Essay Review
Physics from Fisher Information

***D. A. Lavis**[*] **and R. F. Streater**[*]*

B. R. Frieden, *Physics From Fisher Information* (Cambridge: Cambridge University Press), xx + 240 pp., ISBN: 0-521-63167-X, £47.50.

B. R. Frieden uses a single procedure (called *extreme physical information*) with the aim of deriving 'most known physics, from statistical mechanics and thermodynamics to quantum mechanics, the Einstein field equations and quantum gravity'. His method, which is based on Fisher Information, is given a detailed exposition in this book, and we attempt to assess the extent to which he has succeeded in his task.

Keywords: Fisher Information; Extreme Physical Information; Variational Method; Four-Vector.

1. Introduction

Anyone interested in the foundations of statistical mechanics has always had a variety of different and conflicting approaches to choose from (Lavis 1977). Most of these take, explicitly or implicitly, a relative frequency interpretation of probability. A notable exception to this, based on Shannon's work on information theory (Shannon and Weaver 1964), was initiated by Brillouin (1956) and developed by Jaynes (1957). This approach (usually referred to as the *maximum entropy method*) has now been developed to encompass other aspects of statistics and probability theory (Jaynes 1983). It also provides an attractive way to teach statistical mechanics (see e.g. Hobson 1971, Turner and Betts 1974). There are some problems with this approach (Lavis and Milligan 1985), but they are mainly a matter of philosophical taste. An

[*]Department of Mathematics, King's College, Strand, London WC2R 2LS, U.K. (*e-mail*: David.Lavis@kcl.ac.uk and Ray.Streater@kcl.ac.uk)

implication of the underlying Bayesian view of probability is that the entropy of the system is dependent not just on the nature of the system *but also on the experimenter's (inevitably imperfect) knowledge of the system.* Not everyone is comfortable with this.

There are interesting similarities and differences between the work of Jaynes and Frieden, who begins his book with the quote:

> All things physical are information-theoretic in origin and this is a participatory universe $\cdots$. Observer participancy gives rise to information; and information gives rise to physics.

from J. A. Wheeler. His emphasis is, therefore, on *getting* rather than simply *having* information; that is to say on measurement, though whether he really believes that the physics would not be there without the measurement is difficult to say. It is, of course, a standard part of quantum mechanics that measurements cannot, in general, be made without affecting the system, but Frieden seems to be saying rather more than this. On page 2 we read that 'Physics is, after all, the science of measurement. That is, physics is a quantification of *observed* phenomena.' [his italics] and on page 3 he predicates his *extreme physical information* point of view on the proposition that 'all physical theory results from observation: in particular, *imperfect* observation' [his italics]. Frieden's programme is much more ambitious than Jaynes'. While Jaynes, within the area of the foundations of physics, confined himself to statistical mechanics, Frieden claims to be able to derive the fundamental equations of almost all of physics.

In trying to evaluate the success of this project it is reasonable, as in the case of Jaynes, to distinguish questions of philosophical predilection from those of method. We shall mainly concentrate on the latter and within this brief we try to answer two questions:

(1) Does the method of extreme physical information appear to be a promising approach to physics?

(2) Has Frieden provided a persuasive account of its use?

If we were convinced that the answer to (1) were clearly negative we should be wasting our time writing this review. However, it seems to us that there may be something interesting here. It is unfortunate that Frieden's attempt to describe the method is so seriously flawed both in logic and rigour. In particular his understanding of the meaning of scalars, vectors and tensors is very confused, and this is compounded by a rather sketchy notion of the meaning of a 'four-vector'.[1]

[1] Not helped by the rather inconsistent notation for vectors. There seems to be an

2. The Method of 'Extreme Physical Information'

Frieden's approach is based on a version of the variational method and we now give a brief summary of this procedure. Consider a system with variables $\boldsymbol{x} = (x_1, x_2, \dots, x_{M})$ in the space Υ and functions $\boldsymbol{q}(\boldsymbol{x}) = (q_1(\boldsymbol{x}), q_2(\boldsymbol{x}), \dots, q_{N}(\boldsymbol{x}))$. The *Lagrangian density* $\mathcal{L}(\boldsymbol{q}(\boldsymbol{x}))$ is some function of the set q_n and their partial derivatives $q_{nm} = \partial q_n/\partial x_m$ for $n = 1, 2, \dots, N$, $m = 1, 2, \dots, M$ and we define the Lagrangian functional

$$L[\boldsymbol{q}] = \int_{\mathsf{T}} \mathcal{L}(\boldsymbol{q}(\boldsymbol{x}))\,\mathrm{d}^{M}x, \tag{1}$$

where T is the hypercube $x_m^{(0)} \le x_m \le x_m^{(1)}$ in Υ.[2] One way of axiomatizing a physical system for which we have a Lagrangian density $\mathcal{L}(\boldsymbol{q}(\boldsymbol{x}))$ is to assert that the equations of motion are given by finding an extremum[3] for $L[\boldsymbol{q}]$ as the function forms of $q_n(\boldsymbol{x})$, $n = 1, 2, \dots, N$, are changed subject to their values being fixed over the boundaries of T. According to the calculus of variations such an extremum is given when the functions $q_n(\boldsymbol{x})$ satisfy the *Euler-Lagrange equations*

$$\sum_{m=1}^{M} \frac{\partial}{\partial x_m}\left(\frac{\partial \mathcal{L}}{\partial q_{nm}}\right) - \frac{\partial \mathcal{L}}{\partial q_n} = 0. \tag{2}$$

Frieden observes that in most physical systems the Lagrangian density is of the form

$$\mathcal{L}(\boldsymbol{q}(\boldsymbol{x})) = \tfrac{1}{2}\sum_{n=1}^{N} c_n[\boldsymbol{\nabla} q_n(\boldsymbol{x})]^2 + \mathcal{L}_1(\boldsymbol{q}(\boldsymbol{x})), \tag{3}$$

which, from (2), then gives the equations of motion

$$c_n \nabla^2 q_n = \frac{\partial \mathcal{L}_1}{\partial q_n}, \qquad n = 1, 2, \dots, N. \tag{4}$$

Frieden is certainly correct in pointing out (p. 24) that the problem with this approach to physics is in determining the appropriate Lagrangian density.

Most undergraduate courses in dynamics introduce variational methods, if at all, at the end with the form of the Lagrangian density provided by

---

implicit intention to use boldface roman symbols for vectors of dimension $M \ge 4$ and bold italic symbols for three-dimensional vectors, but this convention is not used consistently.

[2] We shall denote the volume element $\mathrm{d}x_1 \times \mathrm{d}x_2 \times \cdots \times \mathrm{d}x_M$ by $\mathrm{d}^M x$. It is to be regretted that Frieden denotes this quantity by $\mathrm{d}\boldsymbol{x}$. The possibility of confusion with the infinitesimal vector element in Υ is obvious, particularly when, as on page 305, he uses $\mathrm{d}\boldsymbol{r}$ as a three-dimensional volume element and $\mathrm{d}\boldsymbol{\sigma}$ as a three-dimensional vector in the same equation.

[3] Normally a minimum.

post hoc reasoning. One of the avowed aims of Frieden's book is to 'present a *systematic approach to deriving* Lagrangians' [his italics] (p. 24). His starting point is to construe the Lagrangian functional as a quantity $K[\boldsymbol{q}]$, which is 'called the 'physical information' of the system' (p. 71) and which has the form

$$K[\boldsymbol{q}] = I[\boldsymbol{q}] - J[\boldsymbol{q}], \tag{5}$$

where $I[\boldsymbol{q}]$ is the *Fisher information*. For one variable x distributed with a probability density function $p(x)$

$$I[p] = \int \frac{[p'(x)]^2}{p(x)} \mathrm{d}x. \tag{6}$$

Since the probability density function $p(x)$ is nonnegative $q(x) = \sqrt{p(x)}$ is a real function and (6) can be re-expressed in the form

$$I[q] = 4 \int [q'(x)]^2 \mathrm{d}x. \tag{7}$$

To within a constant, which can obviously be included, this is of the form required for the square-gradient term in the Lagrangian density and the generalization

$$I[\boldsymbol{q}] = \sum_{n=1}^{N} \int \mathrm{i}_n(\boldsymbol{x})\, \mathrm{d}^M x, \quad \text{where} \quad \mathrm{i}_n(\boldsymbol{x}) = 4N \left|\boldsymbol{\nabla} q_n(\boldsymbol{x})\right|^2, \tag{8}$$

is not something to cause a problem.[4] The difficulty is, of course, to have a plausible argument for the remaining term of the Lagrangian density, or equivalently of the total information, represented respectively in (3) by $\mathcal{L}_1$ and in (5) by J. One line of approach, in the spirit of Jaynes, or of many problems in control theory, would be to try to incorporate this term as an integral constraint. This Frieden rejects as '*ad hoc*' (p. 69). He claims to have a '*natural* way' [his italics] (p. 69), of obtaining J and the bulk of the book is devoted to an attempt to substantiate this claim.

As indicated above Frieden's approach is based on measurement. He interprets $I[\boldsymbol{q}]$ as the amount of information contained in the data collected during the measurement process and $J[\boldsymbol{q}]$ as the *bound information*, defined as the amount of information in the phenomenon. It is therefore reasonable, at first sight, to go along with his argument (p. 71) that

$$J[\boldsymbol{q}] \geq I[\boldsymbol{q}], \tag{9}$$

[4]Although, in the interests of clarity, it may have been better to write $\mathrm{i}(q_n(\boldsymbol{x}))$ rather than $\mathrm{i}_n(\boldsymbol{x})$, the same being the case in (15), below. The factor of N is missing in the first appearance of these formulae on page 72, but has made an unexplained appearance by page 91. In the interests of consistency we include it from the beginning.

since a measuring process cannot extract more information from the phenomenon than is present. The problem with this is that the two quantities are *functionals* not numbers. They are susceptible to numerical evaluation only when the forms of the function $q_n(\boldsymbol{x})$, $n = 1, 2, \ldots, N$ are known. Unless, of course, we were to understand (9) to be true *for all* functions.

Frieden next postulates his '*axiom* 1' (p. 70) that 'perturbed amounts of information' satisfy the formula

$$\delta J[\boldsymbol{q}] = \delta I[\boldsymbol{q}]. \tag{10}$$

Why this should be true, rather than, for example, an inequality like (9), is by no means obvious. But at least, if we interpret, as Frieden does, this formula as a variational principle for $K[\boldsymbol{q}]$ with respect to variation of the functions $q_n(\boldsymbol{x})$, then we have the Euler-Lagrange equations

$$\sum_{m=1}^{M} \frac{\partial}{\partial x_m}\left(\frac{\partial \mathcal{K}}{\partial q_{nm}}\right) - \frac{\partial \mathcal{K}}{\partial q_n} = 0 \tag{11}$$

to solve for the functions $q_n(\boldsymbol{x})$, where

$$K[\boldsymbol{q}] = \int \mathcal{K}(\boldsymbol{q}(\boldsymbol{x}))\, \mathrm{d}^M x. \tag{12}$$

For ease of discussion we denote the functions which satisfy (11) as $q_n^{(\mathrm{E})}$.

The usefulness of axiom 1 depends of having an independent argument for obtaining $\mathcal{J}(\boldsymbol{q}(\boldsymbol{x}))$ and then the functional

$$J[\boldsymbol{q}] = \int \mathcal{J}(\boldsymbol{q}(\boldsymbol{x}))\, \mathrm{d}^M x. \tag{13}$$

Then, of course, the inequality (9), with the functions $q_n(\boldsymbol{x})$ set to the forms $q_n^{(\mathrm{E})}(\boldsymbol{x})$, would be a theorem which needed to be proved. In his attempt to resolve this situation Frieden now makes an assumption (p. 71) which is both crucial to the development and also either completely equivalent to (9) or totally unjustified. He supposes that

$$I[\boldsymbol{q}] - \kappa J[\boldsymbol{q}] = 0, \tag{14}$$

where $\kappa \leq 1$ is a *constant.* This statement can, of course, be interpreted in at least two different ways. If it were to be taken to be true

(i) just for the functions $q_n^{(\mathrm{E})}(\boldsymbol{x})$, derived from (11), then it would be trivially equivalent to (9) (also taken with $q_n(\boldsymbol{x})$ given by $q_n^{(\mathrm{E})}(\boldsymbol{x})$) and in need of proof.

(ii) for all functional forms $q_n(\boldsymbol{x})$, then it doesn't follow from (9) since κ should also then be a functional of $\boldsymbol{q}$ with a value, greater than or equal to unity for all functions $q_n(\boldsymbol{x})$.

In any event (14) is of only limited use.[5] The serious work in most of the book is done by assuming 'a zero-condition on the microscopic level'. This is achieved by two more axioms. *Axiom* 2 asserts that[6]

$$\mathcal{J}(\boldsymbol{q}(\boldsymbol{x})) = \sum_{n=1}^{N} \mathrm{j}_n(\boldsymbol{x}), \tag{15}$$

and *axiom* 3 asserts that

$$\mathrm{i}_n(\boldsymbol{x}) - \kappa\, \mathrm{j}_n(\boldsymbol{x}) = 0, \qquad \text{for all } \boldsymbol{x} \text{ and } n. \tag{16}$$

It is a little difficult to know how to interpret this equation. It is certainly true that it implies (14), in either the sense (i) or (ii), but the converse is far from being the case. If we were to understand (16) as a set of equations to be used to find the appropriate functional forms for $q_n(\boldsymbol{x})$ with a given predetermined form for $\mathrm{j}_n(\boldsymbol{x})$, then, as was pointed out by Kibble (2000), there is no reason to suppose that the resulting forms will agree with those derived from the Euler-Lagrange equations (11).

However, this is not the way that Frieden approaches the problem. He has *no independent method* of defining the functions $\mathrm{j}_n(\boldsymbol{x})$. So he sets out to derive them by some combination of the use of (11) and (14) or (16). He has two different approaches to doing this, which he designates as (a) and (b) (pp. 75–76). These will be discussed with particular examples in Section 4..

3. Invariance Principles

Since Frieden places great emphasis on the use of invariance principles we shall briefly discuss this aspect of the method.

3.1. Using Four-Vectors

On pages 84–89, in a section entitled rather grandly 'Derivation of Lorentz group of transformations', Frieden imposes the condition that the Fisher

[5] In the derivation of the Schrödinger and Klein-Gordon equations.

[6] A similar decomposition of the integrand $\mathcal{I}(\boldsymbol{q}(\boldsymbol{x}))$ of the Fisher information follows from (8).

information given by (8) must be invariant in form for any allowed linear transformation $\boldsymbol{x}' = \boldsymbol{B}\boldsymbol{x}$ in Υ. Since

$$\boldsymbol{\nabla} q_n(\boldsymbol{x}) \;=\; \boldsymbol{B}\boldsymbol{\nabla} q_n'(\boldsymbol{x}'), \qquad \mathrm{d}^3r'\mathrm{d}t' = \det\{\boldsymbol{B}\}\mathrm{d}^3r\mathrm{d}t, \quad \text{with} \quad \mathrm{d}^3r = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z, \tag{17}$$

it follows that the condition is satisfied if $\boldsymbol{B}$ is unitary with $\det\{\boldsymbol{B}\} = 1$. At this point he observes that 'the most well-known rotation matrix solution is the Lorentz transformation' with $\boldsymbol{x} = (\mathrm{i}ct, x, y, z)$

$$\boldsymbol{B} = \begin{pmatrix} \gamma & -(\gamma u/c) & 0 & 0 \\ -(\gamma u/c) & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{18}$$

where $\gamma = 1/\sqrt{1 - u^2/c^2}$. On this basis he makes the wholly unwarranted inference that $\boldsymbol{x}$ must always be a four-vector, that is a vector which transforms according to the Lorentz transformation. On page 89 we read that 'An information density that is not covariant could not have implied the Lorentz transformation'. How are we to understand this statement? If 'covariant' is taken just to mean that $\boldsymbol{x}$ transforms according to a unitary transformation then this does not imply that $\boldsymbol{B}$ is of the form (18), even if we give no physical meaning to u/c. There are unitary transformations with $\det\{\boldsymbol{B}\} = 1$ for all values of M. If on the other hand 'covariant' means that $\boldsymbol{x}$ transforms according to the Lorentz transform then the statement reads like a badly expressed tautology. In any event the whole discussion hardly supports the title of the section. It is difficult to avoid the suspicion that part of his motive is to smuggle into the analysis c as the speed of light. It also gives him an excuse for using vectors with imaginary components, a practice rightly condemned by Kibble (2000).

3.2. Using Fourier Transforms

In Section 3.8 Frieden uses a discussion of an optical measurement device to show (in his opinion) that

(i) *A unitary transformation of coordinates* naturally arises during the course of measurement of many (perhaps all) phenomena.[his italics]

(ii) The method of extreme physical information, as defined in (3.16) [the variational principle, equivalent to our (10)] and (3.18) [our equation (14)] *is implied by* the unitary transformation and the perturbing effect of the measurement. [his italics]

Frieden uses the well-known relationship between Fourier transformation of a light-wave and Fraunhofer diffraction (see e.g. Longhurst 1957). He considers a plane light wave of wavelength λ shining on a particle moving on a line with coordinate x and subject to a potential $V(x)$. The intensity of the light after interacting with the particle is $u(x)$ and, having passed through a single slit, the Fraunhofer diffraction pattern on a screen at distance R is $U(x')$ with the relationship

$$U(x') = \int u(x) \exp\left(\frac{2\pi \mathrm{i} x\, x'}{\lambda R}\right) \mathrm{d}x. \tag{19}$$

This arrangement is taken to be a process of measuring the location of the particle and $u(x)$ is understood as being related to the quantum mechanical probability amplitude ψ in object space with $U(x')$ being related to the quantum mechanical probability amplitude ϕ in measurement space. He now performs the change of variables $\mu = hx'/\lambda R$, with $\psi(x) = u(x)$ $\phi(\mu)\sqrt{h} = U(x')$, so that (19) becomes

$$\phi(\mu) = \frac{1}{\sqrt{2\pi\hbar}} \int \psi(x) \exp\left(\frac{\mathrm{i} x\, \mu}{\hbar}\right) \mathrm{d}x, \tag{20}$$

where $\hbar = h/2\pi$.

This is an outline of Frieden's argument in support of (i), with the unitary transformation being Fourier and $\psi(x)$ the complex-valued amplitude used in defining the Fisher information. The reason for introducing the parameter h is at first sight a little puzzling, particularly as he tells us that 'h is what we call 'Planck's parameter'. It is later found to be a constant; see Sec. 4.1.14.' If h is used in a change of variables it surely *must* be a constant. In seeking further enlightenment we refer to Section 4.1.14, which consists of two and a half lines. It claims to show that $\hbar$ is a 'universal constant' and this is partly because (the extremum value of) 'J is regarded as a universal constant'. We are directed for enlightenment on this point to Section 3.4.14 where we discover this is because J 'can only be a function of whatever free parameters the functional K contains (e.g. h, k, c)... The absolute nature of the scenario suggests, then, that we regard the extreme value of J as a *universal physical constant*.'[7] He then goes on to argue that if J is a function only of c then c is a universal constant because J is a universal constant. This seems to be where we came in. He refers to this argument as 'an unusual 'bootstrap' effect'.

It seems to us that it is not only unusual but invalid. It is difficult to escape the conclusion that the introduction of h, like the earlier introduction

[7] A reference is given, but it consists only of the acknowledgement of discussions with the person referenced, so it provides no supporting evidence.

of c, is a way of covertly including a physical parameter. This impression is reinforced when we read that the new variable μ introduced in the same change of variables 'turns out to be the particle momentum'.[8] On the basis of this discussion the next subsection claims to confirm equation (3.18) [our equation (14)] using unitarity. With $N = 2$ and $\psi(x) = q_1(x) + \mathrm{i}q_2(x)$, (8) gives[9]

$$I[\psi] = 8 \int \left| \frac{\mathrm{d}\psi(x)}{\mathrm{d}x} \right|^2 \mathrm{d}x. \tag{21}$$

The argument is that since measurement is a unitary (Fourier) transformation 'the Fisher information I, expressed equivalently in momentum[10] space, represents the 'bound' information for the particle.' This might lead us to suppose that, using (20),

$$J[\phi] = 8 \int \left| \frac{\mathrm{d}\phi(\mu)}{\mathrm{d}\mu} \right|^2 \mathrm{d}\mu = \frac{8}{\hbar^2} \int x^2 \left|\psi(x)\right|^2 \mathrm{d}x. \tag{22}$$

But this would not lead to (14). Instead Frieden simply substitutes the inverse of (20) into (21) to give

$$J[\phi] = \frac{8}{\hbar^2} \int \mu^2 \left|\phi(\mu)\right|^2 \mathrm{d}\mu. \tag{23}$$

Then, of course, it is trivially obvious that (14) follows with $\kappa = 1$. He now proceeds to use the same trick to 'confirm' the variational formula (10). He defines $J[\phi]$ by (23) and takes the variation $\delta J[\phi]$. He then inserts the Fourier transform into the variation $\delta I[\psi]$ with $I[\psi]$ given by (21). That this leads to $\delta I[\psi] = \delta J[\phi]$ is hardly a surprise.

4. Particular Examples

In this section we consider some of Frieden's accounts of the application of the extreme physical information method to different physical situations, dividing them according to whether he uses his approach (a) or (b). Since (a) is used in the later parts of the book we shall discuss it after considering (b).

[8] To find out how this happens in Frieden's scheme we are referred forward to pages 115–117. This argument is considered below.

[9] For no very obvious reason Frieden drops the factor of 8. Since, however, it reappears at other places in the book we shall leave it in.

[10] He makes free use of μ being the momentum in spite of the fact that this is only 'established' in the next chapter.

4.1. Approach (b)

This method is used in Appendix D for Schrödinger's equation and in Chapter 4 for the Klein-Gordon equation. The 'macroscopic level condition' (14) is first used to *define* $J[\boldsymbol{q}]$ and the required functional forms for $q_n(\boldsymbol{x})$ are then given by the Euler-Lagrange equations. This approach is extended in the latter part of Chapter 4, where the functional used for the Klein-Gordon equation is reused at the microscopic level to obtain the Dirac equation. This allows Frieden to claim that in some circumstances (16) and (11) produce distinct and equally valid sets of equations.

Naively one might suppose that it is straightforward to use (14) to define $J[\boldsymbol{q}]$; we have simply $J[\boldsymbol{q}] = I[\boldsymbol{q}]/\kappa$. However, from (5), $K[\boldsymbol{q}] = (\kappa - 1)I[\boldsymbol{q}]/\kappa$ and, since in all cases of the use of this approach Frieden takes $\kappa = 1$, we then have $K[\boldsymbol{q}] = 0$, for all choices of $\boldsymbol{q}$. This is a conclusion that the author rejects as 'mere tautology' (p. 106) and he has a most ingenious method of circumventing it. This consists in what he conceives to be the use of 'the invariance, or symmetry principle governing each phenomenon' (p. 107). In practice this means using Fourier transformation and other manipulations to ensure that when $J[\boldsymbol{q}]$, obtained initially from (14) (with $\kappa = 1$), is substituted into (5) it has been transmuted so that it is no longer identical to $I[\boldsymbol{q}]$. The way this is done in particular cases will now be discussed.

In Appendix D Frieden sets out to derive Schrödinger's equation for a particle of mass m moving on a straight line. The position of the particle x is a random variable and it is subject to a conservative potential $V(x)$, with the total energy W conserved. With $N = 2$ and the complex wave function $\psi(x) = q_1(x) + \mathrm{i}\, q_2(x)$ we have the Fisher information (21).[11] Using (14) with $\kappa = 1$ he now uses the inverse of the Fourier transform (20) to express the bound information in the form (23). The aim now is to use the variational principle (10) to derive Schrödinger's equation. To do this both $I[\psi]$ and $J[\phi]$ must be functionals of the same function. So ϕ must be transformed to ψ or vice-versa. This would lead, of course, to the inconvenient result that I and J are *identically equal*. To avoid these difficulties Frieden now proceeds to transform the bound information by the following steps. From (23),

$$\begin{aligned} J[\phi] = \frac{4}{\hbar^2}\langle \mu^2 \rangle &= \frac{8m}{\hbar^2}\langle E_{\mathrm{kin}} \rangle = \frac{8m}{\hbar^2}\langle \{W - V(x)\} \rangle \\ &= \frac{8m}{\hbar^2}\int \{W - V(x)\}|\psi(x)|^2 \mathrm{d}x = J[\psi]. \end{aligned} \tag{24}$$

Using the final form for $J[\psi]$ Schrödinger's equation follows from the variational formula. The problem with this argument is that he is implicitly

[11] He leaves the identification of $N = 2$ to the final lines of the appendix, but this does not affect the logic of his argument.

assuming what he is setting out to prove. That $\phi(\mu)$ given by (20) is the momentum-space wave-function is either a consequence of Schrödinger's equation or a supposition used to derive it. In any event $\langle\mu^2\rangle$ is the expected value of the squared momentum leading to expected kinetic energy only if $\psi(x)$ in (20) giving $\phi(\mu)$ is *already* a solution of Schrödinger's equation. So these steps cannot be used in the general functional form.

In Chapter 4 a similar procedure is used to first derive the Klein-Gordon equation and then the Dirac equation. Here $M = 4$ with $\boldsymbol{x} = (\mathrm{i}x, \mathrm{i}y, \mathrm{i}z, ct)$ and N is even with $\psi_n(\boldsymbol{r}, t) = q_n(\boldsymbol{r}, t) + \mathrm{i}q_{n+N/2}(\boldsymbol{r}, t)$, $n = 1, 2, \ldots, N/2$.[12] A crucial element of the derivation is the proposed form of normalisation of the Klein-Gordon wave-function in space-time. For general N, this is given by his equations (4.10) and (4.11) and is equivalent to

$$c\sum_{n=1}^{N/2}\int|\psi_n(\boldsymbol{r}, t)|^2\mathrm{d}^3r\mathrm{d}t = 1. \tag{25}$$

Even when $N = 2$, which is the case he discusses on page 121, this condition is inconsistent with the Klein-Gordon equation. This is most easily seen in momentum-energy space. Equation (25), with $N = 2$, implies that ψ_1, as a function of the four variables $(\boldsymbol{r}, t)$, is square-integrable and, from Parseval's theorem, this is also the case for its Fourier transform ϕ_1, as a function of momentum $\boldsymbol{\mu}$ and energy E. But any such ϕ_1 must (as a distribution) satisfy the momentum-energy space version

$$(E^2 - \boldsymbol{\mu}^2 - m^2)\phi_1(\boldsymbol{\mu}, E) = 0, \tag{26}$$

of the Klein-Gordon equation. So $\phi_1(\boldsymbol{\mu}, E) = 0$ unless $E^2 - \boldsymbol{\mu}^2 = m^2$. There are no square-integrable functions with this property, other than the zero function.

The author's motivation for adopting (25) is contained in his assertion that the conventional normalisation condition is inconsistent. This condition he takes to be (p. 121)

$$\int|\psi_1(\boldsymbol{r}, t)|^2\mathrm{d}^3r = 1. \tag{27}$$

This is simply wrong. The correct normalization is the Wigner form (Segal 1965)

$$\int\left\{\psi_1(\boldsymbol{r}, 0)\omega\psi_1(\boldsymbol{r}, 0) + \dot{\psi}_1(\boldsymbol{r}, 0)\omega^{-1}\dot{\psi}_1(\boldsymbol{r}, 0)\right\}\mathrm{d}^3r = 1, \tag{28}$$

[12] Setting $N = 2$ is adequate for the Klein-Gordon equation but he uses $N = 4$ and $N = 8$ for the Dirac equation.

where ω is the pseudo-differential operator $\sqrt{m^2 - \nabla^2}$.

However, as we proceed through his analysis, other problems also arise. Starting with the Fisher information

$$I[\boldsymbol{\psi}] = 4Nc \sum_{n=1}^{N/2} \int \left\{ \frac{1}{c^2} |\dot{\psi}_n(\boldsymbol{r}, t)|^2 - |\boldsymbol{\nabla}\psi_n(\boldsymbol{r}, t)|^2 \right\} \mathrm{d}^3 r \mathrm{d}t, \tag{29}$$

the same argument is used as for the Schrödinger equation to obtain

$$J[\boldsymbol{\phi}] = \frac{4N}{\hbar^2} \left\langle \frac{E^2}{c^2} - \mu^2 \right\rangle, \tag{30}$$

where under the Fourier transformation $\boldsymbol{r}$ is conjugate to $\boldsymbol{\mu}/\hbar$ and ct to $E/c\hbar$. As in the case of his derivation of Schrödinger's equation we must bear in mind that the right-hand side is still a *functional* of the, as yet, arbitrary set $\phi_n(\boldsymbol{r}, t)$, $n = 1, 2, \ldots, n/2$. We have already discussed the rather curious argument which Frieden uses to come to the conclusion that J is a 'universal constant'. *Even if this were true* it could only be the case for (30) with the ϕ_n satisfying the Klein-Gordon equation. So the argument that since

> the two factors in (4.13) [our equation (30)] are independent, each must be a constant. Then parameter $\hbar$ must be a universal constant (p. 116)

could at most be valid only in this circumstance. Obviously $\langle \mu^2 \rangle$ and $\langle E^2 \rangle$ are constant *for particular wave-functions* since all the variables have been integrated out. This has nothing to do with them being independent and J will also depend on the wave-function chosen: so it cannot be argued from this that $\hbar$ is a universal constant.

Frieden now examines the expectation value term on the right of (30). In spite of the fact that $J[\boldsymbol{\phi}]$ is still a functional he argues that, since the fluctuations of $\boldsymbol{\mu}$ and E change with boundary conditions and $J[\boldsymbol{\phi}]$ is, according to him, a universal constant, then the combination of variables $E^2/c^2 - \mu^2$ must be constant. This has, for him, two fortunate consequences:

(i) By a little dimensional analysis he persuades us that the constant in question is mc enabling him to obtain the standard energy-momentum formula

$$E^2 = c^2\mu^2 + m^2c^4. \tag{31}$$

This in turn allows the identification of E with energy and $\boldsymbol{\mu}$ with momentum.

(ii) By substitution from (31) into (30) and using the (incorrect) normalization condition (25) he obtains

$$I[\boldsymbol{\psi}] = 4N\left(\frac{mc}{\hbar}\right)^2 = J[\boldsymbol{\phi}]. \tag{32}$$

It may seem that discovering that what you thought was a functional of the set ϕ_n has in fact the same constant value for all choices of ϕ_n would be a problem. Not, however, if you are prepared to recycle the normalization condition (25) to give J as a functional of ψ_n in the form

$$J[\boldsymbol{\psi}] = \frac{4Nm^2c^3}{\hbar^2}\sum_{n=1}^{N/2}\int |\psi_n(\boldsymbol{r},t)|^2 \mathrm{d}^3r\mathrm{d}t. \tag{33}$$

It might be argued that $I[\boldsymbol{\psi}]$ is exactly the same, which would lead to difficulties. However, if we choose to stick with the form (29) the variational principle (10) yields the Klein-Gordon equation

$$c^2\hbar^2\nabla^2\psi_n - \hbar^2\ddot{\psi}_n - m^2c^4\psi_n = 0, \tag{34}$$

for each of the $N/2$ components.

In the latter part of Chapter 4 Frieden takes the expressions (29) and (33) for $I[\boldsymbol{\psi}]$ and $J[\boldsymbol{\psi}]$ and uses the microscopic level condition (16), with $\kappa = 1$ to extract the Dirac equation.

4.2. Approach (a)

This method is used in Chapters 5–8, for electromagnetism, the Einstein field equations, classical statistical mechanics and $1/f$ noise, respectively; and in Chapter 9 to draw conclusions concerning the distribution of the magnitudes of the physical constants. It consists in inserting some general functional forms for $\mathrm{j}_n(\boldsymbol{x})$ into (11) and the microscopic level condition (16) and then solving to make these functions and the value of κ explicit. The main problem with this approach is the doubtful pedigree of (16) and some rather serious mathematical errors. In the interests of brevity we shall confine our attention to Chapters 7, 8 and 9.

In Section 7.3 Frieden presents a derivation of the Boltzmann distribution. In this case the only variable of the system is the energy E. He decides at the outset that he prefers the Fisher information to have a negative sign and thinks that he can do this by making the substitution $x = \mathrm{i}E$ into (8), with $M = 1$.[13] It seems to us that his substitution makes the Fisher information

[13] He then uses x for this new variable arguing at the end that it is in fact equivalent to E. We shall use E from the outset.

imaginary not negative, but in the interests of the discussion we go along with his formula which is[14]

$$I[\boldsymbol{q}] = -4\sum_{n=1}^{N}\int \{q_n'(E)\}^2 \mathrm{d}E. \tag{35}$$

He then chooses the general form

$$J[\boldsymbol{q}] = 4\sum_{n=1}^{N}\int J_n(q_n(E))\mathrm{d}E, \tag{36}$$

for the bound information and the Euler-Lagrange equations (11) give

$$q_n''(E) = \frac{1}{2}\frac{\mathrm{d}J_n}{\mathrm{d}q_n}, \qquad n = 1, 2, \ldots, N. \tag{37}$$

Frieden now applies integration by parts to re-express (35) in the form

$$I[\boldsymbol{q}] = 4\sum_{n=1}^{N}\int q_n(E)q_n''(E)\mathrm{d}E. \tag{38}$$

There is a problem with this. At the beginning of the book (p. 5) he tells us that the 'limits of integrals [which he omits] are fixed and, usually, infinite'. But, of course, in this case the energy is 'bounded below but of unlimited size above' (p. 183). This means that if E is in the range $[E_0, \infty)$ he has implicitly assumed that

$$\lim_{E\to\infty} q_n(E)q_n'(E) = 0, \qquad q_n(E_0)q_n'(E_0) = 0. \tag{39}$$

The latter is inconsistent with the Boltzmann distribution. Now, from (14), (36) and (39),

$$I[\boldsymbol{q}] - \kappa J[\boldsymbol{q}] = 4\sum_{n=1}^{N}\int \{q_n(E)q_n''(E) - \kappa J_n(q_n(E))\}\,\mathrm{d}E. \tag{40}$$

As indicated above this method now involves extracting the microscopic level conditions

$$q_n(E)q_n''(E) = \kappa J_n(q_n(E)), \qquad n = 1, 2, \ldots, N, \tag{41}$$

from (40). It is now straightforward to show that[15]

[14] The elusive factor of N has again disappeared.

[15] The method used by Frieden to establish this result leads him to the conclusion that the constants A_n are positive, but this is not necessarily the case since substitution of $J_n(q_n)$ from (42) into each of (37) and (41) leads to (43) for any value of A_n.

$$J_n(q_n) = A_n q_n^{2\kappa}, \qquad (42) \qquad\qquad q_n''(E) = \kappa A_n \{q_n(E)\}^{2\kappa-1}. \qquad (43)$$

Frieden now has the problem of assigning a value to κ. He argues (p. 183) that $\kappa = 1$ because this would lead, via the relationship $q_n^2(E) = p_n(E)$ between the $q_n(E)$ and the probability $p_n(E)$ to $J[\boldsymbol{q}]$ being

> a normalization integral, this represents particularly weak prior information in the sense that any PDF obeys normalization. A phenomenon that only obeys normalization is said to exhibit 'maximum ignorance' in its independent variable.

Imposing the condition that $p_n(E)$ is normalized over $[E_0, \infty)$ with the $p_n(E) \to 0$, as $E \to \infty$, now gives

$$p_n(E) = \beta_n \exp\{-\beta_n(E - E_0)\}, \tag{44}$$

where $\beta_n = 2\sqrt{\kappa A_n}$. Apart from the problem of establishing that the parameters A_n are positive, the difficulty with this argument is the reasoning used to set $\kappa = 1$. Maximum ignorance (otherwise called *the principle of indifference* (Keynes 1949)) leads to a uniform distribution, with respect to some measure, over the probability space, not the Boltzmann distribution. In standard accounts of statistical mechanics (Lavis and Bell 1999) the Boltzmann, or canonical, distribution is that for a system *known* to be in thermal contact with a heat reservoir at temperature T. In the approach of Jaynes (1957), the known information is the expectation value $\langle E \rangle$ of the energy and this is used, together with normalization, to give Lagrangian constraints on the maximum entropy procedure leading to (44). Frieden introduces knowledge of $\langle E \rangle$ only at the point where he needs to establish that $\beta_n = 1/kT$. This he does by appealing to the equipartition theorem.[16] A more elegant account of this is given by Jaynes (1957).

The Maxwell velocity distribution[17] is, of course, a consequence of the Boltzmann distribution when the energy consists of the sum of a kinetic energy, quadratic in the velocities, and a potential energy which is a function only of the configuration variables. In spite of this Frieden presents an independent derivation of the velocity distribution, taking the fact that the results agree as 'a verification of the overall theory' (p. 192). The analysis is very similar to that described above except that: (i) He chooses $M = 3$ with the three variables variously called x_1, x_2, x_3 or x, y, z (p. 189) being the (real) components of the momentum. (ii) For no obvious reason he now

[16] True only when the Hamiltonian is a quadratic function of its variables (Huang 1963), although Frieden chooses to assume, more strongly, that the system is a perfect gas.

[17] Referred to by Frieden as the *Maxwell-Boltzmann* velocity law.

takes the functions J_n to be explicit functions of the vector $\boldsymbol{x} = (x_1, x_2, x_3)$. (iii) He carries over the condition $\kappa = 1$ from his derivation of the Boltzmann distribution. The result of these changes leads to the replacement of (43) by $\nabla^2 q_n(\boldsymbol{x}) = -f_n(\boldsymbol{x}) q_n(\boldsymbol{x})$. The arbitrary functions $f_n(\boldsymbol{x})$, which are a consequence of (ii) are now assumed to be of the form $A_n + B_n\boldsymbol{x}^2$ to order to give solutions expressed in terms of Hermite polynomials. These he argues 'represent stationary solutions *en route* to the Gaussian equilibrium solution' (p. 191), drawing our attention to the fact that such forms have been found as solutions that follow from the Boltzmann transport equation. He finds it 'rather remarkable' (p. 191) that these results can be obtained using his method. Given the sequence of rather arbitrary assumptions he has made to achieve his ends, we do not find it particularly remarkable.

The latter part of Chapter 7 is concerned with the time-dependent behaviour of the *Shannon entropy* for a 'system of one or more particles moving randomly within an enclosure' (p. 194). He takes this quantity to be

$$H(t) = -\int p(\boldsymbol{r}|t) \ln\{p(\boldsymbol{r}|t)\} \mathrm{d}^3 r, \tag{45}$$

where $p(\boldsymbol{r}|t)$ is the 'probability density for finding a particle at position $\boldsymbol{r} = (x, y, z)$ within the enclosure at the known time t' (p. 194). However, this quantity $H(t)$ is *not* the entropy of the assembly of particles. Let the number of particles be $\mathcal{N}$. Then *at equilibrium* the entropy is given (Lavis 1977), to within multiplicative and additive constants, by

$$S = -\int \rho(\boldsymbol{x}, \boldsymbol{\nu}) \ln\{\rho(\boldsymbol{x}, \boldsymbol{\nu})\} \mathrm{d}^M x \mathrm{d}^M \nu, \tag{46}$$

where $\boldsymbol{x}$ is the $M(= 3\mathcal{N})$-dimensional vector of the positions of all the particles and $\boldsymbol{\nu}$ is the corresponding M-dimensional momentum vector. The formula (46) reduces to (45) (at equilibrium) only if the phase space probability density function is uniform in momenta and satisfies a molecular chaos condition in position.

It is, of course, possible to determine the time evolution of entropy-like quantities only when we are given the equation which determines the evolution of the density function. The Boltzmann H-theorem $\mathrm{d}H/\mathrm{d}t \geq 0$, quoted by Frieden on page 194, applies when the distribution function for the momentum, or velocity (*not position as in (45)*) satisfies the Boltzmann transport equation (Huang 1963). If (46) is generalized to non-equilibrium situations then the probability density function $\rho(\boldsymbol{x}, \boldsymbol{\nu}, t)$ satisfies Liouville's equation

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla}_x.(\rho \boldsymbol{X}) + \boldsymbol{\nabla}_\nu.(\rho \boldsymbol{Y}) = 0, \tag{47}$$

where $\dot{\boldsymbol{x}} = \boldsymbol{X}$, $\dot{\boldsymbol{\nu}} = \boldsymbol{Y}$ are the equations governing the flow in phase space. From this it is not difficult to show that when the flow is measure-preserving, that is $\boldsymbol{\nabla}_x.\boldsymbol{X} + \boldsymbol{\nabla}_\nu.\boldsymbol{Y} = 0$, then the form of entropy given by (46) is time-invariant.

This result provides one of the main problems for non-equilibrium statistical mechanics and a number of different proposals have been made for a generalized form of entropy (see e.g. Lavis 1977, Lavis and Milligan 1985, Dougherty 1993, 1994). It doesn't seem to us that Frieden makes any significant contribution to this work. His formula (45), based as it is simply on a probability density $p(\boldsymbol{x}|t)$ which is a function just of the position of one particle is neither the Boltzmann nor the Gibbs form of entropy. Using the equation (7.69) which he gives for 'conservation of flow' it is not possible to say anything about the sign of $\mathrm{d}H/\mathrm{d}t$ and the upper bound which he achieved is valid, at most for $|\mathrm{d}H/\mathrm{d}t|$. In any event, the final simplification presented on page 203 is invalidated by an error in the vector manipulation. The 3×3 matrix $\boldsymbol{\psi}\boldsymbol{\psi}^\dagger$, formed from the three-dimensional column vector $\boldsymbol{\psi}$ and its hermitian conjugate, is replaced by the scalar $\boldsymbol{\psi}^\dagger\boldsymbol{\psi} = |\boldsymbol{\psi}|^2$. This mistake unfortunately invalidates the 'remarkably simple result' contained in equation (7.114).

In Chapter 8 Frieden sets out to show that the power spectrum $S(\omega)$ of an *intrinsic random function*[18] $X(t)$ is of the form $S(\omega) = A/\omega$. After some discussion[19] he concludes that the Fisher information is

$$I[S] = 4\int_0^\Omega \frac{1}{S(\omega)}\mathrm{d}\omega. \tag{48}$$

He then, following the procedure for his approach (a), supposes the bound information is of the form

$$J[S] = 4\int_0^\Omega F(S(\omega),\omega)\mathrm{d}\omega. \tag{49}$$

The variational equation (11) and the microscopic level equation(16) then give

$$\frac{1}{S^2} = -\frac{\partial F}{\partial S}, \qquad (50) \qquad\qquad \frac{1}{S} = \kappa F. \qquad (51)$$

It is obvious by substituting from (51) into (50) that the only solution to this pair of equations is $\kappa = 1$, $F(S,\omega) = 1/S$; and there is no mechanism for determining the form of $S(\omega)$.

[18] $X(t)$ is an intrinsic random function if it is not stationary, but, for all $\tau \geq 0$, $X(t+\tau) - X(t)$ is stationary.

[19] The details of this become irrelevant in the light of subsequent errors.

Frieden, however, thinks he has a method. Substituting for one of the factors S^{-1} in (50) from (51) he obtains

$$\frac{\kappa F}{S} = -\frac{\partial F}{\partial S}, \tag{52}$$

which has the general solution $F(S,\omega) = G(\omega)S^{-\kappa}$. He then uses an argument based on scale invariance to show that $G(\omega) = B\omega^{1-\kappa}$ and substituting into (51) he obtains the final solution

$$F(S,\omega) = A^{\kappa-1}\omega^{1-\kappa}S^{-\kappa}, \qquad (53) \qquad S(\omega) = A\omega^{-1}. \qquad (54)$$

He admits as a 'minor point' (p. 214) that these steps do not lead to a solution for $S(\omega)$ when $\kappa = 1$. What he fails to notice, however, is that (53) satisfies (50) *only when* $\kappa = 1$. The formula (54) is a consequence of the assumed scaling form

$$S(\omega)\mathrm{d}\omega = S(\omega/\alpha)\mathrm{d}\omega/\alpha, \qquad \text{for all } \alpha > 0, \tag{55}$$

for the spectrum and has nothing to do with the application of the method of extreme physical information.

The idea that the universal physical constants are distributed according to some probability law is interesting and, in Chapter 9, Frieden sets out to show that the law is of the form $p(x) = A/x$. To do this he invokes the extreme physical information approach (a) with $\boldsymbol{x} = (\mathrm{i}x, 0, 0, 0)$. He seems to believe that this step means that his coordinates are 'relativistic invariants i.e. four-vectors' (p. 220). He also supposes that this allows him to have the Fisher information with a negative sign. As we have already mentioned in connection with a similar step in Chapter 7, it seems to us that the effect is to make this quantity imaginary. In any event the steps in the implementation of the extreme physical information method are largely irrelevant since his probability law, as in Chapter 8, follows directly from the assumed scaling form for $p(x)$. Of course, the inverse probability law cannot be normalized over all positive x and Frieden decides to limit the range of x to the interval $[1/b, b]$ for some $b > 0$. On normalization this gives

$$p(x) = \frac{1}{2x\ln(b)}, \qquad \text{for } 1/b \leq x \leq b. \tag{56}$$

This choice of limits impose some obvious symmetry properties on the distribution, one of which is that it is evenly weighted about $x = 1$. Frieden claims that 'Why the value 1 should have this significance is a mystery' (p. 226). We find it difficult to share his awe at a result which he has built into his distribution. It may be that there is some physical truth in the application of the distribution (56) to the physical constants. It does not seem to us, however, that the method of extreme physical information makes a contribution to understanding this either in terms of its derivation or interpretation.

5. Conclusions

The application of the ideas of information theory to physics is interesting; and the use of Fisher information to provide the gradient terms in the Lagrangian for a variational procedure is of some importance. The crucial step, however, is to provide in some rational and widely-applicable manner the remaining terms of the Lagrangian. Frieden believes he is able to do this by using the idea of bound information. We have shown, however, that there are errors in his procedure, both at the level of the derivation of the defining equations (14) and (16) and also in particular applications. In the case of the latter we have not discussed all the systems contained in the book. We thought it was more useful to deal in greater depth with some particular cases; similar difficulties can be uncovered in the other chapters.

We regret to say that we find this book to be fundamentally flawed in both its overall concept and mathematical detail. It cannot be read as a textbook providing a valid approach to physics. It could, perhaps, however, be a source of stimulation for some new and interesting work.

References

Brillouin L. (1956) *Science and Information Theory* (New York: Academic Press).

Dougherty J. P. (1993) *Explaining statistical mechanics* Stud. Hist. Phil. Sci. **24**, 843–866.

Dougherty J. P. (1994) *Foundations of non-equilibrium statistical mechanics* Phil. Trans. Roy. Soc. A **346**, 259–305.

Hobson A. (1971) *Concepts in Statistical Mechanics* (New York: Gordon and Breach).

Huang K. (1963) *Statistical Mechanics* (New York: John Wiley).

Jaynes E. T. (1957) *Information theory in statistical mechanics I* Phys. Rev. **106** 620–630; *Information theory in statistical mechanics II* Phys. Rev. **108** 171–190.

Jaynes E. T. (1983) *Papers on Probability, Statistics and Statistical Physics* (Dordrecht: Reidel).

Keynes J. M. (1949) *A Treatise on Probability*, (London: Macmillan).

Kibble T. W. B. (2000) Review of Frieden's book in *Contemporary Physics*, to appear.

Lavis D. A. (1977) *The role of statistical mechanics in classical physics* Brit. J. Phil. Sci. **28** 255–279.

Lavis D. A. and Bell G. M. (1999) *Statistical Mechanics of Lattice Systems 1* (Berlin: Springer).

Lavis D. A. and Milligan P. J. (1985) *Review article: The work of E. T. Jaynes on probability, statistics and statistical physics* Brit. J. Phil. Sci. **36** 193–210.

Longhurst R. S. (1957) *Geometrical and Physical Optics* (London: Longmans).

Segal I. E. (1965) *Mathematical Problems of Relativistic Physics* (New York: A.M.S.).

Shannon C. E. and Weaver W. (1964) *The Mathematical Theory of Communication*, (Urbana: University of Illinois Press).

Turner R. E. and Betts D. S. (1974) *Introductory Statistical Mechanics* (London: Sussex University Press).