosophical arguments about the "meaning of probability". It is the computer printouts, not the philosophy or theorems, that is making the converts.

Lacking the space to present all the technical details now in print, in the following we survey the main new applications and results, with references. The field is presently in a phase of rapid growth, spreading into new fields with the number of workers appearing to double every year; and it is no longer possible for one person to keep up with all that is being done. Therefore the following is only a partial list, of applications known to the writer.

SPECTRUM ANALYSIS

The first breakthrough occurred in 1967, when John Parker Burg produced power spectrum estimates from incomplete geophysical data, by maximizing the entropy of the underlying time series $\{y_1, y_2, \ldots y_n\}$ defined at discrete times $t = 1,2,\ldots$, subject to the constraints of the data. More specifically, one has measured values of the autocovariance

$$R = n^{-1} \sum_t y_t y_{t+k} \tag{10}$$

for $m+1$ lags, $0 \leq k \leq m < n$. The true power spectrum is

$$P(f) = \sum_{k=-n}^{n} R_k \cos 2\pi kf \tag{11}$$

but there is nothing "random" about this. The problem is that $m < n$; our data are incomplete. On frequentist views, one would not see how probability theory is applicable to such a problem; but from our viewpoint we see that this is a classic example of the generalized inverse problem discussed above; the class C of possible spectra is given by (11) in which the R_k are taken equal to the data when $|k| \leq m$, and are arbitrary but for the nonnegativity requirement $P(f) \geq 0$ when $|k| > m$. Picking out a particular "best" estimate $P(f)$ from class C thus amounts to finding the "best" extrapolation of $\{R_k\}$ beyond the data.

Previous to this, power spectra had been estimated by the Blackman-Tukey (1958) method, which led to an estimate, for this problem, of:

$$[\hat{P}(f)]_{BT} = \sum_{k=-m}^{m} R_k W_k \cos(2\pi kf) \tag{12}$$

where W_k is a "taper" or "window" function, typically chosen as

$$W_k = \frac{1}{2}(1 + \cos \pi k/m) \tag{13}$$

which avoids spurious "side-lobes" in $\hat{P}(f)$ by tapering the data smoothly to zero at $k=m$, which would otherwise be an abrupt discontinuity. But it is at once evident that (12) does not lie in the class C of possible solutions; for (12) disagrees with the data at every data point where $W_k \neq 1$. Furthermore, (12) is not in general nonnegative. Therefore we know, not as a plausible inference, but as a demonstrated fact, that a time series with the spectrum (12) could not have produced our data! In addition, the estimate (12) failed to resolve sharp spectrum lines satisfactorily; the window function, in removing the side lobes, threw away half the resolution that we would have had without it.

Burg pointed out these shortcomings of (12) and proceeded to find a spectrum estimate by a totally different argument. Find the probability distribution $p(y_1,\ldots,y_n) = p(y)$ that has maximum entropy

$$S = -\int p(y) \log p(y) d^n y \tag{14}$$

subject to the constraints that the expectations of the R_k agree with the data (10). The resulting generalized canonical distribution

$$p(y) \propto \exp[-\Sigma \lambda_k R_k] \tag{15}$$

is, in view of of (10), a multivariate gaussian for which one can show that the entropy is also, to within an additive constant,

$$S = \int df \log P(f) \tag{16}$$

which is coming to be called the "Burg entropy" although they are substantially the same thing, only expressed in different variables. One could say equally well that he is maximizing (16). By a curious algebraic twist, the power spectrum estimate obtained as an expectation over (15) turns to have the form

$$\hat{P}(f) = \left(\Sigma \lambda_k \cos 2\pi k f\right)^{-1} \quad . \tag{17}$$

Burg used a computation algorithm given in the 1940's by N. Levinson to find the Lagrange multipliers λ_k in (15), (17) that agree with the data.

The resulting computer printouts were a revelation. The maximum-entropy spectra were clean and sharp; spurious artifacts like sidelobes were eliminated, but at the same time the resolution was greatly increased rather than sacrificed. No linear processing of the data could have produced such results.

The field then grew rapidly; to cite a few of the key references, the thorough mathematical analysis in the review article of Smylie, Clarke, and Ulrych (1973) and in the thesis of John Parker Burg (1975) are still required reading. By 1978 the literature had grown to the point where the IEEE issued a special volume of reprints on PME spectrum analysis (Childers, 1978) and Haykin (1979) edited another volume. Currie (1981) demonstrated its application to detection of small geophysical/meteorological effects. Haykin (1982) is a Special Issue of the IEEE Proceedings devoted to spectrum analysis, with numerous articles on the theory and practice of PME methods. Many more are in the Proceedings of the Laramie and Calgary Workshops on Bayesian/Maximum Entropy methods (Smith and Grandy, 1985; Justice, 1985).

IMAGE RECONSTRUCTION

The writer does not know exactly when the first attempt to use PME in image reconstruction occurred. Early discussions by Frieden et al.

(1972) and Ables (1974) gave some theory and computer simulations. Perhaps the major landmark is the article of Gull and Daniell (1978), which has real results supported by a very clear, simple rationale. Image reconstruction was seen as another generalized inverse problem based on (5) as follows. Since anything we can actually compute is digitized, we break the true scene into n picture elements, or "pixels" and imagine that scene generated by distributing N small "quanta" of intensity, the i'th pixel receiving N_i of them. Let $x_i = N_i/N$ be the fraction in the i'th pixel, and denote the resulting scene by $X = \{x_1, x_2, \ldots x_n\}$. It has an entropy S(X) given by (8).

But we do not know these numbers; we have available only a blurred scene consisting of $m < n$ pixels, with intensities d_k given by (5) in which the matrix A is the digitized point-spread function that describes the imperfections of our telescope (we are still not far, in either topic or basic rationale, from Laplace's problem of the mass of Saturn). Given the data $D = \{d_1, \ldots d_m\}$ of the blurred image, the reconstructed estimate of the true scene X that can be realized in the greatest number of ways is the one that maximizes the entropy S(X) subject to the constraints (5). This is a pure generalized inverse problem; the likelihood L(X) is constant on the set of possible scenes consistent with (5), zero elsewhere.

Gull and Daniell gave the resulting computer printouts for some real problems in radio and x-ray astronomy. They were just as impressive as were those of Burg's spectrum analysis. Again, the spurious artifacts of previous linear data analysis methods disappeared, while the resolution and dynamic range improved. This beautifully concise article is also required reading; it can be read and comprehended fully in an hour.

Frieden (1980) gave an equally impressive reconstruction of a galaxy so distant that the original optical photograph appears to the eye as a featureless blob; yet the maximum entropy reconstruction reveals five clearly resolved spiral arms. Having seen this reconstruction, one can go back and look at the original blurred image and see that there was, indeed, evidence for those arms in the data. The variational principle that generates it ensures that the maximum entropy reconstruction cannot show any detail for which there is no evidence in the data that were used as constraints.

In image reconstruction, maximum entropy is doing something much like what a skilled x-ray diagnostician learns to do. Knowing just what to look for, he can perceive details in a blurred picture that are quite invisible to the untrained eye. But at present most maximum entropy algorithms are doing this without being told in advance what to look for. If we can put the kind of prior knowledge that the x-ray diagnostician is using, into the underlying "hypothesis space" on which our entropy is defined, maximum entropy reconstructions will become still better. Two examples, for astronomical scenes about which we already know some features in advance, are given by Horne (1982) and Skilling (1983); we do not yet know how to do this in general.

Again the method spread rapidly to other applications. Gull and Skilling (1984) give many more examples, including reconstruction of a blurred image of an auto license plate for the London Police (whom we understand now have their own in-house maximum entropy facilities, with program supplied by Gull and Skilling).

NOISY DATA

A new feature forced itself upon us in these practical implementations of PME. The "pure PME formalism" illustrated in Eqs. (7)-(9) above has

assumed that data used as constraints are noiseless. With noisy data, pure PME or any other method must inevitably start showing spurious detail which is only an artifact of the noise, but which is fundamentally impossible to remove entirely (since the information needed to distinguish signal from noise is not there). But one can remove most of this by making allowance for noise in the same way that Laplace did; by using the full Bayes' theorem (3) for inference, in which we take the prior probability proportional to the multiplicity factor W in (6). The prior probability of a scene X becomes proportional to exp[NS(X)]. Then the scene with the highest posterior probability is the one that maximizes not S(X) but

$$NS(X) + \log L(X) \tag{18}$$

where L(X) is its likelihood, no longer rectangular, in the light of the probable errors in the data. If the blurred image data (5) have independent gaussian errors of RMS value h, then

$$L(X) = \exp\left\{-\sum_k (d_k - m_k)^2/2h^2\right\} \tag{19}$$

where $\{m_k\}$ are the "mock data", RHS of (5), that we would have obtained if X were the true scene but the noise were absent.

Maximizing (18) replaces the "hard" constraints of pure PME with "soft" constraints determined by the shape of the likelihood function. The reconstruction moves up a little higher on the entropy hill, giving us a slightly smoother picture. In effect, fine details in the data below the noise level, that we would have to accept as real if we knew there was no noise, are reinterpreted as more likely due to noise and ignored. A higher noise level makes softer constraints and a reconstruction showing less detail. Strongly represented features are those for which the details in the data rise above the noise and give strong evidence for a real thing; just Laplace's original significance test, now applied to every pixel.

The question of the proper choice of N has given rise to a great deal of discussion, into which we cannot go here. Some of this--and a great deal more--will appear in Justice (1985).

OTHER APPLICATIONS

After the success of the one-dimensional PME spectrum analysis and two-dimensional image reconstructions, one was encouraged to try reconstructing three-dimensional objects from incomplete data. Medical tomography (Minerbo, 1979) seems to be the first example. But the classical three-dimensional generalized inverse problem, crystallographic structure analysis from x-ray scattering data, had been calling out for such a method for decades.

Here the data are notoriously incomplete, consisting of magnitudes, but not phases, of only a few of the Fourier coefficients of the electron density function d(x). The analogy to Burg's original problem is clear; one must extrapolate the data to higher wavenumbers. But it is more complicated because one must also estimate phases; this is discussed in some detail by Bricogne (1982,1984) and Wilkins et al. (1983,1984). The Proceedings of the recent Orsay workshop (Bricogne, 1985) will have many more articles on it, including one by Skilling on the computation problem.

A PME structure analysis of biological macromolecules is given by Bryan et al. (1983). It appears that this is to become one of the major areas of application in view of the number of biological structures in need of analysis, and the resulting large efforts going into it at several places.

It seems that every success of PME in one area points to a new application in a related area. Sibisi et al. (1983,1984) apply it to estimation of line frequency and decay rate from free decay NMR signals, Livesey et al. (1984) to analysis of EXAFS data, Gburski et al. (1984) to inferring line shapes from a few calculated moments. Mead and Papanicolaou (1984) show that it has advantages over previous methods for estimating mathematical functions from incomplete information, such as the first few terms of a series expansion. Applications in econometrics are beginning to appear.

CONCLUSION

We have tried to present a general survey, rather than the technical details, of a line of thought that has been evolving slowly and painfully, impeded by philosophical differences, for some thirty years. But, as will be apparent from the above, we are now far beyond the stage of philosophical debate. Wherever it has been applied in a competent way to computationally feasible problems, PME has not only succeeded, but led to substantive improvements over previous methods of inference from incomplete information. To acknowledge and represent explicitly the incompleteness of our information is just what these problems had needed.

We are thus encouraged to renew our efforts in the applications originally envisaged in irreversible statistical mechanics and, eventually, to try to see Bohr's and Einstein's views of quantum theory finally unified as an example of the same kind of reasoning. We have as yet no new results on explicit experimental numbers but there are many results of a more general nature (Jaynes, 1978,1980).

It is easy to show that Predictive Statistical Mechanics leads automatically to such known correct results as the Onsager reciprocities, the Kubo formulas for transport coefficients, the Wiener prediction algorithm, and in a short-memory approximation, to Fokker-Planck equations. But these always appear in more generally applicable form; the reciprocities require no assumptions about short memory or regression of spontaneous fluctuations, the Kubo formulas are no longer limited to the quasi-stationary, long-wavelength regime, Fokker-Planck equations hold not only in momentum space, but in any thermodynamic state space, whose coordinates are the macroscopic quantities of interest.

Einstein's view of fluctuations as providing the "driving force" that makes an irreversible process go, and Onsager's view of the entropy gradient in thermodynamic state space as providing the "steering" telling it in which direction to go, appear automatically, but no longer restricted to situations close to equilibrium. A generalized "Bubble Dynamics" shows how a bubble of probability moves up the entropy hill far from equilibrium, at a velocity proportional to

(mean square fluctuation) x (local entropy gradient)

while constantly readjusting its size and shape to the local curvature of the entropy function. At a local loss of entropy convexity, the stabilizing forces are lost and the bubble stretches out leading to a bifurcation or other instability corresponding to a phase change: a time dependent version

of what Gibbs showed in 1873 for static conditions. The various "catastrophes" of Rene Thom appear as consequences of different kinds of local loss of entropy convexity.

Indeed, such results have been appearing in such quantity that we have fallen far behind in getting them written up for publication. But with the new applications outside thermodynamics off to a good start, we can now return to the original goal, with strenuous efforts to correct this.

REFERENCES

J. G. Abels (1974) "Maximum Entropy Spectral Analysis", Astron. Astrophys. Suppl. 15; 383.

G. Bricogne (1982), in Computational Crystallography, D. Dayre, Editor, pp. 258-264; Oxford University Press, New York.

G. Bricogne (1984), "Maximum Entropy and the Foundations of Direct Methods", Acta Cryst. (in press).

G. Bricogne, ed. (1985), Proceedings of the EMBO Workshop on Maximum-Entropy Methods, Orsay, April 24-28, 1984. Applications in crystal and biological macromolecular structure determination from x-ray/neutron scattering data.

R. K. Bryan, M. Bansal, W. Folkhard, C. Nave, and D. A. Marvin, (1983), "Maximum-Entropy calculation of the electron density at 4A resolution of Pf1 filamentous bacteriophage", Proc. Nat. Acad. Sci. U.S.A. 80, 4728.

S. F. Burch, S. F. Gull & J. Skilling (1983), "Image Restoration by a Powerful Maximum Entropy Method", Comp. Vision, Graphics & Image Processing, 23, 118-123.

John Parker Burg (1967), "Maximum Entropy Spectral Analysis", Proc. 37th Annual Meeting, International Society of Exploration Geophysicists, Oklahoma City. Reprinted in Childers (1978).

John Parker Burg (1975), Maximum Entropy Spectral Analysis, Ph.D. Thesis, Stanford University.

D. G. Childers, Editor (1978); Modern Spectrum Analysis, IEEE Press and J. Wiley & Sons, N.Y. A reprint collection.

R. T. Cox (1946), "Probability, Frequency, and Reasonable Expectation", Am. J. Phys. 14, 1-13. This is greatly expanded in R. T. Cox, The Algebra of Probable Inference, Johns Hopkins University Press, Baltimore (1961).

R. G. Currie (1981), "Solar Cycle Signal in Earth Rotation: Nonstationary Behavior", Science, 211, 386.

R. G. Currie (1981), "Evidence for 18.6 year Mn Signal in Temperature and Drought Conditions in North America Since AD 1800", J. Geophys. Res. 86, #C11, pp. 11055-11064.

R. G. Currie (1981), "Solar Cycle Signal in Air Temperature in North America: Amplitude, Gradient, Phase and Distribution", J. Atmos. Sci., 38, 809-818.

G. J. Daniell & S. F. Gull (1980), "Maximum Entropy Algorithm Applied to Image Enhancement", IEE Proc. 127E, 170-172.

A. C. Fabian, R. Willingale, J. P. Pye, S. S. Murray and G. Fabbiano (1980), "The X-Ray Structure and Mass of the Cassiopeia A Supernova Remnant", Mon. Not. Roy. Astr. Soc., 193, 175-188.

B. R. Frieden (1972), "Restoring with Maximum Likelihood and Maximum Entropy", J. Opt. Soc. Am. 62, 511-518.

B. R. Frieden (1980), "Statistical Models for the Image Restoration Problem", Computer Graphics and Image Processing, 12, 40-59.

Z. Gburski, C. G. Gray, and D. E. Sullivan, (1984), Chem. Phys. Lett. 106, 55.

J. Willard Gibbs (1902) Elementary Principles in Statistical Mechanics, reprinted in The Collected Works of J. Willard Gibbs, Vol. 2, by Yale University Press, New Haven, Conn., 1948 and by Dover Publications, Inc., New York, 1960.

S. F. Gull & G. J. Daniel (1978), "Image Reconstruction from Incomplete and Noisy Data", Nature 272, 686-690.

S. F. Gull & G. J. Daniell (1979), "The Maximum Entropy Method", in Image Formation from Coherent Functions in Astronomy, C. van Schooneveld, Ed., D. Reidel Pub. Co.

S. F. Gull & J. Skilling (1983), "The Maximum Entropy Method", IAU/URSI Symposium on Indirect Imaging, Sydney Australia.

S. Gull and J. Skilling (1984), "The Entropy of an Image", in J. A. Roberts, ed., Indirect Imaging, Cambridge U. Press.

S. Haykin, Editor (1979), Nonlinear Methods of Spectral Analysis, Topics in Applied Physics, Vol. 34; Springer-Verlag, New York.

S. Haykin, Editor (1982). The September 1982 IEEE Proceedings, Vol. 70, is a special issue devoted to spectrum analysis.

W. Heisenberg (1958), Daedalus 87, 100.

K. D. Horne (1982), "Eclipse Mapping of Accretion Disks in Cataclysmic Binaries", Ph.D. Thesis, CalTech. PME with "prior prejudice" favoring circular symmetry.

E. T. Jaynes (1967), "Foundations of Probability Theory and Statistical Mechanics", in Delaware Seminar in the Foundations of Physics, M. Bunge, Ed., Springer-Verlag, Berlin.

E. T. Jaynes (1976), "Confidence Intervals vs. Bayesian Intervals", in Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, W. L. Harper and C. A. Hooker, Eds. Reidel Publishing Co., Dordrecht-Holland. Reprinted in Jaynes (1983).

E. T. Jaynes (1978), "Where do we Stand on Maximum Entropy?", in The Maximum Entropy Formalism, R. D. Levine and M. Tribus, Eds., M.I.T. Press, Cambridge Mass. Reprinted in Jaynes, (1983).

E. T. Jaynes (1980), "The Minimum Entropy Production Principle", in Annual Review of Physical Chemistry, S. Rabinovitch, Ed., Annual Reviews, Inc., Palo Alto Calif. Reprinted in Jaynes, (1983).

E. T. Jaynes (1983), Papers on Probability, Statistics, and Statistical Physics, R. D. Rosenkrantz, Editor, D. Reidel Publishing Co., Dordrecht-Holland. Contains reprints of the preceding four articles and nine others.

E. T. Jaynes, (1984), "Prior Information and Ambiguity in Inverse Problems", SIAM-AMS Proceedings, 14, 151-166.

E. T. Jaynes, (1985), "The Evolution of Carnot's Principle", in Bricogne (1985).

H. Jeffreys (1961), Theory of Probability, Oxford University Press.

R. W. Johnson & J. E. Shore (1983), "Minimum-Cross-Entropy Spectral Analysis of Multiple Signals", IEEE Trans. Acoust. Speech & Signal Processing ASSP-31, 574-582.

J. H. Justice, Ed., (1985), Proceedings of the Workshop on Bayesian/Maximum Entropy Methods, University of Calgary, August 1984. In Press.

M. C. Kemp (1980), "Maximum Entropy reconstructions in emission tomography", Medical Radionuclide Imaging, 1, 313-323.

P. S. Laplace (1814), Essai Philosophique sur les Probabilites, Courcier Imprimeur, Paris; reprints of this work and of Laplace's much larger Theorie Analytique des Probabilites are available from Editions Culture et Civilisation, 115, Ave. Gabriel Lebron, 1160 Brussels, Belgium.

A. K. Livesey (1984), "Structural Investigations of Metallic Glasses", Ph.D. Thesis, Cambridge University.

A. K. Livesey & J. Skilling (1984), "Maximum Entropy Theory", Submitted to Acta Cryst.

L. R. Mead & N. Papanicolaou (1984), "Maximum Entropy in the Problem of Moments", J. Math. Phys. 25, 2404-2417.
G. Minerbo (1979), "MENT: A maximum entropy algorithm for reconstructing a source from projection data", Comp. Grap. & Image Processing, 10, 48-68.
L. H. Schick and R. Inguva (1981), "Information Theoretic Processing of Seismic Data", Geophys. Res. Lett. 8, 1199.
P. F. Scott (1981), "A 31 GHz map of W3(OH) with a resolution of 0.3 arsec", Mon. Not. R. Astr. Soc. 194, 25P-29P.
J. E. Shore & R. W. Johnson (1980), "Axiomatic Derivation of Maximum Entropy and the Principle of Minimum Cross-Entropy", IEEE Trans. Inform. Theory IT-26, 26-37.
J. E. Shore (1981), "Minimum Cross-Entropy Spectral Analysis", IEEE Trans. Acoust. Speech & Signal Processing ASSP-29, 230-237.
S. Sabisi (1983), "Two-Dimensional Reconstructions from One-Dimensional Data by Maximum Entropy", Nature 301, 134-136.
S. Sibisi, J. Skilling, R. Brereton, E. D. Laue, J. Staunton (1984), "Maximum Entropy Method in 13C-NMR Spectroscopy" (in preparation for Chemical Communications).
J. Skilling, A. W. Strong, K. Bennett (1979), "Maximum Entropy Image Processing in Gamma-Ray Astronomy", Mon. Not. Roy. Astron. Soc., 187, 145-152.
J. Skilling (1983), "Maximum Entropy Image Reconstruction from Phaseless Fourier Data", presented at Optical Soc. Am. meeting on Signal Recovery with Incomplete Information and Partial Constraints Incline Village, Nevada, Jan. 1983.
J. Skilling and S. Gull (1983), "Algorithms and Applications", in C. R. Smith & W. T. Grandy (1985).
J. Skilling & S. Gull (1984), "The Entropy of an Image", Am. Math. Soc. SIAM Proceedings, Vol. 14.
J. Skilling & R. K. Bryan, (1984), "Image Reconstruction by Maximum-Entropy: General Algorithms", Mon. Not. Roy. Astr. Soc. In Press.
C. R. Smith & W. T. Grandy, Editors (1985); Proceedings of the 1981, 1982 1983 Workshops on Maximum Entropy and Bayesian Methods in Inverse Problems; University of Wyoming, Laramie WY 92071 (2 Volumes; in press).
D. E. Smylie, G. K. C. Clarke, and T. J. Ulrych, (1973), "Analysis of Irregularities in the Earth's Rotation", in Computational Physics, B. A. Bolt, B. Alder, & S. Feinbach, Eds., Academic Press, New York.
S. Steenstrup & S. W. Wilkins (1984), "On Information and Complementarity in Crystal Structure Determination", Acta Cryst. (in press).
R. J. Tuffs (1984), Secular Changes in the Supernova Remnant Cassiopeia A", Ph.D. Thesis, Cambridge University, U.K.
S. W. Wilkins, J. N. Varghese & M. S. Lehmann (1983), "Statistical Geometry. I: A Self-Consistent Approach to the Crystallographic Inversion Problem Based on Information Theory", Acta Cryst. A39, 49-60.
S. W. Wilkins, J. N. Varghese, & S. Steenstrup (1984), "X-Ray Structure Determination using Information Theory--the Statistical Geometric Approach", in Proceedings, IAU/URSI Symposium on Indirect Imaging, Sydney, Australia, Aug. 1983 (in press).
R. Willingale, "Use of the Maximum Entropy Method in X-Ray Astronomy", Mon. Not. R. Astr. Soc. 194, 359-364.