0

Introduction

0.1 Aims of the book

The *primary aim* of this book is to develop a *theory of measurement* that incorporates the observer into the phenomenon under measurement. By this theory, the observer becomes both a collector of data and an activator of the phenomenon that gives rise to the data. These ideas have probably been best stated by J. A. Wheeler (1990; 1994):

> 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.

The measurement theory that will be presented is largely, in fact, a quantification of these ideas. However, the reader might be surprised to find that the "information" that is used is not the usual Shannon or Boltzmann entropy measures, but one that is relatively unknown to physicists, that of R. A. Fisher.

The measurement theory is simply a description of how Fisher information flows from a physical source effect to a data space. It therefore applies to all scenarios where quantitative data from repeatable experiments may be collected. This describes measurement scenarios of physics but, also, of science in general. The theory of measurement is found to define an analytical procedure for deriving all laws of science. The approach is called EPI, for "extreme physical information."

The *secondary aim* of the book is to show, by example, that most existing laws of science fit within the EPI framework. That is, they can be derived by its use. (Many can of course be derived by other approaches, but, apparently, no other single approach can derive *all* of them.) In this way the EPI approach unifies science under an umbrella of measurement and information. It also leads to new insights into how the laws are interrelated and, more importantly, to new laws and to heretofore unknown *analytical expressions* for physical

John A. Wheeler, from a photograph taken *c*. 1970 at Princeton University. Sketch by the author.

constants, such as for the Weinberg angle and Cabibbo angle (Chapter 11) of the weak nuclear interaction.

In this way, the usual ways and the means of science are reversed. The various branches of physics are usually derived from the *top down*, i.e., from some suitable action Lagrangian, which in turn predicts a class of measurements. By comparison, EPI derives science by viewing it, in effect, from the *bottom up*. It is based upon measurements first and foremost.

Measurements are usually regarded as merely random outputs from a particular effect. The EPI approach logically reverses this, tracking from output to input, from the data to the effect. It uses knowledge of the information flow in the measurement process to derive the mathematical form of the physical effect that gives the output measurements. The approach is subject to some caveats.

Caveat 1: The usual aim of theory is to form mathematical models for physical effects. This is our aim as well. Thus, the EPI approach is limited to deriving the *mathematical expression* of physical effects. It does not form, in some way, the physical effects themselves. The latter are presumed always to exist "out there" in some fixed form.

Caveat 2: One does not get something from nothing, and EPI is no exception to this rule. Certain things must be assumed about the unknown effect. One is *knowledge of a source*. The other is knowledge of an appropriate *invariance principle*. For example, in electromagnetic theory (Chapter 5), the source is the charge-current density, and the invariance principle is the equation of continuity of charge flow. Notice that these two pieces of information do not by themselves imply electromagnetic theory. However, they do when used in tandem with EPI.

In this way, an invariance principle plays an *active* role in deriving a physical law. Note that this is the reverse of its passive role in orthodox approaches to physics, which instead regard the invariance principle as a *derived* property from a *known* law. (Noether's theorem is often used for this purpose.) This is a key distinction between the two approaches, and should be kept in mind during the derivations.

How does one know *what* invariance principle to use in describing a given scenario?

Caveat 3: Each application of EPI relies upon the user's ingenuity. EPI is not a rote procedure. It takes some imagination and resourcefulness to apply. However, experience indicates that every invariance principle that is used with EPI yields a valid physical law. The approach is exhaustive in this respect.

During the same years that quantum mechanics was being developed by Schrödinger (1926) and others, the field of classical measurement theory was being developed by R. A. Fisher (1922) and co-workers (see Fisher Box, 1978, for a personal view of his professional life). According to classical measurement theory, the quality of any measurement(s) may be specified by a form of information that has come to be called Fisher information. Since these formative years, the two fields – quantum mechanics and classical measurement theory – have enjoyed huge success in their respective domains of application. Until recent times it had been presumed that the two fields are distinct and independent.

However, the two fields actually have strong overlap. The thesis of this book is that all physical law, from the Dirac equation to the Maxwell–

Boltzmann velocity dispersion law, may be unified under the umbrella of classical measurement theory. In particular, the information aspect of classical measurement theory – Fisher information – is the key to the unification.

Fisher information is part of an overall theory of physical law called the principle of EPI. The unifying aspect of this principle will be shown by example, i.e., by application to the major fields of physics: quantum mechanics, classical electromagnetic theory, statistical mechanics, gravitational theory, etc. The defining paradigm of each such discipline is a wave equation, a field equation, or a distribution function of some sort. These will be derived by use of the EPI principle. A separate chapter is devoted to each such derivation. New effects are found, as well, by the information approach.

Such a unification is, perhaps, long overdue. Physics is often considered the science of measurement. That is, physics is a quantification of *observed* phenomena, and observed phenomena contain noise, or fluctuations. The physical paradigm equations (mentioned above) define the fluctuations or errors from ideal values that occur in such observations. That is, *the physics lies in the fluctuations*. On the other hand, classical Fisher information is a scalar measure of these very physical fluctuations. In this way, Fisher information is intrinsically tied into the laws of fluctuation that define theoretical physics.

EPI theory proposes that all physical theory results from observation: in particular, *imperfect* observation. Thus, EPI is an observer-based theory of physics. We are used to the concept of an imperfect observer in addressing quantum theory, but the imperfect observer does not seem to be terribly important to classical electromagnetic theory, for example, where it is assumed (wrongly) that fields are known exactly. The same comment can be made about the gravitational field of general relativity. What we will show is that, by admitting that any observation is imperfect, one can derive both the Maxwell equations of electromagnetic theory and the Einstein field equations of gravitational theory. The EPI view of these equations is that they are expressions of fluctuation in the values of measured field positions. Hence, the four-positions $(\boldsymbol{r}, t)$ in Maxwell's equations represent, in the EPI interpretation, random excursions from an ideal, or mean, four-position over the field.

Dispensing with the artificiality of an "ideal" observer allows us to reap many benefits for purposes of *understanding* physics. EPI is, more precisely, an expression of the "inability to know" a measured quantity. For example, EPI derives quantum mechanics from the viewpoint that an ideal position cannot be known. We have found, from teaching the material in this book, that students more easily understand quantum mechanics from this viewpoint than from the conventional viewpoint of derivative operators that somehow represent energy or momentum. Furthermore, that *the same* inability to know also leads to the

Maxwell equations when applied to that scenario is even more satisfying. It is, after all, a human desire to find common cause in the phenomena we see.

Unification is also, of course, the major aim of physics, although EPI is probably not the ultimate unification that many physicists seek. Our aim is to propose a *comprehensive* approach to deriving physical laws, based upon a new theory of measurement. Currently, the approach presumes the existence of sources and particles. EPI derives major classes of particles, but not all of them, and does not derive the sources. (See Caveat 2 preceding.) We believe, however, that EPI is a large step in the right direction. Given its successes so far, the sources and remaining particles should eventually follow from these considerations as well.

At this point we want to emphasize *what this book is not about*. This is not a book whose primary emphasis is upon the *ad hoc* construction of Lagrangians and their extremization. That is a well-plowed field. Although we often derive a physical law via the extremization of a Lagrangian integral, the information viewpoint we take leads to other types of solutions as well. Some solutions arise, for example, out of *zeroing* the integral. (See the derivation of the Dirac equation in Chapter 4.) Other laws arise out of a combination of both zeroing and extremizing the integral. Similar remarks may be made about the process by which the Lagrangians are *formed*. The zeroing and extremizing operations actually allow us to *solve for* the Lagrangians of the scenarios (see Chaps. 4–9, and 11). In this way we avoid, to a large degree, the *ad hoc* approach to Lagrange construction that is conventionally taken. This subject is discussed further in Secs. 1.1 and 1.8.8. The rationale for both zeroing and extremizing the integral is developed in Chapter 3. It is one of *information transfer* from phenomenon to data.

The layout of the book is, very briefly, as follows. The current chapter is intended to derive and exemplify mathematical techniques that the reader might not be familiar with. Chapter 1 is an introduction to the concept of Fisher information. This is for single-parameter estimation problems. Chapter 2 generalizes the concept to multidimensional estimation problems, ending with the scalar information form I that will be used thereafter in the applications Chapters 4–11. Chapter 3 introduces the concept of the "bound information" J, leading to the principle of EPI. This is derived from various points of view. Chapters 4–15 apply EPI to various measurement scenarios, in this way deriving the fundamental wave equations and distribution functions of science. Chapter 16 is a chapter-by-chapter summary of the key points made in the development. The reader in a hurry might choose to read this first, to get an idea of the scope of the approach and the phenomena covered.

0.2 Level of approach

The level of physics and mathematics that the reader is presumed to have is that of a senior undergraduate in physics. Calculus, through partial differential equations, and introductory matrix theory are presumed parts of his/her background. Some notions from elementary probability theory are also used. However, since these are intuitive in nature, the appropriate formula is usually just given, with reference to a suitable text as needed.

A cursory scan through the chapters will show that a minimal amount of prior knowledge of physical theory is actually used or needed. In fact, *this is the nature of the information approach taken* and is one of its strengths. The main physical input to each application of the approach is a simple law of invariance that is obeyed by the given phenomenon.

The overall mathematical notation that is used is that of conventional calculus, with additional matrix and vector notation as needed. Tensor notation is only used where it is a "must" – in Chaps. 6 and 11 on classical and quantum relativity, respectively. No extensive operator notation is used; this author believes that specialized notation often hinders comprehension more than it helps the student to understand theory. Sophistication *without* comprehension is definitely not our aim.

A major step of the information principle is the extremization and/or zeroing of a scalar integral. The integral has the form

$$K \equiv \int d\mathbf{x}\,\mathscr{L}[\mathbf{q}, \mathbf{q}', \mathbf{x}], \quad \mathbf{x} \equiv (x_1, \ldots, x_M), \quad d\mathbf{x} \equiv dx_1 \cdots dx_M, \quad \mathbf{q}, \mathbf{x} \text{ real},$$
$$\mathbf{q} \equiv (q_1, \ldots, q_N), \quad q_n \equiv q_n(\mathbf{x}),$$
$$\mathbf{q}'(\mathbf{x}) \equiv \partial q_1/\partial x_1, \partial q_1/\partial x_2, \ldots, \partial q_N/\partial x_M. \tag{0.1}$$

Mathematically, $K \equiv K[\mathbf{q}(\mathbf{x})]$ is a "functional," i.e., a single number that depends upon the values of one or more functions $\mathbf{q}(\mathbf{x})$ continuously over the domain of $\mathbf{x}$. Physically, K has the form of an "action" integral, whose extremization has conventionally been used to derive fundamental laws of physics (Morse and Feshbach, 1953). Statistically, we will find that K is the "physical information" of an overall system consisting of a measurer and a measured quantity. The limits of the integral are fixed and, usually, infinite. The dimension M of $\mathbf{x}$-space is usually 4 (space-time). The functions q_n of $\mathbf{x}$ are probability amplitudes, i.e., functions whose squares are probability densities. The q_n are to be found. They specify the physics of a measurement scenario. Quantity $\mathscr{L}$ is a known function of the q_n, their derivatives with respect to all the x_m, and $\mathbf{x}$. $\mathscr{L}$ is called the "Lagrangian" density (Lagrange, 1788). It also takes on the role of an information density, by our statistical interpretation.

The solution to the problem of extremizing the information K is provided by a mathematical approach called the "calculus of variations." Since the book makes extensive use of this approach, we derive it in the following.

0.3 Calculus of variations

0.3.1 Derivation of Euler–Lagrange equation

We find the answer to the lowest-dimension version $M = N = 1$ of the problem, and then generalize the answer as needed. Consider the problem of finding the single function $q(x)$ that satisfies

$$K = \int_a^b dx\, \mathscr{L}[x,\, q(x),\, q'(x)] = extrem., \qquad q'(x) \equiv dq(x)/dx. \tag{0.2}$$

A well-known example is the case $\mathscr{L} = \frac{1}{2}mq'^2 - V(q)$ of a particle of mass m moving with displacement amplitude q at time $x \equiv t$ in a known field of potential $V(q)$. We will return to this problem below.

Suppose that the solution to the given problem is the function $q(x)$ as shown in Fig. 0.1. Of course, at the endpoints (a, b) the function has the values $q(a)$, $q(b)$, respectively. Consider any finite departure $q_\varepsilon(x, \varepsilon)$ from $q(x)$,

$$q_\varepsilon(x, \varepsilon) = q(x) + \varepsilon\eta(x), \tag{0.3}$$

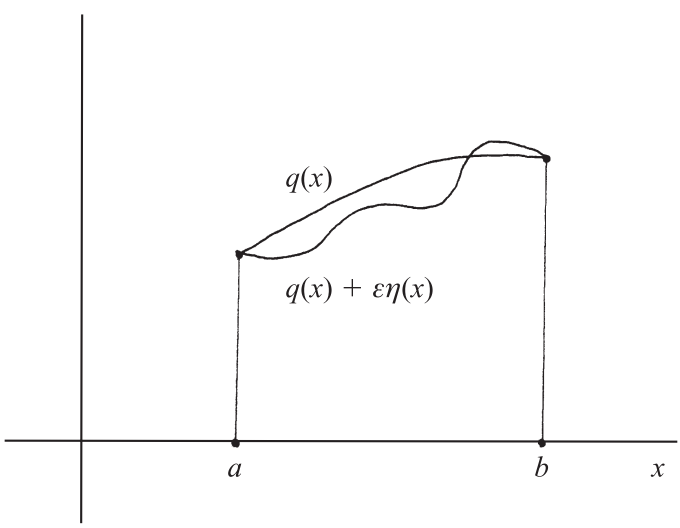

Fig. 0.1. Both the solution $q(x)$ and any perturbation $q(x) + \varepsilon\eta(x)$ from it must pass through the endpoints $x = a$ and $x = b$.

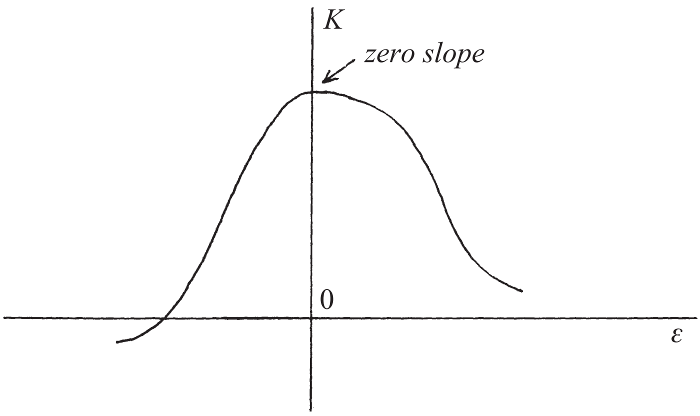

Fig. 0.2. K as a function of perturbation size parameter ε.

with ε a finite number and $\eta(x)$ any perturbing function. Any function $q_\varepsilon(x, \varepsilon)$ must pass through the endpoints so that, from Eq. (0.3),

$$\eta(a) = \eta(b) = 0. \tag{0.4}$$

Equation (0.2) is, with this representation $q_\varepsilon(x, \varepsilon)$ for $q(x)$,

$$K = \int_a^b dx\, \mathscr{L}[x, q_\varepsilon(x, \varepsilon), q'_\varepsilon(x, \varepsilon)] \equiv K(\varepsilon), \tag{0.5}$$

a function of the small parameter ε. (Once x has been integrated out, only the ε-dependence remains.)

We use ordinary calculus to find the solution. By the construction (0.3), $K(\varepsilon)$ attains the extremum value when $\varepsilon = 0$. Since an extremum value is attained there, $K(\varepsilon)$ must have zero slope at $\varepsilon = 0$ as well. That is,

$$\left.\frac{\partial K}{\partial \varepsilon}\right|_{\varepsilon=0} = 0. \tag{0.6}$$

The situation is sketched in Fig. 0.2.

We may evaluate the left-hand side of Eq. (0.6). By Eq. (0.5), $\mathscr{L}$ depends upon ε only through quantities q and q'. Therefore, differentiating Eq. (0.5) gives

$$\frac{\partial K}{\partial \varepsilon} = \int_a^b dx \left[\frac{\partial \mathscr{L}}{\partial q_\varepsilon}\frac{\partial q_\varepsilon}{\partial \varepsilon} + \frac{\partial \mathscr{L}}{\partial q'_\varepsilon}\frac{\partial q'_\varepsilon}{\partial \varepsilon}\right]. \tag{0.7}$$

The second integral is

$$\int_a^b dx\, \frac{\partial \mathscr{L}}{\partial q'_\varepsilon}\frac{\partial^2 q_\varepsilon}{\partial x\, \partial \varepsilon} = \left.\frac{\partial \mathscr{L}}{\partial q'_\varepsilon}\frac{\partial q_\varepsilon}{\partial \varepsilon}\right|_a^b - \int_a^b \frac{\partial q_\varepsilon}{\partial \varepsilon}\frac{d}{dx}\left(\frac{\partial \mathscr{L}}{\partial q'_\varepsilon}\right) dx \tag{0.8}$$

after an integration by parts. (In the usual notation, setting $u = \partial\mathscr{L}/\partial q'_\varepsilon$ and $dv = \partial^2 q_\varepsilon/\partial x\, \partial\varepsilon$.)

We now show that the first right-hand term in Eq. (0.8) is zero. By Eq. (0.3),

$$\frac{\partial q_\varepsilon}{\partial\varepsilon} = \eta(x), \tag{0.9}$$

so that by Eq. (0.4)

$$\left.\frac{\partial q_\varepsilon}{\partial\varepsilon}\right|_b = \left.\frac{\partial q_\varepsilon}{\partial\varepsilon}\right|_a = 0. \tag{0.10}$$

This proves the assertion.

Combining this result with Eq. (0.7) gives

$$\frac{\partial K}{\partial\varepsilon} = \int_a^b dx \left[\frac{\partial\mathscr{L}}{\partial q_\varepsilon}\frac{\partial q_\varepsilon}{\partial\varepsilon} - \frac{\partial q_\varepsilon}{\partial\varepsilon}\frac{d}{dx}\left(\frac{\partial\mathscr{L}}{\partial q'_\varepsilon}\right)\right]. \tag{0.11}$$

Factoring out the common term $\partial q_\varepsilon/\partial\varepsilon$, evaluating (0.11) at $\varepsilon = 0$, and using Eqs. (0.3) and (0.9) give

$$\left.\frac{\partial K}{\partial\varepsilon}\right|_{\varepsilon=0} = \int_a^b dx \left[\frac{\partial\mathscr{L}}{\partial q} - \frac{d}{dx}\left(\frac{\partial\mathscr{L}}{\partial q'}\right)\right]\eta(x). \tag{0.12}$$

By our criterion (0.6) this is to be zero at the solution q. However, the factor $\eta(x)$ is, by hypothesis, arbitrary. The only way the integral can be zero, then, is for the factor in square brackets to be zero at each x, that is,

$$\frac{d}{dx}\left(\frac{\partial\mathscr{L}}{\partial q'}\right) = \frac{\partial\mathscr{L}}{\partial q}. \tag{0.13}$$

This is the celebrated *Euler–Lagrange* solution to the problem. It is a differential equation whose solution clearly depends upon the function $\mathscr{L}$, called the "Lagrangian," for the given problem. Some examples of its use follow.

Example 1: Return to the Lagrangian given below Eq. (0.2) where $x = t$ is the independent variable. We directly compute

$$\frac{\partial\mathscr{L}}{\partial q'} = mq' \quad \text{and} \quad \frac{\partial\mathscr{L}}{\partial q} = -\frac{\partial V}{\partial q}. \tag{0.14}$$

Using this in Eq. (0.13) gives as the solution

$$mq'' = -\frac{\partial V}{\partial q}, \tag{0.15}$$

that is, Newton's second law of motion for the particle.

It may be noted that Newton's law will not be derived in this manner in the text to follow. The EPI principle is covariant, i.e., treats time and space in the same way, whereas the above approach (0.14), (0.15) is not. Instead, the EPI approach will be used to derive the more general Einstein field equation, from which Newton's law follows as a special case (the weak-field limit). Or, see Appendix D.

The reader may well question where this particular Lagrangian came from. The answer is that it was chosen merely because it "works," i.e., leads to Newton's law of motion. It has no prior significance in its own right. This has been a well-known drawback to the use of Lagrangians. The next chapter addresses this problem in detail.

Example 2: What is the shortest path between two points in a plane? The integrated arc length between points $x = a$ and $x = b$ is

$$K = \int_a^b dx\, \mathscr{L}, \quad \mathscr{L} = \sqrt{1 + q'^2}. \tag{0.16}$$

Hence

$$\frac{\partial \mathscr{L}}{\partial q'} = \tfrac{1}{2}(1 + q'^2)^{-1/2} 2q', \quad \frac{\partial \mathscr{L}}{\partial q} = 0 \tag{0.17}$$

here, so that the Euler–Lagrange Eq. (0.13) is

$$\frac{d}{dx}\left(\frac{q'}{\sqrt{1 + q'^2}}\right) = 0. \tag{0.18}$$

The immediate solution is

$$\frac{q'}{\sqrt{1 + q'^2}} = const., \tag{0.19}$$

implying that $q' = const.$, so that $q(x) = Ax + B$, with $A, B = const.$, the equation of a straight line. Hence we have shown that the path of extreme (not necessarily shortest) distance between two fixed points in a plane is a straight line. We will show below that the extremum is a minimum, as intuition suggests.

Example 3: Maximum entropy problems (Jaynes, 1957a; 1957b) have the form

$$\int dx\, \mathscr{L} = max., \quad \mathscr{L} = -p(x) \ln p(x) + \lambda p(x) + \mu p(x) f(x) \tag{0.20}$$

with λ, μ constants and $f(x)$ a known "kernel" function. The first term in the integral defines the "entropy" of a probability density function (PDF) $p(x)$. (Notice we use the notation p in place of q here.) We will say a lot more about the concept of entropy in chapters to follow. Directly