Lagrangians of physics and the game of Fisher-information transfer

B. Roy Frieden
Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

Bernard H. Soffer
Hughes Research Laboratories, 3011 Malibu Canyon Road, Malibu, California 90265
(Received 10 April 1995)

The Lagrangians of physics arise out of a mathematical game between a "smart" measurer and nature (personified by a demon). Each contestant wants to maximize his level of Fisher information I. The game is zero sum, by conservation of information in the closed system. The payoff of the game introduces a variational principle—extreme physical information (EPI)—which fixes both the Lagrangian and the physical constant of each scenario. The EPI approach provides an understanding of the relationship between measurement and physical law. EPI also defines a prescription for constructing Lagrangians. The prior knowledge required for this purpose is a rule of symmetry or conservation that implies a unitary transformation for which I remains invariant. As an example, when applied to the smart measurement of the space-time coordinate of a particle, the symmetry used is that between position-time space and momentum-energy space. Then the unitary transformation is the Fourier one, and EPI derives the following: the equivalence of energy, momentum, and mass; the constancy of Planck's parameter h; and the Lagrangian that implies both the Klein-Gordon equation and the Dirac equation of quantum mechanics.

PACS number(s): 05.40.+j, 03.65.Bz, 12.10.Kt, 89.70.+c

INTRODUCTION

The Lagrangian approach [1] to physics has been utilized now for over 200 years. It is one of the most potent and convenient tools of theory ever invented. However, an enigma of physics is the question of where its Lagrangians come from. It would be nice to justify and derive them from a prior principle, but none seems to exist. Indeed, when a Lagrangian is presented in the literature, it is often with a disclaimer, such as [2] "It usually happens that the differential equations for a given phenomenon are known first, and only later is the Lagrange function found, from which the differential equations can be obtained." Even in a case where the differential equations are *not* known, often candidate Lagrangians are first constructed, to see if "reasonable" differential equations result.

Hence the Lagrange function has been principally a contrivance for getting the correct answer. It is the means to an end—a differential equation—but with no significance in its own right. One of the aims of this article is to show, in fact, that Lagrangians do have prior significance. A second aim is to present a systematic approach to deriving Lagrangians. A third is to clarify the role of the observer in a measurement. These aims will be achieved through use of the concept of Fisher information.

R. A. Fisher (1890–1962) was a researcher whose work is not well known to physicists. He is renowned in the fields of genetics, statistics, and eugenics. Among his pivotal contributions to these fields [3] are the maximum likelihood estimate, the analysis of variance, and a measure of indeterminacy now called "Fisher information." (He also deduced that the famous geneticist G. Mendel had fabricated his famous experimental results with pea plants. They were too regular to be true, statistically.) It will become apparent that his form of information has great utility in physics as well.

Table I shows a list of Lagrangians [2], emphasizing the common presence of a squared-gradient term. In quantum mechanics, this term represents mean kinetic energy, but why mean kinetic energy should be present remains a mystery (Schrödinger called it "incomprehensible" [4]). Moreover, in other fields of physics the term no longer has this meaning. What we will show is that, in general, the squared gradient represents a phenomenon that is natural to all fields, i.e., *information*.

THE ERROR IN A SMART MEASUREMENT, AND DISORDER

The "smart" measurement

Consider the basic problem of estimating a parameter of value θ. The estimate follows from an imperfect observation $y=\theta+x$ of θ, in the presence of random noise x. See Fig. 1. For brevity, this measurement-estimation procedure will be called a "smart measurement" of θ. It results in an estimate $\hat{\theta}$ which is a function $\hat{\theta}(y)$.

The system comprising quantities y, θ, and x is a *closed* one. No other input effects (such as additional noise sources) are assumed present. It will become apparent that the closed nature of the measurement system implies an isolated *physical* system as well.

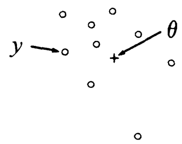

FIG. 1. A smart measurement. Based upon observation y, form an estimate of parameter θ.

Fisher information

This information arises as a measure of the expected error in a smart measurement. Consider the class of "unbiased" estimates, obeying $\langle \hat{\theta}(y) \rangle = \theta$; these are correct "on average." Then, as shown in Appendix A, the mean-square error e^2 in the estimate $\hat{\theta}$ obeys a relation [5]

$$e^2 I \geq 1 \ , \tag{1}$$

where I is called the Fisher information

$$I = \int dx\, p'^2(x)/p(x) \ . \tag{2}$$

The prime denotes a derivative d/dx and the integration limits are infinite. The quantity $p(x)$ denotes the probability density function for x. Equation (1) is called the Cramer-Rao inequality. It expresses reciprocity between the error e and the Fisher information I. The quantity I is thereby a quality metric of the estimation procedure. Since quality increases (e decreases) as I increases, I is called an "information."

A related quantity to I is the Shannon entropy [6] (called Shannon "information" in this paper). Historically, I predates the Shannon form by about 25 years (1922 vs 1948). There are some known relations connecting the two information concepts [7–9] but these are not germane to our purposes.

The analytic properties of the two information measures are quite different. Thus, whereas Shannon's is a *global* measure [of smoothness in $p(x)$], Fisher's is a *local* measure. Hence, when extermized through variation of $p(x)$, Fisher's form gives a differential equation while Shannon's always gives directly the same form of solution, an exponential function. See Appendix A.

Despite these differences, it is shown in Appendix B that I is approximated by a cross-entropy quantity. This implies that I and the *Boltzmann entropy* are, to an extent, related quantities.

TABLE I. Lagrangians for various physical phenomena. Where do these come from, and in particular, why do they all contain a squared gradient term? (WE indicates the wave equation.)

Phenomenon	Lagrangian
Classical Mech.	$\frac{1}{2} m \left[\frac{\partial q}{\partial t} \right]^2 - V$
Flexible string or compressible fluid	$\frac{1}{2} \rho \left[\left[\frac{\partial q}{\partial t} \right]^2 - c^2 \nabla q \cdot \nabla q \right]$
Diffusion eq.	$-\nabla \psi \cdot \nabla \psi^* - \cdots$
Schrödinger WE	$-\frac{\hbar^2}{2m} \nabla \psi \cdot \nabla \psi^* - \cdots$
Klein-Gordon Eq.	$-\frac{\hbar^2}{2m} \nabla \psi \cdot \nabla \psi^* - \cdots$
Elastic WE	$\frac{1}{2} \rho \dot{q}^2 - \cdots$
Electromagnetic eqs.	$4 \sum_{n=1}^{4} \Box q_n \cdot \Box q_n - \cdots$
Dirac eqs.	$-\frac{\hbar^2}{2m} \nabla \psi \cdot \nabla \psi^* - \cdots = 0$
General relativity (eqs. of motion)	$\sum_{m,n=1}^{4} g_{mn}(q(\tau)) \frac{\partial q_m}{\partial \tau} \frac{\partial q_n}{\partial \tau}$, g the metric tensor
Boltzmann law	$4 \left[\frac{\partial q(E)}{\partial E} \right]^2 - \cdots$, $p(E) \equiv q^2(E)$
Maxwell-Boltzmann law	$4 \left[\frac{\partial q(v)}{\partial v} \right]^2 - \cdots$, $p(v) \equiv q^2(v)$
Lorentz transformation (special relativity)	$\partial_i q_n \partial_i q_n$ (invariance of integral)
Helmholtz WE	$-\nabla \psi \cdot \nabla \psi^* - \cdots$

Relation to Heisenberg uncertainty principle

In form, Eq. (1) resembles the Heisenberg principle. This is no mere coincidence. In application of Eq. (1) to estimating the classical position θ of a particle [10], I becomes $4\langle\mu^2\rangle/\hbar^2$, where μ denotes momentum and $\hbar$ is the Planck constant/2π. [Equation (22) derived below is a four-dimensional generalization of this I.] Substitution into (1) directly gives the Heisenberg principle.

Actually, a stronger, "smart" version of the usual principle so results. Our quantity e is the error in any function $\hat{\theta}(y)$ of the data value. This includes the possibility of an optimum function, representing a best estimate of θ. By contrast, in the usual statement of Heisenberg e represents the error in the *direct* observable y, i.e., without the processing step $\hat{\theta}(y)$. Thus, our result is that the spread in any function of the observable position y obeys reciprocity with the spread in momentum. (Analogously, reciprocity follows between any function of momentum and the coordinate y if momentum, instead, is measured.) See also related work [11].

Amplitude form of I

Equation (2) defines the Fisher information I in terms of the probability law $p(x)$. Fisher found that it is often more convenient to work instead with a real "amplitude" function $q(x)$, where

$$p(x)=q^2(x)\ . \tag{3}$$

(Parenthetically, it is intriguing that Fisher used probability amplitudes [12] independently of their use in quantum mechanics. The purpose was to discriminate among population classes.) Using form (3) in (2) immediately gives

$$I=4\int dx\, q'^2(x)\ , \tag{4}$$

of a simpler form than (2) and showing that I simply measures the gradient content in $q(x)$ [and hence in $p(x)$]. This is the origin of the squared gradients in the Table I of Lagrangians, as will become apparent.

I as an entropy

I is actually a measure of the degree of disorder of a system. Strong disorder means a lack of predictability of values of x over its range, i.e., a uniform or "unbiased" probability density function $p(x)$. Such a curve is shown in Fig. 2(b). The curve has small gradient content (if it is physically meaningful, i.e., is piecewise continuous). Then by (4) the Fisher information I is small. Conversely, if a curve $p(x)$ shows bias to particular x values then it exhibits low disorder. See Fig. 2(a). Analytically, the curve will be steeply sloped about these x values, and so the value of I becomes (now) large. The net effect is that I measures the degree of disorder of the system. In other words, I is a form of entropy.

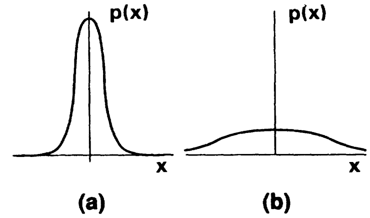

FIG. 2. (a) High gradients, therefore high I. Narrow effective range in x, therefore small disorder. (b) Low gradients, therefore low I. Wide effective range in x, therefore strong disorder. The upshot is that I varies inversely with disorder.

A MEASUREMENT-INDUCED DEFINITION OF PHYSICS?

We found above that a smart measurement of particle position gives rise to one law of physics—the Heisenberg principle. Is there, in fact, a general reaction by nature to a smart measurement? Does each law of physics follow from a particular smart measurement? Related speculations have been made by Wigner [13] and by Wheeler [14]. Physics has been aptly called the "science of measurement" [15]. We show in the following sections that physics *arises* out of measurement, literally.

Fisher information in a vector measurement

A generalization [Eq. (13) below] of the Fisher I of Eq. (2) has many physically useful attributes [16]: additivity to data, independence of the unknown θ, invariance to general coordinate shift or rotation, coordinate covariance, and gauge invariance. The latter means that it contains space and time derivatives so that the usual substitutions

$$\nabla\to\nabla-ie\,\mathbf{A}/c\hbar\ ,\quad \partial/\partial t\to\partial/\partial t+ie\phi/\hbar\ ,\quad i=\sqrt{-1} \tag{5}$$

can be made. $\mathbf{A}$ and ϕ are the usual electromagnetic potentials and e is the particle charge.

The generalized information arises when a vector $\boldsymbol{\theta}$ is to be estimated on the basis of a vector of measurements $\boldsymbol{\rho}=\boldsymbol{\theta}+\mathbf{r}$. Now the information (4) takes the more general form

$$I=4\int d\mathbf{r}\,\nabla q\cdot\nabla q\ . \tag{6}$$

Also, the probability $p(x)$ of Eq. (3) becomes

$$p(\mathbf{r})=q^2(\mathbf{r})\ . \tag{7}$$

The amplitude $q(\mathbf{r})$ may also be expressed in terms of a basis set of "mode functions" $q_n(\mathbf{r})$,

$$q(\mathbf{r})\equiv\sum_{n=1}^{N}q_n(\mathbf{r})\ . \tag{8}$$

Any specific choice for the mode functions will depend upon the specific physical scenario. An ultimate aim of this paper is to develop a general procedure for forming the Lagrangian appropriate to these modes.

Paradigm of the broken urn

We next show that, for a particular isolated system, I tends to be minimal with time [17]. Consider a scenario where many particles fill a small urn. Imagine these to be ideal, point masses that collide elastically and that are not in an exterior force field. We want a smart measurement of their mean horizontal position θ. Accordingly, a particle at horizontal position y is observed, $y=\theta+x$, where x is a random fluctuation from θ. Define the mean-square error $e^2(t)=\langle[\theta-\hat{\theta}(y)]^2\rangle$ due to repeatedly forming estimates $\hat{\theta}(y)$ of θ within a small time interval $(t,t+dt)$. How should e vary with t?

Initially, at $t=0$, the particles are within the small urn. Hence, any observed value y should be near to θ; then, any good estimate $\hat{\theta}(y)$ will likewise be close to θ, and as a result $e^2(0)$ will be small. Next, the walls of the container are broken, so that the particles are free to randomly move away. They will follow, of course, the random walk process which is called Brownian motion [18].

Consider a later time interval $(t,t+dt)$. For Brownian motion, $p(x)$ is Gaussian with a variance $\sigma^2 \propto t$. By Eq. (2), then $I=1/\sigma^2 \propto 1/t$, or I decreases with t. Consequently, as $t\to\infty$, $I\to$minimum.

This point is further clarified in the case when $p(x)$ is Gaussian. We had, then, $I=1/\sigma^2$. But also the entropy H [see Eq. (A1)] obeys $H=\frac{1}{2}+\ln\sqrt{2\pi\sigma^2}$. Eliminating the common parameter σ^2 in the last two expressions gives $I=2\pi\exp(1-2H)$. Clearly, then, as H increases I decreases.

An arrow of time

For the preceding problem involving classical particles, we found that Fisher information I decreases with time after a measurement is made. In fact, in general, I decreases with time after each measurement, as is found below. Hence, Fisher information defines an arrow of time. It points in the direction of *decreasing ability to estimate*. Interestingly, this arrow usually agrees with that of thermodynamic entropy, since it is known [19,20] that measurement is an irreversible process which, accordingly, increases local entropy monotonically. Hence, each arrow points in the direction of increased disorder, if "disorder" is appropriately defined by I or H. Of course, other arrows of time exist as well; see Zeh [36].

We showed above that, for one scenario, I decreases with time. This result is generalized below using a mechanism we call the "information demon" (not to be confused with the Maxwell demon).

THE INFORMATION DEMON

The variational problem under consideration is, so far, $I=$extremum, information I given by Eq. (6). But I is specifically the formal *information received* in the measurement. I therefore does not describe a specific physical scenario. So far, the theory has no room for one. Hence the approach cannot yet be used to derive specific physical laws.

We need to supplement the approach with a term that contains information about the specific physical scenario. How can such a term arise? Whereas Eq. (6) represents information *received* by the user, we have yet to consider an information *payout*. This will be provided by the physical scenario. A clue is provided by the theory of Brillouin [20], according to which thermodynamic entropy and Shannon information are equivalent. By this equivalence, the total entropy plus information for a closed system remains fixed. The information gain (in "bits") by an observer is exactly balanced by the natural entropy change (in cal/K) of the physical system. We now extend this idea to Fisher information I and its physical form J (defined below) taken in a given physical scenario.

Information transfer game

Imagine a contest between an information user called the "observer" and an information giver called the "demon." The observer is the intelligent measure. His aim is to acquire high information I in the measurement, so as to minimize e [see Eq. (1)]. By contrast, the demon represents a physical effect that gives rise to the information. It represents nature's response to the observer. Since the observer and the demon comprise a closed system, information is conserved (as in the previous approach of Brillouin [20]). Consequently, whatever information is gained by the observer is at the expense of the demon. The contest is a zero-sum game [21]. Symmetric with the observer's aims and to preserve the second law [20], the demon's aim is to minimize his information loss. (This symmetry constitutes an aspect of "dualism," as discussed later.)

The game consists of a tactic i by the observer, whose aim is to maximize his acquired information I, and a tactic j by the demon, whose aim is to minimize his payout of information. Although tactics i and j are generally continuous, it is useful to first consider a discrete case. This can be illustrated by a payoff matrix, as in the simple 2×2 game of Table II with $i=1,2$ and $j=1,2$. Each item $I(i,j)$ represents both a payoff to the observer and a payout from the demon. Assume that both players know these payoffs. What tactic should each take?

The demon is to choose the column j that minimizes his loss. If he chooses a column $j=1$, the most he can lose is 5, while if he chooses $j=2$, the most he can lose is 4. Hence he chooses $j=2$. Denote the largest number in column j by $\max_i I(i,j)$. Then his optimum payoff is

$$I=\min_j \max_i I(i,j)=\min_j(5,4)=4 \tag{9}$$

in this example. These information quantities I and $I(i,j)$ of course denote, as well, corresponding gains of information by the observer.

It is important now to give a distinct notation to the payouts of the demon. Call these J and $J(i,j)$, where

TABLE II. A 2×2 payoff matrix $I(i,j)$.

i \ j	1	2
1	1.5	2.0
2	5.0	4.0

$J=I$ and $J(i,j)=I(i,j)$. Hence, (9) may be recast as

$$J=\min_j \max_i J(i,j)=\min_j(5,4)=4 \ . \tag{10}$$

That is, the value 4 is both the optimized payoff to the observer and the optimized payout of the demon. This is for an optimum strategy for the demon, as defined above.

System net information change

Regard the observer and demon as a system. Then $\Delta I = I - J$ is the net information change in the system due to the game. Then from Eqs. (9) and (10)

$$\Delta I=\min_j \max_i I(i,j)-\min_j \max_i J(i,j)$$
$$=\min_j[\max_i I(i,j)-\max_i J(i,j)]=0 \ . \tag{11}$$

The middle equality trivially follows since all elements $I(i,j)=J(i,j)$ identically. The zero in (11) follows from the zero-sum nature of the game.

Designate the maximized $I(i,j)$ over i as $I_{\max}(j)$, and likewise for $J(i,j)$. Then (11) is equivalent to

$$\Delta I=I_{\max}(j)-J_{\max}(j)=\min=0 \ . \tag{12}$$

where the minimum is over j.

PRINCIPLE OF EXTREME PHYSICAL INFORMATION

All that remains is to assign roles for i and j in our information game. These will now be continuous rather than discrete, but the preceding game results still hold. See Fig. 3.

Choice j, of the demon, is to govern the amount of information I to be doled out. This he wants to minimize. But, by the form of Eq. (6), I is minimal when the curves $q_n(\mathbf{r})$ have small gradient content, defining broad and smooth functions. This also represents a situation of high disorder [see Fig. 2(b)]. Hence the minimization in principle (12) is through the shapes of the modes $q_n(\mathbf{r})$. These the demon wants to maximally broaden and smooth out. An ordinary variational problem in the $q_n(\mathbf{r})$ results.

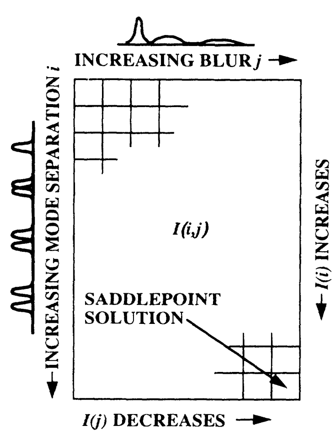

FIG. 3. The information game. The observer bets on a state of mode separation i (a row value). The demon bets on a state of blur j (a column value). The information payoff from the demon to the user is indicated by the item $I(i,j)$ on the playboard. Each player wants to maximize his information. The resulting bets (i,j) will always be at the indicated saddle point.

Characteristic information state

Choice i, of the observer, governs the amount of information to be gained, which he wants to maximize. It is known that, in a situation (8) of multiple modes, I increases as the modes are separated [23]. Hence i is chosen to define a variable mode spacing, which the observer wants to increase until the modes no longer overlap. See Fig. 3. In this mode geometry, information (6) becomes

$$I=4\sum_{n=1}^{N}\int d\mathbf{r}\,\nabla q_n\cdot\nabla q_n \ . \tag{13}$$

This is the trace of the Fisher information matrix [5]. Also probability law (7) becomes

$$p(\mathbf{r})=\sum_{n=1}^{N} q_n^2(\mathbf{r}) \ . \tag{14}$$

In summary, $I_{\max}$ and $J_{\max}$ in (12) correspond to a condition of completely separated modes. A system with such modes is said to be in its "characteristic information state." This defines the ability of the system to convey information over all possible mode positions, and is analogous to the concept of channel capacity (i.e., maximized information) in Shannon information theory. It also has a counterpart in statistical mechanics as an "unmixed" state, for which the entropy H is *minimized.* Thus by either measure H or I, the disorder is minimized in this state.

By Eq. (13) and the preceding, the continuous version of principle (12) is

$$\Delta I=4\sum_{n=1}^{N}\int d\mathbf{r}\,\nabla q_n\cdot\nabla q_n-\int d\mathbf{r}\,F[\mathbf{q}(\mathbf{r}),\mathbf{r}]$$
$$=\text{extremum}=0 \ , \tag{15}$$

through variation of the q_n, where

$$\int d\mathbf{r}\,F[\mathbf{q}(\mathbf{r}),\mathbf{r}]\equiv J \ , \quad \mathbf{q}\equiv(q_1,\ldots,q_N) \ . \tag{16}$$

The functional F is defined below. Since the minimum in principle (12) might sometimes be a point of inflection, we replaced it with an extremum requirement. Equation (15) is called the principle of extreme physical information (EPI), as discussed below.

EPI principle as a consequence of the game

We constructed principle (15) only by analogy, i.e., as a continuous version of the discrete principle (12). It is more important to show that (15) directly follows from the play of the information transfer game. As we discussed, the game index j has the physical significance of being the mode functions $\mathbf{q}$. Then information I and J are *functionals* $I(\mathbf{q})$ of $\mathbf{q}$ [obeying Eqs. (13) and (16)]. During the play of the game, the observer maximizes I and J (since equal) by separating the modes $\mathbf{q}$.

Meanwhile, the demon functionally minimizes I and J through variation of the shapes of the $\mathbf{q}$. Then by the latter operations, $\delta I(\mathbf{q})=\delta J(\mathbf{q})=0$. This implies as well that $\delta(I(\mathbf{q})-J(\mathbf{q}))=0$; or, equivalently, that Eq. (15) is satisfied.

Interpretation of J and F

Recall that J is the payoff of information by nature. Hence, J is an expression of I in terms of the physics of the scenario. Or, J is the expression of I as an invariant quantity under an appropriate transformation. In fact, by the form of Eq. (13) as an inner product, I remains invariant to a general unitary transformation on $\mathbf{q}$. To ensure that J has physical meaning (as required), the unitary transformation to be chosen must be one that defines transform coordinates that have physical meaning.

For example, in the quantum mechanics scenario these are the momentum-energy coordinates that characterize a Fourier (unitary) transformation. Then J has the physical significance of being essentially the mean-square particle momentum (see below).

In any scenario, the particular unitary transformation is suggested by a symmetry, invariance, or conservation property of the scenario. This is an attractive aspect of the approach, since such a property is usually basic to a given scenario. For example, in quantum mechanics that property is the symmetry between position-time space and momentum-energy space. Quantity F, in Eq. (16), then becomes the conjugate to the inner product $\sum_n \nabla q_n \cdot \nabla q_n$ in measurement space. This is generally true as well.

Information exchange processes

Principle (15) states that a physical law $\mathbf{q}(\mathbf{r})$ arises during an exchange of information between nature and a smart measurer. The exchange process is initiated by the measurer, so that the physical law that fixes the measurement arises in reaction to the measurement.

There are, of course, other exchange processes in nature. In each, a physical effect results from the exchange of a theoretical "substance." Thus the exchange of energy (say, potential to kinetic) for a particle gives rise to a change in motion. Or, the exchange of mesons gives rise to a force law. Now we find that the exchange of information gives rise to probability laws and their modes, as well as the constancy of many (all?) fundamental physical constants, and even some *deterministic* laws [16] (e.g., Maxwell's equations and the Einstein gravitational equations of motion).

The information exchange takes place within a closed, or isolated, physical system. This follows, since the information game is solely between the observer and nature. Our previous assumption that information is conserved is based upon this premise.

Principle of extreme physical information

Principle (15) is, specifically one of *extreme exchange* of Fisher information. For brevity, we call quantity $\Delta I = I - J$ "physical information," and the principle (15) "extreme physical information" (EPI). Operationally, the principle requires probability modes $q_n(\mathbf{r})$ to be varied, in the usual manner of a problem in the calculus of variations. Thus the extremum so obtained is a variational extremum, i.e., one whose first variation is zero.

Quantity ΔI as an information distance

The form of the EPI principle (15) requires a solution $q_n(\mathbf{r})$ to simultaneously extremize and zero it. How can this be? These are the properties of an ordinary distance measure. That is, the extreme value of the distance between two points in ordinary (x,y,z) space is zero when it is a minimum. Analogously, here the distance is between quantities I and J in the space of functions $q_n(\mathbf{r})$. Distance measurements of this sort are studied in the field of mathematics called differential geometry, and are called "information distance measures" when applied to probability laws. Hence the physical information ΔI is equivalently an *information distance.*

Let us now examine the two components I and J of ΔI in this light. It is shown in Appendix B that the Fisher I part of ΔI is, approximately, a *cross-entropy* form of information. Now, information entropy (in "bits") and thermodynamic entropy in (cal/K) are equivalent phenomena [20] in a closed system (as here). In this regard, note that the other part J of ΔI must equal I, and by hypothesis is of physical origin. Then it must approximately represent *thermodynamic entropy* (as well as Fisher information). Hence, by these correspondences, principle (15) becomes, approximately, a principle of extreme exchange of cross entropy, or a principle of extreme cross-entropy distance. Clearly, this principle relates to the second law in some sense.

Origin of universal constants: Corollary 1

It is important to emphasize the difference in the roles of I and J. I is always of the general form (6) or its mathematical equivalent. On the other hand, J is the extremized value of I for the particular scenario considered. An extremized value is a constant, and we regard each such as a universal physical constant in particular. In this way, each EPI scenario leads to a value of J that fixes a different universal physical constant. We give an example below.

Origin of invariance principles: Corollary 2

The preceding has an interesting consequence. Since the extremized value for J is to be a universal constant, it must be so regardless of reference frame, choice of gauge, choice of curvilinear coordinate system, etc. These are the usual demands on Lagrangian formulations, and we now see how they naturally arise out of the information approach.

Symmetry of the game

We developed the principle from the standpoint of an optimum strategy for the demon. It is interesting to consider it instead from the standpoint of the observer-measure (with whom we tend to identify). Although his aim is diametrically the opposite of the demon's, it turns out that his optimum payoff I is exactly the same as that defined above for the demon [21]. For example, both arrive at the point $i=2$, $j=2$ in the matrix game of Table II. The reason is that this point is a saddle point in the matrix (minimum in row 2 and maximum in column 2). Likewise, with the designated roles for i and j in our smart measurement game of Fig. 3, the payoff point is another saddle point. The upshot is that the same principle (15) again results.

EPI AS A PROCEDURE FOR FORMING LAGRANGIANS

The foregoing suggests an information-based procedure for forming Lagrangians. First select the coordinates $\mathbf{r}$. These are suggested by the nature of the phenomenon, and should comprise a four-vector (although use of smaller dimensionality is permissible, if less efficient to use; see example below). The choice $\mathbf{r}$ should also be such that, in the absence of fields, a probability distribution function (PDF) $p(\mathbf{r})$ obeying form (14) is physically observable. Then the Fisher information I obeys Eq. (13), its form in measurement space. Next, a property of symmetry or conservation F is used to define an appropriate unitary transformation of I from $\mathbf{r}$ (measurement) space to a physical space. I is now in a form J, which contains appropriate parameters (such as the speed of light c, Planck's constant h, etc.) of the physical space. The Lagrangian is then formed as $(I-J)$. An example of this procedure is given later.

EPI AS A PROCEDURE FOR KNOWLEDGE FORMATION

EPI may be regarded as a gedanken procedure for the synthesis of physical laws. Specifically, EPI is a procedure for codifying the laws that govern our measurements. The human element, in the form of the observer, plays the key role of initiating EPI. The observer makes a smart measurement. This is to convey maximum knowledge (information I) about the phenomenon being measured. It is assumed that the appropriate symmetry and/or conservation relation F preexists the gedanken measurement. Next, the characteristic state of modes q_n is enforced by the observer as a kind of "prepared state," in the language of quantum mechanics. These ground rules activate the information transfer game. The result is the physical law systematizing and governing the smart measurement in the given scenario.

But is one justified in defining physical laws through the use of a purely mental construction of reality, such as by use of the game? Would it not be preferable to show that the laws follow instead from physical effects, i.e., independent of an observer and any of his mental processes? Now, by "physical laws" we mean a collection of *mathematical* relations, such as the Schrödinger wave equation and Maxwell equations. These are to be distinguished from a class of *observable* variables $\mathbf{r}$ (mass, charge, length, time, etc.) which such laws functionally connect. Hence the laws are actually mental constructions, as distinguished from the direct observables that they connect.

It is interesting to compare the variational EPI principle with the maximum entropy (ME) principle [22]. ME has its physical roots in the second law, and has been used (appropriately) to derive the laws of equilibrium statistical mechanics. By contrast, EPI has its roots in smart measurement theory, and derives the laws of statistical mechanics as well, but also those of quantum mechanics and (probably) every other *measurable* phenomenon [23]. The reason for the overlap of the two approaches in equilibrium statistical mechanics is, in fact, that in this field EPI becomes ME. This is shown in Appendix C. Hence, EPI is the generally valid principle; and becomes ME in a limiting case (equilibrium) of a special scenario.

EPI addresses a long-standing problem of measurement theory—the "participatory role" to be played by the observer during the act of measurement. Such a role has been much speculated about [14,24,25]. The usual question addressed is how the act of measurement can be made to fit within an existent quantum theory. By contrast, we show that the measurement procedure *forms* the laws of quantum theory (see below). It also forms other physical laws (see below), depending upon what quantity is being measured. Finally, the "participatory role" of the observer is to make the smart measurement and to play the ensuing game, as described previously.

PAST APPLICATIONS OF THE PRINCIPLE

Earlier versions of the EPI principle have been shown [10,16,23,26,27] to derive a host of physical laws and constants. (The present version of EPI preserves the formalism for constructing Lagrangians that is given in the earlier versions, but in addition provides a theoretical basis—the "game"—for the formalism.) The laws are both statistical and *deterministic* in nature. Statistical laws directly result from the use of the EPI principle (15). Deterministic laws and the constants follow from the use of corollaries 1 and 2, or, alternatively, from a known, underlying statistical effect (e.g., the "cosmological principle" in the derivation of general relativity [16,23].) Examples of derived statistical laws are the Dirac equation of relativistic quantum mechanics [16,23] and the Maxwell-Boltzmann velocity distribution [23]. Examples of derived deterministic laws are the Lorentz transformation group of special relativity [16,23] and the Einstein equations of motion of general relativity [16,23]. Constants that have thus far been fixed by EPI are c, m (mass of electron), e (charge of electron), and h [16]. Each such law arises from the smart measurement of an appropriate parameter $\boldsymbol{\theta}$. Examples of $\boldsymbol{\theta}$ are the ideal space-time position for a particle in the Dirac scenario, or the drift velocity of an urn containing ideal gas particles in the Maxwell-Boltzmann scenario. Again, each resulting physical law arises as if *in reaction to* a smart measurement in the characteristic state.

Some predictions

The preceding applications to known physics may be regarded as verifications of the EPI approach. But EPI is a general prescription for establishing probability laws, deterministic laws, and their associated parameters. EPI has also derived additional laws, i.e., made verifiable predictions. So far, these are as follows.

(a) The rate of increase of Boltzmann entropy is bounded from above, at each instant of time, by a number proportional to the current value of the system Fisher information I [27].

(b) The motion of a relativistic, classical particle gives rise to a flow of information that is proportional to its rest mass [16,23].

(c) Classical turbulence is describable by a probability law on the four-vector $(\rho\mathbf{u}, i\rho)$ where ρ is the particle density and $\mathbf{u}$ is the fluid velocity. This probability law obeys a second-order partial different equation derived from the Euler-Lagrange solution to EPI [28].

(d) A new fundamental particle is predicted [23].

DERIVATION OF RELATIVISTIC QUANTUM MECHANICS

The EPI approach is perhaps best verified by its application to quantum mechanics [16]. The Klein-Gordon equation and Dirac equation will result, as well as the equivalence of mass and energy, and the constancy of the Planck constant h.

From the form of Eq. (6), Fisher information I is the trace of a matrix, and as such is invariant to a rotation of coordinates. Hence it is a relativistically covariant quantity. It follows that all smart measurements must be of four-dimensional quantities [16]. Therefore, let the observer now attempt to estimate the space-time, ideal (classical) position $\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_4)$ of a particle of mass m. The fluctuations from the classical value [29] $\boldsymbol{\theta}$ are designated

$$x_1 = ix, \quad x_2 = iy, \;, \quad x_3 = iz \;, \quad x_4 = ct \;,$$
$$(x_1, x_2, x_3) = i\mathbf{r} \;, \quad i = \sqrt{-1} \;. \tag{17}$$

(The parameter c is presumed fixed as a universal constant, from a prior application of the EPI to the derivation of gravitational phenomena [16].) We seek the probability law $p(\mathbf{r},t)$, and the modes $q_n(\mathbf{r},t)$. It is convenient to pack the real modes $q_n(\mathbf{r},t)$ as new, *complex modes*

$$\psi_n \equiv q_{2n-1} + iq_{2n} \;, \quad n = 1, 2, \ldots, N/2 \;. \tag{18}$$

The $\psi_n = \psi_n(\mathbf{r},t)$ are now the unknowns of the problem.

Note that, when expressed in terms of the ψ_n, probability law p and information I remain in the same form as before. The use of quantities (17) and (18) given directly

$$\sum_{n=1}^{N/2} \psi_n^* \psi_n = \sum_{n=1}^{N} q_n^2 = p(\mathbf{r},t) \tag{19a}$$

by Eq. (14), and

$$4c \sum_{n=1}^{N/2} \int\int d\mathbf{r}\, dt \left[-(\nabla\psi_n)^* \cdot \nabla\psi_n + \left[\frac{1}{c^2}\right] \left[\frac{\partial\psi_n}{\partial t}\right]^* \frac{\partial\psi_n}{\partial t} \right] = I \tag{19b}$$

by Eq. (13). Result (19a) in particular shows that the modes ψ_n are automatically probability amplitudes since p is a probability. The Born assumption to this effect does not have to be made.

By the general approach, we need to find what I equals physically for this scenario. That is, we seek a transformation under which I remains invariant. Because of the inner-product nature of form (19b), this transformation will generally be a unitary transformation. According to plan, in the transformed space I becomes J, the *physical* manifestation of I. This requires the transform space coordinates to have physical significance. Coordinates of momentum and energy have such significance. They characterize the particular unitary transformation called the Fourier transformation.

Quantum Fourier symmetry

Define a momentum-energy space as the Fourier conjugate space to position-time,

$$(i\mathbf{r}, ct) \leftrightarrow (i\boldsymbol{\mu}/\hbar, E/c\hbar) \;, \tag{20a}$$

with

$$\psi_n \leftrightarrow \phi_n \tag{20b}$$

as Fourier transform mates ψ_n and ϕ_n. As a result of (20b),

$$(\nabla\psi_n, \partial\psi_n/\partial t) \leftrightarrow (-i\boldsymbol{\mu}\phi_n/\hbar, iE\phi_n/\hbar) \;. \tag{20c}$$

Note that quantities E and $\boldsymbol{\mu}$ are simply regarded as "coordinates" of the Fourier space. They do not have any prior physical significance. Indeed, any possible relation between "energy" coordinate E and "momentum" coordinates $\boldsymbol{\mu}$ is at this point undefined. It will be derived later as the famous equivalence (26) of energy, mass, and momentum.

Also, h is at first regarded as a parameter that is constant in a particular problem, but which is not necessarily a *universal* constant. EPI theory will later fix it as a universal constant.

It is important that the symmetry relation not be so detailed that, by itself, it implies the thing being sought (here, the Klein-Gordon, and Dirac equations). In fact, relations (20a), (20b) are merely the statement that ψ_n has a Fourier transform. Hence the relation is certainly not overly detailed.

Use of Parseval's theorem

According to this theorem, the integrated area (19b) in position-time space equals a corresponding area in Fourier space, or I remains *invariant* under Fourier (unitary) transformation, as required by EPI. Specifically, (19b) becomes

$$I \equiv J = (4c/\hbar^2) \int \int d\mu \, dE \, P(\mu, E)(-\mu^2 + E^2/c^2) \ , \quad \text{(21a)}$$

where we denote

$$\sum_{n=1}^{N/2} \phi_n^* \phi_n \equiv P(\mu, E) \ . \quad \text{(21b)}$$

Terms in μ^2 and E^2 in (21a) arise out of correspondences (20c) applied twice by the theorem.

From its definition (21b), $P \geq 0$. Also, integrating Eq. (21b) over all (μ, E) gives, by Parseval's theorem, the area under curve $\sum \psi_n^* \psi_n$, which is unity by normalization of $p(\mathbf{r}, t)$. Hence P may be regarded as a probability law in momentum-energy space. Then the right hand side of (21a) is simply an expectation

$$J = \left[\frac{4c}{\hbar^2} \right] \left\langle -\mu^2 + \frac{E^2}{c^2} \right\rangle \ . \quad \text{(22)}$$

This expression has many implications, as follows.

Planck's parameter as a constant

By corollary 1, J is to be a universal constant. Since the two factors in (22) are independent, each must be a constant. In the first factor, a parameter c is already fixed as a universal constant, from the EPI general relativity derivation [16]. Then parameter h must be a universal constant as well.

Equivalence of matter and energy

Consider next the second factor. Now, the fluctuations in E and μ necessarily change from one set of boundary conditions to another. This would make the factor a variable, unless

$$-\mu^2 + \frac{E^2}{c^2} = \text{const} \equiv A^2(m, c) \ , \quad \text{(23)}$$

where A is some function of the rest mass m and the speed of light c (the only other constants of the free-field scenario). Solving for E gives

$$E^2 = c^2 \mu^2 + A^2(m, c) c^2 \ . \quad \text{(24)}$$

By dimensional analysis, the function $A(m, c)$ must obey the relation

$$A = mc \ , \quad \text{(25)}$$

where m is defined to be the mass of the particle. Equation (24) then becomes

$$E^2 = c^2 \mu^2 + m^2 c^4 \ . \quad \text{(26)}$$

This is the familiar equivalence of mass, momentum, and energy. We see that it is a consequence of EPI theory.

Klein-Gordon and Dirac equations

We can now proceed to form the physical information. Putting (26) into (21a) gives

$$J = (4m^2c^3/\hbar^2) \int \int d\mu \, dE \sum_{n=1}^{N/2} \phi_n^* \phi_n$$

$$= (4m^2c^3/\hbar^2) \int \int d\mathbf{r} \, dt \sum_{n=1}^{N/2} \psi_n^* \psi_n \quad \text{(27)}$$

by Parseval's theorem. Then, by Eqs. (19b) and (27), the Fisher information transfer (15) is

$$\Delta I \equiv I - J = 4c \sum_{n=1}^{N/2} \int \int d\mathbf{r} \, dt \left[-(\nabla \psi_n)^* \cdot \nabla \psi_n + \left[\frac{1}{c^2} \right] \left[\frac{\partial \psi_n}{\partial t} \right]^* \left[\frac{\partial \psi_n}{dt} \right] - \frac{m^2 c^2}{\hbar^2} \psi_n^* \psi_n \right] . \quad \text{(28)}$$

According to the EPI principle (15), this is now to be extremized through variation of the ψ_n. The resulting Euler-Lagrange equation is the Klein-Gordon equation of relativistic quantum mechanics [20]. Or factoring the integrand and using a matrix-vector approach gives the Dirac equation [16]. Finally, the electromagnetic potentials $\mathbf{A}$ and ϕ are injected into the theory by invoking corollary 2—in particular, invariance to gauge choice. This is well known to be accomplished by the substitutions (5). When these are made, the results become correct for the particle in the presence of a general electromagnetic field.

Schrödinger wave equation

This may be shown, in the usual way [30], to be the nonrelativistic limit of the Klein-Gordon equation, or of the Dirac equation in the absence of a magnetic field. In this way, the Schrödinger wave equation follows as well from EPI. It cannot be derived directly because EPI is a relativistically covariant theory, while the Schrödinger formulation is not (since it treats space and time differently).

This said, in fact, a *noncovariant* use of EPI does derive the Schrödinger wave equation directly [17]. (EPI is a robust theory.) This entails ignoring the time, using for coordinates $\mathbf{r}$ only the space coordinates of a particle. Then the *time-independent* Schrödinger equation, or the Dirac equation, results. A further drawback of the noncovariant approach is that the mass-energy relation (26) then has to be assumed, rather than derived as here. Also, the constancy of h is not proved. Hence, although EPI can be used with noncovariant coordinates $\mathbf{r}$ to yield a correct probability law $p(\mathbf{r})$, only the full-fledged covariant approach gives the added benefits of a correct time dependence $\psi_n(\mathbf{r}, t)$, predicted connections among the coordinates as in Eq. (26), and a predicted physical constant, such as h in Eq. (22).

Synopsis

Quantum mechanics has been seen to arise from an interplay between nature and an observer who intelligently measures the space-time coordinates of a particle. The interplay takes the form of an information game. The game is an exchange of information between coordinate-time space (the observer's domain) and momentum-energy space (the demon's). The latter space is defined by the specific unitary transformation chosen, here the Fourier one. The payoff of the game fixes the Klein-Gordon or Dirac equation. These are expressions of maximal width for the complex modes ψ_n. The result is a maximally broadened $p(\mathbf{r},t)$, and hence a maximally disordered system.

PHILOSOPHICAL ASPECTS

EPI is an epistemology for eliciting those laws that describe the observed world. EPI also has an essential "twoness" or "dualism" about it. Thus there are two protagonists in the information transfer game with, as a result, two forms I and J of information. Each protagonist aims to maximize his information state: the observer, by placing modes $\mathbf{q}$ in the characteristic (nonoverlap) state; and the demon, by maximally broadening the modes. This amounts to two distinct tactics.

The two protagonists, each employing a distinct tactic, successfully address a long-standing metaphysical question about the relation between "mind" and "matter" (i.e., physical world). "Mind" is represented by the gedanken observer, who makes a smart measurement and demands a maximum of information in the data. He accomplishes this by increasing the mode separation, which is choice i in the information game matrix $I(i,j)$. This constitutes one-half of the dualism, and of the game.

In the physical half of the game, the world responds to measurement by maintaining the most disorder, i.e., returning the least information. This response may be understood as another way of expressing the second law of thermodynamics: Nature acts to maximize the disorder in the data, through the Fisher measure J (not through the usual entropy H). This takes the form of a maximal broadening of the mode widths, corresponding to choice j in the information matrix $I(i,j)$. In this way, the choice (i,j) is made, the game is played, and the dualism is attained.

Two recent conjectures [14,31] touch on the problems of dualism. These provide a suggestive, but nonquantitative, discussion of the foundations of physical theory. EPI theory, by contrast, quantifies the relationship between consciousness, information, and the understanding of physics. We end this paper by describing these conjectures and showing how EPI theory takes them out of the realm of speculation.

The first conjecture is, "All things physical are information-theoretic in origin and this is a participatory universe . . . . Observer participancy gives rise to information; and information gives rise to physics" [14].

In the context of EPI theory "observer participancy" is manifest as the smart measurement. This "gives rise to information" I, whose "origin" is the physical phenomenon as manifest in information J. The transfer of information from form J to I takes place during a game, and the net transfer $(I-J)$ of "information gives rise to physics" through the resulting EPI principle (15).

The second conjecture is, "Matter and consciousness are two realities in themselves, which are capable of mutual interaction" [31].

In the context of EPI theory the observer "consciously" measures, obtaining data at information level I. Corresponding to I is the "matter" form J. These are distinct "'realities in themselves" which "mutually interact" during the information transfer game.

Thus, Fisher information and the EPI principle provide the missing mechanism for substantiating the two conjectures. The *creating mind,* although separate from the world it seeks to explain, engages it in a dynamic and reciprocal relationship that gives rise to physical theory.

APPENDIX A: SOME COMPARISONS OF FISHER'S AND SHANNON'S FORMS OF INFORMATION

Derivation of Fisher information

Fisher information I, Eq. (2), derives readily. We follow Van Trees [5]. Consider the class of estimators $\hat{\theta}(\mathbf{y})$ that are unbiased, obeying

$$\langle \hat{\theta}(\mathbf{y})-\theta \rangle = 0 \equiv \int d\mathbf{y}[\hat{\theta}(\mathbf{y})-\theta]\mathbf{p}(\mathbf{y}|\theta) , \tag{A1}$$

where $p(\mathbf{y}|\theta)$ is the probability density for a vector of data values $\mathbf{y}$ in the presence of one parameter value θ. Differentiate (A1) with respect to θ, giving

$$\int d\mathbf{y}(\hat{\theta}-\theta)\frac{\partial p}{\partial\theta} - \int d\mathbf{y}\, p = 0 . \tag{A2}$$

Use the identity

$$\frac{\partial p}{\partial\theta} = p\frac{\partial \ln p}{\partial\theta} \tag{A3}$$

and the fact that p obeys normalization. Then (A2) becomes

$$\int d\mathbf{y}(\hat{\theta}-\theta)\frac{\partial \ln p}{\partial\theta}p = 1 . \tag{A4}$$

Factoring the integrand gives

$$\int d\mathbf{y}\left[\frac{\partial \ln p}{\partial\theta}\sqrt{p}\right][(\hat{\theta}-\theta)\sqrt{p}\,] = 1 . \tag{A5}$$

Square this equation. Then the Schwarz inequality gives

$$\left[\int d\mathbf{y}\left[\frac{\partial \ln p}{\partial\theta}\right]^2 p\right]\left[\int d\mathbf{y}(\hat{\theta}-\theta)^2 p\right] \geq 1 . \tag{A6}$$

This is Eq. (1) of the paper, called the Cramer-Rao inequality [5]. It links the mean-square error e^2 of estimation [second factor in (A6)] to the Fisher information I (defined as the first factor). In the case $\mathbf{y}=y$ of one data value obeying the additive form $y=\theta+x$, I directly becomes Eq. (2) of the paper. [The additive data form implies that $p(y|\theta)=p(y-\theta)$. This means that $p(y|\theta)$ preserves the same shape irrespective of the size of θ. This is a statement of shift invariance, and corresponds, e.g., to Galilean invariance in nonrelativistic phenomena or Lorentz invariance in relativistic phenomena.]

Comparisons with Shannon's form of entropy

The Shannon entropy H is

$$H=-\int dx\, p(x)\ln p(x)\ . \tag{A7}$$

For our purposes, it is useful to work with the discrete form

$$H=-\Delta x \sum_n p(x_n)\ln p(x_n)\ , \tag{A8}$$

Δx being regarded as very small. The sum (A8) is simply a sum of values of a function $p(x_n)\ln p(x_n)$ over its pixels x_n. The pixels may be summed over in any order; the same sum always results. Graphically, this means that if the curve $p(x_n)$ undergoes a rearrangement of its points $(x_n p(x_n))$, although the shape of the curve will drastically change, the value of H remains constant. H is then said to be a *global measure* of the behavior of $p(x_n)$.

By comparison, the discrete form of Fisher information I is, from Eq. (2),

$$I=\Delta x^{-1}\sum_n \frac{[p(x_{n+1})-p(x_n)]^2}{p(x_n)}\ . \tag{A9}$$

If the curve $p(x_n)$ undergoes a rearrangement of points x_n as above, now the *local* slope values $[p(x_{n+1})-p(x_n)]/\Delta x$ will change drastically, and so the sum (A3) will also change strongly. Discontinuities in $p(x_n)$ will now occur, and these have (in the continuous limit) infinite slopes, so that I in fact will go toward infinitely. Since I is thereby sensitive to local rearrangement of pixels, it is said to have a property of *locality*.

Thus H is a global measure, while I is a local measure, of the behavior of the curve $p(x_n)$. These properties hold in the limit $\Delta x \to 0$, and so apply to the continuous probability density $p(x)$ as well.

This global vs local property has an interesting ramification. Because the integrand of I contains a squared derivative p'^2 or $\nabla q_n \cdot \nabla q_n$ [see Eqs. (2) and (13)], when the integrand is used as part of a Lagrangian the resulting Euler-Lagrange equation will contain second-order derivative terms p'' or $\nabla^2 q_n$ (see Refs. [16,17,23]). That is, a second-order differential equation results. This dovetails with nature, in that the fundamental differential equations that define probability densities or amplitudes in physics are second-order differential equations. Indeed, the thesis of this paper is that the correct such equations result when the EPI agenda is followed.

By contrast, the integrand of H in (A7) does not contain a derivative. Therefore when this integrand is used as part of a Lagrangian the resulting Euler-Lagrange equation will not contain any derivatives: it will be an algebraic equation, with the immediate solution that $p(x)$ is of an exponential form (see Ref. [22]). This is not, then, a differential equation, and hence cannot represent a general physical scenario. The exceptions are those physical effects which happen to be of an exponential form, as in statistical mechanics. (In these cases, I gives the correct solutions anyhow; see Appendix C.)

It follows that, if one or the other of global measure H or local measure I is to be used in a variational principle in order to derive the physical law $p(x)$ describing a *general* scenario, the preference is for the local measure I.

APPENDIX B: APPROXIMATION OF FISHER I BY CROSS ENTROPY K

The accuracy with which an observer can know the Fisher information I for a system is necessarily finite. Since, by definition (2), I obeys

$$I=\int dx\, p'^2/p(x)\ , \tag{B1}$$

to know I one must first know $p(x)$. But an experimental probability density $p(x)$ can only be known through the use of a finite "bin" size Δx, by counting the number of events within discrete intervals $(x_n, x_n+\Delta x)$, $n=1,\ldots,N$. The latter cover the domain of x. This limits the accuracy with which $p(x)$ is known since, in effect, the average value of p over each finite interval $(x_n x_n+\Delta x)$ is computed instead of the instantaneous value.

This would be of little importance if Δx could be made arbitrarily small, but for any physical scenario there will always be a smallest interval Δx permitted in practice. For example, if x is a length, the smallest possible Δx that one could use (in any scenario) is the smallest length, the Planck length of value 10^{-33} cm.

By Eq. (B1), $p'(x)$ must be known as well. With the finite bin size Δx as above, values $p(x_n)$ are calculated, so that each $p'(x)=p'(x_n), n=1,\ldots,N$, is conveniently computed as a first difference

$$p'(x_n)=[p(x_n+\Delta x)-p(x_n)]/\Delta x\ . \tag{B2}$$

Hence the exact Eq. (B1) is necessarily replaced by a finite approximation

$$I\cong \sum_{n=1}^{N} \Delta x \frac{[p(x_n+\Delta x)-p(x_n)]^2}{(\Delta x)^2 p(x_n)}\ . \tag{B3}$$

This is reexpressed as

$$I=\Delta x^{-1}\sum_n p(x_n)\left[\frac{p(x_n+\Delta x)}{p(x_n)}-1\right]^2 . \tag{B4}$$

Now the quantity $p(x_n+\Delta x)/p(x_n)$ is close to unity since Δx is small. Therefore, the quantity in brackets in (B4),

$$p(x_n+\Delta x)/p(x_n)-1\equiv \nu\ , \tag{B5}$$

is small. Now for small ν the expansion

$$\ln(1+\nu)=\nu-\nu^2/2 \tag{B6}$$

holds, or equivalently,

$$\nu^2=2[\nu-\ln(1+\nu)]\ . \tag{B7}$$

Then by Eqs. (B5) and (B7), (B4) becomes

$$I=-2\Delta x^{-1}\sum_n p(x_n)\ln\frac{p(x_n+\Delta x)}{p(x_n)} +2\Delta x^{-1}\sum_n p(x_n+\Delta x)-2\Delta x^{-1}\sum_n p(x_n)\ . \tag{B8}$$

But each of the two far-right sums is Δx^{-1}, by normalization, so that their difference cancels, leaving

$$I \cong -2\Delta x^{-1} \sum_n p(x_n) \ln \frac{p(x_n+\Delta x)}{p(x_n)}$$
$$\equiv -K(p(x_n), p(x_n+\Delta x)) , \quad \text{(B9)}$$

the cross entropy between $p(x_n)$ and $p(x_n+\Delta x)$.

Thus to minimize I is, approximately, to maximize K. Each will give approximately the same solution $p(x_n)$.

In the preceding calculation, the fundamental quantity was I, with K found as an approximation to it. But, of course, the reverse viewpoint may instead be taken. That is, assume instead that K is the fundamental quantity, more specifically, the continuous integral version of (B9). Again, this is to be calculated in the presence of a finite bin size Δx. Then K turns out to be approximated by a sum of terms whose quadratic (in Δx) term is proportional to I [32]. (However, the lower-order terms do not drop out, as they do here. The result is not as "clean.") It follows that to maximize K is to minimize I, since the lower-order terms do not contribute to the variational problem.

So we come full circle. By this (second) viewpoint [33] the fundamental quantity is K, and I exists only as an approximation to *it.*

Which of the two viewpoints is correct? In fact, if the second viewpoint is taken, to solve the variational problem still requires minimizing I. This is because function $p(x+\Delta x)$ is dependent upon function $p(x)$. Both must be varied simultaneously, and the only way to do this is to expand $p(x+\Delta x)$ as a power series in Δx, bringing in functions $p'(x), p''(x)$, etc. The result is a return to the Fisher I problem (which was the first viewpoint). In effect, it is only minimization of I that matters, by *either* viewpoint.

APPENDIX C: TRANSITION FROM EPI TO ME IN STATISTICAL MECHANICS

Represent a general $p(x)$ as

$$p(x)=\exp[g(x)] \quad \text{(C1)}$$

in Eqs. (A1) and (B1). These become

$$H=-\int dx\, p(x)g(x) , \quad I=\int dx\, p(x)g'^2(x) . \quad \text{(C2)}$$

The coordinate x is, as yet, undefined. Later we give it the physical significance of a velocity or an energy.

Regard each of H and I to be extremized subject to the same constraints, including (of course) that of normalization of $p(x)$. (In the context of EPI, the constraints fix J for the scenario.) The usual method of Lagrange multipliers is used to additively tack on the constraint terms.

We are interested in finding a class of solutions $p(x)$ that is common to the two extremization problems. This will occur if H and I are proportional to within a normalization integral,

$$H=-AI-B\int dx\, p(x) . \quad \text{(C3)}$$

Using Eqs. (C2) in (C3) gives a requirement on $g(x)$ that

$$g(x)=Ag'^2(x)+B . \quad \text{(C4)}$$

This simple differential equation has a general solution

$$g(x)=(x-C)^2/4A+B ,$$

where A, B, and C are arbitrary constants.

Hence by (C1) the common solution $p(x)$ is generally in the form of the exponential of a quadratic function. Depending upon the size of A, this includes both the normal solution (for finite A), and the linear exponential solution (for $A\rightarrow-\infty$). Physically, these respectively define the Maxwell-Boltzmann velocity distribution law, for the choice x =velocity; and the Boltzmann energy distribution law, for x =energy. These PDF's are the familiar solutions of equilibrium statistical mechanics.

What we have shown, then, is that equilibrium statistical mechanics is the common meeting ground of the EPI and ME approaches to estimating PDF's. Next, consider the more general circumstance of temporally nonequilibrium statistics. EPI generally seeks such solutions, since EPI is a generally covariant theory, treating time like any other coordinate. By contrast, ME extremization follows from the Boltzmann H theorem result that $dH/dt\geq 0$, which implies that time must approach infinity in order for H to be maximized. Hence, ME is restricted in scope to temporal equilibrium solutions.

The distribution functions of nonequilibrium statistical mechanics are known to obey the Boltzmann transport differential equation. The solution to this is a general superposition of Hermite-Gauss functions [34]. In fact, EPI generates these solutions [35] as subsidiary minima in ΔI, with the absolute minimum attained by the Maxwell-Boltzmann solution. This is under the constraints of normalization and mean energy.

However, under the same constraint inputs, ME only gives the equilibrium (as above), Maxwell-Boltzmann answer [22]; it fails to produce any of the higher-order Hermite-Gauss solutions. (Mathematically, this is because multiple solutions follow from differential equations, which ME cannot produce; also see Appendix A and the second preceding paragraph.) Hence EPI and ME only coincide at the most elemental level of statistical mechanics, that of equilibrium statics. Beyond this level, ME does not apply.

[1] J. L. Lagrange, *Méchanique Analitique* (LaVeuve Desaint, Paris, 1788); republished in *Oeuvres de Lagrange* (Gauthier-Villars, Paris, 1888).

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

[3] R. A. Fisher, *Statistical Methods and Scientific Inference,* 2nd ed. (Oliver and Boyd, London, 1959).

[4] E. Schrödinger, Ann. Phys. (Leipzig) **79**, 361 (1926).

[5] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 1968).

[6] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[7] A. J. Stam, Information Control **2**, 101 (1959).
[8] N. M. Blachman, IEEE Trans. Inf. Theor. **IT-11**, 267 (1965).
[9] B. R. Frieden, *Probability, Statistical Optics and Data Testing,* 2nd ed. (Springer-Verlag, Berlin, 1991), p. 407.
[10] B. R. Frieden, Phys. Lett. A **169**, 123 (1992).
[11] J. M. Levy-Leblond, Phys. Lett. A **111**, 353 (1985).
[12] R. A. Fisher, Ann. Eugenics **12**, 1 (1943).
[13] E. P. Wigner, *Symmetries and Reflections* (Ox Bow Press, Woodbridge, CT, 1979).
[14] J. A. Wheeler, in *Proceedings of the 3rd International Symposium on Foundations of Quantum Mechanics, Tokyo, 1989,* edited by S. Kobayashi, H. Ezawa, Y. Murayama, and S. Nomura (Physical Society of Japan, Tokyo, 1990), p. 354.
[15] This is the positivist viewpoint. See, e.g., Lord Kelvin, *Popular Lectures and Addresses* (MacMillan, London, 1989), Vol. I, Epigraph to Chap. 37. We quote: "When you cannot measure [what you are speaking about], your knowledge is of a meager and unsatisfactory kind: . . . you have scarcely . . . advanced to the stage of science."
[16] B. R. Frieden, in *Advances in Imaging and Electron Physics,* edited by P. W. Hawkes (Academic, Orlando, 1994), Vol. 90, pp. 123–204.
[17] B. R. Frieden, Phys. Rev. A **41**, 4265 (1990).
[18] A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York, 1965).
[19] See J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1955), pp. 379–387. A more modern view is that the erasure of (computer) memory required for *registering* the measurement is the fundamental act that causes entropy to increase. See chapters by Bennett, Zurek, and Lubkin in *Maxwell's Demon, Entropy, Information and Computing,* edited by H. Leff and A. Rex (Princeton University Press, Princeton, NJ, 1990).
[20] L. Brillouin, *Science and Information Theory* (Academic, New York, 1962), pp. 168 and 232.
[21] O. Morgenstern and J. von Neumann, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ, 1947).
[22] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957).
[23] B.R. Frieden, Physica A **198**, 262 (1993).
[24] See chapters by D. Bohm, G. Ludwig, A. Shimony, E. Wigner, and others, in *Foundations of Quantum Mechanics,* edited by B. d'Espagnat (Academic, New York, 1971).
[25] See chapters by F. London and E. Bauer, E. Wigner, and others in *Quantum Theory and Measurement,* edited by J. Wheeler and W. Zurek (Princeton University Press, Princeton, NJ, 1983).
[26] B. R. Frieden and R. J. Hughes, Phys. Rev. E **49**, 2644 (1994).
[27] B. Nikolov and B. R. Frieden, Phys. Rev. E **49**, 4815 (1994).
[28] W. J. Cocke, Steward Observatory, University of Arizona (private communication).
[29] Our measurement approach appears, at first, to be at odds with one form of the Copenhagen interpretation, according to which the parameter θ does not have an objective reality, and therefore cannot be a knowable quantity. See, e.g., B. d'Espagnat, *Reality and the Physicist* (Cambridge University Press, Cambridge, England, 1989), pp. 156 and 157. The most general form of Fisher information I does depend upon θ (see Refs. [3] or [5]). However, in the particular data case $\boldsymbol{\rho}=\boldsymbol{\theta}+\mathbf{r}$ assumed here, the general Fisher form becomes Eq. (13) (see Ref. [16], pp. 135 and 136), and this *does not* depend upon $\boldsymbol{\theta}$. Then since the rest of the EPI procedure, including formation of J and of the Lagrangian $(I-J)$, depends upon this form of I, it is also independent of $\boldsymbol{\theta}$ [see Eq. (15) specifically]. Hence EPI theory *in toto* does not depend upon the objective reality of θ, and satisfies this aspect of the Copenhagen framework. (Opposed to the Copenhagen interpretation is that of Bohm, according to which $\boldsymbol{\theta}$ does have an objective reality. This certainly agrees with the aims of the smart measurer in EPI theory.)
[30] L. I. Schiff, *Quantum Mechanics* (McGraw Hill, New York, 1955), pp. 320 and 330.
[31] This is an interpretation of E. P. Wigner's views by B. d'Espagnat, *Reality and the Physicist* [29], p. 213.
[32] G. V. Vstovsky, Phys. Rev. E **51**, 975 (1995).
[33] In this regard, see E. R. Caianiello, Riv. Nuovo Cimento **15**, 1 (1992).
[34] I. B. Rumer and M. Sh. Ryvkin, *Thermodynamics, Statistical Physics and Kinetics* (Mir, Moscow, 1980), p. 526.
[35] Ref. [17], p. 4271.
[36] H.-D. Zeh, *Physical Basis of the Direction of Time* (Springer-Verlag, Berlin, 1992).