CLEARING UP MYSTERIES – THE ORIGINAL GOAL †

E. T. Jaynes‡
Wayman Crow Professor of Physics
Washington University, St. Louis MO, U.S.A.

Abstract: We show how the character of a scientific theory depends on one's attitude toward probability. Many circumstances seem mysterious or paradoxical to one who thinks that probabilities are real physical properties existing in Nature. But when we adopt the "Bayesian Inference" viewpoint of Harold Jeffreys, paradoxes often become simple platitudes and we have a more powerful tool for useful calculations. This is illustrated by three examples from widely different fields: diffusion in kinetic theory, the Einstein–Podolsky–Rosen (EPR) paradox in quantum theory, and the second law of thermodynamics in biology.

† The opening talk at the 8'th International MAXENT Workshop, St. John's College, Cambridge, England, August 1-5, 1988. In the Proceedings Volume, *Maximum Entropy and Bayesian Methods*, J. Skilling, Editor, Kluwer Academic Publishers, Dordrecht–Holland (1989), pp. 1–27.
‡ Mailing Address: Campus Box #1105, Washington University, 1 Brookings Drive, St. Louis MO 63130.

INTRODUCTORY REMARKS

Our group has the honour to be among the first to use this splendid new Fisher building with its 300 seat auditorium. But perhaps, at a meeting concerned with Bayesian inference, we should clarify which Fisher inspired that name.

St. John's College was founded in the year 1511, its foundress being the Lady Margaret Beaufort. John Fisher was then Chancellor of the University of Cambridge, and after her death he found himself obliged to make heroic efforts to ensure that her wishes were carried out. But for those efforts, made some 480 years ago, St. John's College would not exist today. Historians have suggested that, but for the efforts of John Fisher in holding things together through a turbulent period, the entire University of Cambridge might not exist today.

Although the terms "Bayesian" and "Maximum Entropy" appear prominently in the announcements of our meetings, our efforts are somewhat more general. Stated broadly, we are concerned with this: "What are the theoretically valid, and pragmatically useful, ways of applying probability theory in science?"

The new advances of concern to us flow from the recognition that, in almost all respects that matter, the correct answers were given here in St. John's College some fifty years ago, by Sir Harold Jeffreys. He stated the general philosophy of what scientific inference is, fully and correctly, for the first time; and then proceeded to carry both the mathematical theory and its practical implementation farther than anyone can believe today, who has not studied his works.

The ideas were subtle, and it required a long time for their merit to be appreciated; but we can take satisfaction in knowing that Sir Harold has lived to see a younger generation of scientists eagerly reading, and profiting by, his work. In September 1983 I had a long, delightful conversation over tea with Sir Harold and Lady Jeffreys, and know how pleased they both were.

Important progress is now being made in many areas of science by adopting the viewpoint and extending the methods of Harold Jeffreys. Even those of us who were long since convinced of their theoretical merit are often astonished to discover the amount of numerical improvement over "orthodox" statistical methods, that they can yield when programmed into computers. It is hardly ever small except in trivial problems, and nontrivial cases have come up where they yield orders of magnitude better sensitivity and resolution in extracting information from data.

This means that in some areas, such as magnetic resonance spectroscopy, it is now possible to conduct quantitative study of phenomena which were not accessible to observation at all by the previously used Fourier transform methods of data analysis; old data may have a new lease on life. The technical details of this are to appear in the forthcoming book of G. L. Bretthorst (1988).

Even when the numerical improvement is small, the greater computational efficiency of the Jeffreys methods, which can reduce the dimensionality of a search algorithm by eliminating uninteresting parameters at the start, can mean the difference between what is feasible and what is not, with a given computer. As the complexity of our problems increases, so does the relative advantage of the Jeffreys methods; therefore we think that in the future they will become a practical necessity for all workers in the quantitative sciences.

How fitting it is that this meeting is being held back where these advances started. Our thanks to the Master and Council of St. John's College, who made it possible.

THE MOTIVATION

Probability theory is a versatile tool, which can serve many different purposes. The earliest signs of my own interest in the field involved not data analysis, but recognition that the Jeffreys viewpoint can clear up outstanding mysteries in theoretical physics, by raising our standards of logic. As James Clerk Maxwell wrote over 100 years ago and Harold Jeffreys quoted 50 years ago, probability theory is itself the true logic of science.

The recent emphasis on the data analysis aspect stems from the availability of computers and the failure of "orthodox" statistics to keep up with the needs of science. This created many opportunities for us, about which other speakers will have a great deal to say here. But while pursuing these important applications we should not lose sight of the original goal, which is in a sense even more fundamental to science. Therefore in this opening talk we want to point out a field ripe for exploration by giving three examples, from widely different areas, of how scientific mysteries are cleared up, and paradoxes become platitudes, when we adopt the Jeffreys viewpoint. Once the logic of it is seen, it becomes evident that there are many other mysteries, in all sciences, calling out for the same treatment.

The first example is a simple exercise in kinetic theory that has puzzled generations of physics students: how does one calculate a diffusion coefficient and not get zero? The second concerns the currently interesting Einstein–Podolsky–Rosen paradox and Bell inequality mysteries in quantum theory: do physical influences travel faster than light? The third reexamines the old mystery about whether thermodynamics applies to biology: does the high efficiency of our muscles violate the second law?

DIFFUSION

Think, for definiteness, of a solution of sugar in water, so dilute that each sugar molecule interacts constantly with the surrounding water, but almost never encounters another sugar molecule. At time $t = 0$ the sugar concentration varies with position according to a function $n(x, 0)$. At a later time we expect that these variations will smooth out, and eventually $n(x, t)$ will tend to a uniform distribution.

Since sugar molecules – or as we shall call them, "particles" – are not created or destroyed, it seems natural to think that there must have been a diffusion current, or flux $J(x, t)$ carrying them from the high density regions to the low, so that the change in density with time is accounted for by the conservation law:

$$\frac{\partial n}{\partial t} + \mathrm{div}(J) = 0\,. \tag{1}$$

Phenomenologically, Fick's law relates this to the density gradient:

$$J = -D\,\mathrm{grad}(n) \tag{2}$$

In the case of sugars, the diffusion coefficient D is easy to measure by optical rotation. In Maxwell's great Encyclopaedia Brittanica article on diffusion he quotes the experimental result of Voit for sucrose: $D = 3.65 \times 10^{-5}$ square cm/sec.

Our present problem is: how do we calculate $J(x, t)$ from first principles? Maxwell gave the simple kinetic theory of diffusion in gases, based on the idea of the mean free path. But in a liquid there is no mean free path. Maxwell, who died in 1879, never knew the general theoretical formula for the diffusion coefficient which we now seek, and which applies equally to gases, liquids, and solids.

Only with the work of Einstein in the first decade of this Century were the beginnings made in seeing how to deal with the problem, culminating finally in the correct formula for the diffusion coefficient. But Einstein had to work at it harder than we shall, because he did not have Harold Jeffreys around to show him how to use probability theory.†

It would seem that, given where a particle is now, we should find its velocity v, and summing this over all particles in a small region would give the local flux $J(x, t)$. However, the instantaneous

† As far as we have been able to determine, Jeffreys' view of probability theory was unknown in continental Europe throughout Einstein's lifetime; this was a handicap to Einstein in more ways than one.

 4

velocity of a particle is fluctuating wildly, with a mean–square value given by the Rayleigh–Jeans equipartition law; and that is not the velocity we seek. Superposed on this rapidly fluctuating and reversing thermal velocity, of the order of 100 meters/sec, is a very much slower average drift velocity representing diffusion, which is our present interest.

Given where a particle is now, $x(t)$, its average velocity over a time interval 2τ centered at the present is

$$\bar{v} = \frac{x(t+\tau) - x(t-\tau)}{2\tau} \tag{3}$$

so if we make our best estimate of where the particle will be a time τ in the future that is long on the time scale of thermal fluctuations, and where it was an equal time in the past, we have an estimate of its average slow velocity about the present time. The probability that it will move from $x(t)$ to $y \equiv x(t+\tau)$ in the future is given by some distribution $P(y|x,\tau)$. Its motion is the result of a large number of small increments (encounters with individual water molecules). Therefore the Central Limit Theorem, interpreted with the judgment that scientists develop (but cannot always explain to mathematicians, because it draws on extra information that a mathematician would never use in proving the theorem) tells us that this will have a Gaussian form, and from symmetry the mean displacement is zero:

$$P(y|x,I) = A\exp[-(y-x)^2/2\sigma^2(\tau)] \tag{4}$$

where I stands for the general prior information stated or implied in our formulation of the problem. All the analysis one could make of the dynamics of sugar–water interactions would, in the end, serve only to determine the spreading function $\sigma^2(\tau) = (\delta x)^2$, the expected square of the displacement.

But now our trouble begins; the particle is as likely to be battered to the right as to the left; so from symmetry, the expectation of y is $\langle y \rangle = x$. Now all the equations of motion, however complicated, are at least time–reversal invariant. Therefore for the past position $z \equiv x(t-\tau)$, conventional reasoning says that we should have the same probability distribution (4) which is independent of the sign of τ, and again $\langle z \rangle = x(t)$. Therefore the estimated velocity is zero.

Surely, this must be right, for our particle, interacting only with the surrounding water, has no way of knowing that other sugar molecules are present, much less that there is any density gradient. From the standpoint of dynamics alone (*i.e.*, forces and equations of motion) there is nothing that can give it any tendency to drift to regions of lower rather than higher density. Yet diffusion does happen!

In the face of this dilemma, Einstein was forced to invent strange, roundabout arguments – half theoretical, half phenomenological – in order to get a formula for diffusion. For example, first estimate how the density $n(x,t)$ would be changed a long time in the future by combining the distributions (4) generated by many different particles, then substitute it into the phenomenological diffusion equation that we get by combining (1) and (2); and from that reason backwards to the present time to see what the diffusion flux must have been.

This kind of indirect reasoning has been followed faithfully ever since in treatments of irreversible processes, because it has seemed to be the only thing that works. Attempts to calculate a flux directly at the present time give zero from symmetry, so one resorts to "forward integration" followed by backward reasoning. Yet this puzzles every thoughtful student, who thinks that we ought to be able to solve the problem by direct reasoning: calculate the flux $J(x,t)$ here and now, straight out of the physics of the situation. That symmetry cannot be exactly right; but where is the error in the reasoning?.

Furthermore, instead of our having to assume a phenomenological form, a correct analysis ought to give it automatically; *i.e.*, it should tell us from first principles why it is the density gradient, and not some other function of the density, that matters, and also under what conditions this will be true. Evidently, we have a real mystery here.

Why did our first attempt at direct reasoning fail? Because the problem is not one of physical prediction from the dynamics; it is a problem of inference. The question is not "How do the equations of motion require the particles to move about on the average?" The equations of motion do not require them to move about at all. The question is: "What is the best estimate we can make about how the particles are in fact moving in the present instance, based on all the information we have?" The equations of motion are symmetric in past and future; but our information about the particles is not.

Given the present position of a particle, what can we say about its future position? The zero movement answer above was correct; for predicting where it will be in the future, the knowledge of where it is now makes all prior information about where it might have been in the past irrelevant. But estimating where it was in the past is not a time–reversed mirror image of this, for we have prior knowledge of the varying density of particles in the past. Knowledge of where it is now does not make that prior knowledge irrelevant; and sound logic must take both into account.

Let us restate this in different language. Eq. (4) expresses an average over the class of all possible motions compatible with the dynamics, in which movements to the right and the left have, from symmetry, equal weight. But of course, our particular particle is in fact executing only one of those motions. Our prior information selects out of the class of all possibilities in (4) a smaller class in which our particle is likely to be, in which movements to the right and left do not have equal weight. It is not the dynamics, but the prior information, that breaks the symmetry and leads us to predict a non–zero flux.

While $P(x|z,t)$ is a direct probability, the same function as (4), the probability we now need is $P(z|x,t)$, an inverse probability which requires the use of Bayes' theorem:

$$P(z|x,t,I) = AP(z|I)\,P(x|z,I)\,. \tag{5}$$

The prior probability $P(z|I)$ is clearly proportional to $n(z)$, and so from (3)

$$\log P(z|x,I) = \log n(z) - (z-x)^2/2\sigma^2(\tau) + (\text{const.})\,. \tag{6}$$

Differentiating, the most probable value of the past position z is not x, but

$$\hat{z} = x + \sigma^2 \ \text{grad}(\log n) = x + (\delta x)^2\,\text{grad}(\log n) \tag{7}$$

whereupon, substituting into (3) we estimate the drift velocity to be

$$\bar{v} = -(\delta x)^2/2\tau \ \ \text{grad}(\log n) \tag{8}$$

and our predicted average diffusion flux over the time interval 2τ is

$$J(x,t) = n\bar{v} = -(\delta x)^2/2\tau \ \ \text{grad}(n)\,. \tag{9}$$

Bayes' theorem has given us just Einstein's formula for the diffusion coefficient:

$$D = \frac{(\delta x)^2}{2\tau} \tag{10}$$

and a good deal more. We did not assume that grad(n) was the appropriate phenomenological form; Bayes' theorem told us that automatically. At the same time, it told us the condition for validity of that form; unless $(\delta x)^2$ is proportional to τ, there will be no unique diffusion coefficient, but only a sequence of apparent diffusion coefficients $D(\tau)$ for the average drift over different time intervals 2τ. Then the flux $J(x,t)$ will depend on other properties of $n(x,t)$ than its present gradient, and in place of (2) a more complete Bayesian analysis will give a different phenomenological relation, involving an average of grad(n) over a short time in the past. Thus (9) is only the beginning of the physical predictions that we can extract by Bayesian analysis.

While (8) is the best estimate of the average velocity that we could make from the assumed information, it does not determine the velocity of any one particle very well. But what matters is the prediction of the observable net flux of N particles. In principle we should have calculated the joint posterior distribution for the velocities of N particles, and estimated their sum. But since that distribution factors, the calculation reduces to N repetitions of the above one, and the relative accuracy of the prediction improves like $\sqrt{N}$, the usual rule in probability theory.

In practice, with perhaps $0.001\,M$ sugar solutions, the relevant values of N are of the order of 10^{16}, and the prediction is highly reliable, in the following sense: for the great majority of the N–particle motions consistent with the information used, the flux is very close to the predicted value.

DISCUSSION

The above example may indicate the price that kinetic theory has paid for its failure to comprehend and use the Bayesian methods that Harold Jeffreys gave us 50 years ago, and how many other puzzles need to be reexamined from that viewpoint. The only reason why the fluxes persisted in being zero was failure to put the obviously necessary prior information into the probabilities. But as long as one thinks that probabilities are real physical properties of systems, it seems wrong to modify a probability merely because our state of knowledge has changed.

The idea that probabilities can be used to represent our own information is still foreign to "orthodox" teaching, although the above example shows what one gains by so doing. Orthodoxy does not provide any technical means for taking prior information into account; yet that prior information is often highly cogent, and sound reasoning requires that it be taken into account. In other fields this is considered a platitude; what would you think of a physician who looked only at your present symptoms, and refused to take note of your medical history?

In the next talk, Ray Smith will survey the arguments of George Pólya and Richard Cox indicating the sense in which Bayesian inference is uniquely determined by simple qualitative desiderata of rationality and logical consistency. Here I want only to indicate something about the rationale of their application in real problems.

Conventional training in the physical sciences concentrates attention 100% on physical prediction; the word "inference" was never uttered once in all the science courses I ever took. Therefore, the above example was chosen because its rationale is clear and the actual calculation is utterly trivial; yet its power to yield not only results that previously required more work but also more details about them, is apparent at once.

To appreciate the distinction between physical prediction and inference it is essential to recognize that propositions at two different levels are involved. In physical prediction we are trying to describe the real world; in inference we are describing only our state of knowledge about the world. A philosopher would say that physical prediction operates at the ontological level, inference at the epistemological level. Failure to see the distinction between reality and our knowledge of reality puts us on the Royal Road to Confusion; this usually takes the form of the Mind Projection Fallacy, discussed below.

The confusion proceeds to the following terminal phase: a Bayesian calculation like the above one operates on the epistemological level and gives us only the best predictions that can be made from the information that was used in the calculation. But it is always possible that in the real world there are extra controlling factors of which we were unaware; so our predictions may be wrong. Then one who confuses inference with physical prediction would reject the calculation and the method; but in so doing he would miss the point entirely.

For one who understands the difference between the epistemological and ontological levels, a wrong prediction is not disconcerting; quite the opposite. For how else could we have learned about those unknown factors? It is only when our epistemological predictions fail that we learn new things about the real world; those are just the cases where probability theory is performing its most valuable function. Therefore, to reject a Bayesian calculation because it has given us an incorrect prediction is like disconnecting a fire alarm because that annoying bell keeps ringing. Probability theory is trying to tell us something important, and it behooves us to listen.

THE MIND PROJECTION FALLACY

It is very difficult to get this point across to those who think that in doing probability calculations their equations are describing the real world. But that is claiming something that one could never know to be true; we call it the Mind Projection Fallacy. The analogy is to a movie projector, whereby things that exist only as marks on a tiny strip of film appear to be real objects moving across a large screen. Similarly, we are all under an ego–driven temptation to project our private thoughts out onto the real world, by supposing that the creations of one's own imagination are real properties of Nature, or that one's own ignorance signifies some kind of indecision on the part of Nature.

The current literature of quantum theory is saturated with the Mind Projection Fallacy. Many of us were first told, as undergraduates, about Bose and Fermi statistics by an argument like this: "You and I cannot distinguish between the particles; *therefore* the particles behave differently than if we could." Or the mysteries of the uncertainty principle were explained to us thus: "The momentum of the particle is unknown; *therefore* it has a high kinetic energy." A standard of logic that would be considered a psychiatric disorder in other fields, is the accepted norm in quantum theory. But this is really a form of arrogance, as if one were claiming to control Nature by psychokinesis.

In our more humble view of things, the probability distributions that we use for inference do not describe any property of the world, only a certain state of information about the world. This is not just a philosophical position; it gives us important technical advantages because of the more flexible way we can then use probability theory. In addition to giving us the means to use prior information, it makes an analytical apparatus available for such things as eliminating nuisance parameters, at which orthodox methods are helpless. This is a major reason for the greater computational efficiency of the Jeffreys methods in data analysis.

In our system, a *probability* is a theoretical construct, on the epistemological level, which we assign in order to represent a state of knowledge, or that we calculate from other probabilities according to the rules of probability theory. A *frequency* is a property of the real world, on the ontological level, that we measure or estimate. So for us, probability theory is not an Oracle telling how the world must be; it is a mathematical tool for organizing, and ensuring the consistency of, our own reasoning. But it is from this organized reasoning that we learn whether our state of knowledge is adequate to describe the real world.

This point comes across much more strongly in our next example, where belief that probabilities are real physical properties produces a major quandary for quantum theory, in the EPR paradox. It is so bad that some have concluded, with the usual consistency of quantum theory, that (1) there is no real world, after all, and (2) physical influences travel faster than light.

BACKGROUND OF EPR

8

Quantum Mechanics (QM) is a system of mathematics that was not developed to express any particular physical ideas, in the sense that the mathematics of relativity theory expresses the ideas of Einstein, or that of genetics expresses the ideas of Mendel. Rather, it grew empirically, over about four decades, through a long series of trial–and–error steps. But QM has two difficulties; firstly, like all empirical equations, the process by which it was found gives no clue as to its meaning. QM has the additional difficulty that its predictions are incomplete, since in general it gives only probabilities instead of definite predictions, and it does not indicate what extra information would be required to make definite predictions.

Einstein and Schrödinger saw this incompleteness as a defect calling for correction in some future more complete theory. Niels Bohr tried instead to turn it into a merit by fitting it into his philosophy of complementarity, according to which one can have two different sets of concepts, mutually incompatible, one set meaningful in one situation, the complementary set in another. As several of his early acquaintances have testified (Rozental, 1964), the idea of complementarity had taken control of his mind years before he started to study quantum physics.

Bohr's "Copenhagen Theory" held that, even when the QM state vector gives only probabilities, it is a complete description of reality in the sense that nothing more can ever be known; not because of technological limitations, but as a matter of fundamental principle. It seemed to Einstein that this completeness claim was a gratuitous addition, in no way called for by the facts; and he tried to refute it by inventing thought experiments which would enable one to get more information than Bohr wished us to have. Somehow, the belief has been promulgated that Bohr successfully answered all of Einstein's objections.

But when we examine Bohr's arguments, we find a common logical structure; always they start by postulating that the available measurement apparatus is subject to his "uncertainty" limitations; and then by using only classical physics (essentially, only Liouville's theorem) they come to the conclusion that such an apparatus could not be used for Einstein's purpose. Bohr's foregone conclusion is always assured by his initial postulate, which simply appears out of nowhere. In our view, then, the issue remains open and we must raise our standards of logic before there can be any hope of resolving it.

Leslie Ballentine (1970) analyzed the Bohr and Einstein positions and showed that much of the chanting to the effect that "Bohr won, Einstein lost" is sustained by quoting Einstein's views and attributing them to Bohr. Virtually all physicists who do real quantum–mechanical calculations interpret their results in the sense of Einstein, according to which a pure state represents an ensemble of similarly prepared systems and is thus an incomplete description of an individual system. Bohr's completeness claim has never played any functional role in applications, and in that sense it is indeed gratuitous.

CONFRONTATION OR RECONCILIATION?

Put most briefly, Einstein held that the QM formalism is incomplete and that it is the job of theoretical physics to supply the missing parts; Bohr claimed that there are no missing parts. To most, their positions seemed diametrically opposed; however, if we can understand better what Bohr was trying to say, it is possible to reconcile their positions and believe them both. Each had an important truth to tell us.

But Bohr and Einstein could never understand each other because they were thinking on different levels. When Einstein says QM is incomplete, he means it in the ontological sense; when Bohr says QM is complete, he means it in the epistemological sense. Recognizing this, their statements are no longer contradictory. Indeed, Bohr's vague, puzzling sentences – always groping for the right word, never finding it – emerge from the fog and we see their underlying sense, if we keep in mind that Bohr's thinking is never on the ontological level traditional in physics. Always he is discussing not Nature, but our information about Nature. But physics did not have the vocabulary for expressing ideas on that level, hence the groping.

Paul Dirac, who was also living here in St. John's College at the time when he and Harold Jeffreys were doing their most important work side by side, seems never to have realized what Jeffreys had to offer him: probability theory as the vehicle for expressing epistemological notions quantitatively. It appears to us that, had either Bohr or Dirac understood the work of Jeffreys, the recent history of theoretical physics might have been very different. They would have had the language and technical apparatus with which Bohr's ideas could be stated and worked out precisely without mysticism. Had they done this, and explained clearly the distinction between the ontological and epistemological levels, Einstein would have understood it and accepted it at once.

Needless to say, we consider all of Einstein's reasoning and conclusions correct on his level; but on the other hand we think that Bohr was equally correct on his level, in saying that the act of measurement might perturb the system being measured, placing a limitation on the information we can acquire and therefore on the predictions we are able to make. There is nothing that one could object to in this conjecture, although the burden of proof is on the person who makes it. But we part company from Bohr when this metamorphoses without explanation into a claim that the limitation on the *predictions* of the present QM formalism are also – in exact, minute detail – limitations on the *measurements* that can be made in the laboratory!

Like Einstein, we can see no justification at all for this gratuitous assumption. We need a more orderly division of labour; it is simply not the proper business of theoretical physics to make pronouncements about what can and what cannot be measured in the laboratory, any more than it would be for an experimenter to issue proclamations about what can and cannot be predicted in the theory.

The issue of what kind of limitation on measurement really exists – or indeed, whether any limitation at all exists – is still logically an open question, that belongs to the province of the experimenter; but we may be able to settle it soon in the quantum optics laboratory, thanks to the spectacular recent advances in experimental techniques such as those by H. Walther and coworkers (Rempe et al, 1987) as discussed by Knight (1987) and in the *Scientific American* (June 1987, p. 25).

We believe that to achieve a rational picture of the world it is necessary to set up another clear division of labour within theoretical physics; it is the job of the laws of physics to describe physical causation at the level of ontology, and the job of probability theory to describe human inferences at the level of epistemology. The Copenhagen theory scrambles these very different functions into a nasty omelette in which the distinction between reality and our knowledge of reality is lost.

Although we agree with Bohr that in different circumstances (different states of knowledge) different quantities are predictable, in our view this does not cause the concepts themselves to fade in and out; valid concepts are not mutually incompatible. Therefore, to express precisely the effect of disturbance by measurement, on our information and our ability to predict, is not a philosophical problem calling for complementarity; it is a technical problem calling for probability theory as expounded by Jeffreys, and information theory. Indeed, we know that toward the end of his life, Bohr showed an interest in information theory.

EPR

But to return to the historical account; somehow, many physicists became persuaded that the success of the QM mathematical formalism proved the correctness of Bohr's private philosophy, even though hardly any – even among his disciples – understood what that philosophy was. All the attempts of Einstein, Schrödinger, and others to point out the patent illogic of this were rejected and sneered at; it is a worthy project for future psychologists to explain why.

The Einstein–Podolsky–Rosen (EPR) article of 1935 is Einstein's major effort to explain his objection to the completeness claim by an example that he thought was so forceful that nobody could miss the point. Two systems, S_1 and S_2, that were in interaction in the past are now separated, but they remain jointly in a pure state. Then EPR showed that according to QM an experimenter can measure a quantity q_1 in S_1, whereupon he can predict with certainty the value of q_2 in S_2. But he can equally well decide to measure a quantity p_1 that does not commute with q_1; whereupon he can predict with certainty the value of p_2 in S_2.

The systems can be so far apart that no light signal could have traveled between them in the time interval between the S_1 and S_2 measurements. Therefore, by means that could exert no causal influence on S_2 according to relativity theory, one can predict with certainty either of two noncommuting quantities, q_2 and p_2. EPR concluded that both q_2 and p_2 must have had existence as definite physical quantities before the measurements; but since no QM state vector is capable of representing this, the state vector cannot be the whole story.

Since today some think that merely to verify the correlations experimentally is to refute the EPR argument, let us stress that EPR did not question the existence of the correlations, which are to be expected in a classical theory. Indeed, were the correlations absent, their argument against the QM formalism would have failed. Their complaint was that, with physical causation unavailable, only instantaneous psychokinesis (the experimenter's free–will decision which experiment to do) is left to control distant events, the forcing of S_2 into an eigenstate of either q_2 or p_2. Einstein called this "a spooky kind of action at a distance".

To understand this, we must keep in mind that Einstein's thinking is always on the ontological level; the purpose of the EPR argument was to show that the QM state vector cannot be a representation of the "real physical situation" of a system. Bohr had never claimed that it was, although his strange way of expressing himself often led others to think that he was claiming this.

From his reply to EPR, we find that Bohr's position was like this: "You may decide, of your own free will, which experiment to do. If you do experiment E_1 you will get result R_1. If you do E_2 you will get R_2. Since it is fundamentally impossible to do both on the same system, and the present theory correctly predicts the results of either, how can you say that the theory is incomplete? What more can one ask of a theory?"

While it is easy to understand and agree with this on the epistemological level, the answer that I and many others would give is that we expect a physical theory to do more than merely predict experimental results in the manner of an empirical equation; we want to come down to Einstein's ontological level and understand what is happening when an atom emits light, when a spin enters a Stern–Gerlach magnet, etc. The Copenhagen theory, having no answer to any question of the form: "What is really happening when - - - ?", forbids us to ask such questions and tries to persuade us that it is philosophically naïve to want to know what is happening. But I do want to know, and I do not think this is naïve; and so for me QM is not a physical theory at all, only an empty mathematical shell in which a future theory may, perhaps, be built.

THE BELL INEQUALITIES

John Bell (1964) studied a simple realization of the EPR scenario in which two spin 1/2 particles denoted by A and B were jointly in a pure singlet state (like the ground state of the Helium atom) in the past. This is ionized by a spin–independent interaction and they move far apart, but they remain jointly in a pure singlet state, in which their spins are perfectly anticorrelated.

Each of two experimenters, stationed at A and B, has a Stern–Gerlach apparatus, which he can rotate to any angle. Following Bell's notation, we denote by $P(A|a)$ the probability that spin A will be found up in the direction of the unit vector "a"; and likewise $P(B|b)$ refers to spin B being up in the direction "b". For a singlet state, these are each equal to 1/2 from symmetry. The spooky business appears in the joint probability, which QM gives as

$$P(AB|ab) = \frac{1}{2}\,\sin^2(\theta/2) \tag{11}$$

where $\cos\theta = a \cdot b$. This does not factor in the form $P(AB|ab) = P(A|a)P(B|b)$ as one might expect for independent measurements. We can measure A in any direction we please; whereupon we can predict with certainty the value of B in the same direction.

From this, EPR would naturally conclude that the results of all possible measurements on B were predetermined by the real physical situation at B; *i.e.,* if we find B up in any direction b, then we would have found the same result whether or not the A measurement was made. Bohr would consider this a meaningless statement, since there is no way to verify it or refute it. Also, he would stress that we can measure B in only one direction, whereupon the perturbation of the measurement destroys whatever might have been seen in any other direction. Note that, as always, Bohr is epistemological; the notion of a "real physical situation" is just not in his vocabulary or his thinking.

EPR will then require some hidden variables in addition to the QM state vector to define that "real physical situation" which is to predetermine the results of all measurements on B. Bell, seeking to accommodate them, defines a class of hidden variable theories – call them Bell theories – in which a set of variables denoted collectively by λ also influences the outcomes A and B. It appears to him that the intentions of EPR are expressed in the most general way by writing

$$P(AB|ab) = \int P(A|a,\lambda)\,P(B|b,\lambda)\,p(\lambda)\,d\lambda \tag{12}$$

and he derives some inequalities that must be satisfied by any probability expressible in this form. But the QM probabilities easily violate these inequalities, and therefore they cannot result from any Bell theory. Let us understand at exactly what point in Bell's reasoning the conflict with QM is introduced.

Of course, the fundamentally correct relation according to probability theory would be,

$$P(AB|ab) = \int P(AB|ab\lambda)\,P(\lambda|ab)\,d\lambda\,. \tag{13}$$

But if we grant that knowledge of the experimenters' free choices (a,b) would give us no information about λ: $P(\lambda|ab) = p(\lambda)$ (and in this verbiage we too are being carefully epistemological), then Bell's interpretation of the EPR intentions lies in the factorization

$$P(AB|ab\lambda) = P(A|a\lambda)\,P(B|b\lambda) \tag{14}$$

whereas the fundamentally correct factorization would read:

$$P(AB|ab\lambda) = P(A|Bab\lambda)\,P(B|ab\lambda) = P(A|ab\lambda)\,P(B|Aab\lambda) \tag{15}$$

in which both a, b always appear as conditioning statements. However, Bell (1987) thinks that the EPR demand for locality, in which events at A should not influence events at B when the interval is spacelike, require the form (14). In his words: "It would be very remarkable if b proved to be a causal factor for A, or a for B; *i.e.*, if $P(A|a\lambda)$ depended on b or $P(B|b\lambda)$ depended on a. But according to quantum mechanics, such a dilemma can happen. Moreover, this peculiar long–range influence in question seems to go faster than light".

Note, however, that merely knowing the direction of the A measurement does not change any predictions at B, although it converts the initial pure singlet state into a mixture. It is easy to verify that according to QM, $P(B|ab) = P(B|b) = 1/2$ for all a, b. As we would expect from (15), it is necessary to know also the result of the A measurement before the correlation affects our predictions; according to QM, $P(B|Aab) = (1 - \cos\theta)/2$. Thus while the QM formalism disagrees with Bell's factorization (14), it appears consistent with what we have called the "fundamentally correct" probability relations (perhaps now it is clearer why we said that some of Bohr's ideas could have been expressed precisely in Bayesian terms).

Equation (14) is, therefore, the point where Bell introduces a conflict with QM. Recognizing this, it is evident that one could produce any number of experimental tests where the predictions of QM conflict with various predictions of (14). The particular set of inequalities given by Bell is only one example of this, and not even the most cogent one. We shall leave it as an exercise for the reader to show that, at this point, application of Bayesian principles would have yielded a significance test for (14) that is more powerful than the Bell inequalities.†

Regardless, it seemed to everybody twenty years ago that the stage was set for an experimental test of the issue; perform experiments where the predictions of quantum theory violate the Bell inequalities, and see whether the data violate them also. If so, then all possible local theories – whether causal or not – are demolished in a single stroke, and the Universe runs on psychokinesis. At least, that was the reasoning.

The experiments designed to test this, of which the one of Alain Aspect (1985, 1986) is perhaps the most cogent to date, have with only one exception ended with the verdict "quantum theory confirmed", and accordingly there has been quite a parade of physicists jumping on the bandwagon, declaring publicly that they now believe in psychokinesis. Of course, they do not use that word; but at the 1984 Santa Fe Workshop (Moore & Scully, 1986) more than one was heard to say: "The experimental evidence now forces us to believe that atoms are not real." and nobody rose to question this, although it made me wonder what they thought Alain's apparatus was made of.

Alain Aspect himself has remained admirably level–headed through all this, quite properly challenging us to produce a classical explanation of his experiment; but at the same time refusing to be stampeded into taking an obviously insane position as did some others. The dilemma is not that the QM formalism is giving wrong predictions, but that the current attempts at interpreting that formalism are giving us just that spooky picture of the world that Einstein anticipated and objected to. Of course, those with a penchant for mysticism are delighted.

How do we get out of this? Just as Bell revealed hidden assumptions in von Neumann's argument, so we need to reveal the hidden assumptions in Bell's argument. There are at least two of them, both of which require the Jeffreys viewpoint about probability to recognize:

(1) As his words above show, Bell took it for granted that a conditional probability $P(X|Y)$ expresses a physical causal influence, exerted by Y on X. But we show presently that one cannot even reason correctly in so simple a problem as drawing two balls from Bernoulli's Urn, if he interprets probabilities in this way. Fundamentally, consistency requires that conditional probabilities express *logical* inferences, just as Harold Jeffreys saw. Indeed, this is also the crucial point that Bohr made in his reply to EPR, in words that Bell quoted and which we repeat below.

(2) The class of Bell theories does not include all local hidden variable theories; it appears to us that it excludes just the class of theories that Einstein would have liked most. Again, we need to learn from Jeffreys the distinction between the epistemological probabilities of the QM formalism and the ontological frequencies that we measure in our experiments. A hidden

† As we noted long ago (Jaynes, 1973), In the optical "photon correlation" experiment where according to QM the two photons have parallel polarization, the non–existence of correlations when the polarizers are at a 90 degree angle is a more sensitive (and experimentally simpler) test of QM than are the Bell inequalities.

variable theory need not reproduce the mathematical form of the QM probabilities in the manner of (12) in order to predict the same observable facts that QM does.

The spooky superluminal stuff would follow from Hidden Assumption (1); but that assumption disappears as soon as we recognize, with Jeffreys and Bohr, that what is traveling faster than light is not a physical causal influence, but only a logical inference. Here is Bohr's quoted statement (italics his):

> "Of course there is in a case like that just considered no question of a mechanical disturbance of the system under investigation during the last critical phase of the measuring procedure. But even at this stage there is essentially the question of *an influence on the very conditions which define the possible types of predictions regarding the future behavior of the system.*"

After quoting these words, Bell added: "Indeed I have very little idea what this means." And we must admit that this is a prime example of the cryptic obscurity of Bohr's writings. So – with the benefit of some forty years of contemplating that statement in the context of his other writings – here is our attempt to translate Bohr's statement into plain English:

> "The measurement at A at time t does not change the real physical situation at B; but it changes our state of knowledge about that situation, and therefore it changes the predictions we are able to make about B at some time t'. Since this is a matter of logic rather than physical causation, there is no action at a distance and no difficulty with relativity [also, it does not matter whether t' is before, equal to, or after t]."

Again we see how Bohr's epistemological viewpoint corresponds to Bayesian inference, and could have been expressed precisely in Bayesian terms. However, Bohr could not bring himself to say it as we did, because for him the phrase "real physical situation" was taboo.

But it may seem paradoxical that two different pure states (eigenstates of noncommuting quantities q_2 and p_2) can both represent the same real physical situation; if so, then perhaps the conclusion is that one has learned an important fact about the relation of the QM state vector to reality. This supports the Einstein view of the meaning of a pure state as an ensemble; for in statistical mechanics it is a platitude that the true microstate may appear in two different ensembles, representing two different states of knowledge about the microstate.

BERNOULLI'S URN REVISITED

Define the propositions:

$I \equiv$ "Our urn contains N balls, identical in every respect except that M of them are red, the remaining $N - M$ white. We have no information about the location of particular balls in the urn. They are drawn out blindfolded without replacement."

$R_i \equiv$ "Red on the i'th draw, $i = 1, 2, \ldots$"

Successive draws from the urn are a microcosm of the EPR experiment. For the first draw, given only the prior information I, we have

$$P(R_1|I) = M/N \; . \tag{16}$$

Now if we know that red was found on the first draw, then that changes the contents of the urn for the second:

$$P(R_2|R_1, I) = (M-1)/(N-1) \tag{17}$$

and this conditional probability expresses the causal influence of the first draw on the second, in just the way that Bell assumed.

But suppose we are told only that red was drawn on the second draw; what is now our probability for red on the first draw? Whatever happens on the second draw cannot exert any physical influence on the condition of the urn at the first draw; so presumably one who believes with Bell that a conditional probability expresses a physical causal influence, would say that $P(R_1|R_2, I) = P(R_1|I)$. But this is patently wrong; probability theory requires that

$$P(R_1|R_2, I) = P(R_2|R_1, I)\ . \tag{18}$$

This is particularly obvious in the case $M = 1$; for if we know that the one red ball was taken in the second draw, then it is certain that it could not have been taken in the first.

In (18) the probability on the right expresses a physical causation, that on the left only an inference. Nevertheless, the probabilities are necessarily equal because, although a later draw cannot physically affect conditions at an earlier one, *information* about the result of the second draw has precisely the same effect on our *state of knowledge* about what could have been taken in the first draw, as if their order were reversed.

Eq. (18) is only a special case of a much more general result. The probability of drawing any sequence of red and white balls (the hypergeometric distribution) depends only on the number of red and white balls, not on the order in which they appear; *i.e.*, it is an exchangeable distribution. From this it follows by a simple calculation that for all i and j,

$$P(R_i|I) = P(R_j|I) = M/N \tag{19}$$

That is, just as in QM, merely knowing that other draws have been made does not change our prediction for any specified draw, although it changes the hypothesis space in which the prediction is made; before there is a change in the actual prediction it is necessary to know also the results of other draws. But the joint probability is by the product rule,

$$P(R_i, R_j|I) = P(R_i|R_j, I)\, P(R_j|I) = P(R_j|R_i, I)\, P(R_i|I) \tag{20}$$

and so we have for all i and j,

$$P(R_i|R_j, I) = P(R_j|R_i, I) \tag{21}$$

and again a conditional probability which expresses only an inference is necessarily equal to one that expresses a physical causation. This would be true not only for the hypergeometric distribution, but for any exchangeable distribution. We see from this how far Karl Popper would have got with his "propensity" theory of probability, had he tried to apply it to a few simple problems.

It might be thought that this phenomenon is a peculiarity of probability theory. On the contrary, it remains true even in pure deductive logic; for if A implies B, then not–B implies not–A. But if we tried to interpret "A implies B" as meaning "A is the physical cause of B", we could hardly accept that "not–B is the physical cause of not–A". Because of this lack of contraposition, we cannot in general interpret logical implication as physical causation, any more than we can conditional probability. Elementary facts like this are well understood in economics (Simon & Rescher, 1966; Zellner, 1984); it is high time that they were recognized in theoretical physics.

OTHER HIDDEN – VARIABLE THEORIES

Now consider Hidden Assumption (2). Bell theories make no mention of time variation of the hidden variable λ; but if it is to take over the job formerly performed by the QM state vector ψ, then λ must obey some equations of motion which are to replace the Schrödinger equation.

This is important, because one way for a causal theory to get probability into things is time alternation; for example, in conditions where present QM yields a time independent probability p for spin up, λ would be oscillating in such a way that for a fraction p of the time the result is "up", etc. Indeed, Einstein would have considered this the natural way to obtain the QM probabilities from a causal theory, for in his early papers he defined the "probability of a state" as the fraction of the time in which a system is in that state. But this is a relation between QM and the causal theory of a different nature than is supposed by the form (12).

At first glance, one might object to this statement by saying that Bell theories do not explicitly forbid λ to vary with time. But if it did, then (12) would not reproduce the QM probabilities; it would yield time–dependent probabilities in situations where the QM probabilities are constant. Indeed, if the Bell theory is a truly causal local theory the probabilities given by (12) could take on only the values 0 and 1, and the QM probabilities would be time averages of them.

That time alternation theories differ fundamentally from QM is clear also from the fact that they predict new effects not in QM, that might in principle be observed experimentally, leading to a crucial test. For example, when two spins are perfectly anticorrelated, that would presumably signify that their λ's are oscillating in perfect synchronism so that, for a given result of the A measurement, the exact time interval between the A and B measurements could determine the actual result at B, not merely its QM probability. Then we would be penetrating the fog and observing more than Bohr thought possible. The experiments of H. Walther and coworkers on single atom masers are already showing some resemblance to the technology that would be required to perform such an experiment.

We have shown only that some of the conclusions that have been drawn from the Bell–Aspect work were premature because (1) the spooky stuff was due only to the mistaken assumption that a conditional probability must signify a physical influence, and (2) the Bell arguments do not consider all possible local theories; the Bell inequalities are only limitations on what can be predicted by Bell theories. The Aspect experiment may show that such theories are untenable, but without further analysis it leaves open the status of other local causal theories more to Einstein's liking.

That further analysis is, in fact, already underway. An important part of it has been provided by Steve Gull's "You can't program two independently running computers to emulate the EPR experiment" theorem, which we learned about at this meeting. It seems, at first glance, to be just what we have needed because it could lead to more cogent tests of these issues than did the Bell argument. The suggestion is that some of the QM predictions can be duplicated by local causal theories only by invoking teleological elements as in the Wheeler–Feynman electrodynamics. If so, then a crucial experiment would be to verify the QM predictions in such cases. It is not obvious whether the Aspect experiment serves this purpose.

The implication seems to be that, if the QM predictions continue to be confirmed, we exorcise Bell's superluminal spook only to face Gull's teleological spook. However, we shall not rush to premature judgments. Recalling that it required some 30 years to locate von Neumann's hidden assumptions, and then over 20 years to locate Bell's, it seems reasonable to ask for a little time to search for Gull's, before drawing conclusions and possibly suggesting new experiments.

In this discussion we have not found any conflict between Bohr's position and Bayesian probability theory, which are both at the epistemological level. Nevertheless, differences appear on more detailed examination to be reported elsewhere. Of course, the QM formalism also contains fundamentally important and correct ontological elements; for example, there has to be something physically real in the eigenvalues and matrix elements of the operators from which we obtain detailed predictions of spectral lines. It seems that, to unscramble the epistemological probability statements from the ontological elements we need to find a different formalism, isomorphic in some sense but based on different variables; it was only through some weird mathematical accident that it was possible to find a variable ψ which scrambles them up in the present way.

There is clearly a major, fundamentally important mystery still to be cleared up here; but unless you maintain your faith that there is a rational explanation, you will never find that explanation. For 60 years, acceptance of the Copenhagen interpretation has prevented any further

16

progress in basic understanding of physical law. Harold Jeffreys (1957) put it just right: "Science at any moment does not claim to have explanations of everything; and acceptance of an inadequate explanation discourages search for a good one."

Now let us turn to an area that seems about as different as one could imagine, yet the underlying logic of it hangs on the same point: What happens in the real world depends on physical law and is on the level of ontology. What we can predict depends on our state of knowledge and is necessarily on the level of epistemology. He who confuses reality with his knowledge of reality generates needless artificial mysteries.

THE SECOND LAW IN BIOLOGY

As we learn in elementary thermodynamics, Kelvin's formula for the efficiency of a Carnot heat engine operating between upper and lower temperatures T_1, T_2:

$$\eta \leq 1 - T_2/T_1 \,, \tag{22}$$

with equality if and only if the engine is reversible, expresses a limitation imposed by the second law of thermodynamics. But the world's most universally available source of work – the animal muscle – presents us with a seemingly flagrant violation of that formula.

Our muscles deliver useful work when there is no cold reservoir at hand (on a hot day the ambient temperature is at or above body temperature) and a naïve application of (22) would lead us to predict zero, or even negative efficiency. The observed efficiency of a muscle, defined as

$$\eta \equiv \frac{(\text{work done})}{(\text{work done} + \text{heat generated})}$$

is difficult to measure, and it is difficult to find reliable experimental values with accounts of how the experiments were done. We shall use only the latest value we have located, (Alberts, *et al.* 1983). The heat generated that can be attributed to muscle activity appears to be as low as about 3/7 of the work done; which implies that observed muscle efficiencies can be as high as 70% in favourable conditions, although a Carnot engine would require an upper temperature T_1 of about 1000 K to achieve this. Many authors have wondered how this can be.

The obvious first answer is, of course, that a muscle is not a heat engine. It draws its energy, not from any heat reservoir, but from the activated molecules produced by a chemical reaction. Only when we first allow that primary energy to degrade itself into heat at temperature T_1 – and then extract only that heat for our engine – does the Kelvin efficiency formula (22) apply in its conventional meaning. It appears that our muscles have learned how to capture the primary energy before it has a chance to degrade; but how do we relate this to the second law?

Basic material on muscle structure and energetics of biochemical reactions is given by Squire (1981) and Lehninger (1982), and profusely illustrated by Alberts, *et al* (1983). The source of energy for muscle contraction (and indeed for almost everything a cell does that requires energy) is believed to be hydrolysis of adenosine triphosphate (ATP), for which the reported heat of reaction is $\Delta H = -9.9$ kcal/mol, or 0.43 ev per molecule. This energy is delivered to some kind of "engine" in a muscle fiber, from whence emerges useful work by contraction. The heat generated is carried off by the blood stream, at body temperature, $273 + 37 = 310$ K. Thus the data we have to account for are:

Ambient temperature: 310 K
Source energy: 0.43 ev/molecule
Efficiency: 70%.

We do not attempt to analyze all existing biological knowledge in this field about the details of that engine, although in our conclusions we shall be able to offer some tentative comments on it. Our present concern is with the general physical principles that must govern conversion of chemical energy into mechanical work in any system, equilibrium or nonequilibrium, biological or otherwise, whatever the details of the engine. In the known facts of muscle performance we have some uniquely cogent evidence relevant to this problem.

The status of the second law in biology has long been a mystery. Not only was there a seeming numerical contradiction between muscle efficiency and the second law, but also the general self-organizing power of biological systems seemed to conflict with the "tendency to disorder" philosophy that had become attached to the second law (much as Bohr's philosophy of complementarity had become attached to quantum mechanics). This led, predictably, to a reaction in the direction of vitalism.

In our view, whatever happens in a living cell is just as much a real physical phenomenon as what happens in a steam engine; far from violating physical laws, biological systems exhibit the operation of those laws in their full generality and diversity, that physicists had not considered in the historical development of thermodynamics. Therefore, if biological systems seem to violate conventional statements of the second law, our conclusion is only that the second law needs to be restated more carefully. Our present aim is therefore to find a statement of the second law that reduces to the traditional statements of Clausius and Gibbs in the domain where they were valid, but is general enough to include biological phenomena.

The "tendency to disorder" arguments are too vague to be of any constructive use for this purpose; and in any event it is clear that they must be mistaken and it would be interesting to understand why (we think that Maxwell explained why at the end of the aforementioned article on diffusion). Muscle efficiency will provide our test case, because here we have some quantitative data to account for. But a muscle operates in a nonequilibrium situation, for which no definite second law is to be found in the thermodynamic literature. The conventional second law presupposes thermalisation because temperature and entropy are defined only for states of thermal equilibrium. How do we circumvent this?

Some have thought that it would be a highly difficult theoretical problem, calling for a generalised ergodic theory to include analysis of "mixing" and "chaos". Another school of thought holds that we need a modification of the microscopic equations of motion to circumvent Liouville's theorem (conservation of phase volume in classical Hamiltonian systems, or unitarity in quantum theory), which is thought to be in conflict with the second law.

We suggest, on the contrary, that only very simple physical reasoning is required, and all the clues needed to determine the answer can be found already in the writings of James Clerk Maxwell and J. Willard Gibbs over 100 years ago. Both had perceived the epistemological nature of the second law and we think that, had either lived a few years longer, our generalised second law would long since have been familiar to all scientists. We give the argument in three steps: (a) reinterpret the Kelvin formula, (b) make a more general statement of the second law, (c) test it numerically against muscles.

The observed efficiency of muscles may be more cogent for this purpose than one might at first think. Since animals have evolved the senses of sight, sound, and smell to the limiting sensitivity permitted by physical law, it is only to be expected that they would also have evolved muscle efficiency (which must be of equal survival value) correspondingly. If so, then the maximum observed efficiency of muscles should be not merely a lower bound on the maximum theoretical efficiency we seek, but close to it numerically.

GENERALISED EFFICIENCY FORMULA

Consider the problem first in the simplicity of classical physics, where the Rayleigh–Jeans equipartition law holds. If in the Kelvin formula (22) we replace temperature by what it then amounts to – energy per degree of freedom $E/N = (1/2)\,kT$, it takes the form

$$\eta \le 1 - (E_2/N_2)(N_1/E_1) \tag{23}$$

which does not look like much progress, but by this trivial rewriting we have removed the limitation of thermal equilibrium on our energy source and sink. For "temperature" is defined only for a system in thermal equilibrium, while "energy per degree of freedom" is meaningful not only in thermal equilibrium, but for any small part of a system – such as those activated molecules – which might be far from thermal equilibrium with the surroundings.

One might then question whether such a nonequilibrium interpretation of (22) is valid. We may, however, reason as follows. Although conventional thermodynamics defines temperature and entropy only in equilibrium situations where all translational and vibrational degrees of freedom (microscopic coordinates) have the same average energy, it cannot matter to an engine whether all parts of its energy source are in equilibrium with each other.

Only those degrees of freedom with which the engine interacts can be involved in its efficiency; the engine has no way of knowing whether the others are or are not excited to the same average energy. Therefore, since (23) is unquestionably valid when both reservoirs are in thermal equilibrium, it should be correct more generally, if we take E_2/N_2 and E_1/N_1 to be the average energy in those degrees of freedom with which the engine actually interacts. But while a muscle has a small source reservoir, it has a large sink. Therefore for E_2/N_2 we may take $(1/2)\,kT_2$ at body temperature.

As a check on this reasoning, if the primary energy is concentrated in a single degree of freedom and we can extract it before it spreads at all, then our engine is in effect a "pure mechanism" like a lever. The generalised efficiency (23) then reduces to $1 - kT_2/2E_1$ or, interpreting E_1 as the work delivered to the lever,

$$(\text{Work out}) = (\text{Work in}) - (1/2)\,kT_2\,. \tag{24}$$

The last term is just the mean thermal energy of the lever itself, which cannot be extracted reproducibly by an apparatus that is itself at temperature T_2 or higher. At least, if anyone should succeed in doing this, then he would need only to wait a short time until the lever has absorbed another $(1/2)\,kT_2$ from its surroundings, and repeat – and we would have the perpetual motion machine that the second law holds to be impossible. Thus (24) still expresses a valid second law limitation, and the simple generalisation (23) of Kelvin's formula appears to have a rather wide range of application.

But although these are interesting hints, we are after something more general, which can replace the second law for all purposes, not merely engines. To achieve this we must understand clearly the basic physical reason why there is a second law limitation on processes. We suggest that the fundamental keyword characterizing the second law is not "disorder", but *reproducibility.*

THE REASON FOR IT

The second law arises from a deep interplay between the epistemological macrostate (*i.e.*, the variables like pressure, volume, magnetization that an experimenter measures and which therefore represent our knowledge about the system) and the ontological microstate (the coordinates and momenta of individual atoms, which determine what the system will in fact do). For example, in either a heat engine or a muscle the goal is to recapture energy that is spread in an unknown and uncontrolled way over many microscopic degrees of freedom of the source reservoir, and concentrate it back into a single degree of freedom, the motion of a piston or tendon. The more it has spread, the more difficult it will be to do this.

The basic reason for the "second law" limitation on efficiency is that the engine must work reproducibly; an engine that delivered work only occasionally, by chance (whenever the initial microstate of reservoirs and engine happened to be just right) would be unacceptable in engineering and biology alike.

The initial microstate is unknown because it is not being controlled by any of the imposed macroscopic conditions. The initial microstate might be anywhere in some large phase volume W_i compatible with the initial macrostate M_i; and the engine must still work. It is then Liouville's theorem that places the limitation on what can be done; physical law does not permit us to concentrate the final microstates into a smaller phase volume than W_i and therefore we cannot go reproducibly to any final macrostate M_f whose phase volume W_f is smaller than W_i. The inequality $W_i \leq W_f$ is a necessary condition for any macroscopic process $M_i \to M_f$ to be reproducible for all initial microstates in W_i.

Of course, something may happen by chance that is not reproducible. As a close analogy, we can pump the water from a tank of volume V_1 into a larger tank of volume $V_2 > V_1$, but not into a smaller one of volume $V_3 < V_1$. Therefore any particular tagged water molecule in one tank can be moved reproducibly into a larger tank but not into a smaller one; the probability of success would be something like V_3/V_1. Here the tanks correspond to the macrostates M, their volumes V correspond to phase volumes W, the tagged molecule represents the unknown true microstate, and the fact that the water flow is incompressible corresponds to Liouville's theorem.

Now we know that in classical thermodynamics, as was first noted by Boltzmann, the thermodynamic entropy of an equilibrium macrostate M is given to within an additive constant by $S(M) = k \log W(M)$, where k is Boltzmann's constant. This relation was then stressed by Planck and Einstein, who made important use of it in their research. But the above arguments make it clear that there was no need to restrict this to equilibrium macrostates M. Any macrostate – equilibrium or nonequilibrium – has an entropy $S(M) = k \log W(M)$, where $W(M)$ is the phase volume compatible with the controlled or observed macrovariables X_i (pressure, volume, magnetization, heat flux, electric current, *etc.*) that define M. Then a generalised second law

$$S(\text{initial}) \leq S(\text{final}) \tag{25}$$

follows immediately from Liouville's theorem, as a necessary condition for the change of state $M_i \to M_f$ to be reproducible.

Stated more carefully, we mean "reproducible by an experimenter who can control only the macrovariables $\{X_i\}$ that define the macrostates M". A little thought makes it clear that this proviso was needed already in the classical equilibrium theory, in order to have an air–tight statement of the second law which could not be violated by a clever experimenter. For if Mr. A defines his thermodynamic states by the n macrovariables $\{X_1 \dots X_n\}$ that he is controlling and/or observing, his entropy S_n is a function of those n variables. If now Mr. B, unknown to Mr. A, manipulates a new macrovariable X_{n+1} outside the set that Mr. A is controlling or observing, he can bring about, reproducibly, a change of state for which S_n decreases spontaneously, although S_{n+1} does not. Thus he will appear to Mr. A as a magician who can produce spontaneous violations of the second law, at will.

But now we must face an ambiguity in the definition and meaning of W; it appears to have different aspects. The phase volume $W(X_1 \dots X_n)$ consistent with a given set of extensive macrovariables $\{X_1 \dots X_n\}$ is a definite, calculable quantity which represents on the one hand the degree of control of an experimenter over the microstate, when he can manipulate only those macrovariables; thus W appears ontological. On the other hand, W represents equally well our degree of ignorance about the microstate when we know only those macrovariables and nothing else; and thus it appears epistemological. But as illustrated by the scenario of Mr. A and Mr. B above, it is a matter of free choice on our part which set of macrovariables we shall use to define our macrostates; thus it appears also anthropomorphic! Finally, we have been vague about just how many microscopic degrees of freedom are to be included in W. Then what is the meaning of the second law (25)? Is it an ontological law of physics, an epistemological human prediction, or an anthropomorphic art form?

Part of the answer is that Eq. (25) cannot be an ontological statement (that is, a deductively proved consequence of the laws of physics) because the mere calculation of W makes no use of the equations of motion, which alone determine which macrostate will in fact evolve from a given microstate in W_i. It may be that, because of properties of the equations of motion that we did not use, our experimenter's method of realizing the macrostate M_i would not, in many repetitions, produce all microstates in the volume W_i, only a negligibly small subset of them occupying a phase volume $W' << W_i$. Then the process $M_i \to M_f$ might still be possible reproducibly even though $S_i > S_f$, if $S' \leq S_f$. Conversely, because of special circumstances such as unusual constants of the motion, the process $M_i \to M_f$ might prove to be impossible even though $S_i < S_f$. The second law cannot be proved by deductive reasoning from the information that we actually have.

On the other hand, (25) is always epistemological because it is always true that $W(M)$ measures our degree of ignorance about the microstate when we know only the macrostate M. Thus the original second law and our generalisation (25) of it have the same logical status as Bayesian inference; they represent the best predictions we can make from the information we have. In fact, a refined form of (25) can be derived as an example of Bayesian inference. Therefore the second law works functionally like any other Bayesian inference; the predictions are usually right, indicating that the information used was adequate for the purpose. Only when the predictions are wrong do we learn new things about the ontological laws of physics (such as new constants of the motion).

It is greatly to our advantage to recognize this. By getting our logic straight we not only avoid the Mind Projection Fallacy of claiming more than has been proved, we gain an important technical flexibility in using the second law. Instead of committing the error of supposing that a given physical system has one and only one "true" ontological entropy, we recognize that we could have many different states of knowledge about it, leading to many different entropies referring to the same physical system (as in the scenario of Mr. A and Mr. B above), which can serve many different purposes.

Just as the class of phenomena that an experimenter can evoke from a given system in the laboratory depends on the kind of apparatus he has (which of its macrovariables he can manipulate), so the class of phenomena that we can predict with thermodynamics for a given system depends on the kind of knowledge we have about it. This is not a paradox, but a platitude; indeed, in any scientific question, what we can predict depends, obviously, on what information we have. If we fail to specify what biological information we propose to take into account, then thermodynamics may not be able to give us any useful answer because we have not asked any well posed question.

Even when it does not lead to different final results, taking prior information into account can affect computational efficiency in applying the second law, because it can help us to make a more parsimonious choice of the microvariables that we shall include in W. For it to be generally valid, the entropy in (25) must be, in principle, the total entropy of all systems that take part in the process. But this does not, in practice, determine exactly how much of the outside world is to be included. In a sense everything in the universe is in constant interaction with everything else, and one must decide when to stop including things. Including more than we need is not harmful in the sense of leading to errors, since this only adds the same quantity to both sides of (25). But it can cost us wasted effort in calculating unnecessary details that cancel out of our final results.

At this point the aforementioned flexibility of our methods becomes important. We have already made use of it in the discussion following Eq. (23); now we want to apply that reasoning to phase volumes and to general processes. In a fast process, that happens in a time so short that thermal equilibrium of the whole system is never reached, only the phase volume belonging to those degrees of freedom actually involved in the interactions could be relevant; the second law may be applied in terms of Liouville's theorem in a relatively small subspace of the full one that we use in equilibrium theory. In the application to muscle efficiency, this means that we need calculate only phase volumes corresponding to degrees of freedom that are directly involved in muscle operation; ones that are affected only later, after the muscle contraction is over, may be relevant for the ultimate fate of the heat generated, but they cannot affect its efficiency.

This corresponds to a familiar procedure in treatment of spin systems. Spin–spin relaxation is often orders of magnitude faster than spin–lattice relaxation, so one can consider the microvariables of the spin system as forming a nearly isolated dynamical system in their own right, with a "private" second law of their own. Slichter (1980) shows that this approach enables one to predict masses of details correctly.

In the above we have supposed the classical equipartition law; but our arguments should need modifying only if the engine (the piston or tendon) interacts directly with degrees of freedom for which equipartition fails. In the case of muscles, it appears that the direct interactions are with coordinates of low–frequency vibration modes of large protein molecules. How energy gets transferred from an excited electronic state of ATP to such a vibration mode would remain in the province of quantum theory; but this can be virtually 100% efficient.

QUANTITATIVE DERIVATION

Now we are ready for a specific calculation of muscle efficiency using the above principles. The phase volumes W that we calculate are, of course, functions of the macrovariables that define the macrostates. In the case of a muscle, what is happening is just that energy Q_1 is being abstracted from the source reservoir and energy Q_2 is delivered to the sink, the difference appearing as work. Energy is the only macrovariable being manipulated, so our phase volumes will be functions of source and sink energies. We need not consider a phase volume for the engine, because that is the same at the beginning and end (in cyclic operation, the engine is restored ready to run again). As in conventional statistical mechanics, we introduce the density functions $\rho(E)$, often called structure functions, of source and sink by considering their energies known to some tolerances δE. Thus the phase volumes for source and sink are

$$W_1 = \rho_1(E_1)\,\delta E_1 \tag{26a}$$

$$W_2 = \rho_2(E_2)\,\delta E_2 \tag{26b}$$

Then the initial and final phase volumes are

$$W_i = \rho_1(E_1)\,\rho_2(E_2)\,\delta E_1\,\delta E_2 \tag{27a}$$

$$W_f = \rho_1(E_1 - Q_1)\,\rho_2(E_2 + Q_2)\,\delta E_1\,\delta E_2 \tag{27b}$$

With Q_1 and Q_2 definite quantities, the tolerances δE_1 and δE_2 are the same at the beginning and end, so they cancel out and their values do not matter. The necessary condition of reproducibility $W_i \le W_f$ when we manipulate only energies now becomes:

$$\rho_1(E_1)\,\rho_2(E_2) \le \rho_1(E_1 - Q_1)\,\rho_2(E_2 + Q_2)\,. \tag{28}$$

Let us try to predict the maximum work obtainable by using only this relation (which makes no use of such notions as temperature, equation of state, heat capacity, or reversible operation). Given the energy Q_1 extracted from the source, the maximum work we can get reproducibly is $Q_1 - Q_2$, where from (28), Q_2 is the root of

$$\log \rho_1(E_1) + \log \rho_2(E_2) = \log \rho_1(E_1 - Q_1) + \log \rho_2(E_2 + Q_2)\,. \tag{29}$$

Now vary Q_1; the RHS of (29) remains constant, and $Q_1 - Q_2$ is a maximum when

$$\frac{\partial}{\partial Q_1} \log \rho_1(E_1 - Q_1) + \frac{\partial}{\partial Q_2} \log \rho_2(E_2 + Q_2) = 0 \tag{30}$$

Therefore the maximum efficiency is

$$\eta = \frac{Q_1 - Q_2}{E_1} \tag{31}$$

where Q_1, Q_2 are the simultaneous roots of (29) and (30). Note that this is the "global" efficiency that we need for this application, in which we contemplate extracting as much of the total available energy E_1 as possible, whereas the Kelvin formula (22) is the differential efficiency, holding when the amount of energy Q_1 extracted is small compared to the total energy E_1 in the high temperature reservoir, so that its temperature is not changed appreciably by the operation of the engine.

Now we need to decide on the functions $\rho_1(E_1)$ and $\rho_2(E_2)$. Recall some familiar examples of such functions; for an ideal gas of n particles in volume V,

$$\rho(E) = \frac{V^n\,(2\pi m E)^{3n/2\,-\,1}}{\Gamma(3n/2)}\,. \tag{32}$$

For n classical harmonic oscillators with frequencies $\{\omega_1 \ldots \omega_n\}$,

$$\rho(E) = \frac{(2\pi)^n}{(\Pi_i \omega_i)\,\Gamma(n)}\,E^n\,. \tag{33}$$

In both cases, $\rho(E)$ is proportional to $E^{N/2}$, where N is the number of degrees of freedom of the system. This is approximately true for most systems even in quantum statistics, where N may be regarded as a slowly varying function of E, signifying the effective number of degrees of freedom excited at energy E. So let us take

$$\log \rho_1(E_1) = \frac{N_1}{2} \log E_1 + \text{const.} \tag{34a}$$

$$\log \rho_2(E_2) = \frac{N_2}{2} \log E_2 + \text{const.} \tag{34b}$$

which seems quite realistic for the case of muscles. Eliminating Q_2 from (29), (30), Q_1 is determined from

$$(N_1 + N_2)\ \log\left[\frac{E_1\ -\ Q_1}{E_1}\right] = N_2\ \log\left[\frac{N_1 E_2}{N_2 E_1}\right] \tag{35}$$

and then Q_2 is found from (30). But from (23) we recognize the quantity

$$r \equiv \frac{N_1 E_2}{N_2 E_1} \tag{36}$$

as the analog of (T_2/T_1) in equilibrium theory. Then after some algebra, we find that (31) is

23

$$\eta = 1 + \frac{N_2}{N_1}\, r \;-\; \left[\frac{N_1 + N_2}{N_1}\right] r^{\frac{N_2}{N_1+N_2}} \,. \tag{37}$$

In the case $N_1 = N_2$, this is $(1 - \sqrt{r})^2$, contrasted with Kelvin's differential efficiency $(1 - r)$. Appropriate for muscles is the limiting form as $N_2 \to \infty$, $E_2/N_2 \to (1/2)\, kT_2 =$ const. (the blood stream is very large compared to a muscle fiber). Some care is needed in taking the limit, and (37) then reduces to

$$\eta = 1 - r + r \log r \,. \tag{38}$$

Now everything boils down to the question: what is r for a muscle? As before, let us take for the large sink reservoir, $E_2 = (1/2)\ N_2\ kT_2$, where $T_2 = 310\ K$. The maximum theoretical efficiency surely corresponds to the maximum concentration of primary energy that seems possible in a muscle; the energy of ATP hydrolysis of one molecule is concentrated into a single vibration mode and is captured before it spreads to others. Therefore for the source, let $E_1 = 0.43\ n$ ev, the heat of reaction of n ATP molecules, and $N_1 = 2n$, corresponding to one vibration mode per molecule. This gives

$$r = \frac{310 \times 1.36 \times 10^{-16}}{0.43 \times 1.6 \times 10^{-12}} = 0.062 \,, \tag{39}$$

from which (38) gives

$$\eta = 76.5\% \,. \tag{40}$$

Doubtless, the near agreement with the value reported by Alberts et al (1983) is fortuitous; the existing measurements are too uncertain to draw any real conclusions. But one might have hoped that the maximum theoretical efficiency would come out just above the maximum observed efficiency; and at least that much has been realized. It appears that the information we used was adequate for the purpose, and there is no longer any mystery.

A CHECK

We derived the efficiency formula (38) without assuming any slow reversible operation as conventional thermodynamics does. On the other hand, neither did we assume that it is not slow, so if our derivation is correct, the formula ought to remain valid in the limit when the process is so slow that the conventional theory does apply. To check this, let us apply conventional theory to a small source whose temperature T_1 drops slowly as the engine runs, so we have a sequence of infinitesimal reversible Carnot cycles. Suppose that the sink is so large that T_2 remains constant. Then drawing heat Q_1 from the source, the maximum work we can get according to classical thermodynamics is

$$W(Q_1) = \int_0^{Q_1} \left[1 - \frac{T_2}{T_1(Q)}\right] dQ \,. \tag{41}$$

Now suppose, corresponding to the Rayleigh–Jeans assumptions in our first derivation, that the source has a constant heat capacity C, so that $T_1(Q) = T_1 - Q/C$, where T_1 is the initial source temperature; then $E_1 = CT_1$. The engine will run only as long as $T_1(Q) > T_2$, so the maximum obtainable work is given when the upper limit of integration is $Q_1 = C(T_1 - T_2)$. Making these substitutions, the integral is easily evaluated, with the result

$$W_{max} = C\left[T_1 - T_2 + T_2 \log \frac{T_2}{T_1}\right] \,. \tag{42}$$

Dividing by $E_1 = CT_1$, we recover the result (38) that we derived previously using only phase volume considerations. This confirms that our generalised second law reduces, as it should, to the conventional one when the latter is applicable.

But this conventional "slow, reversible" second law is not applicable to a muscle, because if a muscle operated slowly enough to make its assumptions valid, other degrees of freedom that we have left out of our calculation would take over and thermalise the primary energy, making the muscle useless. It is just to avoid thermalisation that biological processes must take place rapidly, and thus we require a "fast" second law to analyze them.

Our generalisation of the second law not only preserves the dynamics and therefore the Liouville theorem, it preserves the Clausius relation $S_i \le S_f$ and the Boltzmann entropy formula $S = k \log W$; and it even preserves the intuitive meaning of it that was recognized by Boltzmann, Einstein, and Planck. Therefore we have not changed the basic rationale underlying the second law and the Kelvin efficiency rule in any way; we have only opened our eyes to their full meaning.

Far from being in conflict with the second law, Liouville's theorem is the reason for it. Had Liouville's theorem been discovered before the work of Carnot, it appears to us that the second law, in the full generality we have given it, might have been anticipated theoretically without any reference to heat engines; or indeed to the notions of temperature and thermal equilibrium.

We have made no use of the notions of order and disorder. Indeed, as Maxwell noted in the article on diffusion, those terms are only expressions of human aesthetic judgments. But in a well–known work on statistical mechanics (Penrose, 1970) it is stated that "$\cdots$ the letters of the alphabet can be arranged in 26! ways, of which only one is the perfectly ordered arrangement ABC ... XYZ, all the rest having varying degrees of disorder." To suppose that Nature is influenced by what you or I consider "orderly" is an egregious case of the Mind Projection Fallacy.

As a more pertinent example, Nature has decreed that water vapour has a higher entropy than liquid water, although most of us would consider the vapour far more "orderly" in both structure and behavior. The vapour has a higher entropy than the liquid, not because it is less "orderly", but because the microstates compatible with the vapour macrostate occupy a larger phase volume. Thus we cannot understand the second law, in either biology or physics, in terms of intuitive notions of order and disorder. On the other hand, the second law limitation on macroscopic processes is easily understood in objectively meaningful terms, in both biology and physics, as the price we pay for *reproducibility*.

CONCLUSION

As those promised tentative comments on biological information, we see the above as evidence that the energy of ATP hydrolysis is confined to a single vibration mode in striated muscle; if it spread to two modes, then we would have $r = 2 \times 0.062 = 0.124$, and (38) would predict a theoretical maximum efficiency of only 62%. Had the energy spread to ten vibration modes before being recaptured, the predicted efficiency would be only 8%. It appears that animals have indeed evolved muscle efficiency to the maximum that could be realized in a biochemical environment powered by the ATP hydrolysis reaction, although a reaction with a greater ΔH would permit still higher efficiency.

Finally, what was the effective upper temperature T_1 for the muscle? With two degrees of freedom per ATP molecule, this is given by $kT_1 = 0.43$ ev, or

$$T_1 = \frac{0.43 \times 1.6 \times 10^{-12}}{1.36 \times 10^{-16}} = 5060\ K\,. \tag{43}$$

This is startling because it is about the temperature at the surface of the sun! It appears, then, that a muscle is able to work efficiently not because it violates any laws of thermodynamics, but because it is powered by tiny "hot spots" of molecular size, as hot as the sun.†

This shows how far a biological system is from thermal equilibrium in the respects that matter. If one says that the temperature in a living cell is "uniform", he can mean only that it is uniform as registered by a thermometer whose bulb is thousands of times coarser, and whose response is thousands of times slower, than the units that are performing the essential biological functions.

If we examine the current literature of bioenergetics with this in mind, we are struck by the fact that virtually all treatments begin by stating that biological systems are at uniform temperature and the chemical reactions proceed isothermally; then virtually all the discussion is in terms of reaction free energies ΔF or ΔG. Now the free energy change of a reaction is only a fictitious kind of energy, that could in principle be observed in very special circumstances. It is the work made available when the temperature and concentrations are uniform and the reaction proceeds so slowly that it remains at equilibrium with respect to the original temperature and concentration; *i.e.*, when heat can flow in or out of the cell rapidly enough, and the reactants and products can diffuse in and out rapidly enough, to maintain the initial uniformity.

Conditions in a biological process such as nucleotide synthesis are about as far from this as can be imagined, in at least two respects:

(1) A cell may have very few (less than 20) molecules of a given type, and they are not free to diffuse about because of intracellular membranes; thus the uniform concentrations presupposed in the definition of reaction free energies seem not only not realized, but not even meaningful. Lehninger (1982) warns us that this might invalidate conventional thermodynamic treatments.

(2) A reaction is over – the job is done – in a time too short for the notion of "equilibrium" to be applicable. For many reactions the "real" physical energies ΔU, ΔH that have a meaning independently of thermal equilibrium, may be the ones most relevant for biology. Indeed, biochemical processes are now being studied fruitfully with picosecond time resolutions. On the scale of sizes and times that matter, a living cell is never in a state remotely like thermal equilibrium or uniform concentrations.

Recognizing this, we can understand another reason why biological thermodynamics has been puzzling in the past. Conventional free energy thermodynamics is doubtless adequate to describe slow, gross phenomena such as osmotic effects, but it may be irrelevant for biological functions like muscle contraction and protein synthesis, which necessarily, to avoid thermalisation from the surroundings, take place rapidly and on the molecular scale of size.

As our analysis shows, the small scale does not in itself preclude the application of thermodynamics, but attempts to do this could not have succeeded until the above points were recognized and we had a quite different, "fast" statement of the second law. Of course, muscle performance is only a special case of the general problem, but seeing how to apply the second law to muscle behaviour should give a useful clue for other cases.

In these first crude estimates to illustrate the principle, our reasoning was so general – concerning only phase volumes – that we did not need to invoke any particular details of the mechanism of muscle action. Therefore the analysis should apply as well to striated muscle, smooth muscle, flagella, or any other motile structures, and it would be of great interest to have experimental values of their efficiency; have they all managed to evolve down to a single vibration mode to transfer the energy?

However, the myosin bridge mechanism proposed by Sir A. F. Huxley (1957) for striated muscle and described by Squire (1981) and Alberts, *et al.* (1983) appears not only consistent with our

† This suggests some fascinating speculations; from the beginning all life on the earth has been running on the temperature difference between the sun and the earth. Presumably, the first life was powered directly from the sun, instead of using the indirect route of ATP. Then perhaps biological evolution chose ATP as its energy carrier just because its reaction energy duplicated the effects of the original source; and the memory of this has been retained ever since. If so, then a planet with a hotter sun would evolve life with a still higher muscle efficiency.

speculations; it fits in very nicely with them. The bending of that bridge is a degree of freedom that corresponds to a low–frequency vibration mode for which the classical equipartition law would hold, and the relative stiffness and massiveness of the myosin head makes it seem well adapted to resisting rapid thermalisation while transferring its energy into the macroscopic sliding of the actin fiber. We could hardly have asked for a better candidate for our one vibrational mode to receive the ATP hydrolysis energy.

Presumably, our argument could be refined by taking further information of this kind into account, although the observed facts suggest that the final conclusion cannot be very different; *i.e.*, most of that information will be irrelevant for predicting the net efficiency, although it is highly relevant for predicting other details such as force–velocity curves, fatigue, etc.

Having seen this biological mechanism, it is easy to believe that synthesized or extracted macromolecules could do similar things *in vitro*. Indeed, the first step in this direction has been taken already. In the fascinating "myosin motor" of Shimizu (1979) we have a molecular engine operating *in vitro*; not very efficiently, but nevertheless confirming the idea. In time the design of useful anti–Carnot molecular engines (artificial muscles) might become about as systematic and well understood as the design of dyes, drugs, and antibiotics is now.

REFERENCES

B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson (1983), *Molecular Biology of the Cell*, Garland Publishing Co., New York; pp. 550–609.

A. Aspect and P. Grangier (1985), "Tests of Bell's Inequalities", in *Symposium on the Foundations of Modern Physics; 50 years of the EPR Gedankenexperiment*, P. Lahti and P. Mittlestadt, Editors (World Scientific Publishing Co., Singapore).

A. Aspect (1986), "Tests of Bell's Inequalities with Pairs of Low Energy Correlated Photons", in Moore & Scully (1986).

L. E. Ballentine (1970), "The Statistical Interpretation of Quantum Mechanics", Rev. Mod. Phys. **42**, pp. 358–381.

J. S. Bell (1964), "On the EPR Paradox", Physics **1**, pp. 195-200,

J. S. Bell (1966), "On the Problem of Hidden Variables in Quantum Mechanics:, Rev. Mod. Phys. **38**, 447.

J. S. Bell (1987), *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press. Contains reprints of all of Bell's papers on EPR up to 1987.

N. Bohr (1935), "Can Quantum Mechanical Description of Reality be Considered Complete?", Phys. Rev. **48**, 696.

G. L. Bretthorst (1988), *Bayesian Spectrum Analysis and Parameter Estimation,* Springer Lecture Notes in Statistics, Vol. 48.

A. Einstein, B. Podolsky, and N. Rosen (1935), "Can Quantum Mechanical Description of Reality be Considered Complete?", Phys. Rev. **47**, 777.

A. F. Huxley (1957), Prog. Biophys. Chem. **7**, 255.

E. T. Jaynes (1965), "Gibbs vs Boltzmann Entropies", Am. J. Phys. **33**, 391–398. Reprinted in E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkrantz, Editor, D. Reidel Publishing Co., Dordrecht–Holland (1983).

E. T. Jaynes (1973), "Survey of the Present Status of Neoclassical Radiation Theory", in *Proceedings of the 1972 Rochester Conference on Optical Coherence*, L. Mandel & E. Wolf, editors, Pergamon Press, New York.

E. T. Jaynes (1985), "Generalized Scattering", in *Maximum Entropy and Bayesian Methods in Inverse Problems*, C. R. Smith & W. T. Grandy, Editors, D. Reidel Publishing Co., Dordrecht–Holland; pp. 377–398.

E. T. Jaynes (1986), "Predictive Statistical Mechanics", in Moore & Scully (1986); pp. 33–56.

H. Jeffreys (1931), *Scientific Inference*, Cambridge University Press; later editions 1957, 1973.

H. Jeffreys (1939), *Theory of Probability*, Oxford University Press; numerous later editions.

P. Knight (1987), "Single–atom Masers and the Quantum Nature of Light", Nature, **326**, 329.

A. L. Lehninger (1982), *Biochemistry, The Molecular Basis of Cell Structure and Function*, Worth Publishers, Inc., 444 Park Ave. South, New York; p. 383.

G. T. Moore & M. O. Scully, Editors (1986), *Frontiers of Nonequilbrium Statistical Physics*; Proceedings of the NATO Advanced Study Institute, Santa Fe, June 1984; Plenum Press, New York.

O. Penrose (1970), *Foundations of Statistical Mechanics*, Pergamon Press, Oxford; p. 160.

G. Rempe, H. Walther, N. Klein (1987); Phys. Rev. Let **58**, 353

S. Rozental, Editor (1964); *Niels Bohr, His Life and Work as seen by his Friends and Colleagues*, J. Wiley & Sons, Inc., New York.

H. Shimizu (1979), Adv. Biophys. **13**, 195–278.

H. Simon & N. Rescher (1966); "Cause and Counterfactual", Phil. Sci. **33**, 323–340.

C. P. Slichter, (1980), *Principles of Magnetic Resonance*, Springer, New York.

A. Zellner (1984), *Basic Issues in Econometrics*, Univ. Chicago Press; pp. 35–74.