

Review

Estimating a Repeatable Statistical Law by Requiring Its Stability During Observation

B. Roy Frieden

College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA;

E-Mail: roy.frieden@optics.arizona.edu; Tel.: +1-520-621-4904; Fax: +1-520-621-3389

Academic Editor: Takuya Yamano

Received: 11 September 2015 / Accepted: 11 October 2015 / Published: 28 October 2015

Abstract: Consider a statistically-repeatable, shift-invariant system obeying an unknown probability law $p(x) \equiv q^2(x)$. Amplitude $q(x)$ defines a source effect that is to be found. We show that $q(x)$ may be found by considering the flow of Fisher information $J \rightarrow I$ from source effect to observer that occurs during macroscopic observation of the system. Such an observation is irreversible and, hence, incurs a general loss $I - J$ of the information. By requiring stability of the law $q(x)$, as well, it is found to obey a principle $I - J = \min$ of “extreme physical information” (EPI). Information I is the same functional of $q(x)$ for any shift-invariant system, and J is a functional defining a physical source effect that must be known at least approximately. The minimum of EPI implies that $I \approx J$ or received information tends to well-approximate reality. Past applications of EPI to predicting laws of statistical physics, chemistry, biology, economics and social organization are briefly described.

Keywords: Fisher information; extreme physical information; EPI; EPI derived; epistemology; coarse graining; Kullback–Leibler entropy; Plato’s cave; transition from substance to information; cancer growth

PACS classifications: 05.10.-a; 04.20.Fy; 02.30.Xx; 01.70.+w

1. Background

1.1. How Fundamental Is Information?

It is often asserted that all physically-observable effects are expressions of information; or, more precisely, all observed effects are defined by flows of information to the observer. We consider any such effect that is also repeatable statistically, *i.e.*, is consistent. Exactly how is the mathematical form of the effect defined by its observation?

1.2. What about Least Action?

Of course, historically, estimating physical effects in particular has most often been by use of the variational principle of “least action”. As found by H. von Helmholtz [1] and others, there are appropriate action forms $T - V$ (kinetic minus potential energy) for all known fields of physics. The integrand of $T - V$ is called the Lagrange function for the problem. Equating variation of the Lagrangian to zero produces the law. From this point of view, a Lagrangian principle of least action is behind all physics; and consequently, it is the most basic physical effect of all.

However, there is a distinct limitation to its use. Least action requires quantities T and V to be known as energies or energy-dependent terms. Additionally, and unfortunately, not all natural effects can be specified by their energy levels. Examples are the population dynamics of certain biological, social and cosmological systems, including their laws of allometric growth, and many statistical aspects of economics and social organization.

Aside from this practical issue, there is the basic issue of why, in the first place, physical laws need arise only from energy-dependent terms. After all, science is fundamentally the quantification of observed effects, whether of energy or not. Rather, learning through observation is the *sine qua non* of estimating the distribution. Furthermore, even if energy is to be used, why *a priori* the specific form $T - V$? Why not instead, e.g., a principle $T + V = \min$ or even $T/V = \min$? It seems that the $T - V$ form for “action” has no known prior physical significance other than being a device to derive physical laws by “reverse engineering” them; thus the widely-known statement [2] “It usually happens that the differential equations for a given phenomenon are worked out first, and only later is the Lagrange function found, from which the differential equations can be obtained”.

True, the Planck constant h , with units of energy time or momentum position, is called the “action”, but that does not provide the needed rationale for energy-dependent Lagrangians, and of the special form $T - V$.

No less a scientist than Schrodinger, e.g., called it “incomprehensible” [3] why the principle $T - V = \min$ works to derive his wave equation.

1.3. Overall Aim

Consider the epistemological question: How much can be learned about a scientific phenomenon by observing it one or, at most, a finite number of times? In particular, can it be so derived? Again, this is without necessarily using knowledge of its energy aspects T or V (note: however, in some applications

the information I and J turn out to be proportional to T and V). At this point, it is useful to briefly summarize the derivation to follow. The main development follows this synopsis.

1.4. Synopsis of Derivation

The source effect is statistical, obeying an unknown probability law $p(x)$ in fluctuations x , where $p = q^2$, $q = q(x)$, the system amplitude law. The aim is to estimate $q(x)$. The law $q(x)$ is assumed to be statistically repeatable, differentiable and, for simplicity, shift invariant. We require that any such law obeys stability to first-order perturbation of its solution $q(x)$. The effect is observed in the ordinary sense, *i.e.*, macroscopically. Such an observation constitutes “coarse graining” (discussed in Section 2.5, and in Section 3 at its beginning and end), whereby K-L (Kullback–Leibler) entropy is lost. Therefore, since K-L entropy is proportional to Fisher I (see Section 2.4 below), this is lost, as well. Then, calling the maximum possible value of the source effect J , the amount lost is $I - J \equiv \Delta I$. The J value is regarded as intrinsic to the source during its observation (e.g., in space, time, *etc.*) and, so, is called the source information. The above relation $I - J \equiv \Delta I$ holds for losses ΔI down to infinitesimal values, where $\Delta I = dI$. Then, the amount lost is $I - J = dI$. Then, any variation $\delta(I - J) = \delta(dI) = 0$. This implies that $I - J = \text{extremum}$. Then, its solution $q(x)$ obeys first-order stability, as previously required. Finally, it is shown that, owing to the special forms obeyed by functionals I and J , the solution causes the extreme value of the loss $I - J$ to be a minimum. The result is the EPI principle:

$$I - J = \min \quad (1)$$

We return to the general development.

2. An Information-Based Alternative

Is there, then, an alternative variational approach to least action that does have prior significance and, in particular, does not necessarily require knowledge of energies? In fact, we find that generally disregarding energy considerations and, instead, taking an epistemological viewpoint results in the principle of extreme physical information (EPI).

It is shown below that any currently known physical system (or effect) has a well-defined value J for its maximum Fisher I value. The J value is regarded as intrinsic to the source and, so, is called the source information. This is found to hold, e.g., for all shift-invariant phenomena that ultimately have a physical basis. Principle (1) has been applied in [4–10] to derive wave equations of quantum mechanics, basic thermodynamics and modern cell biology. Also derived have been the phenomena of statistical physics and other sciences [6,11–19]. This includes the laws of cancer growth [6,13], the near-ubiquitous occurrence of statistical power laws [19] in science, including the quarter-power laws of biological and cosmological allometry [6,19], the de Broglie wave hypothesis (now, no longer a mere hypothesis) [17], thermodynamics using Fisher information in place of entropy [6,7], the Euler equation [14] of the chemical density functional and the laws of optimum economic investment [11] and of population dynamics [6,8,18] of animate or inanimate systems.

These derivations are empirical evidence for the validity of EPI, but do not constitute its derivation. Our aim here is to show how EPI naturally arises out of a quest for knowledge.

2.1. Data Information of Fisher

The approach arises from classical information theory, invented largely by R.A. Fisher [20,21] circa 1920 (note: this is not C.E. Shannon's information, devised for other purposes, in about 1945). Let a be an unknown parameter of a shift-invariant system obeying consistent, *i.e.*, repeatable, statistics. It is desired to estimate a from its measurement y . Consider, e.g., any such physical system characterized by a particle in motion whose Ehrenfest-mean position, or momentum, or energy, *etc.*, a is unknown. This is measured, as a value $y = a + x$, with x a fluctuation characteristic of the measured system. For simplicity of analysis, the detector is assumed to be perfect, not contributing extra fluctuation to the measurement. Then, fluctuations x describe the statistics of the system.

2.2. Shift Invariance, Use of Amplitudes $q(x)$

For simplicity, assume a single measurement y of parameter a in a shift-invariant system. Thus, the likelihood law obeys $p_Y(y|a) \equiv p(x)$, $x \equiv y - a$, on random fluctuations x from parameter a . In general, an estimate of a is formed as some chosen function of the datum y . A possible choice of the estimate is y itself.

Let the law $p(x)$ have an amplitude $q = \sqrt{p}$, $q = q(x)$. Amplitude function $q(x)$ is any statistically-repeatable law on fluctuations x from unknown parameter a . These fluctuations are assumed to define the observed effect, *i.e.*, its physics is defined by the fluctuations x from the ideal measurement value $y = a$. This directly expresses the epistemological nature of the approach.

More generally, the system can be in some fixed *complex* amplitude state $\psi(x)$, usually denoted as $\psi_n(x)$ in quantum mechanical problems, where n is an integer, its eigenstate. In the analysis, any value of n may be present. However, the system is, for now, not assumed to have a quantum nature. Moreover, the value of n is not central to the analysis, since the purpose of the measurement is to know a , not n . Thus, for simplicity, n is suppressed, and $\psi_n(x)$ is represented as a real function $\psi_n(x) \equiv q(x)$.

2.3. Definition of the Information

The concept of the Fisher information about the parameter a contained in a single observation is utilized, as below, to understand how accurately a can be known from that measurement. The information is most conveniently represented in terms of amplitude function $q(x)$, rather than $p(x)$, as: [6,20–22]

$$I = 4 \int dx (dq/dx)^2. \quad (2)$$

This assumes shift invariance, as discussed at the outset of Section 1.4. Furthermore, the integration is taken over the entire range of the random variable x . The Fisher I defines how well a can be known in the rms sense (see Equation (16)). The rms error turns out to be minimal for a class of probability laws $p(x)$ (see just above and below Equation (16)).

2.4. Local Nature, Relation to Kullback–Leibler Measure

This information measure (2) is called “local”, since it is sensitive to the fluctuations dq in differentially-neighboring (distance dx) values of $q(x)$. The sensitivity is particularly strong, since it

goes as the square of the fluctuations. Hence, I has long been considered a “measure of roughness” of the data [23]. It results as well that $I \propto H_{KL}(p(x)/p(x+dx))$ [6,24], the latter the Kullback–Leibler (K-L) cross entropy “distance” between $p(x)$ and its differentially-displaced neighbor function $p(x+dx)$.

Implicit to Equation (2) is that quantities a, x, y are scalar numbers. However, more generally, they may be multidimensional. In such cases, the expression (2) for I is easily generalized by the usual replacements $x \rightarrow \mathbf{x}$, $dx \rightarrow d\mathbf{x}$ in the integrand for a vector $\mathbf{x} = x_1, \dots, x_N$ of data fluctuations. Furthermore, the integrand $(dq/dx)^2$ is replaced [6,12] by the sum $\sum_n (\nabla q_n)^2$, where ∇ is the gradient operator, for a system that can be in many different states $n = 1, \dots, N$. Finally, as defined in Equation (12), in terms of complex amplitudes ψ_n , $n = 1, \dots, N$, the information becomes $I = 4N \int d\mathbf{x} \sum_n \nabla \psi_n^* \cdot \nabla \psi_n$ in problems where the \mathbf{x} -space is continuous. As usual, $*$ denotes the complex conjugate. This is called the “Fisher channel capacity” [6], since it is an upper bound to the actual Fisher information [6,12]. However, to keep the analysis simple, we first merely assume the one-dimensional single-state case (2) of the information.

By K. Popper’s criterion of negation, any effect that is claimed to be physical (as opposed to, say, metaphysical) must make predictions that can be falsified. To falsify the effect requires, at the very least, its accurate observation in a well-defined state a . How large can information I be obtained by a single observation or, more simply, how much can one know?

2.5. Plato’s Parable of the Cave

The ancient Greeks gave much thought to how much one can know. Plato’s parable of the cave is meant to illustrate this. In Plato’s parable, people who have spent their entire lives chained inside a cave are allowed to obtain information about events outside only by viewing the shadows they cast on a wall of the cave. Of course, their view of reality outside is highly limited and imperfect, since, for one thing, it consists of but two-dimensional projections of the three-dimensional objects outside. Thus, when a group of people arrive outside, they are perceived inside as but two-dimensional shadows. This perception is intrinsically imperfect. In modern terms, Plato’s conclusion is that the information I acquired by the cave people inside is generally less than that, called level J , that would perfectly specify the outside source. We will see in Section 3 that this is actually an example of the known phenomenon of coarse graining. Plato had the right idea.

2.6. Generic Nature of I

Assuming flat space, I has the same integral form (2) for all such observed effects. It is a generic measure of the degree of randomness in data. Note from the form of (2) that the more rapidly $q(x)$ oscillates, the larger will be I and the more random will be the data it specifies. The generic nature of I is further manifest in its property $I = \max$. (by Equations (6) and (11) below), which holds for all statistically-repeatable phenomena $q(x)$.

What is not generic is the source information J defining the particular system effect. In Plato’s parable above, it has the role of defining the ideal level of information, e.g., the level required to fully specify the “source people” outside the cave. Information J of EPI is further taken up in Sections 3.2 and 3.3 below.

Since the 1920s, the integral (2) for I has been mainly used to define how well the system parameter a can be estimated (see below). This is mere use as a diagnostic. However, our ultimate aim is not to merely estimate the parameter a , but, also, to estimate the system: *i.e.*, the functional form of either $q(x)$ or $p(x)$.

The integral I will likewise prove indispensable for this purpose. At this point, we emphasize that the above integral form (2) for I is, in particular, not generally dependent on prior knowledge of system energy. Thus, neither will formation of the estimate of $q(x)$ or $p(x)$. This is in comparison with the like use of the principle of least action, which always requires knowledge of system energies and, so, rules out its use in scenarios where energy information is not present (as discussed later).

3. Observation Is a Generally Lossy Process

Central to our derivation of EPI is the following effect. Consider the measurement of a state parameter a , say a particle mass. The “Platonic ideal” (illustrated by his parable above) is that, generally speaking, perfection exists, but that man seldom observes it. In fact, according to modern measurement theory, any macroscopic measurement of a parameter is intrinsically lossy, *i.e.*, irreversible in nature. *“This is because it is impossible in principle that an observer inside the universe can reverse the universe. The real universe as observed by objects inside is irreversible, and measurements are permanent records for these observers.”* [25]. In order to produce a definite reading of the state a of the system, the detector in use interacts with the system. It is this very interaction, causing an irreversible exchange of information and energy with those of the system, that gives rise to the output measurement. Its irreversible nature amounts to a lossy, or “coarse-grained” [6,26–28], process. Assume such a measurement to be made of a system with ideal Fisher information level J .

This, by definition, causes the observed data to suffer a loss of K-L entropy and, hence, by the relation $I \propto H_{KL}$ previously found in Section 2.4, a random loss of Fisher information:

$$\Delta I \leq 0 \quad (3)$$

from its source value J . Thus, with generally:

$$I = J + \Delta I, \quad (4)$$

by Equation (3)

$$I \leq J. \quad (5)$$

A macroscopic observation cannot give more information than pre-exists at the source.

The unknown amplitude $q(x)$ and the measurement process are assumed to be statistically repeatable. This implies that the given source information, value J is the upper bound to the acquired information I that is acquired even in repeated measurements y of the system. In summary, the observed effect has maximum information of value:

$$I_{\max} = J. \quad (6)$$

(Note: If such repeatability does not hold, then there will be a different J value at each observation and, resultingly, a different EPI solution $q(x)$. However, it will still be valid for that observation.)

Given the loss (3) of information, the data can no longer perfectly define parameter a . Thus, Inequality (5) proves, in effect, a modern version of Plato's parable: the observers in the cave can only see imperfect shadows of the true reality (here, a) outside.

Coarse graining also demarks the transition from a quantum to classical universe [29,30]. This corresponds to the transition from efficiency $\kappa = 1$ to $\kappa = \frac{1}{2}$ as discussed in Sections 4.1.1 and 4.1.2.

3.1. Physics as Transition from Substance to Observation

In contrast with the intrinsic or source information J , the observer collects a level of information I about a in data collected from the system. He/she is at the receiving end of a basic flow of information:

$$J \rightarrow I, \text{ or } \textit{source} \rightarrow \textit{observation} \quad (7)$$

over a channel. As an example, such information flow occurs during its transfer (say, by photons in a microscope) from a physical source (in the microscope slide) at information level J and in parameter state a , to the observed level I in data space (on the observer's retina). Information level J is defined by the physics of the specimen on the slide, e.g., an amoeba.

The "source", with physical substance, defines the ideal (maximum, as shown in Equation (6)) possible level J of Fisher information I . This information is taken to perfectly define it. By contrast, the observation is the generally imperfect (by Equation (5)) measurement of this physical substance. The direction of the arrows in Equation (7) defines the direction of increasing time, indicating that the physical source (existence or "being") gives rise to ("becomes") the observation, however imperfect.

3.2. Assumed Form for Information J

The flow (7) is of information from source effect to observer. By comparison, information J is taken to result from another flow, this one of some physical quantity internal to the source system (examples in Section 3.3). This physical flow obeys a general form:

$$J \equiv \int dx j[q(x), s(x)]. \quad (8)$$

Here, j is some continuous function of its arguments $q(x), s(x)$. In particular, $s(x)$ defines the physical source flow, as exemplified above (in particular, not necessarily energy). Note also that J does not depend explicitly on the gradient $q'(x)$. Such explicit dependence is reserved for information I , as defined in the generic form (2). If J depended on $q'(x)$, there would no longer be a definite distinction between "source" and "observation" in the information flow Equation (7).

3.3. Examples of Source Information J

In providing a complete description of system state a , information J is intrinsic to the source system. Therefore, J must be expressible in terms of its defining physical properties. For example, in describing: (1) quantum observation of the position of a particle [6,12], J is proportional to the square of the particle's mass; (2) cell growth, J increases with reproductive fitness [6,8,18]; (3) the growth of investment capital in econophysics, J increases as the expected value of the production function [11];

(4) cancer growth, J increases with cancer mass [6,13]; (5) the growth of competing populations, J is proportional to the mean-squared fitness over the populations [6,18]. Notice that in applications (2)–(5), there is no explicit dependence on energy, kinetic or potential, or on their difference, the “action”.

However, although the functional J always has the form Equation (8), there is presently no simple rule for forming it in new scenarios. Thus, the particular functions $s(x)$ and $j(q, s)$ are not known *a priori*. However, there is a systematic approach to finding the functionals. This is the Peirce approach outlined in Section 5.2.

3.4. Need for Stability

Laws of nature $q(x)$ should tend to be invariant, e.g., not varying from place to place in the universe. On this basis, the estimate of $q(x)$ should obey zero variation in some function \mathcal{F} of the two information characteristics I, J of the channel (7) $J \rightarrow I$. That is, $q(x)$ should obey a variational principle $\delta\mathcal{F}(I, J) = 0$. What function $\mathcal{F}(I, J)$ should have this invariance? Equation (4) holds down to infinitesimal values, where $\Delta I = dI$. Then, (4) becomes of the form $I - J = dI$. Then, the variation $\delta(I - J) = \delta(dI) = 0$. This is of the required variational form $\delta\mathcal{F}(I, J) = 0$ with $\mathcal{F}(I, J) = I - J$. Thus, it is the simple loss of information that should be stable.

However, an extremum can be a maximum, minimum or point of inflection. Which one is found next.

4. Derivation of EPI

We are now in a position to derive the EPI principle $I - J = \min$. We had from the preceding:

$$\delta(I - J) = 0. \quad (9)$$

This serves to stabilize the information loss $I - J$ and, hence, the EPI solution $q(x)$ *per se*. It also implies a constrained extremum problem $I - J = \text{extremum}$. Here, the constraint is the knowledge that the source information obeys Equation (8). Denote the integrand of $I - J$ as the Lagrangian \mathcal{L} . Then, by Equations (2) and (8),

$$\mathcal{L} = 4q'^2(x) - j[q(x), s(x)]. \quad (10)$$

(Note that the usual Lagrange multiplier λ is absorbed into j , so that the usual second requirement $\partial\mathcal{L}/\partial\lambda = 0$ of the constrained Lagrange approach is not enforced.) In particular, the extremum in principle $I - J = \text{extremum}$ must be shown to obey form (1), i.e., be a minimum. The nature of the extremum is determined by the use of Legendre’s condition, as follows.

Legendre’s condition states that the extremum is a minimum if $\partial^2\mathcal{L}/\partial q'^2 > 0$ (or a maximum if < 0). Notice that only the first right-hand term in Equation (10) depends on q' . Then, directly, differentiating it twice gives $\partial^2\mathcal{L}/\partial q'^2 = +8 > 0$. Therefore, the extremum is a minimum, giving the EPI principle $I - J = \min$.

Furthermore, by Equation (5), it must be that $I = \kappa J$ with $0 \leq \kappa \leq 1$. Therefore, in summary,

$$I - J = \min., \text{ where } I = \kappa J, 0 \leq \kappa \leq 1. \quad (11)$$

This two-faceted principle is called that of extreme physical information or EPI (as in “EPI stemology”) QED.

4.1. Efficiency Constant κ

Coefficient $\kappa = I/J$ measures the efficiency of the information transfer to the observer about system parameter a . That the transfer generally loses information agrees with Plato's thesis. This loss simply followed, in Equations (3)–(5), from the irreversible nature of making a measurement. However, that the loss is a minimum is a surprisingly optimistic result. As we saw, this followed because of the fundamentally different structure of data information functional I from source functional J .

Of course, for the purposes of gaining knowledge, it would be best if the received information obeyed $I = J$, i.e., $\kappa = 1$. This obeys the equality sign in condition (5), and represents the highest allowed value for I . As shown below, this holds when observing quantum systems. Furthermore, value $\kappa = 1/2$ occurs when observing classical systems. These are shown next.

We first specialize to the case of a one-dimensional quantum system. As discussed at the beginning of Section 2.4, it generalizes for an N -state, complex system ψ_n , $n = 1, \dots, N/2$ to:

$$I = 4N \sum_{n=1}^{N/2} \int dx |d\psi_n/dx|^2 \quad (12)$$

and is called the “Fisher channel capacity”. The real and imaginary parts of each complex amplitude $\psi_n(x)$ are formed out of successive real amplitudes q_{2n-1}, q_{2n} by the rule $\psi_1 \propto q_1 + iq_2$, $\psi_2 \propto q_3 + iq_4$, etc. Specifically,

$$\psi_n(x) \equiv N^{-1/2}(q_{2n-1} + iq_{2n}), i = \sqrt{-1}, n = 1, \dots, N/2. \quad (13)$$

Using Definition (13) in Equation (12) gives:

$$I = 4 \sum_{n=1}^N \int dx (dq_n/dx)^2. \quad (14)$$

This is an obvious generalization of Definition (2), which assumed only the $N = 1$ state to be present. Here, the N information contributions $(dq_n/dx)^2$ are independent and, hence, add by the rules of elementary probability theory.

4.1.1. Why Quantum Cases Have $\kappa = 1$

Either Form (12) or (14) is a sum of squares defining a so-called L_2 length. In quantum phenomena, the transition $J \rightarrow I$ in (7) is via a unitary transformation (in particular, the Fourier one) [6,12] from momentum-energy space to position-time space $\mathbf{x} = (x, y, z, ct)$ space. A unitary transformation preserves an L_2 length, here information J . Hence, the transformation (7) $J \rightarrow I$ of information obeys $I = J$, or $\kappa = 1$. Furthermore, here, the quantity $j \equiv j[\psi(\mathbf{x}), s(\mathbf{x})]$ defining ideal information J via Equation (8) obeys $j = 4N(mc/\hbar)^2$, where m is the particle mass, N is the rank (e.g., $N = 4$ for a nonrelativistic case) of its wave functions ψ_n , $n = 1, \dots, N/2$ and c and \hbar are the usual physical constants. Then, by $E = mc^2$, quantum information $I = J \propto E^2$, the square of the energy. This indicates a new physical property for energy E in the relativistic quantum scenario: both the source and data levels of Fisher information exist interchangeably with the squared quantum energy; or “knowledge” and energy are interchangeable, analogous with the familiar relation between matter and energy in relativity.

4.1.2. Why Classical Cases Have $\kappa = 1/2$

We found in the preceding that $I = J$ in the general quantum scenario. Equation (14) shows that N terms q_n contribute to the value of I .

Next, inverting Equation (13) gives, respectively odd and even indexed q_n in terms of corresponding ψ_n values, as:

$$q_{2n-1} = \sqrt{N} \bmod[\psi_n(x)] \cos \varphi_n(x), \text{ and } q_{2n} = \sqrt{N} \bmod[\psi_n(x)] \sin \varphi_n(x). \quad (15)$$

The $\varphi_n(x)$ are phase functions. We now evaluate Equation (15) as the classical limit is taken. In this limit, the eigenfunctions $\psi_n(x)$ are so crowded together, that each phase function $\varphi_n(x) = 0$ continuously over all x . Then, each $\sin \varphi_n(x) = 0$. Then, by Equation (15), every even component $q_{2n}(x) = 0$.

How do these special values effect the sum (14) for I ? By the preceding, none of its even-indexed amplitudes q_{2n} contribute to that sum. Only their odd-indexed mates q_{2n-1} do. Then, of the total N possible contributions to the sum (14) for I , only $N/2$ contribute. This has a useful consequence.

Then, in particular, I must also lack $N/2$ terms in the sum (14) defining its maximum value J . Then, with $I = \kappa J$, efficiency constant $\kappa = 1/2$.

In fact, by direct applications [6] of EPI, this is found to exactly hold for systems obeying either classical mechanics or classical electromagnetics.

5. Discussion

5.1. In What Sense Is “Everything” Information?

We return to the issue raised at the outset of the paper, that everything we see is effectively information. This paper describes the extent to which Fisher information fills this role. Principle (11) establishes that all demonstrably known physical effects arise out of classical observation, *i.e.*, a transition (7) of information-based substance, or ideal being, into becoming information-based data. This is a model for quantifying the role of observation, *i.e.*, epistemology. Observations serve to transform one form of information (of physical form J) into another (Fisher form I). The observed universe runs on the transformation.

In particular, the EPI principle (11) defines a practical variational approach to finding $q(x)$, the natural law governing the data. The minimum is found by use of the Euler–Lagrange solution [6] to the problem.

Note that EPI actually consists of two conditions (11). The second, $I = \kappa J$, has no counterpart in the familiar least action approach (which is mathematically analogous to the first condition (11)). Actually, the two conditions (11) give practical advantage over least action in cases $\kappa < 1$ (these are important, since, as mentioned, cases $\kappa = 1/2$ define all classical physics [6]). There, it allows unknown law $q(x)$ to be found as the simultaneous solution [6,13] to the two conditions (11). This describes the scenario of deductive knowledge briefly described in Section 5.2 below.

In comparison with deriving classical physics, in deriving quantum wave equations [6,12] the EPI solution obeys $\kappa = 1$, or perfect transfer of information from source to observer (as mentioned, this ignores possible noise fluctuation from the detector, assumed perfect). This assumes unitary evolution of

the system; such evolution allows the source information to be detected without degradation. This also defines a scenario of abductive knowledge described in Section 5.2 below.

A second benefit of the EPI answer is to provide an optimum estimate of the unknown state parameter a . The logic here is that the parameter a will be best estimated if, first of all, the system $q(x)$ “behind it” is best known, as well. Nature evidently allows realization of these benefits: the EPI principle works [4,6,11–19]; and the best estimate (in the sense of absolute minimum mean-square error) of a is often achievable [5,6,12,22]. This is for systems $p(x)$ that are called “efficient”, e.g., those in the exponential family. How good is the best estimate of a ?

An answer is provided by the Cramer–Rao inequality [5,6,22]: The mean-square error e^2 in determining parameter a from the datum y obeys:

$$e^2 \geq 1/I. \quad (16)$$

This assumes unbiased estimation, analogous to an unbiased experiment (generally finite error in each individual estimate, but an average error of zero over an infinity of them). The equality in Equation (16) holds for an efficient system (defined above). However, since J is generally finite, even in the ideal scenario $I = J$ of no information loss I is finite. Then, by the (now) equality (16), there is still finite error e in the estimate. The efficient estimate is not perfect. The Platonic ideal holds here on the level of information, *i.e.*, $I = J$, but not for the estimate of the parameter. However, even so, by (11), since $\kappa = 1$, information I is maximized, so that by (16) the best (smallest possible) error e is minimized. Thus, reality cannot be perfectly known, but it can be known with minimal error.

Finally, the EPI approach naturally provides the so-called “quantum potential” used by Madelung, and later D. Bohm, to derive the Schrodinger wave equation. This is discussed in depth in [31].

5.2. Forming Functional J : Three Stages of Prior Knowledge

In real applications of EPI, its outputs $q(x)$ or $\psi_n(\mathbf{x})$ are correct to varying degrees. Where does error creep in? Functional I is always of the form (2) or (12), depending on dimensionality, so this is not a source of error. By comparison, in any new problem, the challenge is in forming the source information functional J . This is a form of prior knowledge of the unknown effect. The accuracy of the EPI prediction $q(x)$ or $\psi_n(\mathbf{x})$ depends directly on the accuracy of this prior knowledge. Degrees of accuracy called, in descending order, abductive, deductive and inductive, were defined by the philosopher-statistician C.S. Peirce [32].

For example, if the system obeys unitarity (Section 5.1) efficiency $\kappa = 1$, no information is lost, and the outputs, the equations of quantum mechanics, are 100% correct. This form of prior knowledge is called “abductive”.

“Deductive” knowledge of J occurs when the prior knowledge is of a simple invariance, such as an equation of continuity of flow [6]. This gives an EPI output that is accurate over only a limited range of parameters, such as in the nonrelativistic limit.

Finally, “inductive” knowledge gives outputs that follow from constraining extremization of I , such that the outputs $q(x)$ merely obey known empirical averages, such as first and second moments. See [33,34] for details. Here, functional J expresses the averages as ordinary Lagrange constraints. This gives an EPI output that is merely constrained to be smooth (*i.e.*, “unbiased” statistically).

An alternative approach with strong promise of finding useful functionals J falls under the rubric of machine learning [35–37]. There is a large literature on this overall approach. Its general aim is to construct algorithms that can learn from massive amounts of data to the extent that reliably predictive models, say functionals F, G , can be formed. In analogy with our problem (11), these can be combined in the specific combination $F - G = \text{minimum}$. Functionals F, G would be established by processing all of the data. If F turns out to be proportional to I , then G can be identified as proportional to J . Thus, the EPI approach might eventually find its strongest application in this role, which accomplishes a strong decrease in the need for the analytic modeling of massive amounts of data.

In summary, EPI does not depend on knowledge of system energy for its use. Hence, it applies to a much wider scope of phenomena than does least action. These include economic, biological and social phenomena whose energy effects are either unknown or inapplicable. The need for such a more widely-applicable approach has long been sought [38–42]. In this respect, the approach has a shared goal with that of machine learning, and we expect this to ultimately allow source functional J to be found by the use of massive amounts of data.

6. Conclusions

Why does observation tend to agree with reality? We show that the macroscopic observation of any well-defined, unknown statistical effect displays maximum Fisher information about the effect. Of course, as a limitation, because the observation is irreversible, the acquired level I of information cannot exceed the level J of the source effect. Nevertheless the loss is minimal, $I - J = \text{min}$. Thus, the “physical” as versus “observed” aspects of any macroscopically observed reality tend to agree. This is comforting, but far from a trivial result: it states that although man cannot observe reality perfectly, he can in fact observe it with minimal error. This has allowed man to increasingly understand his universe. This is especially apt for quantum effects, whose observation incurs no information loss, *i.e.*, $I - J = 0$. By comparison, when observing classical effects such as mechanics or electromagnetism, where phase information is lacking, half the initial level J of information is missing from the acquired information I , that is, the acquired information obeys $I = (1/2)J$.

The second noteworthy aspect of the principle $I - J = \text{min}$. arises from the fact that I and J are often both known functionals of the probability amplitudes governing the observed effect. Then the principle becomes a variational problem which may be solved, via a straightforward Euler–Lagrange equation, for the probability amplitudes. In this way the wave equations of quantum mechanics, electromagnetism, and many other effects of physics, chemistry, biology, economics and social organization have been derived. Among new notable results are the law of growth of in situ cancer growth, and derivation of what has heretofore been considered the de Broglie wave hypothesis: the wave nature of matter. As we see, this ultimately follows mainly from the irreversible nature of observation.

Acknowledgments

The author acknowledges Pablo Zegers for suggesting the above uses of machine learning in application to quantification of the massive amounts of astronomical data currently being made available by modern observatories in the Andes mountains.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Cahan, D. *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science*; University of California Press: Berkeley, CA, USA, 1993.
2. Morse, P.M.; Feshbach, H. *Methods of Theoretical Physics, Part I*; McGraw-Hill: NY, USA, 1953; p. 278.
3. Schrodinger, E. Quantization as a problem of proper values (Part 1). *Annalen der Physik* **1926**, *79*, 361–376.
4. Frieden, B.R. Fisher information as the basis for the Schrodinger wave equation. *Am. J. Phys.* **1989**, *57*, 1004–1008.
5. Frieden, B.R. Fisher information and uncertainty complementarity. *Phys. Lett. A* **1992**, *169*, 123–130.
6. Frieden, B.R. *Science from Fisher Information*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.
7. Plastino, A.; Plastino, A.R. Information and thermal physics. In *Exploratory Data Analysis Using Fisher Information*; Frieden, B.R., Gatenby, R.A., Eds.; Springer: London, UK, 2007; pp. 119–154.
8. Frank, S.A. Natural selection maximizes Fisher information. *J. Evol. Biol.* **2009**, *22*, 231–244.
9. Frieden, B.R.; Gatenby, R.A. Cell development obeys maximum Fisher information. *Front. Biosci.* **2013**, *5*, 1017–1032.
10. Gatenby, R.A.; Frieden, B.R. Coulomb Interactions between Cytoplasmic Electric Fields and Phosphorylated Messenger Proteins Optimize Information Flow in Cells. *PLoS One* **2010**, *5*, doi:10.1371/journal.pone.0012084.
11. Hawkins, R.J.; Aoki, M.; Frieden, B.R. Asymmetric information and macroeconomic dynamics. *Physica A* **2010**, *389*, 3565–3571.
12. Frieden, B.R.; Soffer, B.H. Lagrangians of physics, and the game of Fisher-information transfer. *Phys. Rev. E* **1995**, *52*, 2274–2286.
13. Gatenby, R.A.; Frieden, B.R. Application of information theory and extreme physical information to carcinogenesis. *Cancer Res.* **2002**, *62*, 3675–3684.
14. Nagy, Á. Spin virial theorem in the time-dependent density-functional theory. *J. Chem. Phys.* **2003**, *119*, 9401–9405.
15. Frieden, B.R.; Gatenby, R.A. Power laws of complex systems from extreme physical information. *Phys. Rev. E* **2005**, *72*, 1–10.
16. Frieden, B.R.; Soffer, B.H. Information-theoretic significance of the Wigner distribution. *Phys. Rev. A* **2006**, *74*, 1–8.
17. Frieden, B.R.; Soffer, B.H.; De Broglie's wave hypothesis from Fisher information. *Physica A* **2009**, *388*, 1315–1330.
18. Frieden, B.R.; Plastino, A.; Soffer, B.H. Population genetics from an information perspective. *J. Theor. Biol.* **2001**, *208*, 49–64.

19. Frieden, B.R.; Soffer, B.H. Parallel information phenomena of biology and astrophysics. In *Exploratory Data Analysis Using Fisher Information*; Frieden, B.R., Gatenby, R.A., Eds.; Springer: London, UK, 2007; pp. 155–172.
20. Fisher, R.A. On the Mathematical Foundations of Theoretical Statistics. *Phil. Trans. R. Soc. Lond. A* **1922**, 222, 309–368.
21. Fisher, R.A. *Statistical Methods and Scientific Inference*; Oliver and Boyd: Edinburgh, UK, 1956.
22. Van Trees, H.L. *Detection, Estimation and Modulation Theory, Part I*; Wiley: New York, NY, USA, 1968.
23. Good, I.J. A nonparametric roughness penalty for probability densities. *Nature* **1971**, 229, 29–30.
24. Savage, L.J. *Foundations of Statistics*; Dover: Englewood Cliffs, NJ, USA, 1972; p. 236.
25. Neumaier, A. What constitutes an observation/measurement in QM? 2012. Available online: <http://physics.stackexchange.com/questions/43406/what-constitutes-an-observation-measurement-in-qm> (accessed on 22 October 2015).
26. Čencov, N.N. *Statistical Decision Rules and Optimal Inferences*; American Mathematical Society: Providence, RI, USA, 1982.
27. Ohya, M.; Petz, D. *Quantum Entropy and its Use*; Springer: Berlin, Germany, 2004.
28. Frieden, B.R.; Hawkins, R.J. Quantifying system order for full and partial coarse graining. *Phys. Rev. E* **2010**, 82, 1–8.
29. Luo, S.-L. Logarithm versus square root: Comparing quantum Fisher information. *Commun. Theor. Phys.* **2007**, 47, 597–600.
30. Gibilisco, P.; Hiai, F.; Petz, D. Quantum covariance, quantum Fisher information and the uncertainty relations. *IEEE Trans. Inform. Theory* **2009**, 55, 439–443.
31. Carroll, R. *On the Quantum Potential*; Arima: Suffolk, UK, 2007.
32. Santaella, L. The Development of Peirce's Three Types of Reasoning: Abduction, Deduction, and Induction. In Proceedings of the 6th Congress of the IASS, Guadalajara, Mexico, 13–19 July, 1999.
33. Frieden, B.R. Introduction to Fisher Information: Its Origin, Uses and Predictions. In *Exploratory Data Analysis Using Fisher Information*; Frieden, B.R., Gatenby, R.A., Eds.; Springer: London, UK, 2007.
34. Flego, S.P.; Plastino, A.R.; Plastino, A. Direct Fisher inference of the quartic oscillator's eigenvalues. *J. Modern Phys.* **2011**, 2, 1390–1396.
35. Zegers, P. Some New Results on the Architecture, Training Process, and Estimation Error Bounds for Learning Machines. Ph.D. Thesis, The University of Arizona, Tucson, AZ, USA, 2002.
36. Zegers, P. Fisher information properties. *Entropy* **2015**, 17, 4918–4939.
37. Amari, S.I. Natural gradient works efficiently in learning. *Neural Comput.* **1998**, 10, 251–276.
38. Von Bertalanffy, L. *General Systems Theory*; George Braziller Inc.: New York, NY, USA, 1969.
39. Crow, J.F.; Kimura, M. *An Introduction to Population Genetics*; Burgess Publishing: Minneapolis, MN, USA, 1970.
40. Hayes, W. *Max Ludwig Henning Delbruck*; National Academy Press: Washington, DC, USA, 1993.
41. Yolles, M.I.; Frieden, B.R. A metahistorical information theory of social change: The theory. *J. Organ. Transf. Soc. Chang.* **2005**, 2, 103–136.

42. Yolles, M.I.; Frieden, B.R. A metahistorical information theory of social change: An application. *J. Organ. Transf. Soc. Chang.* **2005**, *2*, 137–151.

© 2015 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).