0

Introduction

0.1 Aim of the book

The overall 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 physical 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.

During the same years that quantum mechanics was being developed by Schroedinger (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. And until recent times it has 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 meas-

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

urement theory. In particular, the information aspect of 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 extreme physical information (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 either 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, after all, the science of measurement. That is, physics is a quantification of *observed* phenomena.

And observed phenomena contain noise, or fluctuations. The physical paradigm equations (defined 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 (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 reaps many benefits for purposes of *understanding* physics. EPI is, more precisely, an expression of the 'inability to know' a measured quantity. For example, quantum mechanics is derived from the viewpoint of the inability to know an ideal position. 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. 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 Chap. 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 Chap. 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. Chap. 1 is an introduction to the concept of Fisher information. This is for single-parameter estimation problems. Chap. 2 generalizes the concept to multidimensional estimation problems, ending with the scalar information form I that will be used thereafter in the applications Chaps. 4–11. Chap. 3 introduces the concept of the 'bound information' J, leading to the principle of extreme physical information (EPI). This is derived from various points of view. Chaps. 4–11 apply EPI to various measurement scenarios, in this way deriving the fundamental wave equations and distribution functions of physics. Chap. 12 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. But 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

$$K \equiv \int d\mathbf{x} \mathscr{L}[\mathbf{q}, \mathbf{q}', \mathbf{x}], \ \mathbf{x} \equiv (x_1, \ldots, x_M), \ d\mathbf{x} \equiv dx_1 \cdots dx_M, \quad \mathbf{q}, \quad \mathbf{x} \text{ real},$$

$$\mathbf{q} \equiv (q_1, \ldots, q_N), \ q_n \equiv q_n(\mathbf{x}), \ \mathbf{q}'(\mathbf{x}) \equiv \partial q_1/\partial x_1, \partial q_1/\partial x_2, \ldots, \partial q_N/\partial x_M. \tag{0.1}$$

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., 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

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

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 $q_0(x)$ as shown in Fig. 0.1. Of course at the endpoints (a, b) the function has the values $q_0(a)$, $q_0(b)$, respectively. Consider any finite departure $q(x, \varepsilon)$ from $q_0(x)$,

$$q(x, \varepsilon) = q_0(x) + \varepsilon\eta(x), \tag{0.3}$$

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

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

Eq. (0.2) is, with this representation $q(x, \varepsilon)$ for $q(x)$,

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

a function of the small parameter ε. (Once x is 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

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

The second integral is

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

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

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

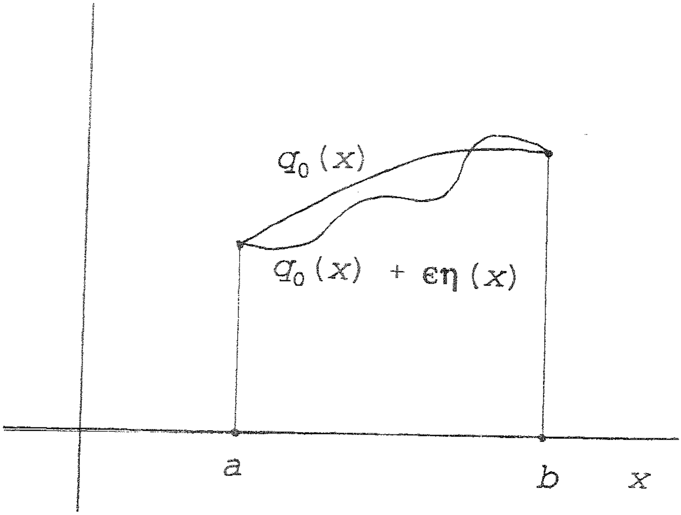

Fig. 0.1. Both the solution $q_0(x)$ and any perturbation $q_0(x) + \varepsilon\eta(x)$ from it must pass through the endpoints $x = a$ and $x = b$.

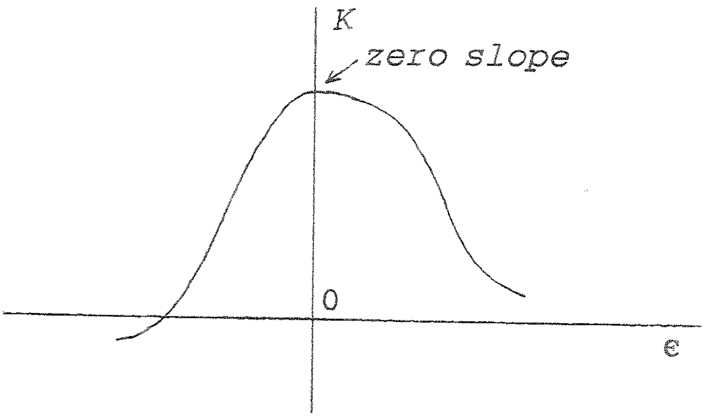

Fig. 0.2. K as a function of perturbation size parameter ε.

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

so that by Eq. (0.4)

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

This proves the assertion.

Combining this result with Eq. (0.7) gives

$$\left.\frac{\partial K}{\partial \varepsilon}\right|_{\varepsilon=0} = \int_a^b dx \left[\frac{\partial \mathscr{L}}{\partial q}\frac{\partial q}{\partial \varepsilon} - \frac{\partial q}{\partial \varepsilon}\frac{d}{dx}\left(\frac{\partial \mathscr{L}}{\partial q'}\right)\right]_{\varepsilon=0}. \tag{0.11}$$

Factoring out the common term $\partial q/\partial \varepsilon$, evaluating it at $\varepsilon = 0$ and using Eq. (0.9) gives

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

By our criterion (0.6) this is to be zero at the solution q. But 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

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

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

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

that is, Newton's 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 laws follow as a special case (the weak-field limit).

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.

1

What is Fisher information?

Knowledge of Fisher information is not part of the educational background of most physicists. Why should a physicist bother to learn about this concept? Surely the (related) concept of entropy is sufficient to describe the degree of disorder of a given phenomenon. These important questions may be answered as follows.

(a) The point made about entropy is true, but does not go far enough. Why not seek a measure of disorder whose variation *derives* the phenomenon? The concept of entropy cannot do this, for reasons discussed in Sec. 1.3. Fisher information will turn out to be the appropriate measure of disorder for this purpose.

(b) Why should a physicist bother to learn this concept? Aside from the partial answer in (a): (i) Fisher information is a *simple* and intuitive concept. As theories go, it is quite elementary. To understand it does not require mathematics beyond differential equations. Even no prior knowledge of statistics is needed: this is easy enough to learn 'on the fly'. The derivation of the defining property of Fisher information, in Sec. 1.2.3, is readily understood. (ii) The subject has very little specialized jargon or notation. The beginner does not need a glossary of terms and symbols to aid in its understanding. (iii) Most importantly, once understood, the concept gives strong payoff – one might call it 'phenomen-all' – in scope of application. It's simply worth learning.

Fisher information has two basic roles to play in theory. First, it is a measure of the ability to estimate a parameter; this makes it a cornerstone of the statistical field of study called parameter estimation. Second, it is a measure of the state of disorder of a system or phenomenon. As will be seen, this makes it a cornerstone of physical theory.

Before starting the study of Fisher information, we take a temporary detour into a subject that will provide some immediate physical motivation for it.

Ronald A. Fisher, 1929, from a photograph taken in honor of his election to Fellow of the Royal Society. Sketch by the author.

1.1 On Lagrangians

The Lagrangian approach (Lagrange, 1788) to physics has been utilized now for over 200 years. It is one of the most potent and convenient tools of theory ever invented. One well-known proponent of its use (Feynman and Hibbs, 1965) calls it 'most elegant'. However, an enigma of physics is the question of where its Lagrangians come from. It would be nice to justify and derive them from a prior principle, but none seems to exist. Indeed, when a Lagrangian is presented in the literature, it is often with a disclaimer, such as (Morse and Feshbach, 1953) 'It usually happens that the differential equations for a given phenomenon are known first, and only later is the Lagrange function found, from which the differential equations can be obtained.' Even in a case where the differential equations are *not* known, often candidate Lagrangians are first constructed, to see if 'reasonable' differential equations result.

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

R. A. Fisher (1890–1962) was a researcher whose work is not well-known to physicists. He is renowned in the fields of genetics, statistics and eugenics. Among his pivotal contributions to these fields (Fisher, 1959) are the maximum likelihood estimate, the analysis of variance, and a measure of indeterminacy now called 'Fisher information.' (He also found it likely that the famous geneticist Gregor Mendel contrived the 'data' in his famous pea plant experiments. They were too regular to be true, statistically.) It will become apparent that his form of information has great utility in physics as well.

Table 1.1 shows a list of Lagrangians (most from Morse and Feshbach, 1953), emphasizing the common presence of a squared-gradient term. In quantum mechanics, this term represents mean kinetic energy, but why mean kinetic energy should be present remains a mystery: Schroedinger called it 'incomprehensible' (Schroedinger, 1926).

Historical note: As will become evident below, *Schroedinger's mysterious Lagrangian term was simply Fisher's data information.* May we presume from this that Schroedinger and Fisher, despite developing their famous theories nearly simultaneously, and with basically just the English channel between them, never communicated? If they had, it would seem that the mystery should have been quickly dispelled. This is an enigma.

What we will show is that, in general, the squared gradient represents a phenomenon that is natural to all fields, i.e., *information*. In particular, it is the amount of Fisher information residing in a variety of data called *intrinsic data*.

Table 1.1. *Lagrangians for various physical phenomena. Where do these come from and, in particular, why do they all contain a squared gradient term? (Reprinted from Frieden and Soffer, 1995.)*

Phenomenon	Lagrangian
Classical Mech.	$\frac{1}{2}m\left(\frac{\partial q}{\partial t}\right)^2 - V$
Flexible String or Compressible Fluid	$\frac{1}{2}\rho\left[\left(\frac{\partial q}{\partial t}\right)^2 - c^2\nabla q\cdot\nabla q\right]$
Diffusion Eq.	$-\nabla\psi\cdot\nabla\psi^* - \ldots$
Schrödinger W. E.	$-\frac{\hbar^2}{2m}\nabla\psi\cdot\nabla\psi^* - \ldots$
Klein–Gordon Eq.	$-\frac{\hbar^2}{2m}\nabla\psi\cdot\nabla\psi^* - \ldots$
Elastic W. E.	$\frac{1}{2}\rho\dot{q}^2 - \ldots$
Electromagnetic Eqs.	$4\sum_{n=1}^{4}\Box q_n\cdot\Box q_n - \ldots$
Dirac Eqs.	$-\frac{\hbar^2}{2m}\nabla\psi\cdot\nabla\psi^* - \ldots = 0$
General Relativity (Eqs. of motion)	$\sum_{m,n=1}^{4} g_{mn}(q(\tau))\frac{\partial q_m}{\partial\tau}\frac{\partial q_n}{\partial\tau}$ ↑ metric tensor
Boltzmann Law	$4\left(\frac{\partial q(\mathrm{E})}{\partial \mathrm{E}}\right)^2 - \ldots,\ p(\mathrm{E}) \equiv q^2(\mathrm{E})$
Maxwell–Boltzmann Law	$4\left(\frac{\partial q(v)}{\partial v}\right)^2 - \ldots,\ p(v) \equiv q^2(v)$
Lorentz Transformation (special relativity)	$\partial_i q_n \partial_i q_n$ (invariance of integral)
Helmholtz W. E.	$-\nabla\psi\cdot\nabla\psi^* - \ldots$

The remaining terms of the Lagrangian will be seen to arise out of the information residing in the *phenomenon* that is under measurement. Thus, all Lagrangians consist entirely of two forms of Fisher information – data information and phenomenological information.

The concept of Fisher information is a natural outgrowth of classical measurement theory, as follows.

1.2 Classical measurement theory

1.2.1 The 'smart' measurement

Consider the basic problem of estimating a single parameter of a system (or phenomenon) from knowledge of some measurements. See Fig. 1.1. Let the parameter have value θ, and let there be N data values $y_1, \ldots, y_N \equiv \mathbf{y}$ in vector notation, at hand. The system is specified by a conditional probability law $p(\mathbf{y}|\theta)$ called the 'likelihood law'.

The data obey $\mathbf{y} = \theta + \mathbf{x}$, where the $x_1, \ldots, x_N \equiv \mathbf{x}$ are added noise values. The data are used in an estimation principle to form an estimate of θ which is an *optimal* function $\hat{\theta}(\mathbf{y})$ of all the data; e.g., the function might be the sample mean $N^{-1}\sum_n y_n$. The overall measurement procedure is 'smart' in that $\hat{\theta}(\mathbf{y})$ is on average a better estimate of θ than is any one of the data observables.

The noise $\mathbf{x}$ is assumed to be *intrinsic* to the parameter θ under measurement. For example, θ and $\mathbf{x}$ might be, respectively, the ideal position and quantum fluctuations of a particle. Data $\mathbf{y}$ are, correspondingly, called *intrinsic data*. No additional noise effects, such as noise of detection, are assumed present here. (We later allow for such additional noise in Sec. 3.8 and Chap. 10.) The system consisting of quantities $\mathbf{y}$, θ, $\mathbf{x}$ is a *closed*, or physically isolated, one.

1.2.2 Fisher information

This information arises as a measure of the expected error in a smart measurement. Consider the class of 'unbiased' estimates, obeying $\langle \hat{\theta}(\mathbf{y}) \rangle = \theta$; these are correct 'on average'. The mean-square error e^2 in such an estimate $\hat{\theta}$ obeys a relation (Van Trees, 1968; Cover and Thomas, 1991)

$$e^2 I \geqslant 1, \tag{1.1}$$

where I is called the Fisher 'information'. In a particular case of interest $N = 1$ (see below), this becomes

$$I = \int dx\, p'^2(x)/p(x), \; p' \equiv dp/dx. \tag{1.2}$$

(Throughout the book, integration limits are infinite unless otherwise specified.) Quantity $p(x)$ denotes the probability density function for the noise value

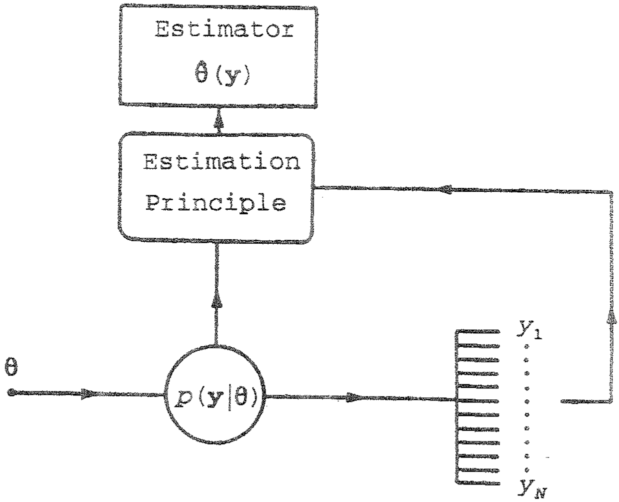

Fig. 1.1. The parameter estimation problem of classical statistics. An unknown but fixed parameter value θ causes intrinsic data y through random sampling of a likelihood law $p(\mathbf{y}|\theta)$. Then, the random likelihood law and the data are used to form the estimator $\theta(\mathbf{y})$ via an estimation principle. (Reprinted from Frieden, 1991, by permission of Springer-Verlag Publishing Co.)

x. If $p(x)$ is Gaussian, then $I = 1/\sigma^2$ with σ^2 the variance (see derivation in Sec. 8.3.1).

Eq. (1.1) is called the Cramer–Rao inequality. It expresses *reciprocity* between the mean-square error e^2 and the Fisher information I in the intrinsic data. Hence, it is an expression of *intrinsic* uncertainties, i.e., in the absence of outside sources of noise. It will be shown at Eq. (4.53) that the reciprocity relation goes over into the Heisenberg uncertainty principle, in the case of a single measurement of a particle position value θ. Again, this ignores the possibility of noise of detection, which would add in additional uncertainties to the relation (Arthurs and Goodman, 1988; Martens and de Muynck, 1991).

The Cramer–Rao inequality (1.1) shows that estimation quality increases (e decreases) as I increases. Therefore, I is a quality metric of the estimation procedure. This is the essential reason why I is called an 'information'. Eqs. (1.1) and (1.2) derive quite easily, shown next.

1.2.3 Derivation

We follow Van Trees (1968). Consider the class of estimators $\hat{\theta}(\mathbf{y})$ that are unbiased, obeying

$$\langle\hat{\theta}(\mathbf{y}) - \theta\rangle \equiv \int d\mathbf{y}[\hat{\theta}(\mathbf{y}) - \theta]p(\mathbf{y}|\theta) = 0. \tag{1.3}$$

Probability density function (PDF) $p(\mathbf{y}|\theta)$ describes the fluctuations in data values $\mathbf{y}$ in the presence of the parameter value θ. PDF $p(\mathbf{y}|\theta)$ is called the 'likelihood law'. Differentiate Eq. (1.3) $\partial/\partial\theta$, giving

$$\int d\mathbf{y}(\hat{\theta} - \theta)\frac{\partial p}{\partial\theta} - \int d\mathbf{y}\, p = 0. \tag{1.4}$$

Use the identity

$$\frac{\partial p}{\partial\theta} = p\frac{\partial \ln p}{\partial\theta} \tag{1.5}$$

and the fact that p obeys normalization. Then Eq. (1.4) becomes

$$\int d\mathbf{y}(\hat{\theta} - \theta)\frac{\partial \ln p}{\partial\theta}p = 1. \tag{1.6}$$

Factor the integrand as

$$\int d\mathbf{y}\left[\frac{\partial \ln p}{\partial\theta}\sqrt{p}\right][(\hat{\theta} - \theta)\sqrt{p}] = 1. \tag{1.7}$$

Square the equation. Then the Schwarz inequality gives

$$\left[\int d\mathbf{y}\left(\frac{\partial \ln p}{\partial\theta}\right)^2 p\right]\left[\int d\mathbf{y}(\hat{\theta} - \theta)^2 p\right] \geqslant 1. \tag{1.8}$$

The left-most factor is defined to be the Fisher information I,

$$I \equiv I(\theta) \equiv \int d\mathbf{y}\left(\frac{\partial \ln p}{\partial\theta}\right)^2 p,\ p \equiv p(\mathbf{y}|\theta), \tag{1.9}$$

while the second factor exactly defines the mean-squared error e^2,

$$e^2 \equiv \int d\mathbf{y}[\hat{\theta}(\mathbf{y}) - \theta]^2 p. \tag{1.10}$$

This proves Eq. (1.1).

It is noted that $I = I(\theta)$ in Eq. (1.9), i.e., in general I depends upon the (fixed) value of parameter θ. But note the following important exception to this rule.

1.2.4 *Important case of shift invariance*

Suppose that there is only $N = 1$ data value taken so that $p(\mathbf{y}|\theta) = p(y|\theta)$. Also, suppose that the PDF obeys a property

$$p(y|\theta) = p(y - \theta). \tag{1.11}$$

This means that the fluctuations in y from θ are invariant to the size of θ, a kind of shift invariance. (This becomes an expression of *Galilean invariance* when random variables y and θ are 3-vectors instead.) Using condition (1.11) and identity (1.5) in Eq. (1.9) gives

$$I = \int dy \left[\frac{\partial p(y-\theta)}{\partial (y-\theta)} \right]^2 \bigg/ p(y-\theta), \tag{1.12}$$

since $\partial/\partial\theta = -\partial/\partial(y-\theta)$. Parameter θ is regarded as fixed (see above), so that a change of variable $x = y - \theta$ gives $dx = dy$. Equation (1.12) then becomes Eq. (1.2), as required. Note that I no longer depends upon θ. This is convenient since θ was unknown.

1.3 Comparisons of Fisher information with Shannon's form of entropy

A related quantity to I is the Shannon entropy (Shannon, 1948) H (called Shannon 'information' in this book). This has the form

$$H \equiv -\int dx p(x) \ln p(x). \tag{1.13}$$

Like I, H is a functional of an underlying probability density function (PDF) $p(x)$. Historically, I predates the Shannon form by about 25 years (1922 vs. 1948). There are some known relations connecting the two information concepts (Stam, 1959; Blachman, 1965; Frieden, 1991) but these are not germane to our purposes. H can be, but is not always, the thermodynamic, Boltzmann entropy.

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

1.3.1 Global vs. local nature

For our purposes, it is useful to work with a discrete form of Eq. (1.13),

$$H = -\Delta x \sum_n p(x_n) \ln p(x_n) \equiv \delta H, \; \Delta x \to 0. \tag{1.14}$$

(Notation δH emphasizes that Eq. (1.14) represents an *increment* in information.) Of course, the sum in Eq. (1.14) may be taken in any order. Graphically, this means that if the curve $p(x_n)$ undergoes a rearrangement of its points $(x_n, p(x_n))$, although the shape of the curve will drastically change the value of

H remains constant. H is then said to be a *global* measure of the behavior of $p(x_n)$.

By comparison, the discrete form of Fisher information I is, from Eq. (1.2),

$$I = \Delta x^{-1} \sum_n \frac{[p(x_{n+1}) - p(x_n)]^2}{p(x_n)}. \tag{1.15}$$

If the curve $p(x_n)$ undergoes a rearrangement of points x_n as above, discontinuities in $p(x_n)$ will now occur. Hence the local slope values $[p(x_{n+1}) - p(x_n)]/\Delta x$ will change drastically, and so the sum (1.15) will also change strongly. Since I is thereby sensitive to local rearrangement of points, it is said to have a property of *locality*.

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

This global vs. local property has an interesting ramification. Because the integrand of I contains a squared derivative p'^2 (see Eq. (1.2)), when the integrand is used as part of a Lagrangian the resulting Euler–Lagrange equation will contain second-order derivative terms p''. Hence, a second-order differential equation results (see Eq. (0.25)). This dovetails with nature, in that the major fundamental differential equations that define probability densities or amplitudes in physics are second-order differential equations. Indeed, the thesis of this book is that the correct differential equations result when the information I-based EPI principle of Chap. 3 is followed.

By contrast, the integrand of H in (1.13) does not contain a derivative. Therefore, when this integrand is used as part of a Lagrangian the resulting Euler–Lagrange equation will not contain any derivatives (see Eq. (0.22)); it will be an algebraic equation, with the immediate solution that $p(x)$ has the exponential form Eq. (0.22) (Jaynes, 1957a,b). This is not, then, a differential equation, and hence cannot represent a general physical scenario. The exceptions are those distributions which happen *to be* of an exponential form, as in statistical mechanics. (In these cases, I gives the correct solutions anyhow; see Chap. 7.)

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

As all of the preceding discussion implies, H and I are two distinct functionals of $p(x)$. However, quite the contrary is true in comparing I with an entropy that is closely related to H, namely, the Kullback–Leibler entropy. This is discussed in Sec. 1.4.

tion I must have a *physical equivalence* in terms of the measured phenomenon (as information H and physical entropy H_B were related in the preceding example). Represent by J the physical effect, or manifestation, of I. Thus, J is the information that is intrinsic (or, 'bound') to the phenomenon under measurement. This is why we call J the 'bound' Fisher information. Computationally, it is to be evaluated – not directly by (2.19) – but equivalently in terms of physical parameters of the scenario (more on this later). The precise tie-in between I and J is quantified as follows.

3.3.2 *Information transferral effects*

Suppose that a real measurement of a four-parameter $\boldsymbol{\theta}$ is made. The parameter is any physical characteristic of a system. The system is necessarily *perturbed* (Sec. 3.8, Chap. 10) by the measurement. More specifically, its amplitude functions $\mathbf{q}(\mathbf{x})$ are perturbed. An analysis follows.

As at Eq. (0.3), denote the perturbations in $\mathbf{q}(\mathbf{x})$ as the vector $\varepsilon\boldsymbol{\eta}(\mathbf{x})$, $\varepsilon \to 0$, with the $\boldsymbol{\eta}(\mathbf{x})$ arbitrary perturbation functions. By Eq. (2.19), the perturbations in $\mathbf{q}(\mathbf{x})$ cause a perturbation δI in the intrinsic information (see Eq. (0.41)). Since J will likewise be a functional of the $\mathbf{q}(\mathbf{x})$, a perturbation δJ also results. Is there a relation between δI and δJ?

We next find such a relation, and others, for informations I and J. This is by the use of an *axiomatic model* for the measurement process. There are three axioms as given below.

Prior to measurement of parameter $\boldsymbol{\theta}$, the system has a bound information level J. Then a real measurement $\bar{\mathbf{y}}$ is initiated – say, by shining light upon the system. This causes an information transformation or transition $J \to I$ to take place, where I is the intrinsic information of the system. The measurement perturbs the system (Sec. 3.8), so that informations I and J are perturbed by amounts δI and δJ. How are the four informations related?

First, by the correspondence (3.11) and the Brillouin effect (3.5), I and J should obey

$$J \geqslant I. \tag{3.12}$$

This suggests the possibility of some loss of information during the information transition. We also postulate the perturbed amounts of information to obey

$$\textit{axiom } 1\colon \delta I = \delta J. \tag{3.13}$$

This is a *conservation law*, one of conservation of information change. It is a predicted new effect. Its validity is demonstrated many times over in this book; as a vital part of the information approach that is used to derive the fundamental physical laws in Chaps. 4–9 and 11. It also will be seen to occur whenever the measurement space has a conjugate space that connects with it by unitary transformation (Sec. 3.8).

The basic premises (3.12), (3.13) are further discussed in Secs. 3.4.5 and 3.4.7, and are exemplified in Sec. 3.8. We next build on these premises.

3.4 Principle of extreme physical information (EPI)

3.4.1 Physical information K and variational principle

Eq. (3.13) is a statement that

$$\delta I - \delta J \equiv \delta(I - J) = 0. \tag{3.14}$$

Define a new quantity.

$$K \equiv I - J. \tag{3.15}$$

Quantity K is called the 'physical information' of the system. As with I and J, it is a functional of the amplitudes $\mathbf{q}(\mathbf{x})$. By its definition and Eq. (3.12), K is a loss, or defect, of information. The zero in Eq. (3.14) implies that this defect of information is the same over a *range* of perturbations $\varepsilon\boldsymbol{\eta}(\mathbf{x})$. Eqs. (3.14) and (3.15) are the basis for the EPI principle, Eqs. (3.16) and (3.18), as shown next.

Combining Eqs. (3.14) and (3.15) gives $\delta K = 0$ or

$$K = I - J = Extrem. \tag{3.16}$$

This is a variational principle for finding the $\mathbf{q}$ (see Sec. 0.3). It states that at the solution $\mathbf{q}$ the information K is an extremum. As was postulated at Eq. (3.13), the variational principle is in reaction to a measurement. Variational principle (3.16) is one-half of our overall information approach.

3.4.2 Zero-conditions, on micro- and macrolevel

Eq. (3.15) shows that generally

$$I \neq J. \tag{3.17}$$

In fact, from inequality (3.12), a 'lossy' situation exists: the information I in the data never exceeds the amount J in the phenomenon,

$$I = \kappa J \text{ or } I - \kappa J = 0, \ \kappa \leqslant 1. \tag{3.18}$$

This may be shown to follow from the I-theorem; see Sec. 3.4.7. Eq. (3.18) comprises the second half of our overall information approach.

Interestingly, only the values $\kappa = 1/2$ or 1 have occurred in those physical-law derivations that fix κ at specific values (see subsequent chapters). These verify that $I \leqslant J$ or, equivalently, $\kappa \leqslant 1$ in (3.18). The nature of κ is taken up further in Sec. 3.4.5.

Returning to the question of information J, Eq. (3.18) provides a partial answer. It shows that J is proportional to I, in numerical value. However, as mentioned before, in contrast to I its *functional form* will depend upon physical parameters of the scenario (e.g., c or $\hbar$) and not solely upon the amplitude functions $\mathbf{q}$. This is further taken up in Sec. 3.4.5.

Equation (3.18) is a zero-condition on the *macroscopic* level, i.e., in terms of the information in physical observables (the data). On the other hand, (3.18) may be derived from viewing the information transfer procedure on the *microscopic* level. Postulate the existence of *information densities* $i_n(\mathbf{x})$ and $j_n(\mathbf{x})$ such that

$$\textit{axiom 2: } I \equiv \int d\mathbf{x} \sum_n i_n(\mathbf{x}) \text{ and } J \equiv \int d\mathbf{x} \sum_n j_n(\mathbf{x}), \tag{3.19}$$

$$\text{where } i_n(\mathbf{x}) = 4\nabla q_n \cdot \nabla q_n$$

by Eq. (2.19). Also, $j_n(\mathbf{x})$ is a function that depends upon the particular scenario (see subsequent chapters). In contrast with I and J, densities $i_n(\mathbf{x})$ and $j_n(\mathbf{x})$ exist on the microscopic level, i.e., within local intervals $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$.

Let us demand a zero-condition on the microscopic level,

$$\textit{axiom 3: } i_n(\mathbf{x}) - \kappa j_n(\mathbf{x}) = 0, \text{ all } \mathbf{x}, n. \tag{3.20}$$

Summing and integrating this $d\mathbf{x}$, and using Eqs. (3.19), verifies Eq. (3.18), which was our aim.

Condition (3.20) will often be used to implement physical-law derivations in later chapters. It is to be used in the sense that quantities $[i_n(\mathbf{x}) - \kappa j_n(\mathbf{x})]$ comprise the *net* integrands in Eq. (3.19) after possible partial integrations are performed. By $[i_n(\mathbf{x}) - \kappa j_n(\mathbf{x})]$ we mean the terms that remain.

Information densities $i_n(\mathbf{x})$ and $j_n(\mathbf{x})$ have the added physical significance of defining the field dynamics of the scenario; see Sec. 3.7.3.

3.4.3 EPI principle

In summary, the information model consists of the axioms (3.13), (3.19) and (3.20). As we saw, these imply the variational principle (3.16) and the zero-condition (3.18). In fact, *the variational principle and the zero-condition comprise the overall principle we will use to derive major physical laws*. Since, by Eq. (3.16), physical information K is extremized the principle is called 'extreme physical information' or EPI. In some derivations, the microlevel condition Eq. (3.20) (which, we saw, implied condition (3.18)) will be used as well.

value. On this basis, the number of amplitude components q needed to describe a given physical phenomenon could conceivably be of unlimited value. However, the value of N will also describe the number of mathematical solutions *that are sufficient* for satisfying EPI conditions (3.16) and (3.18). Typically, this is the number of independent solutions to a set of Euler–Lagrange equations. In any physical circumstance this will be a well-defined, finite number. For example, $N = 2$ for quantum mechanics as described by the Klein–Gordon equation, or, $N = 8$ for the Dirac equation (see Chap. 4). In this way, each physical phenomenon will be specified by a definite, finite number N of amplitude functions.

3.4.11 *A mathematical game*

In certain scenarios the EPI principle may be regarded as the workings of a 'mathematical game' (Morgenstern and von Neumann, 1947). In fact the simplest such game arises. This is a game with *discrete*, deterministic moves i and j defined below. The transition to continuous values of i and j may be made without changing the essential results.

Those two arch-rivals A and B are vying for a scarce commodity (information). See Table 3.1. Player A can make either of the row value moves $i = 1$ or 2; and player B can make either of the column value moves $j = 1$ or 2. Each move pair (i, j) constitutes a play of the game. Each such move pair defines a location in the table, and a resulting payout of the commodity to A, *at the expense of* B. ('Thus A is happy, and B is not.') For example, the move (1, 2) denotes a gain by A of 2.0 and a loss by B of 2.0. Such a game is called 'zero-sum', since at any move the total gain of both players is zero. Assume that both players know the payoff table. What are their respective optimum moves?

Consider player A. If he chooses a row $i = 1$ then he can gain either 3.0 or 2.0. Since these are at the expense of B, B will choose a column $j = 2$ so as to minimize his loss to 2.0. So, A knows that if he chooses $i = 1$ he will gain 2.0. Or, if A chooses the row $i = 2$ then he will gain, in the same way, 4.0. Since this exhausts his choices, to maximize his gain he chooses $i = 2$, with the certainty that he will gain 4.0.

Next, consider player B. If he chooses a column $j = 1$ then he will pay out either 3.0 or 5.0. But then A will choose $i = 2$ so as to maximize his gain to 5.0. So, if B chooses $j = 1$ he will lose 5.0. Or, if B chooses column $j = 2$ then he will lose, in the same way, 4.0. Since this exhausts his choices, to minimize his loss he chooses $j = 2$, with the certainty that he will lose 4.0.

Notice that the resulting optimum play of the game, from *either* the viewpoint of A or B, is the move (2, 2). This is the extreme lower-right item in the

Table 3.1. *A* 2×2 *payoff matrix.*

i \ j	1	2
1	3.0	2.0
2	5.0	4.0

table. This item results because the point (2, 2) is locally a *saddle point* in the payouts, i.e., a local minimum in j and a maximum in i. Also, note that there is nothing random about this game; every time it is played the result will be the same. Such a game is called 'fixed-point'.

3.4.12 EPI as a game of knowledge acquisition

We next construct a mathematical model of the EPI process that has the form of such a mathematical game (Frieden and Soffer, 1995). Note that this game model is an explanatory device, and not a distinct, physical derivation of EPI. Also, delegating human-like attributes to the players is merely part of the anthropomorphic model taken. The 'game that is played' is merely a descriptive device. EPI is a physical process (Chap. 10).

The game model can also be regarded as an epistemological model of EPI. It will show that EPI arises, in certain circumstances, *as if it were* the result of a quest for information and knowledge.

In many problems the Fisher coordinates are mixed real and imaginary quantities. Denote by $\mathbf{x}$ the subset of real coordinates of the problem. We want to deal, in this section, with problems having only real coordinates. Thus, let the amplitudes $\mathbf{q}(\mathbf{x})$ of the problem represent either the fully dimensioned amplitudes for an all-real case, or the *marginal* probability amplitudes in the real $\mathbf{x}$ for a mixed case. Eq. (2.19) shows that, for real coordinates $\mathbf{x}$, I monotonically *decreases* as the amplitudes $\mathbf{q}$ are monotonically broadened or blurred. This effect should hold for any fixed state of correlation in the intrinsic data.

On the other hand, we found (Sec. 2.3.2) that the form Eq. (2.19) for I represents a model scenario of *maximized* information due to efficient estimation and independent data. The latter effect is illustrated by the following example.

Example Suppose that there are two Fisher variables (x, y) and these obey a Gaussian bivariant PDF, with a common variance σ^2, a *general correlation coefficient* ρ, and a common mean value θ. Regard the latter as the unknown parameter to be estimated. Then the Fisher information Eq. (1.9) gives $I = 2\sigma^{-2}(1+\rho)^{-1}$. This shows how the information in the variables (x, y) depends upon their degree of correlation. Suppose that the variables are initially correlated, with $\rho > 0$. Then as $\rho \to 0$, $I \to 2/\sigma^2 = \text{max. in } \rho$. As a check, this is twice the information in a single variable (see Sec. 1.2.2), as is required by the additivity of independent data (Sec. 2.4.1) in the case of independent data. Hence, independent data have maximal I values.

This effect holds in the presence of any fixed state of blur, or half-width, of the joint amplitude function $\mathbf{q}(\mathbf{x})$ of all the data fluctuations.

We have, then, determined the qualitative dependence of I upon the data correlation and the state of blur. This is summarized in Fig. 3.2. As in Table 3.1, the vertical and horizontal coordinates are designated by i and j, respectively, although they are continuous variables here. Coordinate i increases with the degree of independence (by any measure) of the intrinsic data, and each coordinate j represents a possible trial solution $\mathbf{q}(\mathbf{x})$ to EPI in the particular sequence defined below. Since, as we saw, the I values increase with increasing independence, they must increase with increasing i.

Now, the solution $\mathbf{q} = \mathbf{q}_0$ to EPI is given by the simultaneous solution to Eqs. (3.16) and (3.18). Also, Eq. (2.19) for I presumes maximal independence (vertical coordinate) of the intrinsic data (Sec. 2.1.3). Then the EPI solution $\mathbf{q}_0$ is represented by a particular coordinate value j located along the *bottom row* of Fig. 3.2.

The trial solutions $\mathbf{q}$ are sequenced according to the coordinate j as follows. Let each $q_n(\mathbf{x})$ function of a given vector $\mathbf{q}$ monotonically decrease in blur from the solution 'point' $j \to \mathbf{q}_0(\mathbf{x})$ on the far right to an initial sharp state at the far left, corresponding to $j = 1$. Then, by the form Eq. (2.19) of I, the values I decrease along that row as j *increases* to the right. On the other hand, we found before that, for a given j, values I increase with increasing i (independence) values. Then the solution point $\mathbf{q}_0$ designates a local maximum in i but a minimum in j, i.e., a *saddle point*.

We found, by analysis of the game in Table 3.1, that a saddle point represents the outcome of a zero-sum mathematical game. Hence, the EPI solution point $\mathbf{q}_0$ represents the outcome of such a game. Here, the commodity that the two players are vying for is the amount of Fisher information I in a trial solution $\mathbf{q}$. Then the game is one of Fisher 'information hoarding' between a player A, who controls the vertical axis choice of *correlation*, and a player B, who controls the horizontal axis choice of *the degree of blur*. Who are the opposing players?

The choice of correlation is made by the *observer* (Sec. 2.1.3) in a prior scenario, so this identifies player A. Player B is associated with the degree of

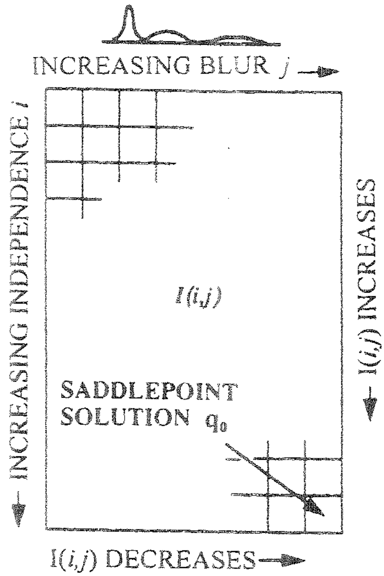

Fig. 3.2. The information game. The observer bets on a state of independence i (a row value). The information demon bets on a state of blur j (a column value). The payoff from the demon to the observer is the information at the saddlepoint, defining the EPI solution q_0. (Reprinted from Frieden and Soffer, 1995.)

blur. This is obviously out of the hands of the observer. Therefore it is adjusted by 'nature'. This may be further verified by the zero-sum nature of the game. By Eqs. (3.13) or (3.18) we notice that any information I that is acquired by the observer is at the expense of the phenomenon under measurement. Hence, player B is identified to be the phenomenon, or, nature in general. Then, according to the rules of such a game, nature has the 'aim' of increasing the degree of blur.

Hence, for real coordinates $\mathbf{x}$ the EPI principle represents a game of information hoarding between the observer and nature. The observer wants to maximize I while nature wants to minimize it. Each different physical scenario, defined by its dependence $J[\mathbf{q}]$, defines a particular play of the game, leading to a different saddle point solution $\mathbf{q}_0$ along the bottom row of Fig. 3.2. This is the game-theory aspect of EPI that we sought.

Such real coordinates $\mathbf{x}$ occur in the Maxwell–Boltzmann law derivation in Chap. 7, in the time-independent Schroedinger wave equation derivation in Appendix D, and in the $1/f$ power-law derivation in Chap. 8. These EPI derivations are, then, equivalent to a play of the knowledge acquisition game.

It is interesting that, because the information transfer efficiency factor $\kappa \leqslant 1$ in Eq. (3.18), $I \leqslant J$ or $K \leqslant 0$. This indicates that, of the two protagonists, nature always wins or at least breaks even in the play of the game. This effect agrees in spirit with Brillouin's equivalence principle Eq. (3.5). That is, in acquiring data information there is an inevitable *larger* loss of phenomenological ('bound' in his terminology) information. You can't win.

Exercise

The scope of application of the game can be widened. Suppose that a particle has a complex amplitude function $\psi(r, t) \equiv q_1(r, t) + iq_2(r, t)$, by the use of Eq. (2.24). Suppose that the particle is subjected to a conservative force field, due to a time-independent potential function $V(r)$. Then energy is conserved, and we may regard the particle as being in a definite energy state E. It results that the amplitude function separates, as $\psi(r, t) = \mathrm{u}(r) \exp(iEt/\hbar)$. Suppose that all time fluctuations t are constrained to lie within a *finite*, albeit large, interval $(-\Delta T, \Delta T)$. Show that, under these circumstances, the time-coordinate contribution to information I is merely an additive constant, and that the space coordinates give rise to positive information terms, so that the game may be played with the space coordinates. Hence, the knowledge game applies, more broadly, to the scenario of a quantum mechanical particle in a conservative force field.

3.4.13 The information 'demon'

The solution q_0 to the game defines a physical law (e.g., the Maxwell–Boltzmann law). The preceding shows, then, that a physical law is the result of a lossy information transfer game between an observer and the phenomenon under observation. Since we can sympathize with the observer's aims of increased information and knowledge, and since he always loses the game to the phenomenon (or at best breaks even), it seems fitting to regard the phenomenon as an all-powerful, but malevolent, force: an information 'demon'. (Note: this demon is not the Maxwell demon.)

In summary, for real Fisher coordinates $\mathbf{x}$ the EPI process amounts to carrying through a game. The game is a zero-sum contest between the observer and the information demon for a limited resource I of intrinsic information. The demon generally wins.

We note with interest that the existence of such a demon implies the I-theorem, Eq. (1.30). (This is the converse of the proof in Sec. 3.4.7.) Information $I - J = K$ represents the change of Fisher information ΔI due to the EPI process $J \rightarrow I$ that takes place over an interaction time interval Δt. Since the game is played and the demon wins, i.e., $K \leqslant 0$, and since $\Delta I = K$, necessarily $\Delta I \leqslant 0$. Then since $\Delta t \geqslant 0$, the I-theorem follows.

In general, the re-expression of a $J[\phi(\mu)]$ in X-space hinges on two effects: (i) that $I[\phi(\mu)]$ is a statistical average, and statistical averages can be re-expressed in various spaces; and (ii) the quantity being averaged – μ^2 – has an equivalent form in X-space, here, essentially the kinetic energy. Such a pair of effects also hold in the derivation of quantum gravitational effects in Chap. 11.

It is not obvious that the 'equivalence' effect (ii) holds in general, i.e., for any unitary transformation of physical coordinates. Effect (ii) is not satisfied, e.g., by merely re-transforming $I[\phi(\mu)]$ back to $I[\psi(\mathbf{x})]$ via the known unitary transformation. In this case the two functionals I and J become identical and, so, information $K = 0$ identically (for *all* choices of amplitude functions $\mathbf{q}$). This is a mere tautology.

Instead, equivalence effect (ii) must be a distinct, *physical input* into the problem, as in Appendix D and Chap. 4 (Secs. 4.1.15, 4.1.16).

However, in some measurement scenarios there is not an obvious unitary transformation connecting X-space with another space. Such cases occur in Chaps. 5–9. But even in such cases there is still an invariance principle of some kind that is obeyed by the measured particle. An example is continuity of flow (see later chapters). In this situation the EPI procedure may not be directly deduced, as here, but rather is taken to rest upon the axiomatic approach of Secs. 3.3.2 and 3.4.2.

As will be seen, the EPI process is shaped, or constrained, by the form of the particular invariance principle for the scenario. This may be shown explicitly for phenomena that obey the knowledge 'game' of Fig. 3.2. All EPI solutions lie along the bottom row, and each phenomenon has a generally different solution point along that row. Each such solution point is defined by its invariance principle.

Obviously, there is an intimate connection between statistical unitary transformations and the EPI approach. More work needs to be done on exploring the various physical unitary transformations, and their implications via the EPI principle.

3.9 EPI as a state of knowledge

According to EPI, there is a hierarchy of *physical knowledge* present. At *the top* are:

(A) the Fisher I-theorem (Sec. 1.8.2), which states that I, like entropy H_B, is a physical entity that *monotonically* changes with time and, also, can be transferred, or can 'flow', from one system to another (Sec. 1.8.10);

(B) the concept of a level J of Fisher information that is intrinsic to, or 'bound' to, each phenomenon (Secs. 3.3.1, 3.4.5); and
(C) the invariance, or symmetry, principle (Sec. 3.4.5) governing each phenomenon.

The laws (A)–(C), which we call the 'top laws', exist prior to, or independent of, any explicit measurements. They can possibly be *verified* (or nullified) by measurement, but that's another matter.

At the second rung down the knowledge ladder are the *three axioms*:

(i) Conservation of information perturbation, Eq. (3.13), during a measurement;
(ii) Eq. (3.19) defining information densities $i_n(\mathbf{x})$, $j_n(\mathbf{x})$ on the microlevel; and
(iii) Eq. (3.20) governing the efficiency of information transition, on the microlevel, from phenomenon to intrinsic data.

At the third rung down the ladder is the EPI principle. This follows (as we found) from either the axioms or from the existence of a physically meaningful unitary transformation space.

Finally, at the fourth rung down the ladder, is the carrying through of EPI as a *calculation*. This requires the EPI principle, as augmented by top law (C). The output of the calculation is the law governing formation of the amplitudes $\mathbf{q}$ for that scenario. For example, in Chap. 4 it is the Klein–Gordon 'law' governing formation of the amplitude ψ.

The question of what should be regarded as the laws of physics is of interest. Should they, e.g., be the 'top' laws (A)–(C) mentioned above, or, as is conventionally assumed, the output laws, such as the Klein–Gordon equation? We can expect, and the chapters ahead will verify, that some invariance principles (C) do double (or more) duty in implying physical laws. For example, the continuity of flow condition is used by EPI to derive both Maxwell's equations (Chap. 5) and the Einstein field equations (Chap. 6). Therefore, there are more physical laws than there are invariance conditions (C) for their derivation. Clearly it is desirable to have to make the fewest assumptions about nature. On this basis, the EPI output laws can be regarded as subsidiary to the top laws. They are also subsidiary in being subject to a contingency situation – measurement – for their existence, as is clarified next.

3.10 EPI as a physical process

The *physical picture* that is provided by EPI should also be considered. We postulate that if real data are at hand, they must have been caused by a *physical process*. The EPI view is that an output law is part of an ongoing physical process that includes the measurement step as its activator. (In this sense, the measurement 'creates' the probability law from which it is sampled. Imagine that!)

The measurement must be a real one upon a real object, say, a particle. The measurement physically activates the three axioms (or the unitary transformation) and, subsequently, EPI as a continuation of the process. In the *absence* of a real measurement upon a real object, the process is not activated so that the output law does not *physically* occur. (This does not prevent us from computationally *using* the form of the output law, e.g., the Schroedinger wave equation, to predict future, or past, states of an unmeasured, hypothetical entity. We are here restricting attention to physical processes, not states of knowledge as in Sec. 3.9.)

The output law continues as a physical process until another measurement is made. This re-initializes the state of the particle; etc. This is a continuing physical process punctuated and refreshed by step-like jolts due to new measurements. The new measurements act as unpredictable, discontinuous, irreversible, instantaneous operations upon the object, somewhat like so many *deus ex machina* activities. Chap. 10 and Sec. 11.2.16 clarify these effects.

Since EPI output laws only physically occur as reactions to measurement they are subsidiary to the top laws (A)–(C), which exist as absolutes, i.e., whether or not measurements take place. On this basis, the *real* laws of physics are, again, the top laws.

Many of the preceding ideas were developed jointly with B. H. Soffer.

We mentioned, above, that the initiation of a measurement creates the probability law from which the data value will be sampled. That is, it locally creates the physics of the observed phenomenon. This view regards reality as being perpetuated by requests for knowledge. It adds a new, creative dimension to the nominally passive act of observation. A traditional view of reality called *logical positivism* holds that all statements other than those describing or predicting observations are meaningless. Creative observation goes one step further, stating that the observations are, themselves, meaningless except insofar as they *create local physics*.

Making a measurement is a quantitative way of asking a question. The idea of measurement begetting phenomenon seems to be the physical counterpart to the adage that a well-posed *mathematical* problem, or question, contains the seeds of its solution. It is interesting to consider whether asking a qualitative question, as well, leads in some sense to a physical phenomenon (partially addressed on pp. 250–2).

This view of measurement has some strange ramifications. For example, in the well-known Schroedinger's cat experiment, it is now *observation* of the cat that either kills it or endows it with life. Or, in the many-worlds theory of

Everett (1973), whereby each new observation occurs in a new world, the new world is now *created* by the observation (see also Sec. 11.2.16).

3.11 On applications of EPI

In each of the following chapters, the EPI principle is applied to a different measurement scenario. Each such scenario leads to the derivation of a different physical law. The ordering of the chapters is, in the main, arbitrary so that they may be read in any order. However, the chapters are grouped as to similarity of approach or of application.

The flow of operations in each chapter's derivation follows those in Fig. 3.4. A parameter $\boldsymbol{\theta}$ is chosen to be measured. The measurement is to be carried through with an instrument that has a given 'instrument function' (Sec. 3.8, Chap. 10). The measurement is initiated. The measurement process interferes, and interacts, with the phenomenon governing the parameter. This results in the perturbation of all the probability amplitudes $\mathbf{q}$ describing the phenomenon in the input (object) space to the instrument.

The phenomenon is identified by a suitable invariance principle. The principle should, by Wheeler's proposal of Sec. 0.1, be identified by the *internal processes of the measuring instrument*. An example was the unitary transformation suggested by the optical device in Sec. 3.8. An alternative to a unitary transformation is a property of continuity of flow for *the sources*. This could likewise be implied by the operation of a measuring device that obeys continuity of flow. The invariance principle is the only physical input to the procedure and, ultimately, allows the bound information J to be solved for.

The continuity of flow and unitary transformation principles are, respectively, invariance principles of the non-equality and equality type. These are designated as types (a) and (b), respectively, in Sec. 3.4.6. Type (a) principles give rise to a unique EPI solution, while type (b) principles give rise to two distinct EPI solutions. It is interesting that type (b) scenarios only occur for quantum phenomena (Chaps. 4, 10, 11 and Appendix D). All other phenomena that are derived in this book are of type (a).

The perturbed probability amplitudes $\mathbf{q}$ perturb, in turn, the channel capacity I (through defining Eq. (2.19)) and information J (through Eq. (3.13)). This activates the steps (3.13)–(3.20) defining the EPI process.

The EPI solutions define the phenomenon in the *input space* to the measuring instrument. Solutions at the output, or measurement, space must be obtained by other means. As examples: the output solution is obtained by a simple convolution of the EPI solution with the instrument function (Eqs. (3.51), (10.26b)).