PHYSICAL REVIEW E **82**, 066117 (2010)

Quantifying system order for full and partial coarse graining

B. Roy Frieden and Raymond J. Hawkins
College of Optical Sciences, University of Arizona, Tucson, Arizona 85721, USA
(Received 19 October 2010; published 23 December 2010)

We show that Fisher information I and its weighted versions effectively measure the order R of a large class of shift-invariant physical systems. This result follows from the assumption that R decreases under small perturbations caused by a coarse graining of the system. The form found for R is generally unitless, which allows the order for different phenomena to be compared objectively. The monotonic contraction properties of R and I in time imply that they are entropies, in addition to their usual status as information. This removes the need for data, and therefore an observer, in physical derivations based upon their use. Thus, this recognizes complementary scenarios to the participatory observer of Wheeler, where (now) physical phenomena can occur in the absence of an observer. Simple applications of the new order measure R are discussed.

DOI: 10.1103/PhysRevE.82.066117 PACS number(s): 89.90.+n, 05.20.−y, 05.40.−a, 05.90.+m

I. MOTIVATION

The concept of *disorder*, as measured by the Shannon or Boltzmann entropy H of a system, has a long and important history. In this paper, we consider a complementary concept—the level of *order*, or complexity, in the system. What, quantitatively, is system order? Numerous authors have considered this question to be equivalent to the statement of the second law of thermodynamics, that *dis*order must increase. On this basis order must decrease, where order is in some sense the opposite of disorder. See, for example [1]. Important properties of order that arise in specific applications are often mentioned, such as its spontaneous, coherent, or statistical natures [2], but the term itself is never quantified mathematically. It is, for example, inadequate to simply define order as the "lack of disorder." The word "lack" has ambiguous meaning. As examples, with disorder measured as usual by the entropy H, is the order to be defined as $C-H$, with C some constant, C/H, $\exp(-CH)$ or a cross-entropy? For all these *ad hoc* constructions a small disorder (here H) gives a large order; or a large H gives a small order; as required. That a multiplicity of possible expressions for order can be so easily formed suggests the need for a quantification of the matter. This paper carries through on the task, deriving the mathematical form for order from first principles.

It is common to regard *Fisher information I* as the order in some sense. Our particular motivation is the many recent derivations of physical laws using the Fisher measure [3–5]. Indeed, the famous Čencov inequality [6–9], as taken up below, defines a Fisher metric of order. Hence, one suspects that, whatever the measure of order may be, it should increase with Fisher I. Indeed it is found later to increase linearly with I.

II. DEFINING CONCEPTS

Consider a one-dimensional system of extension $(b-a), b>a$, whose general coordinate x (a length, or energy, etc.) obeys $a \le x \le b$. The system is defined by its *known*, discrete probabilities $p(x_i) \equiv P(x_i \le \zeta \le x_i+\Delta x)$, where $x_i = a+(i-1)\Delta x$ and $\Delta x = (b-a)/(N-1)$, $i=1,\ldots,N$, N large. Each interval $(x_i, x_i+\Delta x)$ defines a *pixel* of the system. With N large, Δx is sufficiently small to be regarded as a differential dx. It follows that the $p(x_i)$ are very small. For the purpose of conveniently evaluating sums as integrals, we later work with a probability *density* function (PDF)

$$\rho(x_i) \equiv p(x_i)/\Delta x, \tag{1}$$

and, in the limit, we have

$$\Delta x \to dx, \quad x_i \to x,$$

$$p(x_i) \to p(x), \quad \rho(x_i) \to \rho(x), \tag{2}$$

for $a \le x \le b$, and, for order R when it is introduced below, $\Delta R \to dR$. Since the probability law $p(x_i)$ or the PDF $\rho(x_i)$ define the system, we use these terms interchangeably with the term "system."

Both $p(x_i)$ and $\rho(x)$ obey normalization,

$$\sum_{i=1}^{N} p(x_i) = 1, \quad \int_a^b dx \rho(x) = 1, \text{ and } \Delta \sum_{i=1}^{N} p(x_i) = 0 \tag{3}$$

after perturbation of the $p(x_i)$. Note: when no upper limits are shown in sums below they are assumed to be value $i=N$, that is $x_i = x = b$, depending upon whether the discrete or continuous case is taken.

A. Shift invariance

We assume that the system obeys shift invariance in its probability law $p(x)$ and PDF $\rho(x)$. Then the Fisher information I obeys the usual shift-invariant forms [3–5]

$$I = \sum_{i=1}^{N} \frac{1}{p_i}\left(\frac{dp_i}{dx}\right)^2 = \int_a^b dx \frac{1}{\rho(x)}\left(\frac{d\rho(x)}{dx}\right)^2 \ge 0 \tag{4}$$

in the discrete or continuous coordinate representations. Equations (4) directly show that the Fisher information I in either representation obeys positivity. For the sake of brevity and directness, we use the term "positivity" rather than "non-negativity" to describe a quantity whose value is greater than or equal to zero.

B. Use of amplitudes

We also find it convenient to work with real amplitude values $q(x_i) \equiv q_i$, defined in terms of probability values $p(x_i)$ as

$$p(x_i) \equiv q^2(x_i) = q_i^2,$$

$$\Delta p_i = 2q_i \Delta q_i,$$

$$\Delta q_i = \Delta p_i / (2\sqrt{p_i}), \quad I = 4\sum_{i=1}^{N} \left(\frac{dq_i}{dx}\right)^2, \tag{5}$$

the latter by use of Eq. (4). Note that this definition allows an ambiguity in $\pm$ sign for q_i. To remove possible ambiguity in the resulting sequence of q_i, $i=1,\ldots,N$ we assume that one definite sign is decided upon over all x_i. For simplicity, let it be the + sign.

III. FORMAL DEFINITION OF ORDER

Let the order R in a system obey positivity as defined above and depend upon the system amplitude law in some unknown way

$$R \equiv R[q(x_1), \ldots, q(x_N)] \equiv R(\boldsymbol{q}) \geq 0, \tag{6}$$

using vector notation. The $\boldsymbol{q}$ are assumed to be known and fixed, defining the system.

Our aim is to find the function $R(\boldsymbol{q})$. Note that this problem *differs* fundamentally from one in which R is *known* to be, for example, a specific functional of the $\boldsymbol{q}$ and their derivatives $dq(x)/dx|_{x=x_i}$, $i=1,\ldots,N$. (The latter is the case in Lagrangian problems.) No such specific knowledge of functional form is present here. Our aim is to find it. The solution $R(\boldsymbol{q})$ is found, at Eq. (26), to depend in particular upon the first differences $q(x_i) - q(x_{i-1}) \equiv \Delta q(x_i) \to (dq/dx)dx$ of the $\boldsymbol{q}$, after the continuous limit [Eq. (2)] is taken. In the following two subsections we give intuitive reasons for representing the order R as a function $R(\boldsymbol{q})$ of all the amplitudes *across* it in x_i.

A. Extension aspect

Intuitively, the amount of "order" in a system is the degree to which its amplitude law $q(x_i)$ contains a large amount of high-frequency structure. The amount of this structure can, for example, be measured by the total gradient content $(\Delta q(x_i)/\Delta x)^2$ over its extension $(b-a)$. Thus, a larger system extension should provide more order. Intuitively, in an apartment building, there is more order, in the form of structural detail, in two stories than in one. In fact our answer [Eq. (26)] for R gives four times the amount. Also, by the last Eq. (5), the Fisher information I [Eq. (4)] is a direct measure of the gradient content so, again, R should increase with I.

B. High-frequency aspect

Also, the *density* of the order, i.e., the amount of order per Nyquist sampling point for a band-limited PDF, intuitively represents the information in the system. An example is an optical system of large aperture used in conjunction with an array of solid-state detectors. This gives a large amount of high-frequency image order detail per Nyquist interval, i.e., high information content. Such an image contains high levels of both information and order. We next proceed with development of the approach.

C. Effect of perturbations

Consider the affect upon R of perturbing amplitudes $\boldsymbol{q}$ by small amounts $\boldsymbol{\Delta q}$. By Eq. (6) the order must likewise generally change. Depending upon the form of the order measure R, *and* of the perturbations (which can be of either sign), R could go up or down. Using ordinary Taylor series to second order in $\boldsymbol{\Delta q}$, this change is

$$\Delta R \equiv R(\boldsymbol{q} + \boldsymbol{\Delta q}) - R(\boldsymbol{q}),$$

$$= \boldsymbol{\Delta q}^{\mathsf{T}} \nabla_q R + 2^{-1} \boldsymbol{\Delta q}^{\mathsf{T}} \boldsymbol{M} \boldsymbol{\Delta q} + \cdots, \tag{7}$$

where $\boldsymbol{\Delta q}^{\mathsf{T}} \nabla_q R \equiv \Sigma_i (\partial R/\partial q_i) \Delta q_i$ with $\nabla_q R$ the gradient of R, and $\boldsymbol{M}$ the Hessian matrix of elements $M_{i,j} \equiv \partial^2 R/\partial q_i \partial q_j$. As usual T denotes the transpose. We retain terms of the series only out to second order in the Δq_i since these are small. In summary, ΔR is the change in order at a given *hyperspace* point $\boldsymbol{q}$ due to a random displacement $\boldsymbol{\Delta} q$ of the entire $\boldsymbol{q}$ system.

For later use, the above Eqs. (7) for ΔR may be restated as

$$\Delta R = \Delta R_1 + \Delta R_2,$$

$$\Delta R_1 \equiv \boldsymbol{\Delta q}^{\mathsf{T}} \nabla_q R,$$

$$\Delta R_2 \equiv 2^{-1} \boldsymbol{\Delta q}^{\mathsf{T}} \boldsymbol{M} \boldsymbol{\Delta q},$$

$$\boldsymbol{M} \equiv \{M_{i,j}\} = \left\{ \frac{\partial^2 R}{\partial q_i \, \partial q_j} \right\}. \tag{8}$$

D. Hermitian and Riemannian properties

From the last Eq. (8), since partial derivatives may be taken in any order, matrix $\boldsymbol{M}$ is symmetric. Also, the order R is defined as real. Therefore, $\boldsymbol{M}$ is Hermitian. and its eigenvalues are positive,

$$\lambda_i \geq 0, \quad i = 1, \ldots, N. \tag{9}$$

The λ_i are explicit functions of the $M_{i,j}$, so ultimately the $\lambda_i = \lambda_i[R(\boldsymbol{q})] = \lambda_i(\boldsymbol{q})$, implicit functions of the probability amplitudes.

At this point one is tempted to use the Čencov inequality [6–9], which states that the only Hermitian measure of order that monotonically decreases after a coarse graining is the Fisher-Rao metric [cf. Eqs. (12) and (16) below]. (Note that we use the terms "perturbation" and "coarse graining" interchangeably.) By this approach, the metric of the total change in order ΔR, and R itself, become simply proportional to the Fisher-Rao metric. Now, metric $\boldsymbol{M}$ is indeed Hermitian, so that this approach would seem to suffice. However, metric $\boldsymbol{M}$, that of ΔR_2, is only one part of the total metric for this problem. By Eqs. (8), there is the additional contribution from ΔR_1. Now, one may accomplish the additional ΔR_1 sum by adding a term $(\partial R/\partial q_i)\Delta q_i^{-1}$ to the ith diagonal term of $\boldsymbol{M}$, $i=1,\ldots,N$ in the third Eq. (8). These added terms also formally maintain Hermiticity for the total metric since they only add real numbers to the diagonal elements of the Hermitian metric $\boldsymbol{M}$. However, the overall scenario is one of very small perturbations $\Delta q_i \to 0$, and in this limit the added terms approach infinity. Hence the total metric is not generally well-defined Hermitian, voiding the approach.

IV. POSTULATE OF THE MEASURE

As in the preceding, let $R(\boldsymbol{q})$ represent the measure of order in a system of amplitude law $\boldsymbol{q}$. *The order is defined to decrease under coarse graining*. What order has this property? Let the system of amplitude values $q(x_i)$ be perturbed by a second system. The latter is an effective "observer;" either in the familiar "data taking" sense or, more generally, as any physical system *interacting* with the first. For example, the second system might be the outside environment of the first. We assume that all perturbations Δp, Δq, ΔR, etc., resulting from the interactive coarse graining are small. Since the resulting system order R is generally decreased, if the perturbations take place over a small time interval $\Delta t>0$ then $\Delta R \le 0$ over that interval. Or, ipso facto, the change ΔR over the corresponding *negative* time increment $-\Delta t$ is positive, $\Delta R \ge 0$. That is, looking backward in time, the order increases. Since it is mathematically simpler to work with such positive changes, the *analysis* is carried through over this *negative time increment*. However, of course all *applications* and interpretations of the results assume the usual positive time increments. In summary we postulate that R be a function of the $\mathbf{q}$ that satisfies

$$\Delta R \ge 0 \quad \text{for} \quad \Delta t < 0. \tag{10}$$

Čencov's famous inequality will be used to satisfy the requirement of decrease in order including the effects of perturbations out to second order in the probabilities. The answer for $R(\boldsymbol{q})$ turns out to be unique, from a heuristic viewpoint.

A. Transformation of changes

We saw in Sec. III D above that $\boldsymbol{M}$ is Hermitian. It is well known [10] that each N-dimensional point $\mathbf{q}$ may be shifted by a small vector amount $\boldsymbol{\Delta}\mathbf{q}'$ to a new point $\mathbf{q}'$ via an orthogonal, Hermitian matrix $[\mathbf{B}]$ with a special property. This is that the linearly transformed changes

$$\boldsymbol{\Delta}\mathbf{q} \equiv [\mathbf{B}]\boldsymbol{\Delta}\mathbf{q}' \quad \text{give} \quad \Delta R_2 \equiv 2^{-1}\sum_i \lambda_i \Delta q_i'^2. \tag{11}$$

The transformation preserves the length ΔR_2 both before and after its use. Then by definition (8) of ΔR_2

$$\frac{\partial^2 R}{\partial q_i'\, \partial q_j'} = \lambda_i \delta_{ij}, \quad \frac{\partial R}{\partial q_i'} = \lambda_i q_i' + C_i, \tag{12}$$

where δ_{ij} is the Kronecker delta. The second equation follows an integration of the first. We may now re-express the order R and its changes ΔR_1 and ΔR_2 in the shifted system. We show below that, effectively, the $C_i \equiv C(x_i)=0$.

Thus, by Eq. (11), for the new system of changes $\Delta q_i'$ the resulting second-order total change ΔR_2 lacks cross-term products $\Delta q_i' \Delta q_j'$, $i \ne j$. Note that the λ_i depend upon the original cross-term coefficients $\partial^2 R/\partial q_i \partial q_j$ for all i,j. In terms of the new changes $\boldsymbol{\Delta}\mathbf{q}'$ using Eqs. (11) and (12) we have

$$\Delta R = \Delta R_1 + \Delta R_2,$$

$$\Delta R_1 \equiv \boldsymbol{\Delta q}'^{\mathsf{T}} \nabla_q R \equiv \sum_i \Delta q_i'(\lambda_i q_i' + C_i),$$

$$\Delta R_2 = 2^{-1}\sum_i \lambda_i \Delta q_i'^2. \tag{13}$$

The transformation replaces the effectively two-dimensional problem [Eq. (8)] with a much simpler one of a single dimension.

As mentioned at Eq. (10), over a *negative* time increment the order R is *required to increase*,

$$\Delta R = \Delta R_1 + \Delta R_2 \ge 0 \quad \text{for} \quad \Delta t < 0 \tag{14}$$

to second order in changes Δq_i. It is convenient to first regard pixel length Δx and change ΔR as finite, and then take their continuous limits [Eq. (2)]. Hence, we now ask, what order measure R obeys property (14)?

By Eq. (13), the requirement (14) becomes

$$\Delta R \equiv \Delta R_1 + \Delta R_2,$$

$$= \sum_i (\lambda_i q_i' + C_i)\Delta q_i' + 2^{-1}\sum_i \lambda_i \Delta q_i'^2 \ge 0, \tag{15}$$

for $\Delta t<0$. As was discussed, this requirement is equivalent to requiring a loss, or contraction, $-\Delta R$ of order in the usual *positive* time direction $\Delta t>0$.

But Eq. (15), like its antecedent (8), is but a power series expansion for ΔR. The difference is that, whereas (8) was a power series in the original changes Δq_i, which contained all second-derivative cross terms, the transformation (11) to new changes $\Delta q_i'$ has eliminated all cross terms.

B. Čencov's inequality

Now we note that, because the eigenvalues λ_i obey positivity (9) the *second*-order change ΔR_2 will be positive. That is, no matter the cause of the perturbations Δq_i, change ΔR_2 always contributes toward the required positivity of Eq. (15). However, it is the sum $\Delta R_1+\Delta R_2 \equiv \Delta R$ that is to be positive, so that the effect of the chosen λ_i upon ΔR_1 must also be considered. For some choices this might be negative, for example, opposing the required positivity of the sum.

Is there a set of λ_i that, in any scenario of coarse graining, gives $\Delta R_1+\Delta R_2 \ge 0$? Čencov's inequality states that for a Hermitian metric, such as $\boldsymbol{M}$, the required eigenvalues are

$$\lambda_i = 1, i = 1, \ldots, N. \tag{16}$$

However, $\boldsymbol{M}$ is the metric for ΔR_2 and, as discussed below Eq. (9), the overall sum $\Delta R_1+\Delta R_2$ does not have a well-defined Hermitian metric. Hence, we simply evaluate the sum under condition (16), so as to test whether it does indeed obey the required positivity.

C. Order increase due to coarse graining

By Eq. (16), requirement (15) becomes

$$\Delta R = \sum_i (q_i' + C_i)\Delta q_i' + 2^{-1}\sum_i \Delta q_i^2 \geq 0, \tag{17}$$

after noting that by transformation (11), $\boldsymbol{\Delta q}^{\mathrm{T}}\boldsymbol{\Delta q} \equiv [[\mathbf{B}]\boldsymbol{\Delta q}']^{\mathrm{T}}[[\mathbf{B}]\boldsymbol{\Delta q}'] = \boldsymbol{\Delta q}'^{\mathrm{T}}[\mathbf{B}]^{\mathrm{T}}[\mathbf{B}]\boldsymbol{\Delta q}' = \boldsymbol{\Delta q}'^{\mathrm{T}}[\mathbf{B}]^{-1}[\mathbf{B}] \times\boldsymbol{\Delta q}' = \boldsymbol{\Delta q}'^{\mathrm{T}}\boldsymbol{\Delta q}'$. The second equality is a matrix identity, and the third equality is a property of the orthogonal matrix $[\mathbf{B}]$. Then by Eqs. (3)–(5), Eq. (17) becomes

$$\Delta R = \Delta x \sum_i C_i \frac{\Delta q_i'}{\Delta x} + 2^{-1}\Delta x^2 \sum_i \left(\frac{\Delta q_i}{\Delta x}\right)^2,$$

$$\to \int dx C(x)\frac{dq_i'}{dx} + 8^{-1}\Delta x^2 \int dx \frac{1}{\rho}\left(\frac{d\rho}{dx}\right)^2,$$

$$= \int dx' C(x')\frac{dq_i'}{dx'} + 8^{-1}\Delta x^2 I \to dR(x). \tag{18}$$

We also used Eq. (1) and the continuous limit (2). Note that, the new dummy integration variable x' replaces x, so as to clarify that ΔR depends upon x (only) through the second-order small quantity Δx^2 in Eq. (18).

V. *R* AS FUNCTION OF *x*

Since the system is fixed, information I in Eq. (18) is a fixed constant, i.e., not a function of x. Then Eq. (18) is effectively an expansion out to second order in changes dx of the function $dR(x)$. Then likewise $R=R(x)$. This is not surprising since defining Eq. (6) expresses R as a function of the amplitudes $q(x_i)$ over the entire range of x.

The order function $R(x)$ measures the *local* order pointwise at coordinate x. Let $R(x)$ be analytic. Then we may expand $R(x)$ in power series about the point $x=a$. Keeping only terms out to second order,

$$R(x) = R(a) + (x-a)\left.\frac{dR}{dx}\right|_{x=a} + 2^{-1}(x-a)^2 \left.\frac{d^2R}{dx^2}\right|_{x=a}. \tag{19}$$

This assumes that $x-a \equiv \Delta x \to dx$ is small. Then from Eq. (19)

$$\Delta R(x) \equiv R(x) - R(a) = \Delta x \left.\frac{dR}{dx}\right|_{x=a} + 2^{-1}\Delta x^2 \left.\frac{d^2R}{dx^2}\right|_{x=a}, \tag{20}$$

or

$$\Delta R = \Delta x \left.\frac{dR}{dx}\right|_{x=a} + 2^{-1}\Delta x^2 \left.\frac{d^2R}{dx^2}\right|_{x=a}. \tag{21}$$

Then this must equal the final Eq. (18). We note that the term $\int dx' C(x')dq'(x')/dx'$ has implicitly a multiplier $(\Delta x)^0=1$, i.e., is a constant in x (as distinguished from the renamed variable x' in the integral). By comparison, Eq. (21) has no constant term $(\Delta x)^0$. Also, Eq. (18) has no term in $(\Delta x)^1 = \Delta x$. Then matching corresponding coefficients of powers of Δx in Eqs. (18) and (21) gives

$$\int dx' C(x')\frac{dq'(x')}{dx'} \equiv 0,$$

$$\left.\frac{dR}{dx}\right|_{x=a} = 0,$$

$$2^{-1}\left.\frac{d^2R}{dx^2}\right|_{x=a} = 8^{-1}I. \tag{22}$$

The first relation follows, for example, if the arbitrary integration constants C_i are chosen to be zero.

Then by Eqs. (20) and (22), Eqs. (18) and (19) become, respectively,

$$\Delta R(x) = 8^{-1}(\Delta x)^2 I, \quad \Delta t < 0, \quad \text{and}$$

$$R(x) = R(a) + 8^{-1}(x-a)^2 I. \tag{23}$$

Since $I \geq 0$, the former shows that, as required by Eq. (17), $\Delta R(x) \geq 0$ is now *achieved*. The latter shows that, to quadratic changes in x, the local order $R(x)$ must increase. Thus, $R(x)$ measures the order out to extension x of the system. This is over the system interval (a,x). These imply that finite "order" is only associated with finite system extension, so that the amount of order within interval (a,a) is zero,

$$R(a) \equiv 0. \tag{24}$$

By the same token, the total amount of order R in the system must be the amount included over its entire extension (a,b),

$$R \equiv R(b). \tag{25}$$

Then from the preceding three equations

$$R \equiv R(b) = 8^{-1}(b-a)^2 I. \tag{26}$$

This is the answer for *total* (as distinguished from local) system order that we sought.

Taking a first change of Eq. (26) gives $\Delta R = 8^{-1}(b-a)^2\Delta I$. Then requirement (10) shows that not only does the order R decrease after a coarse graining but so, likewise, does the information I,

$$\Delta I \leq 0 \ \text{ for } \ \Delta t > 0,$$

or, conversely

$$\Delta I \geq 0 \ \text{ for } \ \Delta t < 0, \tag{27}$$

i.e., looking backward in time.

Is the measure unique?

We next show that, at least on heuristic grounds, this expression for $R(\mathbf{q})$, is unique (a formal proof is outside the scope of this paper). In the Taylor-series expansion (8) for ΔR, contribution ΔR_1 is first order in the small changes $\boldsymbol{\Delta q}$

while ΔR_2 is second order. Therefore, if ΔR_1 is nonzero, it will dominate the total ΔR. Next, consider any set of coefficients $\nabla_q R$ and changes $\mathbf{\Delta q}$ that satisfy the positivity requirement (10). The requirement is to hold for any $\mathbf{\Delta q}$ and a *general* set of amplitudes $\mathbf{q}$. Then it should hold if the $\mathbf{\Delta q}$ are replaced by their negatives. (We note that normalization requirement (3) will still be obeyed since, by $\Delta p_i = 2q_i \Delta q_i$, the resulting $\mathbf{\Delta p}$ simply change sign, so that their sum remains zero.) But since ΔR_1 is linear in the $\mathbf{\Delta q}$ it must now be negative, making ΔR negative and, hence, violating requirement (10). The only way this result can be avoided is if all components of the gradient $\nabla_q R$ are zero, for a general set of amplitudes $\mathbf{q}$. By Eq. (8) this requires $\Delta R_1 = 0$. Now, this does hold in the special case where amplitudes $\mathbf{q}$ are those defining an equilibrium point, where R is extremized. However, this possibility is ruled out by our requirement that the amplitudes be *general*, i.e., not limited to being at an equilibrium point. An alternative route to positivity must be sought.

The alternative we found is to instead rotate all changes $\mathbf{\Delta q}$ by a small amount, in Eqs. (11)–(15), use the Čencov choice of eigenvalues [Eq. (16)]—which are unique according to Čencov's theorem since $\boldsymbol{M}$ is Hermitian—and then proceed on to Eq. (18). This satisfies the requirement $\Delta R \geq 0$ since we later show, at Eqs. (22), that the integral including function $C(x') = 0$ effectively. On this basis our final expression (26) for R is unique.

VI. DISCUSSION AND SUMMARY

Coarse graining is often used for describing the transition from a quantum variable to its observed value or, more grandly, from a quantum universe to a classical one [11–13]. Another example is that of Voronoi tesselation [14], whereby mesoscopic particle-based fluid models are assumed to be coarse-grained representations of underlying microscopic fluids [15].

By postulating at Eq. (10) that order should decrease under coarse graining, we have quantified the notion of the total order R [Eq. (26)] in a system. This basically traces to the Hermitian nature of the second-derivative matrix $\boldsymbol{M}$ defined in Eq. (8). This property guarantees positive eigenvalues λ_i (9) in a transformed system (11) of changes $\mathbf{\Delta q}'$ in amplitude. The Čencov choice [Eq. (16)] of unit eigenvalues and the use of Eqs. (11)–(17) gives an intermediary result (18). Here ΔR is the sum of an integral of an unknown function $C(x')$ and a term linear in Fisher information I. Then in Eqs. (19)–(22) the term in $C(x')$ is found to be effectively zero, so that (18) gives a simple proportionality $\Delta R \propto I$. Finally, the Taylor expansion (19) used in conjunction with results (22) gives the order (26) as $R = 8^{-1}(b-a)^2 I$.

Dividing Eq. (18) by Eqs. (26), and again using Eqs. (22) gives the relative change in order due to coarse graining, as $\Delta R(x)/R = -[\Delta x/(b-a)]^2$, a loss in the forward-time direction. The coarser the subdivision Δx is the larger is the relative loss in order.

A. Partial graining

In the Appendix, we show that partial coarse graining results in a smaller loss of order than does full coarse graining. The partial nature of the graining is enforced by imposing K linear constraints upon the amplitudes $q'(x)$. The constraints are described in Eq. (A1) by linear weight kernels $F_k(x)$, $k = 1, \ldots, K$. As is well known (and used) in statistics, the existence of constraints tends to control the wildest perturbations in the amplitudes, in effect regularizing them to the coarse graining process. For example, if there are M linear constraint relations among its contributing variables, a $\chi^2(x)$ variable with N total contributions has a *reduced* order $N-M$, rather than N. Also, the lower the order is the *less* spread out, or random, are the fluctuations x [16]. The resulting change in partial order is given in Eq. (A11). It shows that the order change ΔR is negative, as required, but not so negative as in the absence of constraints. This satisfies intuition, as discussed. Also, comparing the result (A12) for the partial order R with result (26) for the full R (in the absence of constraints) shows that the partial order value is the unconstrained value reduced by a sum of weighted Fisher information $I(F_k)$. Thus, the constraints reduce the absolute level of the order. This follows because Eqs. (A1) constrain the fluctuations in the system amplitudes $q'(x)$. Fluctuations define structure, and structure defines order, so that the total amount of order is reduced in this way. See also the examples below.

B. Unitless nature

Since from Eq. (4) the information I has units of x^{-2}, the result (26) shows that the order R is *unitless*. This allows for great flexibility in comparing levels of order change. These can be for different systems (say, a bacterium vs a water molecule) and, even, for different parameters (say, time vs position in a given system or in different systems). Of course the other well-known measure of disorder, the entropy, is unitless. As with thermodynamic entropy, the measure also has the benefit of consisting entirely of macroscopic parameters.

C. Fractal property

Equation (26) is the principle result of the paper. It gives the total order measure as proportional to the shift-invariant Fisher information and to the square of the system extension. The distinction between order R and information I is conveniently seen when the system is degraded by coarse graining defined by a simple linear magnification $y = mx$ of the coordinate x. By Jacobian transformation, the information $I(y)$ in the magnified system is $I(x)/m^2$. Thus, for a stretch $m > 1$ the information goes down. By comparison, the order is $R(x) = (b-a)^2 I(x)$ in the original system, and the stretched system has an order $R(y) = [m(b-a)]^2 I(y) = [m(b-a)]^2 I(x) m^{-2} = R(x)$ once again. *A simple magnification does not affect the order*. For the order to increase after magnification, the amount of detailed structure in the system has to increase as well. This becomes evident from the following examples.

D. Examples

The contrasting properties of order R and information I are shown by the following examples. Suppose that the prob-

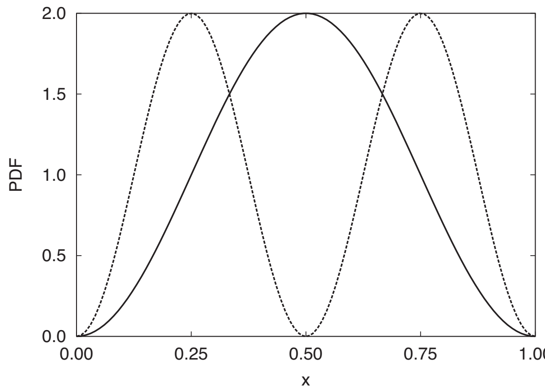

FIG. 1. Plots of system PDFs $\rho(x)=(2/b)\sin^2(n\pi x/b)$ for $n=1,2$.

ability density (PDF) curve $\rho(x)=(2/b)\sin^2(n\pi x/b)$, over extension $0\le x\le b$. This represents a raised sinusoid with n ripples or lobes, with $n=1$ or 2 or, …. Using the second Eq. (4) gives a Fisher $I\equiv I(n)=(2n\pi/b)^2$. The information increases as the square of the number of ripples *per unit length* of the system. Thus, the information measures the concentration or density of detail. How does this compare with the level of order? The use of Eq. (26) gives the order as $R\equiv R(n)=b^2I(n)=(2n\pi)^2$ This is unitless, as proven generally above. It is also independent of the absolute extension b of the system. Thus the order only depends upon the total number n of ripples across it. Examples are shown in Figs. 1 and 2.

In Fig. 1, the solid curve is for $n=1$, the dotted for $n=2$. They are defined over the same support interval (0,1). The order values are $R(1)=4\pi^2$ and $R(2)=16\pi^2$ units, depending quadratically upon the number n of ripples. In Fig. 2, the two PDF curves have an equal number $n=2$ of ripples but different support intervals, as indicated. Although they look quite different, on the basis of order $R(2)=(2\cdot 2\pi)^2=16\pi^2$ units *for each*. That $R(n)$ does not depend upon b also verifies the preceding subsection, by which order is independent of any linear stretch of the system. To change the level of order requires, by the result $R(n)=(2n\pi)^2$, changing the *number n* of local "details" (here, number of ripples) in the system. This is evident in Fig. 1, where it quadruples when n doubles. Evidently, order is sensitive to the number of details across the support interval, rather than to the magnitudes of their local slopes. Conversely, for *fixed* support intervals, order is as sensitive to local slope values as is information I.

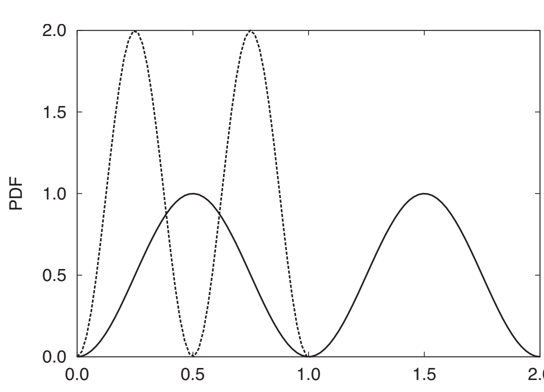

FIG. 2. Plots of system PDFs for $n=2$ ripples over two support intervals.

E. Order and Fisher information as *entropies*

The general results [Eqs. (26) and (27)] indicate that Fisher I and order R both show monotonic loss with (forward) time. Quantities that show such monotonic behavior are typically regarded as entropies. Thus, R and I are entropies. This is compared with the usual interpretation [16] of I as solely a level of information *in data*. It replaces this with a property of *the system* that produces the data, namely, its level of order.

Physically, Eqs. (26) and (27) thereby free all past (e.g., [3,4]) and future applications of Fisher I to physics from the requirement that I represent solely the level of *information in data*. It is only when we attempt to observe phenomena that data are taken. In fact there is no mention of data in the foregoing analysis. Rather, it focuses attention on *system function*, through its PDF $\rho(x)$. In this way a requirement on system data—its output—is replaced by one on system function—its time-dependent order R. This obeys the monotonic time evolution (10), which defines decreasing order with time.

F. Nonparticipatory phenomena

With data now no longer required as the basis for the above applications, there is no longer a need for an *observer*. This is most important, since of course most phenomena occur *un*observed. For example, while a quantum system is not being observed the temporal evolution of its wave function follows the Schrödinger wave equation. The results of this paper therefore now allow this to be derived, using the same mathematics [3–5] as before, but without requiring observing the effect (specifically, the four-position of a particle).

Unobserved variables likewise occur as the *intrinsic variables* of a thermodynamic system, where extrinsic variables such as temperature and pressure are instead used. In fact, by expressing Fisher information I in terms of these extrinsic variables one recovers the familiar laws of thermodynamics including its Legendre transformation and concavity properties [17,18].

Of course there are also physical effects that *do* depend in large part upon *the presence*, and nature, of an observer. A well-known example is the "quantum Zeno effect" [19]. These effects are observer dependent in the sense of Wheeler's [20] "participatory phenomena." In these cases the Fisher I is interpreted as in the past, i.e., as a property of data.

In summary, past derivations of physical laws through variation of Fisher I by use of the principle of Extreme Physical Information [3,4] need no longer be regarded as requiring an observer. Rather, one needs only *a second system* that effectively *coarse grains* the first [see text above (10)], perturbing the first's variables x_i through interaction with them. An example is interaction via two-particle interaction potentials in the Hartree approximation. The result is perturbation of the Fisher *entropy*, in accordance with Eqs. (27). A recent example [21] of such Fisher-based derivations is of what is usually considered the basis for quantum theory, i.e., the de Broglie quantum wave "hypothesis." This derivation then predicts quantum effects, even before their observation. That is, the hypothesis is proven independent of any measurement of waves, and from purely classical considerations of physics and information.

ACKNOWLEDGMENTS

B.R.F. would like to acknowledge stimulating discussions on the essential role of order in cosmology and biological systems with, respectively, Dr. William R. Stoeger of the University of Arizona and Dr. Robert A. Gatenby of the Moffitt Cancer Center, University of South Florida.

APPENDIX: PARTIAL OR CONSTRAINED COARSE GRAINING

The preceding derivation was for the presence of full blown coarse graining. On the other hand, if the coarse graining is constrained the fluctuations $\mathbf{\Delta q}$ will be likewise constrained, and therefore of smaller amplitude. This describes a scenario of partial coarse graining. Interesting discussions of partial coarse graining may be found in [22,23]. There are also notable examples of such graining in (i) analyzing biological whole-genome gene expression, and (ii) mesoscopic nonequilibrium thermodynamics (MNET): an extension of the equilibrium thermodynamics of small systems.

Intuitively, the presence of such constraints represents positive information. Thus it should act to preserve order during the coarse graining. Then, although the order R should still decrease in the positive time direction, it should do so by a smaller magnitude than in their absence. The previous, unconstrained answer [Eq. (26)] should be replaced by one that reflects this effect. This is verified as follows.

Suppose the coarse graining is subject to K weighted constraints $\langle F_k\rangle, k=1,\ldots,K$

$$\sum_{i=1}^{N} F_k(x_i)p'(x_i) = \sum_{i=1}^{N} F_k(x_i)q'^2(x_i) = \langle F_k\rangle,$$

$$\text{with } \sum_{k=1}^{K} F_k(x_i) \equiv S_i \leq 1, \tag{A1}$$

where $i=1,\ldots,N$ and all weights $F_k(x_i)\geq 0$. These constraints supplement mere normalization [Eqs. (3)]. Notice that, for any initial set of weights $F_k(x_i)$ at a fixed x_i, the last inequality $S_i\leq 1$ can always be made to hold by dividing through the weights, and the original constraint values $\langle F_k\rangle$, by a suitable constant. To be definite, let this be the smallest number that satisfies the requirement.

As with normalization, the $\langle F_k\rangle$ are assumed to be maintained after the coarse graining, so that

$$\Delta\langle F_k\rangle = 0, \quad k=1,\ldots,K. \tag{A2}$$

Possible examples are $F_k(x_i)=(x_i-a)^k(b-a)^{-k}$, for an arbitrary K. (Notice that these F_k are unitless.) Here S_i in Eq. (A1) is the sum $S_i=b(1-b)^{-1}(1-b^K)$ of a geometric series, with $b\equiv(i-1)/(N-1)$. Then $b\leq 1$, so that all $S_i\leq 1$ as required. This is without the need for any division by a constant defined below (A1).

Taking a differential of Eq. (A1) gives

$$\Delta\sum_i F_k(x_i)p'(x_i) = \Delta\langle F_k\rangle \equiv 0 = 2\sum_i F_{ki}q_i'\Delta q_i' \tag{A3}$$

in briefer notation. The zero results since the constraints are maintained during the perturbation process. We now replace the Čencov choice [Eq. (16)] by a new choice

$$\lambda_i \equiv 1-\sum_k F_{ki} \geq 0, i=1,\ldots,N. \tag{A4}$$

The inequality is enforced by the last Eq. (A1).

The new choice [Eq. (A4)] again zeros ΔR_1, as shown next. (In this way the total $\Delta R_1+\Delta R_2$ obeys the required positivity, as before.)

1. Quantity ΔR_1

Equation (12) now becomes

$$\frac{\partial^2 R}{\partial q_i'^2} = \lambda_i = 1-\sum_k F_{ki},$$

$$\text{so that } \frac{\partial R}{\partial q_i'} = q_i'\left(1-\sum_k F_{ki}\right) + C_i, \tag{A5}$$

with $\partial^2 R/\partial q_i'\,\partial q_j'=0$ for $i\neq j$ as before.

Substitution of $\partial R/\partial q_i'$ from (A5) into definition (8) of ΔR_1 gives

$$\Delta R_1 = \sum_i q_i'\Delta q_i'\left(1-\sum_k F_{ki}\right) + \sum_i C_i\Delta q_i'. \tag{A6}$$

Next, using constraint Eqs. (3), (5), and (A3) gives

$$\Delta R_1 = \sum_i C_i\Delta q_i'. \tag{A7}$$

2. Change ΔR_2

Using choice (A4) in the last Eq. (11) gives

$$\Delta R_2 \equiv 2^{-1}\Delta x^2\sum_i \Delta q_i'^2\left(1-\sum_k F_{ki}\right) \geq 0, \tag{A8}$$

since of course $\Delta q_i'^2 \equiv (\Delta q_i')^2 \geq 0$, and since the last Eq. (A1) required $\sum_k F_{ki}\leq 1$.

3. Transition to weighted information

Then since $\Delta R=\Delta R_1+\Delta R_2$ the last two equations give

$$\Delta R = \sum_i C_i\Delta q_i' + 2^{-1}\Delta x^2\sum_i \Delta q_i'^2\left(1-\sum_k F_{ki}\right),$$

$$\to \int dx'\,C(x')(dq'/dx') + 8^{-1}\Delta x^2\left(I-\sum_k I(F_k)\right), \tag{A9}$$

using the last Eq. (5), and where $\Delta t\leq 0$ as usual. The $I(F_k)$

are the *weighted Fisher information* [24]

$$I(F_k) \equiv \sum_i F_{ki}\Delta q_i^2 \to \int dx F_k(x)\frac{1}{\rho}\left(\frac{d\rho}{dx}\right)^2 \quad (A10)$$

in the continuous limit (2). In data processing [24], each information $I(F_k)$ allows optimal parameter estimation in the sense of minimizing the F_k^{-1}-weighted mean-squared error of estimation.

4. Final expression for partial order

Use of the analysis [Eqs. (19)–(22)] again gives $\int dx' C(x')(dq'/dx')=0$, so that Eq. (A9) becomes

$$\Delta R(x) = 8^{-1}\Delta x^2\left[I - \sum_k I(F_k)\right] \geq 0, \ \text{for} \ \Delta t \leq 0. \quad (A11)$$

Then steps analogous to Eqs. (23)–(26) give for the constrained order

$$R = 8^{-1}(b-a)^2\left[I - \sum_k I(F_k)\right]. \quad (A12)$$

Equation (A12) shows that R is again unitless, as it was for the full coarse graining case. This is provided the constraint kernels F_k are chosen unitless as in the example above.

As following Eq. (26), taking a differential of Eq. (A12) shows that the $I-\Sigma_k I(F_k)\equiv I_p$ are entropies, in that they change monotonically with time, as also is the order R. Additionally, all information quantities in Eq. (A11) obey positivity. Then each extra information constraint $I(F_k)$ decreases the gain in order from the answer [Eq. (23)] for ΔR in the absence of constraints. This is in the negative time direction, as indicated. Hence, in the *positive* time direction it allows the order change ΔR to be negative, as required, but not so negative as in the absence of constraints. This satisfies the intuition expressed at the beginning of this appendix.

Equation (A11) seems to allow a paradox, whereby with proper choice of constraint kernels $F_k(x)$ the local loss of order $\Delta R(x)$ can be made to be zero. However, Eq. (A12) shows that this can only be accomplished in the trivial case where the total system order R is zero to begin with.

[1] S. J. Jorgensen and Y. M. Svirezhev, *Towards a Thermodynamic Theory for Ecological Systems* (Elsevier, New York, 2004).
[2] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (Elsevier, Oxford, UK, 1993).
[3] B. R. Frieden, *Science from Fisher Information: A Unification* (Cambridge University Press, Cambridge, England, 2004).
[4] *Exploratory Data Analysis Using Fisher Information*, edited by B. R. Frieden and R. A. Gatenby (Springer-Verlag, London, 2007).
[5] M. Reginatto, Phys. Rev. A **58**, 1775 (1998). **60**, 1730 (1999).
[6] N. N. Čencov, *Statistical Decision Rules and Optimal Inference, Translations of Mathematical Monographs* (American Mathematical Society, Providence, 1982), Vol. 53.
[7] L. L. Campbell, Inf. Sci. (N.Y.) **35**, 199 (1985).
[8] L. L. Campbell, Proc. Am. Math. Soc. **98**, 135 (1986).
[9] F. Hansen, Proc. Natl. Acad. Sci. U.S.A. **105**, 9909 (2008).
[10] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, England, 1985).
[11] J. J. Halliwell, Contemp. Phys. **46**, 93 (2005).
[12] S.-L. Luo, Commun. Theor. Phys. **47**, 597 (2007).
[13] P. Gibilisco, F. Hiai, and D. Petz, IEEE Trans. Inf. Theory **55**, 439 (2009).
[14] Q. Du, V. Faber, and M. Gunzburger, SIAM Rev. **41**, 637 (1999).
[15] A. Eriksson, M. N. Jacobi, J. Nyström, and K. Tunstrøm, EPL **86**, 44001 (2009).
[16] B. R. Frieden, *Probability, Statistical Optics and Data Testing* (Springer-Verlag, New York, 2001).
[17] B. R. Frieden, A. Plastino, A. R. Plastino, and B. H. Soffer, Phys. Rev. E **60**, 48 (1999).
[18] S. P. Flego, B. R. Frieden, A. Plastino, A. R. Plastino, and B. H. Soffer, Phys. Rev. E **68**, 016105 (2003).
[19] B. Misra and E. C. G. Sudarshan, J. Math. Phys. **18**, 756 (1977).
[20] J. A. Wheeler, in *Physical Origins of Time Asymmetry*, edited by J. J. Halliwell, J. Perez-Mercader, and W. H. Zurek (Cambridge University Press, Cambridge, England, 1994), pp. 1–29.
[21] B. R. Frieden and B. H. Soffer, Physica A **388**, 1315 (2009).
[22] E. H. Lemus, J. Non-Equilib. Thermodyn. **34**, 371 (2009).
[23] D. Jou, C. Pérez-García, L. S. García-Colín, M. Lopez de Haro, and R. F. Rodríguez, Phys. Rev. A **31**, 2502 (1985).
[24] B. R. Frieden and B. H. Soffer, Phys. Lett. A **374**, 3895 (2010).