An Unexpected Union—
Physics and Fisher Information

Physics from Fisher Information: A Unification. *By B. Roy Frieden, Cambridge University Press, New York, 1998, 328 pages, $74.95.*

First printed in 1998, and twice reprinted in 1999, the book under review is certainly being talked about. One suspects that the ideas it contains—if not the book itself—will continue to be discussed for generations to come. The exposition begins with a quote from John Archibald Wheeler to the effect that "observer participancy gives rise to information, and information gives rise to physics."

The idea that observers are participants in the events they observe has been around at least since Heisenberg, but Frieden has exploited it more systemati-

BOOK REVIEW

By James Case

cally than any predecessor. To do so, he formulates a single unifying principle—that of extreme physical information (EPI)— from which he derives Maxwell's equations, the Einstein field equations, the Dirac and Klein–Gordon equations of quantum mechanics, various laws of statistical physics, and even a few previously undiscovered laws governing nearly incompressible turbulent fluid flows. Although EPI is obviously a unification, Frieden doubts that it will ever become the long sought "theory of everything."

Fisher information is named for its inventor, R.A. Fisher (1890–1962), a British biostatistician who was among the first to develop and employ—when and as the need arose during his work in genetics and eugenics—such methods as maximum likelihood estimation, the analysis of variance, and the design of experiments. He also pointed out that Gregor Mendel had probably falsified the "data" in his famous pea-plant experiments, which seem too clean to be the result of any natural process.

Ronald A. Fisher, 1929, as sketched by B. Roy Frieden, author of the book under review. The sketch, which appears in Physics from Fisher Information: A Unification, *was done from a photograph taken at the time of Fisher's election as a Fellow of the Royal Society.*

Fisher's idea was that attempts to measure physical quantities—such as the time required for the winner of a 100-yard dash to reach the finish line—are invariably frustrated by "noise." That's why multiple stopwatches are ordinarily employed. Moreover, the quantity of information concerning the actual elapsed time contained in such a sample varies directly with the degree to which the sample measurements cluster about a common value. Thus, if y_n denotes the figure (time) recorded on device (stopwatch) $n \in \{1, ..., N\}$ and if $y_n = \theta + x_n$, where θ is the "true" elapsed time and $x_1, \ldots, x_n$ represent random errors, then the quantity I of information concerning θ contained in the vector $\mathbf{y} = (y_1, \ldots, y_N)$ of observations should be inversely related to a (scalar) measure of the dispersion of the elements of $\mathbf{y}$ about θ.

Accordingly, if $p(\mathbf{y}|\theta)$ is the N-dimensional probability density function describing the deviation of $\mathbf{y}$ from $\theta\mathbf{1}$ and $\theta^0(\mathbf{y})$ denotes an unbiased estimator of θ, the expected error

$$E\left[\theta^0(\mathbf{y}) - \theta\right] =$$
$$\int\left[\theta^0(\mathbf{y}) - \theta\right] p(\mathbf{y}|\theta)\, d\mathbf{y} \tag{1}$$

vanishes almost surely. Meanwhile, the mean square error

$$e^2 = \int\left[\theta^0(\mathbf{y}) - \theta\right]^2 p(\mathbf{y}|\theta)\, d\mathbf{y} \tag{2}$$

remains positive a.s. Fisher information $I = I(\theta)$ is defined to be

$$I = \int\left[\partial \ln p(\mathbf{y}|\theta) / \partial\theta\right]^2 p(\mathbf{y}|\theta)\, d\mathbf{y}. \tag{3}$$

I satisfies the consequence of the Cauchy–Schwartz inequality known as the Cramer–Rao inequality: $Ie^2 \geq 1$. If $\mathbf{N} = 1$ and $p(y|\theta) = p(y - \theta)$, (3) reduces to

$$\int \left\{ \left[\partial p(y - \theta) / \partial (y - \theta) \right]^2 / p(y - \theta) \right\} dy =$$

$$\int \left\{ \left[p'(x) \right]^2 / p(x) \right\} dx = I. \tag{4}$$

Moreover, the substitution $p(x) = [q(x)]^2$ yields the even simpler form

$$I = 4 \int \left[q'(x) \right]^2 dx. \tag{5}$$

The introduction of "real probability amplitudes" $q(x)$ in place of probability density functions $p(x)$ simplifies many calculations in modern physics. Moreover, since the integrability of $p(x)$ is equivalent to the square integrability of $q(x)$, the substitution allows the analysis to take place in Hilbert space, a change of venue that von Neumann welcomed.

In higher dimensions, the quantity $[q'(x)]^2$ becomes $\nabla q \cdot \nabla q$ or, after the introduction of complex probability amplitudes, $\nabla \psi \cdot \nabla \psi^*$, where * denotes complex conjugation. Frieden presents an entire table of Lagrangian functions, gleaned from branches of physics as diverse as classical and fluid mechanics, electro- and thermodynamics, quantum theory, and general relativity, in which such forms appear. This suggests that an approach via Fisher information could furnish at least a partial answer to a long unanswered question: Where do Lagrangians come from? Frieden quotes a standard text[†] to the effect that Lagrangians are often plucked out of thin air, simply because the associated Euler–Lagrange equations are the expected equations of motion. There is, as yet, no general explanation for the remaining terms present in the Lagrangians of Frieden's table. The success of the EPI technique requires an ability to account for the additional terms that occur in particular branches of physics.

If an experiment is designed to estimate several different parameter vectors θ_n from measurements $\mathbf{y_n} = \theta_n + \mathbf{x_n}$, the appropriate generalizations of (5) are

$$I = 4 \int \sum_n \nabla q_n \cdot \nabla q_n \, d\mathbf{x}$$

$$= 4 \int \sum_n \nabla \psi_n \cdot \nabla \psi_n^* \, d\mathbf{x}, \tag{6}$$

with either real or complex probability amplitudes. As long as N (the largest n) is even, complex amplitudes can be formed from real ones by writing $i = (-1)^{1/2}$ and $\psi_n = q_{2n-1} + iq_{2n}$, $n = 1, \ldots, N/2$. Alternatively, the q_n's can be recovered from the ψ_n's by separating the real and imaginary parts. The choice between real and complex amplitudes is strictly a matter of convenience. It is also con-venient to define $\mathbf{q(x)}$ as $(q_1(\mathbf{x}), \ldots, q_n(\mathbf{x}))$, and $\psi(\mathbf{x})$ as $(\psi_1,(\mathbf{x}), \ldots, \psi_{N/2}(\mathbf{x}))$.

It follows from (4) that if the random deviations x follow a normal or Bernoulli distribution about θ, then I is inversely proportional to their variance. If x is logistically distributed, so that $p(x) = 1/(1 + \exp(-ax))$, then $I = a/2$. Or, if the deviations are 2-vectors drawn from a bivariate normal distribution in which the components share a common mean θ, a common variance σ^2, and a covariance ρ, then I becomes $2/\sigma^2 (1 + \rho)$. The information content of two variables is thus a decreasing function of their degree of correlation—yet another reason why experimenters should and typically do make the extra effort to observe independent random variables whenever possible.

The form of the foregoing result leads Frieden to speculate that every experiment may constitute a (zero-sum) game against nature, in which the experimenter strives to maximize the Fisher information obtained, while a hostile nature (information demon) strives to minimize it by introducing unexpected correlations and the like. It seems an interesting—if as yet undeveloped—idea.

Frieden compares Fisher information with the Shannon, Boltzmann, and Kullback–Liebler definitions of entropy, which likewise represent attempts to identify useful scalar measures of information. The comparison is most easily made with Shannon entropy $H = \int [\ln p(x)] p(x) dx$, and the discrete approximation

$$\delta H \equiv -\Delta x^{-1} \sum_n p(x_n) \ln p(x_n), \tag{7}$$

which converges to H as $\Delta x \to 0$. δH differs from the discrete form

$$\delta I \equiv -\Delta x^{-1} \sum_n \left[p(x_{n+1}) - p(x_n) \right]^2 / p(x_n) \tag{8}$$

of (4) in that the probabilities $p(\cdot)$ can be reassigned among the carriers $x_1, \ldots, x_n$ without altering the value of (7). Not so (8). It matters for Fisher information I—but not for Shannon information H—whether the likeliest carriers are clustered in space. It is this feature that uniquely qualifies I to illuminate Lagrangian physics.

I, like H, can be viewed as a measure of physical disorder. When H coincides with traditional Boltzmann entropy H_B, as it often does, the second law of thermodynamics stipulates that $dH(t)/dt$ is never negative. And in a variety of other circumstances—among them the situation in which $p(x,t)$ obeys a Fokker–Planck equation of the form $p_t = [D_1(x,t)p]_x + [D_2(x,t)p]_{xx}$—it turns out that $dI(t)/dt$ is never positive. It is also possible to define temperature in terms of Fisher information, and to deduce a version of the perfect

_____

[†] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part I,* McGraw-Hill, New York, 1953.

gas law incorporating that precise definition of temperature, which suggests that it may eventually prove possible to recover all of thermodynamics from the notion of Fisher information. That, of course, remains to be done.

According to (4) and (5), Fisher information I is a real-valued functional defined on the space of PDFs $p(\cdot)$, or probability amplitudes $q(\cdot)$. In order to derive the laws of physics from Fisher information, one must find a second such functional J that, when subtracted from I, furnishes the missing terms in the Lagrangians of Frieden's table. In what follows, I will always have the form (6), while J will have a form appropriate to the particular branch of physics under consideration. Moreover, I will always represent the information in the data obtained via experiment, while J will represent the information content of the physical situation being explored. It makes sense, therefore, to suppose that I can never exceed J, to define an efficiency $\kappa = I/J$, and to observe that κ has always turned out to be either 1 or 1/2 in every situation (branch of physics) that Frieden and his co-workers have explored to date.

An experiment, however sensitive, must always perturb the probability amplitudes $\mathbf{q}(\mathbf{x})$ of the system under observation by an amount $\delta\mathbf{q}$. The consequent perturbations δI and δJ represent quantities of information transferred from nature (the system) to the experimenter. Frieden proposes

$$\text{axiom 1: } \delta I = \delta J.$$

This is a conservation law concerning the transfers of information possible in certain more or less idealized experiments. Its validity, according to Frieden, is demonstrated many times over in his book and is the key to his EPI approach to the understanding of physical reality. The axiom can actually be confirmed whenever the "measurement space" has a conjugate space connected to it by a unitary transformation. The author augments the foregoing conceptual axiom with two additional axioms of a more technical nature.

To consider the effects on the functional K of a small perturbation $\delta\mathbf{q}(\mathbf{x})$ of its argument $\mathbf{q}(\mathbf{x})$, we write $K = I - J$. Then, by axiom 1, we can write

$$\delta K = \delta(I - J) = \delta I - \delta J = 0, \tag{9}$$

and ask for which probability amplitudes $q(x)$ equation (9) holds. The answer is naturally expressed in terms of Euler–Lagrange equations from the calculus of variations, a process best understood by seeing it in action.

Consider a (small) particle of mass m moving on a line about its rest position θ. If the particle is subject to a conservative scalar potential field $V(x)$, its total energy W is conserved. In terms of complex wave functions $\psi_n(x)$, the expression (6) for I becomes

$$I = 4N\sum_n \int |d\psi_n(x)/dx|^2 \, dx, \tag{10}$$

the summation extending from $n = 1$ to $n = N/2$. The particle in question will have not only a position x relative to its rest position θ, but a momentum μ relative to its rest momentum 0. Moreover, in the same way the observations $x_n = y_n - \theta$ are drawn from probability distributions $p_n(x) = [q_n(x)]^2$ corresponding to complex amplitudes $\psi_n(x)$, observed values μ_n would be drawn from distributions $\pi_n(\mu)$ corresponding to complex amplitudes $\phi_n(\mu)$. Again, we write $\phi = (\phi_1, \ldots, \phi_{N/2})$ and $\psi = \psi_1, \ldots, \psi_{N/2}$. Frieden concludes on physical grounds—by analysis of an experimental apparatus that could in theory be used to perform the required measurements—that $J[\phi] = I[\psi]$. Moreover, the ϕ_n and ψ_n are Fourier transforms of one another, so that

$$\psi_n(x) = \left(1/\sqrt{2\pi\hbar}\right)$$
$$\int \phi_n(\mu)\exp(-i\mu x/\hbar)d\mu. \tag{11}$$

When the resulting expressions for $\psi_n'(x)$ are substituted in (10),

$$I = \left(4N/\hbar^2\right)\int \mu^2 \sum_n |\phi_n(\mu)|^2 = J. \tag{12}$$

He then identifies the summation in (12) with the probability distribution $\pi(\mu)$ mentioned earlier, so that

$$J = \left(4N/\hbar^2\right)E\left[\mu^2\right]. \tag{13}$$

Finally, he makes the specifically nonrelativistic approximation that the kinetic energy $\mathcal{E}$ of the particle is $\mu^2/2m$, so that

$$J = \left(8Nm/\hbar^2\right)E[\mathcal{E}]$$
$$= \left(8Nm/\hbar^2\right)E[W - V(x)]$$
$$= \left(8Nm/\hbar^2\right)\int [W - V(x)]p(x)\, dx \tag{14}$$
$$= \left(8Nm/\hbar^2\right)\int [W - V(x)]\sum_n |\psi_n(x)|^2 \, dx.$$

Since *I* and *J* have now been expressed in terms of the quantities $\psi_n(x)$, $K = I - J$ can be combined into a single integral of the form $K = \int \mathcal{L}(\psi_n', \psi_n, x)\, dx$, and the associated Euler–Lagrange equations written:

$$\psi_n''(x) + \left(2m/\hbar^2\right)\left[W - V(x)\right]\psi_n(x) = 0,$$
$$n = 1,\ldots,N/2, \tag{15}$$

the time-independent Schrödinger equation. This is the (one-dimensional, nonrelativistic) approximation of the (covariant, relativistic) Klein–Gordon equation, which can be derived—without approximation—in similar fashion. The unitary nature of the Fourier transformation played an essential role in the foregoing derivation, particularly that of equation (12), where it justified the conclusion that $\kappa = I/J = 1$ in the present theory. That means that quantum mechanics is "efficient" in the sense that the underlying experiments are capable of extracting all available information. Many other physical theories, including electromagnetic theory and gravitational theory, yield only $\kappa = 1/2$.

Frieden derives a version of Heis-enberg's uncertainty principle from the Cramer–Rao inequality, and contrasts the result with the standard version obtained via Fourier duality. He remarks that the EPI version is stronger than the standard one in the sense that it implies the other, whereas the converse is untrue. The standard result applies only to uncertainties that exist before any measurements are taken, while the EPI version also applies to uncertainties that remain afterward.

The material described here is contained in the book's first four chapters and in Appendix D. The next six chapters present further applications of the EPI method. Statistical mechanics, like quantum mechanics, is found to be efficient in the sense that $\kappa = I/J = 1$, while most other theories are only halfway efficient. Frieden finds this unsurprising—quantized versions of most physical theories, including gravitation and electromagnetism, have not been developed, although he expects that they will be found in due course. Indeed, he and others are currently applying the EPI method to a study of quantum gravity—in steps analogous to those that led to the Schrödinger equation presented here—and, in higher dimensions, to the Klein–Gordon equation. Feynman path integrals seem to emerge naturally in this ongoing investigation, which seems destined to culminate in a vector wave equation of gravity.

Also under attack are various problems in turbulence. Probability laws of fluid density and velocity have been found for low-velocity flow in a nearly incompressible fluid. The new laws agree well with published data arising from detailed simulations of such turbulence. Frieden also remarks that the EPI method applies most naturally to the laws of physics that are expressed in the form of (differential) field equations. He sees no reason, however, why the method could not be extended to cover laws concerning the sources of the fields in question. The exposition ends as it began, quoting John Archibald Wheeler's words on observer participation. It seems safe to conclude, all in all, that the unexpected union between physics and Fisher information will prove both lasting and fruitful.

James Case writes from Baltimore, Maryland.