Statistical Science
1993, Vol. 8, No. 4, 433–457

Quantum Statistical Inference

James D. Malley and John Hornstein

Abstract. The three main points of this article are:

1. Quantum mechanical data differ from conventional data: for example, joint distributions usually cannot be defined conventionally;
2. rigorous methods have been developed for analyzing such data; the methods often use quantum-consistent analogs of classical statistical procedures;
3. with these procedures, statisticians, both data-analytic and more theoretically oriented, can become active participants in many new and emerging areas of science and biotechnology.

In the physical realm described by quantum mechanics, many conventional statistical and probabilistic assumptions no longer hold. Probabilistic ideas are central to quantum theory but the standard Kolmogorov axioms are not uniformly applicable. Studying such phenomena requires an altered model for sample spaces, for random variables and for inference and decision making. The appropriate decision theory has been in development since the mid-1960s. It is both mathematically and statistically rigorous and conforms to the requirements of the known physical results.

This article provides a tour of the structure and current applications of quantum-consistent statistical inference and decision theory. It presents examples, outlines the theory and considers applications and open problems.

Certain central concepts of quantum theory are more clearly apprehended in terms of the quantum-consistent statistical decision theory. For example, the Heisenberg uncertainty principle can be obtained as a consequence of the quantum version of the Cramér–Rao inequality. This places concepts of statistical estimation and decision theory, and thus the statistician, at the center of the quantum measurement process.

Quantum statistical inference offers considerable scope for participation by the statistical community, in both applications and foundational questions.

Key words and phrases: Quantum mechanics, Heisenberg uncertainty, joint distribution, Hilbert space, self-adjoint operator, spectral measure, probability-operator measure, decision theory, Bayesian inference, de Finetti representation theorem, Cramér–Rao inequality.

1. INTRODUCTION

One of the principal scientific discoveries of the twentieth century is that much of the behavior of matter and radiation cannot be described by the Newtonian mechanics of particles or the classical theory of fields, even as amended by relativity. Matter and radiation must instead be described by *quantum mechanics* (QM). Of course, both Newtonian and relativistic methods, free of QM considerations, each have appropriate domains of applicability in which they act as sufficiently accurate approximations. But there are extensive domains where quantum fluctuations dominate, and these domains motivate what follows.

QM is a subject both vast and deep, with an astonishing range of applications. It yields many counterin-

James D. Malley is a Research Mathematical Statistician with the National Institutes of Health, Division of Computer Research & Technology, Distributed Systems Branch, Building 12A, Room 3053, Bethesda, Maryland 20892. John Hornstein is a Research Physicist at the Remote Sensing Physics Branch of the Naval Research Laboratory in Washington, DC.

tuitive but well-verified predictions. In QM the chance character of an outcome is not a matter of ignorance or lack of specification on the part of the observer. The source of randomness in QM is neither measurement error nor incompleteness of the observer's knowledge: no cleverly crafted models or alternate error structures can avoid this essential randomness. In QM some outcomes violate the Kolmogorov laws of probability. This can be empirically verified in terms of observed frequencies.

On the other hand, QM effects figure prominently in many practical domains where the machinery of statistical inference and experimental design could provide optimal decisions or more efficient estimates. Examples include medical imaging, spectroscopy, optical communication and signal detection and remote sensing. Additional examples occur in many parts of applied physics, astronomy, chemistry and molecular biology.

In designing experiments and analyzing data when QM effects are important, a question therefore arises: Do statistical methods need to be modified? In this paper we argue that they do and, furthermore, that appropriate modifications may be based on a body of existing techniques that we call methods for *quantum statistical inference* (QSI).

We now mention some potential application areas of QSI. Quantum noise must be taken into account in analyzing spectroscopic data when the probed species or the probing species are scarce, or if they rarely interact when both are present. Typical instances are the spectroscopic detection or estimation of trace species, spectroscopy with small fluxes of probing radiation, and nonlinear spectroscopy. Important special cases include: the astronomy of faint sources, or of neutrinos or gravitational radiation; much experimental work on elementary particles; and the new field of spectroscopy with trapped single ions or atoms. Other classes of examples arise in: control and communication using weak optical or particle beams, as in deep-space optical communication; laser ranging of astronomical bodies; the new fields of molecular electronics and electronics using superlattices, quantum wells, and quantum dots; and the manipulation of isolated trapped electrons, ions, atoms, molecules or biological organisms. One recently developing area that appears primed for a rigorous splicing of quantum theory and statistics is *quantum cryptography:* see *Nature* (October 1991, p. 384), *Science* (7 August 1992, p. 752), *Discover* (September 1992, p. 92) and *Physics Today* (November 1992, p. 21).

Of special interest to us is the rapidly developing field of light-based imaging devices and detection methods and analysis techniques using other forms of radiation, particularly as applied to biotechnology. Possible biomedical applications of quantum-consistent statistical decision theory could include: reduced-dose PET scans; real-time, laser-based reduced-illumination confocal microscopy of living cells and tissue; and bioluminescent molecular tagging and enhanced chemiluminescence, which would allow the study of biological processes at the level of individual molecules. A brief survey of this very rapidly growing field of optical biotechnology, at the molecular and atomic level, is given in *Nature* (3 December 1992, p. 493). That quantum mechanical events at the biomolecular level are more generally entering into the scientist's toolkit is shown by recent success in modeling electron quantum tunneling pathways in proteins: *Science* (11 December 1992, p. 1740).

QSI has already been applied to a variety of important problems in optical detection and communication, to the design of interferometric fiber-optic sensors and to the optimal use of optical data, for example, to resolve closely spaced sources; see Personick (1971), Helstrom (1976), Yuen and Shapiro (1980), Shapiro (1980), Nagaoka (1990), and more recently Shapiro and Shepard (1991). To give some background, we note that QSI was developed primarily by electrical engineers interested in communications systems. Their efforts were a response to the invention of the maser and then the laser. These devices led to amplifiers and oscillators where quantum fluctuations are an important source of noise. Lasers offered the prospect of high performance optical communications, where again quantum noise is significant. It became necessary to calculate the detection and estimation performance of systems having significant quantum noise, and so QSI was born. Using QSI improvement of a full order of magnitude has been observed. Electrical engineers and workers in quantum optics still constitute the bulk of its practitioners.

Two of the researchers working on these and other applications have also been the principal developers of the new methods: Carl Helstrom, who can be credited with founding the field of quantum statistical inference, beginning in the late 1960s, and Aleksandr Holevo, who has made, and continues to make, major contributions, both theoretical and practical.

Helstrom and Holevo have each written complete, rather different, books on QSI (Helstrom, 1976; Holevo, 1982). However, both texts were addressed to an audience of physicists and quantum knowledgeable electrical engineers. The present article seeks to motivate and outline QSI for a primarily statistical audience, one not assumed to be conversant with either the basics of quantum theory or the (ultimately required) mathematics of spectral measures. Since this article is, to our knowledge, the first to attempt this bridging for the statistical community, the reader may be alternately intrigued, confused and challenged: it is for these reasons that the lengthy and more technical Appendix was created.

The need to review QM also presents an opportunity

to outline the essentials of quantum theory in a way that avoids some common misinterpretations of the theory. Key ideas of statistics can sometimes prove useful in deconstructing these misinterpretations, while other, very comfortable and familiar ones may present serious obstacles to understanding QM and QSI. Our broader goal is to systematically compare quantum and plausible classical procedures for the same detection or estimation problem, in an attempt to choose between them. This has lead us to our *conjecture of local equivalence for quantum inference*. The conjecture asserts that: in many, perhaps all, applications the results of QSI can be obtained by ordinary statistical inference, if the set of outcomes and their probabilities are computed according to the rules and axioms of QM, rather than the Kolmogorov axioms applied uncritically. (See Item A1.)

More directly, QSI materially advances quantum theory itself by capitalizing on the fact that QSI has its roots in statistics. For example, one major accomplishment of QSI has been to show that mathematically rigorous quantum analogs of Cramér–Rao lower bounds for variances of estimators lead to generalized *uncertainty relations* on the products of variances of estimators. A special case is the Heisenberg uncertainty relation on position and momentum. These analogs are vigorous enough to resolve long-standing controversies about other uncertainty relations, for example those pairing time with energy, or phase with number of quanta (see Item A2). Moreover, the mathematical strength of these analogs yields strictly sharper bounds that may be obtainable in experimental practice.

QSI consolidates classical statistical inference and quantum probabilities into a unified, rigorous scheme: it gracefully mediates transactions between quantum theory and statistical inference. QSI demonstrates that a useful and consistent statistical decision theory can be developed even when, for example, joint distributions do not exist. Thus, our central claim: *QSI materially increases the sum total of useful probability models.*

The paper is organized as follows. Whereas this first section dealt with history and motivation, Section 2 begins actual study of quantum statistical inference (QSI) by presenting our basic example. More complete statements on basic physics, quantum theory, underlying (often highly nontrivial) mathematics and additional remarks are found in the Appendix. Section 3 deals with Heisenberg uncertainty and its intersection with QSI. Section 4 covers the basics of QSI. Section 5 looks at current and potential applications of QSI, and Section 6 is a summary and conclusion.

Again, much important material that resists easy domestication appears in the Appendix, and references in the text proper are of the form, for example, (See Item A12), which directs the reader to Item 12 in the Appendix.

2. CHALLENGES TO CLASSICAL STATISTICAL DECISION THEORY

2.1 Some Essential Physics

Before presenting our main example of QSI in action it is first necessary to sketch some basic physics (in this section) and then very briefly outline the structure of quantum theory (in the next section).

Many types of quantum objects possess a property called *spin*. Classically, spin is associated with, say, a child's top. For quantum objects however, spin is an essentially nonclassical degree of freedom: a quantum spinning object cannot be consistently described as a spinning classical rigid body, nor as having an angular momentum due to the motion of internal constituents. A good introduction to quantum spin can be found in the well-known lecture notes of Feynman, Leighton and Sands (1965, pp. 5-1 to 6-14). In this paper we mostly work with particles having "spin 1/2," and apart from a constant, spin is assumed to take only two values: $+1$ (or "up") and -1 (or "down"). (See Item A3.)

EXAMPLES. Using just the physics introduced above we show, at the basic probability event level, that quantum phenomena do not necessarily satisfy the usual Kolmogorov axioms. This is done in two ways. The first can be accomplished with just two pairs of polaroid lens sunglasses. We urge the reader to actually try this before proceeding: one of the most counterintuitive features of quantum data is immediately apparent.

However, since this counterexample to the familiar rules of probability has a classical (not strictly quantum) "explanation," we also consider a second version of this three-filter experiment that has *only* a quantum mechanical explanation. This example requires laboratory equipment, but is discussed fully in Feynman, Leighton and Sands (1965, Vol. 3, Chap. 5).

1. Obtain three pieces of polaroid plastic – for example, just pop out the lenses from two pairs of inexpensive sunglasses. Hold two of the lenses at right angles to each other: (essentially) no light passes through (a small amount does, in fact, since the lenses are not perfect filters). Next, hold two of the lens together with one of them turned at 45° with respect to the first. Some light gets through: less than the amount passing through just one lens, but much more than with two placed at right angles to each.

   Now, hold two of the lens at right angles. Let A be the event that light gets through the first lens (essentially with probability 1), and let C be the event that light gets through the second lens by itself (also with probability 1). We have already observed that the probability of $A \cap C$ is 0. Let B be the event that light gets through the third lens, by itself, when it is turned through

45° (again, essentially equal to 1). We have seen that with two lenses arranged at 45° with respect to each other some light gets through, so the probability of $A \cap B$ is not 0.

Finally, with two of the lens set at right angles (no light getting through) put the third lens *between* these two, turning it at 45° with respect to the first. Some light now gets through!

We have therefore observed that

$$\Pr(A \cap B \cap C) > \Pr(A \cap C),$$

and this is *close* to a hand-held, home-experiment violation of the rules for probability and events. It not *exactly* so, since the experiment is not fully quantum, has a classical physics "explanation," and the stated probabilities are, more precisely, light intensities. If the experiment were done at the level of counting *single* photons, it would be fully quantum. As it is, the experiment is a macroscopic consequence of quantum phenomena. Moreover, it is a *precise, hand-held analogy for a fully quantum experiment*, which we now consider.

2. The second example, a quantum version of the lens experiment above, is discussed in detail in Feynman, Leighton and Sands (1965, Vol. 3, Chap. 6). It uses particles with quantum spin 1 and uses physics laboratory equipment, in particular gadgets called Stern–Gerlach devices. These are described below. For the reader's convenience this experiment is outlined in Item A4, but the reader still can perform the "thought experiment" of using three such devices in place of the polaroid lenses used above. The result is a clean violation, at the probability event level, of the basic Kolmogorov axioms: it is an experimentally observed fact of nature.

A more precise dissection of the differences between the two examples is as follows. In the first, the light shining through is obeying a rule that gives the intensity of the transmitted light as the squared cosine of the included angle between any pairing of the filters: it is not strictly speaking a probability rule, at the macroscopic level, but is a consequence of a probability rule at the single photon level. In the second example, the squared-cosine law *is* a probability rule and using spin 1/2 particles and Stern-Gerlach devices, the probability of getting through with spin up, say, obeys this same squared-cosine rule; we derive this below. The important point is that, in QM probability calculations, one begins with a complex-valued *amplitude*, and the squared modulus of this is the required probability. And, loosely speaking, amplitudes add, but probabilities do not. This is all congenially covered in Feynman, Leighton and Sands (1965, Vol. 3, Chaps. 1 through 6; see especially the Summary on pp. 1–10].

2.2 Some Essential Quantum Theory

We narrow our focus to exactly two of the several axioms of QM (see Item A4). In ordinary probability theory, an observable corresponds to a random variable X. The expected value of X is the sum (or integral) of ρX, where ρ is a probability density function. In quantum mechanics X and ρ are replaced with matrices (or more generally, operators).

The *first axiom* of QM is that a quantum system is fully characterized by a density operator, which is a certain linear transformation of a complex-valued vector space. Recall that an Hermitian matrix is one that is equal to its own complex-conjugate transpose. Then, the Hermitian linear operator ρ acting on the space $\mathcal{H}$ is said to be a *density operator* if it is positive definite and its Hilbert space trace equals one; it is always assumed that $\mathcal{H}$ has a positive definite inner product.

The density operator is the quantum counterpart of a conventional *statistical density*, and the trace function does duty as the quantum analog for *statistical expectation*. That is, $\mathrm{tr}[\rho X]$ replaces $E(X)$: where the trace function uses a quantum density operator ρ and Hilbert space operator X, the standard expectation $E(X)$ is taken using a classical density function ρ and classical random variable X.

If the density operator ρ is the projection onto a one-dimensional subspace of $\mathcal{H}$ then the system is said to be in a *pure state*; otherwise it is said to be in a *mixed state* (equivalent language: ρ denotes a mixture state). One verifies that ρ is a pure state if and only if $\rho^2 = \rho$. (See Item A5.) Also, it is frequently convenient to describe a quantum system with density operator ρ as simply a system in the state ρ.

EXAMPLE 1. As a first example of a density operator, consider the study of quantum spin. For a spin 1/2 particle whose spin is "along" a direction in the (x, y) plane, at an angle θ from the x-axis, the density operator can be shown to be:

$$\rho_\theta = \begin{bmatrix} 1 & e^{-i\theta} \\ e^{i\theta} & 1 \end{bmatrix}.$$

This is a density operator since it is Hermitian, positive definite and $\mathrm{tr}(\rho_\theta) = 1$. Moreover, it is the density operator of a pure state, since $(\rho_\theta)^2 = \rho_\theta$. The Hilbert space $\mathcal{H}$ associated with the spin states is a two-dimensional vector space over the complex numbers. (See Item A6.)

EXAMPLE 2. As a basic example of how infinite-dimensional Hilbert spaces occur naturally in applications, consider a single electron constrained to move

along the x-axis. The Hilbert space $\mathcal{H}$ for this system is isomorphic to $\mathcal{L}^2$, the space of equivalence classes of square-integrable, complex-valued functions of a real variable x, with integration over Lebesgue measure.

For any $\varphi(x)$ and $\psi(x) \in \mathcal{H}$ the inner product is given by

$$(\varphi, \psi) = \int_{-\infty}^{+\infty} \varphi^*(x)\psi(x)dx$$

for $\varphi^*(x)$ denoting the complex conjugate of $\varphi(x)$.

The density operator is not the only operator acting on the vector space: many, but not all, quantities of interest are associated with Hermitian operators. (See Item A7.)

The role of these and other operators on the vector space become clearer in the discussion of the second axiom of QM. The second axiom for QM ties together the space $\mathcal{H}$ to the outcomes of experiments. It is assumed that the experimental devices themselves and their associated meters, registers, displays, and so on, are classically described systems, and that the measurements are real-valued. (This highly nontrivial cabal of assumptions is still the subject of much debate in the literature on foundations; it remains unchallenged on the practical level.)

The *second axiom* is the link to the space $\mathcal{H}$, and is made as follows: each quantum measurement is associated with a family of operators on the space $\mathcal{H}$. It is assumed that the outcome of the measurement lies in some Borel measurable set B in $\mathcal{R}^n$, and that the value of the observed outcome is an eigenvalue of one of these operators. Precisely *which* operator is given by a link to the set B via a *probability-operator measure* (p.o.m.): for a given set B, a Hermitian operator $X(B)$ on $\mathcal{H}$ is defined such that

$$\Pr(\text{observed value of } X \text{ lies in } B) = \text{tr}[\rho X(B)],$$

where tr, as above, is the trace function on the operators of $\mathcal{H}$. Usually the p.o.m., that is, the entire family of operators associated with the quantum measurement, is denoted by just X (in a very convenient, often used and often confusing abuse of notation).

The second axiom for QM, the coupling just displayed between probability, densities and observables, we refer to as *the trace-rule for probabilities*. (See Item A8.) At this point we emphasize that it will often not be obvious to the reader how ρ and $X(B)$, for all Borel sets B, are derived in a given problem. Practicing quantum mechanists have though, over time, developed a number of useful intuitions and *correspondence rules* that help here, as the axioms of QM do not include explicit, universal recipes for generating ρ and $X(B)$, for all Borel sets B.

Observe also that the quantum measurement scheme as defined does not conform to the usual statistical declaration that measurement of an observable yields $X = x$ in B. That is, a distinctive feature of random quantities in QM is: the p.o.m. X is defined on Borel sets B, but yields operators on $\mathcal{H}$, and probability is attached to an outcome [= an eigenvalue of $X(B)$] via the trace-rule on $\mathcal{H}$. Classically, a random variable X is defined pointwise and returns real or complex values, with probability attached to an outcome via a density function on a Boolean sample space. We also note that a pure state does *not* meaningfully correspond to a point mass density. Pure state systems in general do not guarantee sharp, repeatable measured outcomes. The trace-rule, even with the added assumption that $X(B)$ is a projection operator associated with a single point $B = \{x\}$, still returns only the probability that the measurement yields the eigenvalue x. Again, quantum randomness is not a removable singularity of the axioms.

This QM replacement of the definition of a "random variable" by a collection of operators on a Hilbert space is at the mathematical core of many of the properties of quantum observables which trigger the inappropriateness of the universality of the usual Kolmogorov rules for probability spaces. This arduous shift, and its consequences, has been pursued at length (see, e.g., Gudder, 1979, 1988), and we must exile essentially all of the details to the intrepid reader and the genuinely formidable mathematical literature. (See Item A9.)

Continuing to illustrate the notation and ideas of QM more simply, let the quantum system be described by a pure state, so the density operator ρ is the projection onto a one-dimensional space, specified by a unit vector $\psi \in \mathcal{H}$: $\rho = \psi\psi^*$. Letting $X(\lambda_i)$ be the projection operator assigned by the observable X to the single point λ_i on the real line, it can be shown that

$$\begin{aligned}\Pr(\text{outcome } \lambda_i \text{ observed}) &= \text{tr}[\rho X(\lambda_i)] \\ &= (\psi, X(\lambda_i)\psi) = \psi^* X(\lambda_i)\psi.\end{aligned}$$

(see Item A10.)

Next, let p_i denote the probability of observing λ_i, one of a finite set of eigenvalues. The *mean* and *variance* for the operator X are defined, respectively, by

$$E(X) = \sum \lambda_i \text{tr}[\rho X(\lambda_i)] = \sum \lambda_i p_i,$$

$$\text{var}(X) = E(X^2) - [E(X)]^2.$$

To further simplify the discussion, the main examples discussed below use only the machinery for the case of operators with discrete eigenvalues, *proper* eigenvectors (see Item A9) and a finite-dimensional space $\mathcal{H}$, and usually only pure states of the quantum system are considered. Then, the expressions for expectation and variance simplify, respectively, to

$$E(X) = \sum \lambda_i (\psi, X(\lambda_i)\psi) = \sum \lambda_i p_i,$$

$$\text{var}(X) = \sum (\lambda_i)^2 p_i - [\sum \lambda_i p_i]^2.$$

With the present restrictions on the measurement and the state of the system, the formulae above exactly parallel the familiar ones of classical statistics. (See Items A11 and A12.)

REMARK. This is probably the best place to consider the assertion that "everything is quantized," which is often (and incorrectly) made in popularized treatments of QM. The discussion comprises part of our effort to deconstruct and demystify some quantum-provoked misrepresentations.

The phrase in question may have one of (at least) three meanings:

1. "All quantities are discrete." This is the version that is most relevant to the present discussion and also the most persistent in popular QM accounts, and it is also false. As discussed earlier, the position operator usually has a continuous spectrum (has only *generalized* eigenvectors), and the momentum and energy operators often have a partly or wholly continuous spectrum.
2. "All quantities are represented by Hermitian operators on Hilbert space." This version is false. Quantities such as time, phase and other angles do not correspond to Hermitian operators. Those that do are called quantum *observables*, while the quantities just mentioned are much more rigorously treated, in QSI, as *parameters* of the quantum process under observation. Also, the *systems approach* of Shapiro and Shepard (1991) shows how some experimental outcomes can be rigorously treated using non-Hermitian operators.
3. "Every system must be subject to the laws of QM, so any system in particular must obey the uncertainty principle." This version is correct, since if even *one* system were not subject to the uncertainty principle (discussed below), it would in principle be possible to use that system to construct a violation of the uncertainty principle for those systems that were ostensibly subject to it.

2.3 Coins of the Quantum Realm

Enough of the essential physics and quantum theory is now at hand to permit looking at an experimental example that sharply illuminates the differences between classical inference and QSI.

EXAMPLE: QUANTUM COIN TOSSING. Consider tossing a biased coin. The bias is known to be one of two different amounts. Table 1 gives the probabilities of heads and tails for the two versions of the coin.

The prior probability ξ of bias state 1 is known to the decision maker; state 2 has a prior of $1 - \xi$.

Choose a 0-1 loss structure for the decision problem: assign loss 1 to an incorrect decision, and assign loss 0 to a correct decision.

TABLE 1
Probabilities of heads and tails for two versions of a coin

	Heads	Tails
Bias state 1	0.4	0.6
Bias state 2	0.8	0.2

Given the opportunity to toss the coin exactly once, and then decide between state 1 or 2, it is a standard exercise to calculate the minimum Bayes risk rule (see, e.g., DeGroot, 1970, Example 1, pp. 140-141). The classical rule is given in Table 2.

A quantum version of this coin-tossing problem is as follows: consider sending an uncharged (neutral) atom through an apparatus called a Stern-Gerlach device; see Figure 1 below, and also Feynman, Leighton and Sands (1965, pp. 5-1, 5-17). The particle passes through an intense, inhomogeneous magnetic field created by the two poles of a large, specially shaped magnet. If the atom has a net spin, it will be deflected slightly toward one of the poles. The signed magnitude of the deflection is proportional to the spin component along the direction joining the two poles.

An atom with spin 1/2 (meaning, it takes the values $+1$ or -1) exits the device with only one of two possible deflections: a small amount up ($+1$) or an equal amount down (-1), assuming that the two poles in the Stern-Gerlach device are arranged vertically, as in Figure 1. This exemplifies the quantization of spin angular momentum, as discussed earlier: only two outcomes are possible. If the particle has not been prefiltered or selected in some way (by an initializing Stern-Gerlach device say) then the probability that it leaves in the "up" state is 1/2, and for the "down" state, also 1/2.

Stern-Gerlach devices provide a nice example of the QM calculation of probabilities. The calculation is relevant to quantum coin tossing and at the same time provides a context for discussing a distinctly quantum mechanical pitfall that must be avoided when discussing conditional probabilities.

Thus, consider two Stern-Gerlach devices, SG1 and SG2, such that only those atoms that emerge in the "up" state relative to SG1 are allowed to enter SG2. SG2 has the same axis as SG1, but its "up" direction makes an angle of θ relative to that of SG2. The atoms

TABLE 2
Classical rule for probability of heads and tails in the coin-tossing problem

For	$0 \le \xi < 1/3$	always decide state 2
For	$1/3 \le \xi < 3/4$	decide state 2 if heads observed decide state 1 if tails observed
For	$3/4 \le \xi \le 1$	always decide state 1

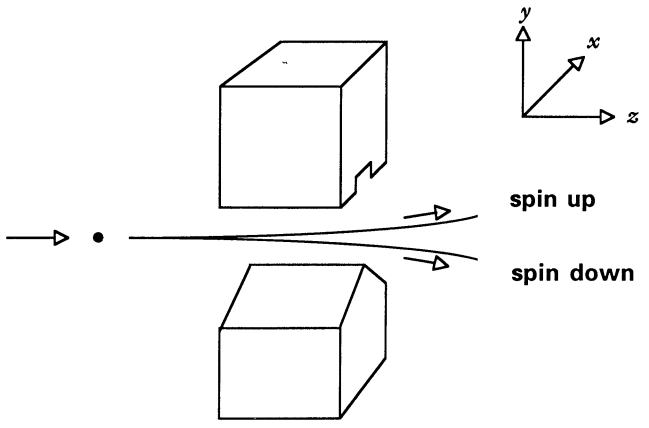

FIG. 1. *Quantum version of the coin-tossing problem. An uncharged (neutral) atom is sent through a Stern-Gerlach device.*

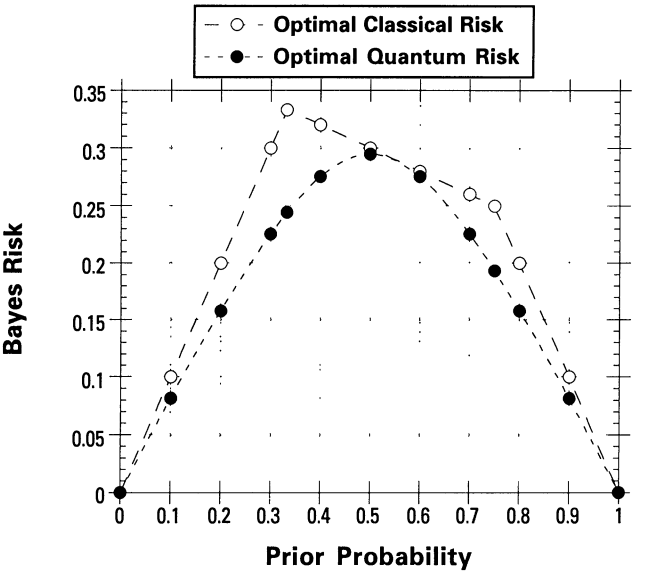

FIG. 2. *Risk curves for classical (with fixed SG2 at $\theta = 0$) and quantum-informed (SG2 tunable) rules.*

entering SG2 will emerge in the "up" state (relative to SG2) with a probability given by

$$\Pr(\text{spin up from SG2, given spin up from SG1}) = \cos^2\left(\frac{1}{2}\theta\right).$$

This can be explicitly verified by using the trace-rule for probabilities. The density operator for a spin 1/2 particle whose spin is "up" along a direction in the (x, y)-plane and making an angle θ with the x-axis is

$$\rho_\theta = \frac{1}{2}\begin{bmatrix} 1 & e^{-i\theta} \\ e^{i\theta} & 1 \end{bmatrix}.$$

Now, the vector

$$\psi = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is the eigenvector corresponding to eigenvalue $+1$ ("up") for the "spin x" operator so that the projection operator associated with $X(+1)$ is

$$\psi\psi^* = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

(See Item A7.) Hence the trace-rule for probabilities gives

$$\operatorname{tr}[\rho_\theta X(+1)] = \frac{1}{4}\operatorname{tr}\left[\begin{bmatrix} 1 & e^{i\theta} \\ e^{-i\theta} & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right] = \cos^2\left(\frac{1}{2}\theta\right),$$

as required. (See Item A13; the form for the displacement operator ρ_θ and the projection operator associated with $X(+1)$ are not immediately derivable from anything we have said so far.)

Suppose now that SG1 is hidden from the observer, who can take measurements only on the atom exiting SG2. This observer thus only notes whether the deflection of the emerging atom corresponds to "SG2 up" or to "SG2 down" ("heads" or "tails"). It is further assumed that SG1 is turned through an angle θ_1 or θ_2 with respect to SG2, with a probability of ξ that the setting is θ_1 (and $1 - \xi$ that it is θ_2).

Choose the angles θ_1 and θ_2 so that, according to the probability rule given above, the observer stationed at the exit ports of SG2 observes heads or tails with exactly the probabilities given in Table 1. That is, let

$$\theta_1 = 2 \arccos(\sqrt{0.4}) = 101.54°,$$

$$\theta_2 = 2 \arccos(\sqrt{0.8}) = 53.13°.$$

The minimum Bayes risk quantum decision rule for the coin-tossing problem is derived in detail below [using the method outlined in Helstrom (1976, pp. 106–108)]. We explicitly assume that the experimenter is allowed, for each value of the prior, to also search for the optimal angle to turn the device SG2, although for an experimenter who is not quantum-informed, it may not be obvious that turning SG2 could lead to significant changes in the decision rule and the subsequent risk. The optimal Bayes quantum-informed rule *automatically* yields an optimal setting for SG2—more on this below.

We observe that the quantum rule does as least as well as the classical rule (and in fact, at all but one point, does strictly better), when the classical rule does not take possible advantage of optimally setting SG2; later we show that this holds even if a classical rule is used in conjunction with an optimal setting of SG2. The risk curves for these two procedures, classical (with fixed SG2 at $\theta = 0$) and quantum-informed (SG2 tunable), are shown in Figure 2. Figure 2 shows how the optimal minimum Bayes risk quantum rule performs

relative to the classical rule. Some noteworthy differences between the two rules are:

a. As shown below (Section 4), the quantum rule always depends on the data, while the classical rule does not do so, in both the upper and lower ranges of the prior: see Table 2. Notice also that the quantum rule is always symmetric about the prior value one-half, and the classical rule is not in general. This may seem puzzling at first, but aside from checking the explicit equations and confirming this invariance of the quantum Bayes risk (but not the Bayes rule itself!) under the transformation

$$\text{prior} \rightarrow (1 - \text{prior}),$$

we advance two, perhaps more convincing, arguments for this symmetry.

Thus, observe that the only physical invariant in the problem is the *included angle* between the possible settings of the SG1 and SG2 devices. As the prior sweeps between 0 and 1, we would anticipate this invariance to be respected by the risk function for any reasonable decision rule. This is the case for the quantum risk, and not the case for the classical, nonquantum risk. Hence, in fact, we are led at the outset to reject the classical rule as "unphysical," and thus, on the basis of physical intuition, confirm the observed symmetry of the quantum Bayes risk curve. The second argument for symmetry of the Bayes risk curve follows from:

b. The quantum rule works, in effect, *by tossing the quantum coin differently, and then making the decision between* θ_1 *and* θ_2: the optimal Bayes quantum rule is equivalent to measuring spin up or down along a direction different from that used by the classical one. In fact, the rule smoothly incorporates both design and optimal decision making into a *single*-step Bayes optimal procedure, and this is obtained in an essentially closed-form solution. Namely, letting

$$\gamma = \xi/(1 - \xi),$$

the quantum measurement then is performed with SG2 turned, with respect to vertical ($\theta = 0$), at an angle θ implicitly given by

$$\tan\theta = -\frac{\sin\theta_2 - \gamma \sin\theta_1}{\cos\theta_2 - \gamma\cos\theta_1}.$$

Hence for example at $\xi = 1/3$, where the classical rule does worst, the optimal setting for SG2 would be $\theta = 49.23°$.

A plot of the risk functions for this setting of SG2 is given in Figure 3. (See Item A14.)

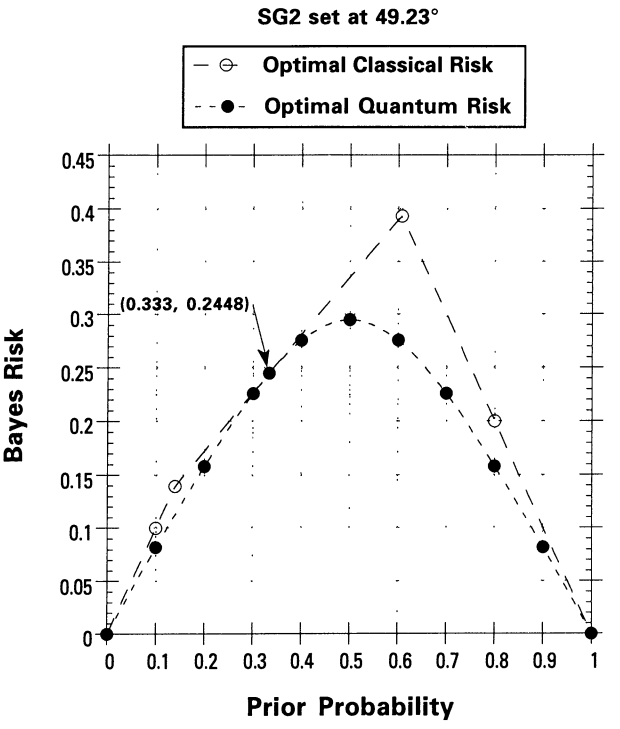

FIG. 3. *Plot of risk functions with SG2 at* $\theta = 49.23°$.

c. Note that it is far from obvious that setting SG2 at an angle of 49.23° results in the minimum Bayes risk procedure for $\xi = 1/3$. Thus, the quantum-consistent decision theory provides specific, essentially closed-form, practical guidance for designing optimal experiments and detection devices.

d. The example illustrates one other essential difference between the classical and quantum decision rules.

To obtain the minimum risk at each value of the prior, the classical rule uses a conventional random variable and Borel σ-algebra. The quantum rule is constructed using operators on a certain Hilbert space, where plausible candidates for "events" might be the (closed) subspaces of the Hilbert space. However, it can be verified that these candidate "events" do not form a Boolean algebra, as would be required for a purely classical approach; see for example Jauch (1968, pp. 26-27).

This distinction is not just the infliction upon the reader of one more mathematical entanglement: Hilbert spaces and operators behave very differently from Boolean algebras and conventional random variables, and these differences imply profound distinctions about measurement, and by implication, decision making and estimation, in the quantum world.

2.4 More Quantum Coin Tossing: The Nonexistence of a Joint Distribution

There are many arguments, both experimental and theoretical, that prove the nonexistence of certain joint distributions; see, for example Gudder (1988, pp. 175–178) and Shimony (1989, pp. 385–386). For completeness we include a sketch of the proof appearing in Gudder (1988): see Item A13.

Recently there has been a resurgence in proofs (= mathematically rigorous, and real, or at least physically plausible, experiments) involving triples of outcomes, obtained by measuring particle spin, say, along three different directions (at three different times), that cannot have any conventional joint probability distribution. Hence, lingering with just simple spin systems as described above has led to increasingly rigorous challenges to the universal existence of joint distributions. And in fact much more has emerged from these recent, multiple derivations; see especially Mermin (1990a,b,c), Greenberger et al. (1990), Peres (1990) and Fivel (1991). We highlight two of these developments:

1. It can be shown in some examples that the QM calculations end with a value $+1$ for the outcome, but ordinary probabilistic reasoning (which assumes the existence of joint distributions) gives a value -1. These calculations are exact and not subject to any additional quantum or classical randomness.
2. The newer derivations sharply underscore the problem of assigning any measurement value ("dynamic attribute") to quantum objects in the absence of actually having performed the required measurement. The problem is that *it appears experimentally and mathematically not always meaningful to ask that any assignment of future possible, mutually exclusive outcomes must be consistent.* See especially Mermin (1990a, pp. 171–173).

Thus, these newer and much sharper (but conceptually still simple) experiments raise serious questions for any classical (non-Bayesian) statistical procedure that, in the quantum world, argues from the assignments and associated likelihoods of any outcomes that "could have occurred but were not observed, given the data at hand." In QM, any non-Bayesian procedure is evidently confronted with a serious, *experimentally driven* questioning of underlying rationale.

We leave this issue for future study by the reader and ourselves. It is surely a research topic positively luxuriant with new statistical possibilities and novel inferential consequences and fuels the question: *Must frequentist-based statistical inference be found logically indefensible, and hence for foundational reasons, be excluded from the quantum world?* (See Item A15.)

3. THE HEISENBERG UNCERTAINTY PRINCIPLE

The Heisenberg uncertainty principle is one of the most famous features of QM. (See Item A16.) Our basic reference is Beltrametti and Cassinelli (1981), and we adopt their use of A, B, C and so on, for quantum observables.

The statistical moments of a quantum observable depend upon the state ρ of the system. In this section the dependence of moments upon state will be indicated explicitly:

$$E(A|\rho) = \text{expectation of } A \text{ for system in state } \rho$$

$$V(A|\rho) = \text{variance of } A \text{ for system in state } \rho.$$

Although the outcomes of measurements on QM systems need not be discrete for all types of observables, for each observable there is always at least one state for which the spread of observed values around the mean $E(A|\rho)$ is as small as desired.

That is, given any $\varepsilon > 0$, there always exists some unit vector such that $V(A|\rho) \leq \varepsilon$; see Beltrametti and Cassinelli (1981, p. 19). In words, in QM the measurements of a single observable A can be made as accurately as desired. (It is, however, not part of the QM formalism itself to describe how a particular observable is to be measured.)

A basic property of Hilbert spaces, the Cauchy–Schwarz inequality, is now used. For any pair of Hermitian operators A, B and every vector ψ (in the intersection of the domains of definition for Hilbert space operators A and B) and associated pure state $\rho = \psi\psi^*$, the inequality gives

$$V(A|\rho) \times V(B|\rho) \geq \frac{1}{4} \left| (A\psi, B\psi) - (B\psi, A\psi) \right|^2.$$

See Holevo (1982, pp. 87–88) for a full derivation.

For the inequality above one of two things must be true:

i. For any $\varepsilon > 0$ there exists *some* state vector ψ and a state $\rho = \psi\psi^*$ such that the product of variances on the left-hand side is less than ε; *or*
ii. there exists some $\varepsilon > 0$ such that for *every* state vector ψ and corresponding state ρ the product of variances is greater than ε, and the left-hand side is bounded strictly away from 0.

Case (ii) is known as *Heisenberg's uncertainty principle.* For Case (ii) the lower limit must be nonzero for *all* states of the system. (See Item A17.)

REMARK. We note that the Heisenberg lower limit, for position and momentum, rigorously applies to the following novel, but perfectly valid experiment: prepare two identical quantum systems, in two different cities, on two different days. Taking position measurements repeatedly in one of the cities yields a spread

of values; similarly for the momentum measurements taken in the other city on the other day. Heisenberg uncertainty then asserts that the product of these separately sampled variances still has a strict, nonzero lower limit.

In words, one series of measurements is not meaningfully "disturbing" the measurement of the other series: there is no causal interaction at work in Heisenberg uncertainty.

This deconstruction of the "disturbance" model leads us to consider other interpretations, sound and unsound. (See Item A18.) We take Niels Bohr's formally neutral position that there is no "why" or "explanation" beneath the rules and results of QM, in particular Heisenberg uncertainty. Recent research points to still more emphatic, simply stated, yet rigorous statements of this point of view: see especially Peres (1990) and Mermin (1990c).

4. QSI: QUANTUM STATISTICAL INFERENCE

4.1 An Outline of Things to Come

This section carries classical statistical decision theory forward into the quantum regime. Standard references for decision theory in the classical mode are Ferguson (1967) and DeGroot (1970). References for the quantum-consistent version are Holevo (1973; 1982, Chaps. 4 and 6) and Helstrom (1976, Chaps. 3, 4 and 7).

The focus here is on problems whose decision spaces are finite. Parameter estimation (and the associated, often infinite-dimensional decision spaces) is considered later, but only in terms of linking the Heisenberg uncertainty principle to the Cramér–Rao inequality. A fuller treatment of estimation approximately parallels the finite decision space theory; see Malley and Hornstein (1992).

First, the basic ideas of classical decision theory are reviewed. Then the quantum analog of the theory is described, using operators and observables in place of random variables. Since decision theory may require using a family of functions (called *detection operators*), it is necessary to verify that joint distributions of all the functions are meaningful where they are required (otherwise joint measurement may be problematic, as shown by the above discussion of the uncertainty principle). Also, Yuen, Kennedy and Lax (1975) show how the *family* of detection operators, and hence the decision rule, is directly definable in terms of a *single* p.o.m.: the individual detection operators are found by integrating the p.o.m. over the separate prior (classical) densities.

Theoretical quantum loss and utility functions can still be calculated from the operators, even when joint distributions, which would have allowed data-driven estimates of loss and utility functions, are not in principle available.

To repeat, without joint distributions, many optimal measurements, when expressed as families of detection operators, cannot be assumed to be realizable. This remains a significant problem for many applications of QSI. While original work for Shapiro and his students [personal communication; see also Shapiro and Shepard, 1991] has shown how some key, optimal measurements of this type can (in the limit) be experimentally realized, we view this enduring problem as incentive for a more intense collaboration between physicists and statisticians. (See Item A19.)

4.2 Review of Classical Statistical Decision Theory

A standard basis for statistical decision theory has three parts:

i. A set Θ of possible "states" of nature, referred to as the *parameter space*;
ii. a set $\mathfrak{A}$ of actions (or decisions) available to the statistician, referred to as the *action space*;
iii. a *loss function* $L(\theta, a)$ defined on $\Theta \times \mathfrak{A}$.

For given Borel measures on the product space $\Theta \times \mathfrak{A}$, it is assumed that L is real-valued and Borel measurable. Quotation marks about the "states" in (i) distinguish these parameter space states from the states of a quantum system; this distinction disappears (i.e., becomes redundant) after the quantum states themselves have been parameterized by QSI.

A statistical decision problem is a specification of a triple $(\Theta, \mathfrak{A}, L)$ and an experiment $\mathcal{E}$ which involves a (classical) random observable X. (X here indicates an observable that is measured and should not be confused with the earlier usage of X to denote a spectral measure.) It is assumed that X takes values in a Borel space, $\mathfrak{X}$, called the sample space, and it is also assumed that the distribution, $\mathcal{F}$, of X depends on the states of nature

$$\mathcal{F} = \mathcal{F}(X, \theta).$$

On the basis of an outcome of the experiment $\mathcal{E}$, the value of X is inferred to be x. The decision maker then chooses an action $d(x) \in \mathfrak{A}$; if the decision rule has no randomized component, the function d maps the Borel sample space $\mathfrak{X}$ into $\mathfrak{A}$. Next, the loss function is evaluated, yielding the random variable $L(\theta, d(x))$, which depends on θ, and whose value is not generally known. The expected value of L, given the observed X, and taken conditionally on θ, is called the *Bayes risk function* for the problem.

Denoting the risk by $\mathfrak{R}$:

$$\mathfrak{R}(\theta, d) = E\big(L(\theta, d(x)) \,|\, \theta\big)$$
$$= \int L(\theta, d(x))\, d\mathcal{F}(x, \theta).$$

$\mathfrak{R}$ therefore represents the average loss to the decision maker, when the true state of nature is θ, using the decision rule d, and an observable X generated by the experiment. It is assumed that the risk exists and is finite for d, and all $\theta \in \Theta$.

4.3 The Essentials of QSI

A good bridge into QSI is provided by the binary decision problem; see Helstrom (1976, pp. 106–113). Many of the essential features of QSI are already displayed in the quantum-consistent resolution of this simply posed problem. (See Item A20.)

Here, the binary decision problem is first discussed using a Bayes approach and then using the Neyman–Pearson approach. The QSI Bayes procedure is used to explicitly solve the quantum coin tossing example given above in Section 2.

The steps in the QSI Bayes solution are shown, but the optimality conditions are stated here without derivations. The derivation of the QSI Neyman–Pearson approach for the problem presents primarily a computational change, without any new conceptual difficulties (aside from the possible problem of using a frequentist approach in the quantum domain at all).

4.4 The Bayes Approach

We suppose the quantum mechanical system is characterized by one of two density operators ρ_0 and ρ_1, which correspond to the two competing hypotheses (states of nature). It is not assumed that these density operators commute: they need not have a joint distribution or be jointly measurable. All operators in this discussion act on a Hilbert space of states of the physical system whose density operator is either ρ_0 or ρ_1.

Prior probabilities for the two states are ζ_0 and ζ_1, with $\zeta_0 + \zeta_1 = 1$, and a cost matrix $\mathrm{C} = [C_{ij}]$ is assumed, with C_{ij} being the cost of making decision i when the true density operator is j, for $i, j = 0$ or 1.

Helstrom (1976, Chap. 4) and Yuen, Kennedy and Lax (1975) show that to solve the binary decision problem we must find two nonnegative Hermitian detection operators Π_0 and Π_1. The practical interpretation of these operators is this: we observe an outcome, and think of the operators Π_0 and Π_1 as two bells. If Π_0 rings, we choose ρ_0 as our guess of the state of nature; if Π_1 rings, we choose ρ_1. Note that the operators are constrained by the condition

$$\Pi_0 + \Pi_1 = \mathbf{I}.$$

This implies

$$\Pi_0(\Pi_0 + \Pi_1) = (\Pi_0 + \Pi_1)\Pi_0,$$

which yields

$$\Pi_0\Pi_1 = \Pi_1\Pi_0.$$

Thus, these detection operators do commute, and are also bounded: they therefore have a joint distribution [a result from functional analysis; see Holevo (1982, p. 71)] and so are in principle jointly measureable. Also, since the desired solution is a constrained optimum (the detectors must sum to the identity), it is in general necessary to introduce an appropriate Lagrange operator Γ. However, Holevo (1973) (see also Helstrom, 1976, p. 93) has shown that the optimality equations, including the Lagrange condition, can be replaced by a single set of symmetrical conditions

$$\Pi_k(\mathrm{W}_m - \mathrm{W}_k)\Pi_m = 0, \qquad \text{for all pairs } k, m; \quad 0 \le k, m \le 1,$$

where

$$W_i = \zeta_0 C_{i0}\rho_0 + \zeta_1 C_{i1}\rho_1, \quad \text{for } i = 0,1$$

for priors ζ_i and density operators $\rho_i, i = 0, 1$. Next, define

$$\lambda = \frac{\zeta_0(C_{10} - C_{00})}{\zeta_1(C_{01} - C_{11})}.$$

This ratio also arises in classical theory, where it yields the decision-boundary value of the likelihood ratio.

Consider the operator $\tau = \rho_1 - \lambda\rho_0$. The optimal detection operators Π_0 and Π_1 are derived from the spectral resolution of τ and its decomposition into a weighted sum of projection operators.

Following Helstrom (1976, pp. 107–108) the sought-for optimal Π_1 is shown to be the projection operator onto the subspace spanned by the eigenvectors corresponding to positive eigenvalues of τ, while Π_0 projects onto the subspace spanned by the eigenvectors corresponding to negative eigenvalues of τ.

If τ has zero as an eigenvalue, the subspace spanned by any eigenvectors corresponding to the zero eigenvalue contributes nothing to the Bayes cost: a random decision can be made when the observed value of τ is zero, with any probability of accepting the null hypothesis, without changing the Bayes cost. (See Item A21.)

EXAMPLE. The materials are now available to explicitly derive the QSI minimum Bayes risk rule given in Section 2.

Using just the density operator for spin measurements given in Section 2, and the squared-cosine rule for probabilities given in the example there, proceed as follows.

First, calculate angles θ_1 and θ_2 which reproduce the probabilities displayed in Table 1. As stated in the example of Section 2, these values are

$$\theta_1 = 101.54^\circ \quad \text{and} \quad \theta_2 = 53.13^\circ.$$

Next let $\rho_1 = \rho(\theta_1)$ and $\rho_2 = \rho(\theta_2)$ be the two required density operators. The optimal QSI operator is equiva-

lent to projecting onto the positive and negative eigenspaces of

$$\tau = \rho_2 - \gamma\rho_1 \quad \text{where } \gamma = \xi/(1-\xi).$$

For each value of γ, find the positive eigenvalues of the operator $\tau = \tau(\gamma)$. For the angles given above, τ has exactly one positive eigenvalue $\eta^+ = \eta^+(\xi)$. (See Item A22.)

From Helstrom (1976), and using the 0–1 loss structure of the problem, the minimum Bayes risk is then

$$\mathcal{R}_{\min}(\xi) = (1 - \xi)(1 - \eta^+(\xi)).$$

The (symmetric) quadratic risk curves appearing in Figures 2 and 3 are the plots of $\mathcal{R}_{\min}$ against ξ.

Again, the classical risk curves displayed are the easily derived polygonal lines as shown; see for example DeGroot (1970). (See Item A23.)

4.5 The Neyman-Pearson Approach

For the binary decision problem, a Neyman–Pearson test can be performed using an operator of the same form as that used for the Bayes approach; only the constant γ is modified. (We proceed here as if a quantum-consistent frequentist approach is indeed viable; see our comments above.)

Start with an operator of the form $\tau = \rho_1 - \gamma\rho_0$. Let Π_1, the detection operator, be given by

$$\Pi_1 = \Pi_+ + \nu\Pi_{0+},$$

where Π_0 is the projection onto the space of positive eigenvectors of τ, Π is the projection onto the space of eigenvectors associated with eigenvalue 0 of τ and ν is a constant to be determined, $0 \le \nu \le 1$.

The Neyman–Pearson test operates by rejecting the null hypothesis whenever an eigenvalue of Π_1 is the observed outcome, and where γ and ν are adjusted to obtain the desired size of the test.

As discussed in Helstrom (1976, pp. 110–111), and indeed in the classical context as well, if the (convex downward) operating characteristic curve of the test is continuous as a function of γ, then a size α test can be found for any α, $0 < \alpha < 1$. In this case ν is set to 0. If the curve is polygonal then γ is found first, as the slope of the curve on which the desired size is found, and then ν is adjusted accordingly. A nonzero value of ν results in a randomized strategy.

In either case ($\nu = 0$ or $\nu \neq 0$) the power of the test is then found from

$$1 - \beta = \text{power of the test} = \operatorname{tr}[\rho_1\Pi_1],$$

for β the probability of falsely rejecting the null hypothesis. (See Item A24.)

4.6 Heisenberg Uncertainty and Cramér-Rao

The machinery developed so far permits a condensed demonstration of the relation between the Heisenberg uncertainty principle and the Cramér–Rao inequality. The discussion largely follows Helstrom (1976), but includes some important technical details, some of which are discussed by Holevo (1982), and then illustrates the connection for the case of position and momentum operators.

Just as in the simple spin problem considered here, quantum-consistent optimal estimation begins by parameterizing the density operator. It is important to note that the parameterization amounts only to selecting a family of operators of interest and does not impose any restrictions on the underlying Hilbert space $\mathcal{H}$ nor on its operators.

Suppose θ is to be estimated by observing the outcome of a quantum-mechanical operator X; write $\hat{\theta}$ for the value of the outcome when X is measured. The bias, $b(\theta)$, of the estimator X is given by

$$b(\theta) = E_\theta(\hat{\theta}) - \theta = \operatorname{tr}[\rho_\theta X] - \theta \operatorname{tr}[\rho_\theta]$$
$$= \operatorname{tr}[\rho_\theta(X - \theta)],$$

since $\operatorname{tr}[\rho_\theta] = 1$.

In the classical Cramér–Rao inequality the logarithmic derivative of the probability density plays a prominent role. Analogously it is useful to introduce quantum operators which act as logarithmic derivatives of the density operator. Here only the *symmetric logarithmic derivative* is considered, but one-sided (right and left) logarithmic derivatives are also useful; see Holevo (1982, pp. 257, 278), Helstrom (1976, pp. 116, 283–285).

First, note that

$$\operatorname{tr}\left[\frac{\partial}{\partial\theta}(\rho_\theta)X\right]$$

defines a continuous linear functional on the Hilbert space of all square-summable operators X on $\mathcal{H}$. [The partial derivative of the density operator ρ_θ is assumed to be sufficiently regular: it is defined as a strong-operator limit on $\mathcal{H}$; see Holevo (1982, pp. 82, 257).] Then by Riesz's theorem for continuous linear functionals (see, e.g., Prugovečki, 1981, p. 184) there exists an Hermitian operator L_θ such that

$$\frac{\partial}{\partial\theta}(\rho_\theta) = \frac{1}{2}(\rho_\theta L_\theta + L_\theta\rho_\theta).$$

The operator L_θ serves as the quantum analog to the classical logarithmic derivative

$$\frac{\partial(\log\rho_\theta)}{\partial\theta}.$$

A calculation entirely along very classical lines then yields

$$\operatorname{var}_\theta(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)^2 = \operatorname{tr}[\rho_\theta(X - \theta)^2]$$
$$\ge [1 + b'(\theta)]^2 / \{\operatorname{tr}[\rho_\theta(L_\theta)^2]\}.$$

This is one version of the univariate quantum Cramér-Rao inequality. The reader may notice the connection with the classical result in that the denominator above is precisely

$$\mathrm{tr}[\rho_\theta (L_\theta)^2] = E_\theta\left[\frac{\partial \log \rho_\theta}{\partial \theta}\right]^2.$$

An alternative form of the inequality is sometimes more useful. Using

$$\mathrm{tr}\left[\frac{\partial}{\partial \theta}(\rho_\theta)\right] = 0,$$

gives, finally,

$$\mathrm{var}_\theta(X) \times \mathrm{var}_\theta(L_\theta) \geq \left[\frac{\partial}{\partial \theta}\{E_\theta(X)\}\right]^2.$$

The practical utility of this result now depends on being able to calculate the logarithmic derivative L_θ. For the important special class of *Gaussian states* (see Holevo, 1982, Chap. 5) the derivatives are explicitly known. Otherwise, there is an inequality for the variance of the symmetric logarithmic derivative that circumvents the need to calculate the logarithmic derivative itself; see Holevo (1982, p. 265) or Malley and Hornstein (1992).

More importantly, the quantum Cramér-Rao inequality just derived, using the symmetric logarithmic derivative, provides in principle *more* information than the usual Heisenberg uncertainty relation, in that the lower bound above is greater than the Heisenberg bound, and strictly so, unless the density operator represents a pure state: see Holevo (1982, p. 265). In other words, the QSI approach here yields strict improvements over the more familiar quantum uncertainty rule, and the door is open for investigation of when physical realizations of the quantum Cramér-Rao rule can be obtained. Such results would improve current experimental practice.

The multiparameter quantum-consistent form of the Cramér-Rao inequality is also a fairly straightforward reworking of the classical argument: see Holevo (1982, pp. 274-277) and Helstrom (1976, pp. 266-270). Only the required definitions and the conclusion need be given here.

Let the density operator be parameterized by a vector

$$\theta = (\theta_1, \ldots, \theta_m)$$

with

$$\rho_\theta = \rho(\theta_1, \ldots, \theta_m).$$

Letting X_i, $1 \leq i \leq m$, be an unbiased operator for θ_i, form the covariance matrix

$$\mathfrak{B} = [\mathfrak{B}_{ij}], \qquad \text{for } \mathfrak{B}_{ij} = E[(X_i - \theta_i)(X_j - \theta_j)].$$

Next, define the m symmetric logarithmic derivatives implicitly by

$$\partial \rho_\theta / \partial \theta_k = \tfrac{1}{2}(\rho_\theta L_k + L_k \rho_\theta), \qquad \text{for } 1 \leq k \leq m.$$

The QSI version of *Fisher information* is then given by

$$\mathfrak{A} = [\mathfrak{A}_{ij}], \qquad \text{for } \mathfrak{A}_{ij} = \tfrac{1}{2}\mathrm{tr}[\rho(L_i L_j + L_j L_i)].$$

The multiparameter QSI Cramér-Rao inequality is, finally,

$$\mathfrak{B} \geq \mathfrak{A}^{-1}.$$

4.7 Example: Position and Momentum

The univariate quantum-consistent Cramér-Rao inequality given above is almost in a useful form for dealing with position and momentum, say.

Any particular application needs three additional items: the proper choice of the operator X, specification of the correct parameterized density operators and then calculation of the logarithmic derivative. Given these, the quantum Cramér-Rao inequality can be applied to estimating the distance by which a state has been rigidly displaced from a reference state (see Malley and Hornstein, 1992). In this instance the *QSI version of the Cramér-Rao inequality implies the Heisenberg uncertainty principle for position and momentum.*

4.8 QSI and Parameter Estimation

A rigorous analysis of quantum parameter estimation requires a systematic use of *spectral measures* for operators having a continuous spectrum. (See Item A8.) Quantum analogs of many classical estimation strategies have now been derived for a wide variety of problems of physical importance. These include quantized Bayes estimators (under various loss functions), maximum likelihood estimation, uniform minimum variance unbiased estimation, minimax and generalized least squares; see Holevo (1982, 1973) and Helstrom (1976, especially Chap. 8) (see Item A25).

5. CURRENT AND POTENTIAL APPLICATIONS OF QSI

There have been four areas in which QSI has already been applied:

(i) Quantum analogs of classical results in mathematical statistics;
(ii) problem-specific quantum communication and detection problems;
(iii) quantum information theory;
(iv) foundational questions in quantum theory.

We briefly discuss these areas and then consider other potential applications of QSI.

5.1 Quantum Analogs of Classical Statistical Results

Helstrom (1967) developed the QSI version of minimum mean-squared error estimation, and this is the source of the Cramér–Rao inequality discussed in Section 4. It has been generalized in several ways.

Personick (1970, 1971) extended the result to conditional-mean minimum mean-squared error, by allowing the parameter vector for the density operator to be a (classical) random variable, with a known distribution. He also obtained quantum versions of the Bhattacharyya and the Barankin lower bounds. The results were then applied to a wide range of communication and signal detection problems, often with substantial gains in detection power and sensitivity.

The QSI analog of minimax estimation was considered by Holevo; see, for example, Holevo (1982, p. 191). Sekiguchi, Hirota and Nakagawa (1990) developed details of the QSI quantum minimax rule, applying it to general quantum state signal estimation. A simplified method of calculation was proposed, employing a geometric interpretation.

Nagaoka (1990) produced a new lower bound of the "Cramér–Rao type" that improves on the bounds described in Section 4 and showed that the new bound maximizes all lower bounds of the Cramér–Rao type. The calculations involve the symmetric logarithmic derivative, but the optimal solution cannot be obtained explicitly except in special cases.

Shapiro, Shepard and Wong (1989) and Shapiro and Shepard (1991), addressed the question of optimum phase-shift estimation of a single mode of a radiation field, under an average-photon-number constraint, finding that the (quantum optics) Susskind–Glogower phase operator is (in the limit) precisely the QSI maximum likelihood estimator for all input states.

Holevo (1991) is among a group of QSI researchers that is systematically developing the field of *quantum stochastic processes*. This yields a consistent, rigorous theory of continuous-time, quantum measurements and an associated quantum-consistent Itô stochastic calculus. This work can be traced from the references in Holevo (1991) and Belavkin (1991); more recently see Accardi and Koroliuk (1992).

5.2 Quantum Communication and Detection Problems

A large and flourishing literature exists on applications of QSI to communication problems in which quantum noise is important. The precise details of the problems are beyond the scope of this article as they assume considerable knowledge of the physics and electrical engineering of communication, detection and receiver design.

Still, a partial list of work in this area must include: Shapiro (1980), Yuen and Shapiro (1980), Shapiro and Wagner (1984), Helstrom, Charbit and Bendjaballah (1987), Charbit, Bendjaballah and Helstrom (1989) and Bendjaballah, Hirota and Reynaud (1991).

Helstrom and Holevo have both made many contributions to the practical applications of QSI, many of which are discussed in detail in their respective books, Helstrom (1976) and Holevo (1982).

The theory of *optimal control of quantum systems* is being developed, with the help of QSI. This theory should eventually deepen both the theory and practice of quantum measurements. For the current status of quantum control theory see Butkovskiy and Samoilenko (1990).

5.3 QSI and Quantum Information Theory

While this article has emphasized statistical decision theory and estimation, there is also a quantum-consistent counterpart to information theory. Gabor (1950) first discussed the need for such a development, but Holevo (1973) was probably the first to move in this direction, and since then Ingarden (1976), Davies (1978), Ohya (1983) and Vourdas (1990) have made contributions. Recent work in this field can be found in Bendjaballah, Hirota and Reynaud (1991).

We also observe that the work on quantum information theory overlaps conceptually with that of Blankenbecler and Partovi (1988), Jaynes (1989), Levine (1989) and Schumacher (1990), who investigate how the notion of maximum entropy can clarify or account for various practical as well as theoretical features of quantum theory.

5.4 Quantum Theory Foundations

The papers of Jaynes (1989) and Levine (1989) also contain discussions of how the quantum state and the density operator may have a Bayesian interpretation. In particular they advance the idea that since one interpretation of the density operator is that it does not specify an objective element of reality but instead quantifies our knowledge of a quantum system, then it is reasonable, perhaps indeed necessary, to view the quantum state as a subjective element of QM. A precise and complete formulation of this outlook would have far-reaching consequences for QM as well as Bayes theory.

That a Bayesian outlook is evidently of fundamental importance in QM is established by the remarkable connection that exists between the long-standing problem of what one means by "identical" particles, the de Finetti representation theorem and the idea of exchangeability. The details appear in van Fraassen (1991, pp. 413 ff.).

The subjective interpretation of quantum states is quite controversial. In particular, it has been vigorously contested by Margenau (1963); Popper (1968, p. 225; 1982, pp. 75, 104–118) and Ballentine (1970).

An alternative formulation of how Bayesian methods

could apply in QM is developed in the work of Gudder (1988), his coworkers and many others, on *quantum logic*. It is still a matter of further research precisely which components of this approach represent truly new insights and problem resolutions, and which (not unimportantly) represent only new derivations of results obtainable by careful use of existing Bayesian methods. Finally, of course, a truly novel derivation and point of view may suggest truly new results.

An unusual line of related research involves a quantum logic analog for maximum likelihood estimation. An intriguing argument is made for this approach, which is then used, for example, to disentangle the very classical statistical problem posed by Simpson's paradox; see Gudder (1988). On the other hand, an initial evaluation suggests that here, for example, a careful, nonquantum Bayes argument leads to the same result (personal communication: Larry Wasserman and Mark Schervish). Evidently the notion of exchangeability enters into this discussion as well: see van Fraassen (1991, pp. 63–65).

Recall that at several points the present article has promoted the view advanced by Helstrom and Holevo that the quantum measurement process is usefully viewed as the process of quantum statistical estimation and decision making. In this regard, Ozawa (1991) has investigated how measurements of position and momentum can be made so as to reduce the error in position, say, at the explicit expense of precision in measuring momentum, with the aim of constructing experiments, for example, to detect gravity waves (assuming them to exist). The product of the measurement variances is still, of course, bounded from below by the Heisenberg uncertainty principle.

The design of an experiment for the deliberate tradeoff of error variances for this pair of variables is important in the field of quantum optics, with the production and applications of *squeezed light* (and other forms of *nonclassical light*), and is a growth industry in itself; see Yamamoto et al. (1990) for a recent review. These novel and powerful uses of classical and nonclassical light in a wide range of applications are part of what has been called the *photonic revolution*. The area is another example of where a classical statistical approach (the design of experiments) could be profitably and rather directly carried forward by statisticians into the quantum domain. It is also another example of the *systems approach* to QSI advanced by Shapiro and Shepard (1991): here one works to optimally control the (input) quantum state as well as the optimal detection scheme.

5.5 Potential Applications and New Directions

The Introduction mentioned several areas where QSI could potentially have a practical or theoretical impact. In each area, only a sustained collaboration between the QSI-informed statistician and the subject-matter specialist (physicist, molecular biologist, quantum optics expert, etc.) will yield useful results. Technically, the problem usually reduces to determining the correct, parameterized density operator and then finding explicit solutions for the optimal decision rule or estimator. Both halves of the problem are usually highly nontrivial.

Embedded in the problem of identifying explicit optimal solutions is that of devising realizations of the optimal operators. A useful first step in this direction would be a means whereby the optimal operator could be assumed to be a projection-valued operator. In general only by invoking a result of Naimark (see Helstrom, 1976; Holevo, 1982) can this step be guaranteed. The theorem introduces an ancillary Hilbert space and extended density operator, and so may be more a change in venue, rather than a full solution to the realization problem. However, introduction of an ancillary Hilbert space can also be exactly what is needed: see Shapiro, Shepard and Wong (1989) and Shapiro and Shepard (1991).

In another direction, Holevo (1973; 1982, pp. 298–302) considers linear unbiased quantum measurement, obtaining quantum analogs for uniform minimum variance unbiased estimates (QSI linear UMVUEs). We note the strong similarity between these results and those of Zyskind, Seely and Rao (see Zyskind, 1967; Seely and Zyskind, 1971; Rao, 1973; Malley, 1986) and we speculate that many of these classical results have natural quantum equivalents, whose derivations would closely parallel the classical ones.

We further note that the results mentioned in subsection 4.6 often yield solutions that are known to be not realizable as Hermitian operators on the original Hilbert space, and Naimark's results must be drawn into the argument to obtain operators that are at least Hermitian on an enlarged Hilbert space system.

This latter result does not truly resolve the problem of physical realizability, of course, but we *conjecture* that a quantum version of the expectation–maximization (EM) algorithm, or a quantum version of the Gibbs sampler, applied to the larger space, could be employed [for background see, e.g., Gelfand and Smith (1990)]. Either of these would treat part of the data as "missing" and the added parameters (specifying the density operator on the larger Hilbert space) are then extraneous parameters to be estimated (or for which samples from the correct posteriors are generated). These methods would produce all required parameter estimates or posteriors, without having to construct new laboratory devices, similar, for example to use of the EM algorithm that produces efficient estimates of all parameters (original or added) without having to actually impute missing values. This is an area, therefore, for which

statisticians could materially contribute to experimental practice in the quantum domain. (See Item A27.)

Finally, there remains the task of validating our *conjecture of equivalence* for quantum statistical inference enunciated in Section 1. This will entail increasing the precision of the statement of the conjecture and extending its range. The conjecture was verified here only for a special two-state decision problem. Other decision problems and a variety of estimation problems need to be investigated to confirm the conjecture, or characterize its range of applicability, and possibly elevate it to the status of a *principle* of equivalence for quantum statistical inference.

6. CONCLUSION

Quantum statistical inference is an established, active, although not widely known, synthesis of quantum physics and statistical methods. The example presented in Section 2 illustrates the value and novel challenges of QSI. The QSI treatment of the spin decision problem, the discussion of quantum estimation in Section 4 and the applications cited in Section 5 show that a quantum-consistent framework for decision making and inference is already available and is both often practical and potentially far-reaching.

Our introduction of the principle of local equivalence for quantum statistical inference is a plausible extension of earlier ideas (see, e.g., Feynman, 1951), and admits a more precise statement because of the technical context provided by QSI. Also, a wide range of familiar statistical methods now have precise equivalents or near equivalents in the quantum domain.

A number of statistical inference problems have also been discussed; some have been resolved here, many others remain open. Solution of some of the open problems would extend the practical utility of QSI. For example, in subsection 4.6 it was noted that the QSI derivation of Heisenberg uncertainty results in a quantum Cramér–Rao lower bound that is for mixture states strictly sharper than the usual derivations. Finding physical realizations of the relevant operators could thus sharpen experimental practice. Or our conjecture described in subsection 5.5 about usé of a quantized version of the EM algorithm or of the Gibbs sampler could avoid the realization problem altogether. And in the wake of the recent results of Mermin and others, components of our current understanding of statistical practice appear to be significantly challenged. Thus, must frequentist methods be embargoed from use in the quantum domain?

Perhaps most significant of all is the new vantage point provided by QSI: *the process of quantum measurement is basically quantum-consistent statistical estimation and decision making.*

This viewpoint, coupled with QSI's efficient synthesis of quantum mechanics and statistical theory, should bring varied and significant benefits to both statisticians and scientists. Included in the possible benefits to scientists we seek, for example, to bring the established methods of the rapidly developing field of quantum optics, and its use of nonclassical light, to bear on sharply improving a variety of biomedical techniques of detection and estimation, as outlined in the Introduction.

APPENDIX

A1. This equivalence does not undermine the practical importance of QSI, and instead validates it, by showing that its conclusions are essentially unique. Thus, any practical application that invokes equivalence, instead of QSI – that is, ordinary statistical inference, but with the correct quantum space of outcomes and probabilities – will usually require a tedious and unstructured search for optimal decision rules and estimators, only to end with a result that may have already been provided by QSI.

This hybrid approach – classical statistical inference, but with outcomes and probabilities calculated from quantum mechanics – is the normative practice in physics and quantum electronics. According to "equivalence," QSI substantiates the standard practice when it includes a search for the optimal decision rule or estimator. QSI offers improvements as compared to the standard practice by avoiding an *unstructured* search for the optimal rule or estimator.

A2. Historically the traditional uncertainty relations were applicable only to variables that are represented quantum mechanically by Hermitian operators; they were therefore not applicable to time, phase angles and other angles. Embracing an approach standard in classical statistics, QSI treats time and angles as *parameters* of the apparatus which produced the objects to be measured. By considering such quantities as parameters to be estimated, or about which decisions must be made, QSI is able to finally secure mathematically sound uncertainty relations for these important quantities.

A3. We briefly discuss some further details of spin. Operators (such as spin) have only a discrete set of eigenvalues and are said to have a *pure-point spectrum:* only these discrete values will be observed. For the spin 1/2 particle, the operator representing the component of the spin angular momentum along the direction defined by a unit vector n is $+(1/2)(\sigma \cdot n)$, where $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ is a "vector" of three Hermitian operators. The standard matrix representation for the operators is given below, but it is important to note here that the eigenvalues of $\sigma \cdot n$ are $+1$ and -1 for *any* direction n. The value observed for any component of σ for an electron, say, must then be one of the two values $+1$ or -1: for any choice of n, these are the only eigenvalues of the operators $\sigma \cdot n$.

A4. Given the more detailed introduction to spin

above, let us look again at the three filter experiment. We use spin "one" systems that assume only the values $+1, 0$ and -1 (that is, "up," "no deflection" and "down").

We follow Feynman, Leighton and Sands (1965, Vol. 3, Chap. 5, pp. 5-6 to 5-10). The experiment consists of a sequence of three Stern-Gerlach devices, each set to split up a beam of spin one particles (e.g., silver atoms) into three channels, labeled $+1, 0, -1$. Each particle has a certain probability of passing through one of the devices in the chain when one or more of that device's output channels are blocked off. This event, the passage of a given particle through one of the devices, has by itself a classical probability of occurrence. This is assumed to be arbitrarily well-approximated by a large sample observed frequency.

Using a slight variation of Feynman's notation (which is not used elsewhere in this paper but which we hope makes the connection with Feynman easier), call the three devices, in sequential order, S, T and S^*. Devices S and S^* are set to have the same magnetic orientation, say vertically, while the device T, situated between S and S^*, is turned at some angle from the vertical, say 45°. Now, looking at device S, say, write $(+1|S)$ to denote that device S has been set to pass only $+1$ particles, its other outputs being blocked off. Also, write, for example, $(+1, 0|T)$ to denote that device T is set to allow passage of both $+1$ and 0 particles, but blocks passage of -1 particles.

Consider setting the channel filters on the devices in various ways, writing for example,

$$(+1|S) \to (+1|T) \to (0|S^*)$$

to represent the passage of a particle through the three devices, with the first two set to allow passage of only $+1$ particles, and the third only allows passage of spin state 0 (no deflection) particles. Actual experiment then demonstrates that the event represented by

$$(+1|S) \to (0|S^*)$$

has probability zero: no particles will be observed leaving the third device. Probability zero is also obtained for particle passage in the arrangement

$$(+1|S) \to (+1, 0, -1|T) \to (0|S^*),$$

while the experiment with the arrangement

$$(+1|S) \to (0|T) \to (0|S^*)$$

will be observed to yield a strictly nonzero probability.

At this point we have observed something unexpected: putting a more absorbing filter in the middle lets more particles pass through. We now make this more explicitly a violation of the Kolmogorov axioms for probability event spaces.

Thus let A be the event that a particle leaves the system represented by

$$(+1|S) \to (0|S^*),$$

and let B be the event that a particle passes through the arrangement

$$(+1|S) \to (0|T).$$

Note that the set-theoretic complement of B, B^c is represented by passage through

$$(+1|S) \to (+1, -1|T)$$

(remember that the notation means that the first device in this arrangement just given blocks off 0 particles), while passage through $(+1, 0, -1|T)$ given passage through $(+1|S)$ is the sure event, denoted by I.

Then, $B \cup B^c = I$ and classically we must have

$$\begin{aligned} A &= A \cap I \\ &= A \cap (B \cup B^c) \\ &= (A \cap B) \cup (A \cap B^c) \end{aligned}$$

by the usual distributive rule, and since the two events on the right are disjoint, probability should sum. Yet by experiment we know that the probabilities of each of the events on the right are strictly nonzero, while we also must have the probability on the left, $\Pr(A) = 0$. This is therefore an experimentally obtainable violation of one basic rule for classical probability event spaces.

A5. Useful references for the physics are Messiah (1961), Gottfried (1966), Schiff (1968) and Cohen-Tannoudji, Diu and Laloë (1977). Dirac (1958) is a good follow-up to any of these. Useful references for the industrial-strength mathematical details of QM are von Neumann (1932, 1955), Prugovečki (1981) and Beltrametti and Cassinelli (1981). None of this is easy reading, but Beltrametti and Cassinelli (1981) evidently provide the most succinct summary.

Concerning alternate formulations of QM, we quote from van Fraassen (1991):

> In the 1960s and 1970s quantum logic became a highly developed mathematical subject. The inspiration was in part the apparent, tacit agreement to develop quantum theory itself starting with the most general notion of a theoretical model which could possibly fit any data the future might bring. The postulates of quantum logic narrowed down this class of models step by step, and eventually the representation theorems showed that the state-space was constrained to be a Hilbert space. As the term "logic" shows, the project was subject to very high standards of rigour. This gives it a special value for the project of interpretation of quantum mechanics. For the more or less axiomatic development allows us to introduce the interpretational additions also at each relevant point, with absolute clarity. In my opinion, any philosophical view on contemporary physics should eventually measure itself by this standard. (pp. 229-300)

A6. The QM literature normally uses (the Dirac-introduced) angle brackets $\langle\psi|\zeta\rangle$ for the Hilbert space inner product, and correspondingly indicates a vector by $|\zeta\rangle$ and the associated dual vector by $\langle\zeta|$. The correspondence between direct and dual vectors is the usual natural one induced by the inner product. Thus, in the QM literature the projection operator onto the one-dimensional subspace spanned by $|\zeta\rangle$ is denoted by $|\zeta\rangle\langle\zeta|$.

A7: Example 1*. As mentioned earlier, the spin angular momentum of a spin 1/2 particle is described by operators σ_x, σ_y and σ_z.

In a convenient complex basis in the spin subspace, the operators corresponding to the x, y and z components of the spin vector are 1/2 σ_x, 1/2 σ_y, and 1/2 σ_z, where

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

are called the Pauli spin matrices. As an example of the convenience of the spin matrices, observe that measuring the component of spin along a direction lying in the (x, y) plane, at an angle θ from the x-axis, corresponds to the operator

$$\sigma_\theta = (\cos\ \theta)\sigma_x + (\sin\ \theta)\sigma_y = \begin{bmatrix} 0 & e^{-i\theta} \\ e^{i\theta} & 0 \end{bmatrix}.$$

Example 2*. In the infinite dimensional Hilbert space needed for an electron moving along the x-axis, say, the operators corresponding to position and momentum are: multiplication by x, and apart from a purely imaginary constant factor, differentiation by x, respectively.

Let X and P denote the operator associated with position and momentum, respectively. Then

$$X\varphi(x) = x\varphi(x),$$

$$P\varphi(x) = (-i\hbar)\frac{\partial}{\partial x}[\varphi(x)].$$

Note that X and P do not commute:

$$PX - XP = (-i\hbar)\mathbf{I}$$

for $\mathbf{I}$ the identity operator on $\mathcal{H}$. The noncommutativity of the position and momentum operators, and this "commutation relation" just given, figures prominently in the later discussion of the Heisenberg uncertainty principle.

A8. The observable X denotes a probability-operator *spectral measure* because the map (in a standard abuse of notation)

$$B \to \mathrm{tr}[\rho X(B)]$$

can be shown to define a real-valued measure on $\mathcal{R}^n$.

Mathematically, necessary and sufficient conditions for the existence of the required measures (p.o.m.'s) is a deep result due to Gleason (1957); see also Beltrametti and Cassinelli (1981).

It is also a feature of a p.o.m., X, that

(i) $X(\cup B_i) = \Sigma X(B_i)$, *for any sequence* B_1, *and* $B_2, \ldots$ *of pairwise disjoint Borel sets in* $\mathcal{R}^n$;

(ii) $X(\mathcal{R}^n) = \mathbf{I}$ *the identity operator on* $\mathcal{H}$.

These properties are said to define a *resolution of the identity*. If the p.o.m. maps Borel sets to projection operators which are orthogonal whenever the corresponding Borel sets are disjoint, then the p.o.m. is an *orthogonal resolution of the identity*. One of the major achievements of functional analysis, the *spectral resolution theorem*, applies to the orthogonal p.o.m.'s; the theorem is crucial to the mathematical foundations of QSI, but is not discussed here; see Yoshida (1980) and Akhiezer and Glazman (1981).

It is traditional in QM to assume that every observable induces (is associated with) an orthogonal p.o.m. For many problems, however, QSI permits or requires the use of nonorthogonal resolutions of the identity. The spectral resolution theorem is then applied to a larger Hilbert space, such that the operator extension of the original p.o.m. is a more tractable orthogonal p.o.m. This key, purely mathematical fact, is an early functional analysis result of Naimark; details appear in both Holevo (1982) and Helstrom (1976). It achieves its most thorough-going expression and utilization in the systems approach to QSI of Shapiro and Shepard (1991).

A9. When $\mathcal{H}$ is infinite dimensional still other serious complications may arise. The p.o.m. X may have an infinite number of eigenvalues or no eigenvectors at all: then one absolutely needs the machinery of the spectral resolution theorem for self-adjoint operators. So armed, the usual notion of eigenvalues can be carried over intact, while the notion of eigenvector is replaced by that of a *generalized eigenvector.*

We note that the use of eigenvalue decompositions via generalized eigenvectors may be rendered more intuitive by thinking of them as linear functionals on $\mathcal{H}$, or by regarding them as a symbolic shorthand for certain Lebesgue-Stieltjes integrals. Good references are Bohm (1986), Hoskins (1979), Zauderer (1983) and Shilov (1968). The profusion of references is offered as an aid to the reader, because generalized eigenvectors, working in concert with the angle bracket notation mentioned earlier, leads to a very convenient functional calculus, one widely encountered in physics, engineering and applied mathematics.

A10. On the other hand, when $\mathcal{H}$ is infinite dimensional, it often happens that the operator X has no proper (= nongeneralized) eigenvectors; the operators corresponding to the position and momentum of a

quantum system are usually of this type. Then the last equality above, expressing the probability of an outcome, is represented using an integral over $\mathcal{R}^n$ with respect to the *spectral measure* induced on $\mathcal{R}^n$ by X; see Prugovečki (1981, Chap. 3, Sec. 5) and Yosida (1980).

A11. In general, however, quantum mechanical calculations of probability, means, variances and so on, and the evaluation of functions of measurements, have two features that distinguish them from familiar calculations in classical mathematical statistics: first, the trace-rule for probability, and second, the technicalities introduced for not-purely discrete quantum measurements and (possibly infinite dimensional) Hilbert spaces $\mathcal{H}$.

A12. The trace-rule for probabilities contains a very important hidden moral: for an operator A, a given set of probabilities may be attached to the Borel sets $S \in \mathcal{R}^n$ by an infinite family of (generally nonunique) projection operators. This critical point is easy to prove rigorously using only basic linear algebra: see Beltrametti and Cassinelli (1981, p. 9).

Thus a mixed state ρ can be written as

$$\rho = \sum \alpha_i \mathcal{P}_i^{(1)} = \sum \beta_j \mathcal{P}_j^{(2)}$$

for distinct sets of positive constants $\{\alpha_i\}, \{\beta_j\}$ and distinct density operators

$$\{\mathcal{P}_i^{(1)}\}, \quad \{\mathcal{P}_j^{(2)}\}$$

(projections onto one-dimensional spaces spanned by ξ_i and ψ_j).

Both decompositions yield the same probabilities (via the trace-rule) for the outcomes associated with the given observable A. In fact, given a set $\{\alpha_i\}$ there is an infinite collection of sets of constants $\{\beta_j\}$ which give the same probabilities as $\{\alpha_i\}$ when the trace-rule is applied.

Therefore, for an arbitrary state ρ it is not possible to meaningfully (consistently) interpret the constants α_i or β_j as "probability weights," such that the quantum system is "in" the state i, defined by the density (projection operator) $\mathcal{P}_i$ (with probability α_i) or "in" state j (with probability β_j). Taking the trace in effect removes any "memory" or "trace" of the projectors used in an expansion for the density. Stating this important point somewhat differently, the interpretation of a single α_i depends on the entire set $\mathcal{P}^{(1)}$ of projectors and not on the projector $\mathcal{P}_i^{(1)}$ alone.

In view of the above, QM conspicuously resists any easy probabilistic interpretation of the coefficients occurring in a particular decomposition of a state. Using the notion from classical statistics this is summarized as: *a density operator is identifiable if and only if it is a pure state.*

On the other hand, it is still possible to envision the coefficients $\{\alpha_i\}$ as probabilities for a compound experiment, as follows. For each pure state appearing in a decomposition of a (mixture) state, imagine constructing separate machines or experiments such that the entity to be studied is known to be in that pure state.

Next, perform a secondary experiment that consists of selecting each primary experiment with a probability equal to the positive constant weight that is associated with each pure state. Making the given measurement on this two-stage experiment generates outcomes with the required probabilities for the observable in question.

But any other decomposition of the state, and its attached list of coefficients and pure states, amalgamated as above, also leads to the same results. So although the final results *may* be conceived as a two-stage experiment (indeed any member of an infinite family of such experiments), with an interpretation of the coefficients as probability weights for that *particular* two-stage experiment, one cannot in QM attach any intrinsic significance to the constants in decompositions of nonpure states.

A13. While the stated probability equation is experimentally valid, there is a nontrivial conceptual shift embedded in its calculation from the axioms of QM.

The equation gives the probability of an outcome associated with a repeated measurement: first measuring spin up with SG1 and then measuring spin up with SG2. It is not defined via a classical conditional probability, one that entails a joint distribution for the two spin measurements.

The nonexistence of a conventional joint distribution for this problem is, now, after nearly seventy years of research, standard and repeatedly verified by experimental study. We thus only briefly discuss this below; see Gudder (1988, p. 54) and Helstrom (1976, p. 66), for a more complete discussion of joint and conditional distributions and repeated measurements in QM. The problem is related to the study of *quantum mechanical stochastic processes*, an area currently under active investigation: see Section 5.

For the convenience of the reader we insert an adaptation of a result appearing in Gudder (1988, pp. 180–181) that proves the nonexistence of joint distributions for, in particular, spin 1/2 systems. Two easily derived results are needed. The first is:

Result 1. For real constants p, q and r, suppose that $0 \le p, q, r \le 1$. Then the events A, B and C can exist in a conventional (Boolean sigma algebra) probability space, with $\Pr(A \cap B) = p$, $\Pr(A \cap C) = q$ and $\Pr(B^c \cap C) = r$, if and only if $q \le p + r \le 1$.

For a single Stern–Gerlach device consider making three spin measurements in the x–y-plane in the directions u, v and w. Or, equivalently, consider making three spin measurements in succession, using three differently oriented Stern–Gerlach devices. Write A for

spin up in the u direction, B for spin up in the v direction, and C for spin up in the w direction; similarly write $-u$ for direction $(\pi - u)$, so measuring spin in this direction is equivalent to measuring spin down in the u direction. Also then, for example, B^c is the event spin down (spin $= -1$) in the v direction. With this notation *Result* 1 is restated as

$$\Pr(u \cap w) \le \Pr(u \cap v) + \Pr(-v \cap w),$$

and hence also

$$\Pr(-u \cap w) \le \Pr(-u \cap v) + \Pr(-v \cap w).$$

By physical symmetry the probabilities above are functions only of the included angle between the directions in their arguments. Denote the angle between u and v by θ_1, between v and w by θ_2 and between w and u by θ_3. and write $g(\theta_1) = \Pr(u \cap v)$ and so on. Then

$$g(\pi - \theta_1) + g(\theta_1) = \Pr(-u \cap v) + \Pr(u \cap v) = \frac{1}{2},$$

with a similar result for θ_2, and θ_3.

Thus

$$g(\pi - \theta_i) = \frac{1}{2} - g(\theta_i), \quad \text{for } i = 1, 2, 3.$$

The second probability inequality stated above now yields

$$\frac{1}{2} - g(\theta_3) \le \left[\frac{1}{2} - g(\theta_1)\right] + \left[\frac{1}{2} - g(\theta_2)\right],$$

and therefore

$$g(\theta_1) + g(\theta_2) - g(\theta_3) \le \frac{1}{2}.$$

Following Gudder (1988, p. 181), an induction argument gives:

Result 2. $g(\pi/n) \le (1/2)(1-1/n)$, for $n = 1, 2, \ldots$.

On the other hand, assuming the existence of a joint probability for measuring pairs of spins, and recalling the squared-cosine rule for the conditional probability described in Example 1 of subsection 2.3 (the quantum coin-tossing example), leads to

$$g(\theta_1) = \Pr(u \cap v) = \Pr(u)\Pr(v|u) = \frac{1}{2} \times \cos^2\left(\frac{1}{2}\theta_1\right).$$

To see this, recall that the probability of measuring spin up with an SG device for a spin 1/2 system is 1/2 for a particle having an initial arbitrary spin orientation. The joint probability is then assumed to be found from the usual product rule: the product of Pr(spin up observed first) and Pr(spin up observed in the new direction, given spin up first observed).

Taking $\theta_1 = \theta_2 = \pi/3$, and $\theta_3 = 2\pi/3$, we find that

$$p = \Pr(u \cap v) = g(\theta_1) = 3/8,$$

$$q = \Pr(u \cap w) = g(\theta_2) = 1/8,$$

$$r = \Pr(-v \cap w) = g(\pi - \theta_2) = 1/2 - g(\theta_2) = 1/8,$$

so that

$$q = 1/8 \le 4/8 = 1/8 + 3/8 = p + r.$$

Hence the condition of *Result* 1 is satisfied. Yet

$$\frac{1}{8} = g\left(\frac{2\pi}{3}\right) = g\left(\pi - \frac{\pi}{3}\right) = \frac{1}{2} - g\left(\frac{\pi}{3}\right)$$
$$\ge \frac{1}{2} - \frac{1}{2}\left(1 - \frac{1}{3}\right) = \frac{1}{6},$$

and this is a contradiction.

Experimental results therefore show that a classical Borel sample space treatment of the three spin outcomes is not possible. Equivalently, no conventional joint distribution probability is possible on this space of three spin outcomes that is consistent with all the marginal distributions. In still other language, the contradiction just obtained is a consequence of the noncommutativity of the three quantum spin operators.

A14. The quantum rule can also be experimentally implemented as follows.

Combine the output beams of SG2 into a single beam, doing this "coherently." Basically, this means the beams are blended into one without making any observation about up or down: see Feynman, Leighton and Sands (1965, p. 6-4). Then, send the single beam to a third Stern–Gerlach device, SG3, and arrange for SG3 to be aligned at the angle θ given above. Finally, observe "up" or "down" as the output of SG3, and decide for bias state 1 if "up" occurs.

Remarkably, coherently combining the output of SG2, and then using an observation from SG3, has the effect of eliminating entirely any quantum effect of SG2. Hence the quantum observer makes no use of the classical (unrotated) SG2 observation and uses an observation obtained for an angle θ and associated state that is neither bias state 1 nor 2.

Thus using either a properly aligned SG3 or a rotated SG2 and a *recalculated* classical decision rule will give the minimum Bayes risk (= 0.2448) at the prior $\xi = 1/3$. This is a working example of what we call the *conjecture of local equivalence for quantum inference*, which is more carefully stated as: *given the proper experiment and the associated quantum mechanical probability assignments, the classical rule maker can derive a decision rule with the same Bayes risk, at a given value of the prior, as the quantum rule maker.*

For this coin-tossing problem, the classical rule maker matches the optimal QSI risk only locally: all rules obtained classically for this problem have polygonal risk curves that are, at best, tangent to and everywhere else above the QSI curve. The QSI procedure is the lower envelope of all the classical rules that achieve the lowest Bayes risk at a given value of the prior. In effect, the QSI procedure smoothly collates all the searches (over all possible settings of SG2) for the optimal, classical Bayes rule: it is not required that we parse the problem into one of first estimating the

optimal amount to turn SG2 (or SG3), and then apply an optimal decision rule. Referring to Figure 2, observe that the classical risk is exactly the quantum optimal risk at *some* value of the prior ξ_0 between 0.5 and 0.6.

The conjecture of equivalence is therefore a local and, for now, narrowly defined statement: all probabilities must be either specified or determined from quantum theory, before the classical rule maker can uncover the optimal angle for setting SG2. No single classical rule for the problem above will match – for all values of the prior – the quadratic optimal risk procedure for this problem so succinctly derived by QSI.

A15. The structure of many important physical theories is an assemblage of three parts (see, e.g., Prugovečki, 1981; the corresponding three parts of QM are:

(i) a *Hilbert space* as the mathematical setting;
(ii) observables or measurements, represented by *operators* on that space, together with a rigorous link between operators and outcomes;
(iii) a differential equation called *Schrödinger's equation*, which expresses the time evolution of a quantum system.

A16. Actually, there are many Heisenberg uncertainty principles: one for each pair of "incompatible" observables; see Prugovečki (1981, p. 260) for the definition of "compatible" observables and Beltrametti and Cassinelli (1981, pp. 26–28) for the related, but mathematically distinct, idea of "complementary" observables.

A17. Necessary (only) conditions for Case (ii) to obtain include:

(a) The operators A and B do not commute;
(b) neither A nor B is bounded;
(c) A and B have no point spectrum (i.e., the induced Borel measures are nonatomic), or, if either one of them has a nonempty point spectrum, then its eigenvectors lie outside the domain of definition of the other.

When A and B commute, the right-hand side of the inequality above vanishes. It is then possible to find states in which both A and B have values (possibly distinct) assumed with probability 1.

On the other hand two noncommuting operators (bounded or unbounded) can still have one or more eigenvectors in common: they simply cannot have all their eigenvectors in common. That is, noncommutativity is *not* equivalent to a positive lower bound for all states, and thus some care is required in equating noncommutativity with Heisenberg uncertainty.

Also, if a pair of observables do have a joint distribution then the right-hand side of the equation above has no positive lower limit, and arbitrarily precise joint measurement is possible. A partial converse to this result exists; see Louton (1975) for a careful study of the relationship between the possibility of joint measurement and the vanishing of the lower limit in the inequality.

A18. The uncertainty principle has been especially targeted for interpretation in conflicting, inaccurate ways. These constitute, however, only some of the serious obstacles to understanding the unique challenges that quantum mechanical data pose for statistical inference, and we confront some of them here.

One such argument has been proposed in an attempt to refute the nonexistence of joint probabilities. Start with many instances (experimental realizations) of a quantum system in a given state. For some of the systems, measure first the position and afterward the momentum; for the others, measure the momentum first and then the position. List the paired observations. It is sometimes claimed that the list constitutes an empirical joint distribution of position and momentum.

But the claim is false, because the second measurement in each pair was not performed on a system in the same state as the first, or equivalently, it requires measurement on a different quantum object. Instead of constituting an empirical joint distribution for a single population of systems, the list instead contains two empirical conditional distributions, which pertain to two distinct populations; see Gudder (1988).

There have been still other attempts to domesticate the uncertainty principle and the nonexistence of certain joint distributions, that is, to find familiar phenomena – often outside of physics – that mimic quantum mechanics but that have less stressful, more satisfying explanations. But so far the search for analogs has not borne fruit, and this search has now extended over more than seventy years.

It is instructive to look at two of these other attempts, and in what ways they fail. The *first* approach tries to understand quantum phenomena by seeing them as merely a manifestation of "measurement order effects," as in: "measuring A before B gives a different result than measuring B before A." The noncommutativity of certain pairs of quantum measurements is then taken to be the sole cause of the distinctive features of QM.

But in this "order effect" model of QM, conditional probabilities play a major role: keeping track of conditional probabilities is all (it is asserted) that is needed to make the conceptual switch from classical to quantum physics. However, in any conventional statistical structure, conditional probabilities are defined in terms of joint probabilities, and, as seen earlier, in QM some of the pertinent joint probabilities simply do not exist.

The *second* approach seeks out familiar macroscopic arenas where joint probabilities do not exist. It is argued that a joint probability is not humanly determinable for one or another everyday phenomenon.

But simple ignorance of a joint probability is not sufficient to make the case: to provide an "everyday

experience" model of quantum phenomena it is necessary to show that merely assuming the existence of a joint probability for the relevant pair of variables must inexorably lead to a mathematical contradiction. No non-QM phenomenon meeting this stringent test has yet been found.

To sum up, the uncertainty principle and the nonexistence of certain joint distributions have thus proven mathematically sound, physically valid but, finally, basically impervious to easy interpretation, let alone along classical statistical lines.

This may be news to some, or just wrong for others, but it is basically "old business" that conventional physics and conventional statistical models cannot be blended into a tame and docile interpretation because (1) values of quantum variables do not in general exist prior to measurement and because of (2) the nonexistence of joint distributions.

A19. Thus, to be current and honest, some invocations of QSI still lead only as far as sets of operator equations, which only implicitly define the optimal solutions, and for which it is not yet known how to physically realize the optimal solutions.

Pursuing this important point a little further, realizability is an issue in QSI but not in classical statistical inference because of the novel character of quantum data. Thus, choosing to measure some quantities excludes having data on others, at least for the same set of "events." In classical statistical inference the existence of joint distributions guarantees the existence of a set of comprehensive data providing all the types of information that will be needed by all conceivable decision rules (although not all rules will inquire about all the types of information available).

QSI decision rules thus yield a *taxonomy of equivalence classes*, each requiring its own type of data, which may be incompatible with the types of data required by the others. It is then possible that a particular decision rule will require an impossible combination of types of data. Since the structure of QSI correctly incorporates QM, it is plausible that such cases will not remain unresolved, but the question is not definitively settled.

A20. The more general problem of deciding between more than two alternatives is a natural extension of the binary problem; this M-ary decision problem is treated in Yuen, Kennedy and Lax (1975), and of course Helstrom (1976).

A21. More precisely, let η_i be the eigenvector of τ corresponding to positive eigenvalue η_i. (This abuse of notation can be conveniently avoided by using Dirac's delta function notation, but to avoid overloading the reader with novelties that is not done here.) Following Helstrom (1976, Chap. 4) the optimum detection operators are

$$\Pi_0 = \sum_{\eta_i<0} \eta_i \eta_i^* \quad \text{and} \quad \Pi_1 = \sum_{\eta_i>0} \eta_i \eta_i^*.$$

The optimal rule is: decide state 0 whenever a measurement of the observable corresponding to Π_0 yields the value $+1$, and decide state 1 when the observable corresponding to Π_1 has the value $+1$. These operators are orthogonal and in each case if one observable is $+1$ then the other must assume the value 0.

The minimum Bayes risk $\mathcal{R}_{\min}$ is found to be

$$\mathcal{R}_{\min} = \operatorname{tr} \Gamma = \zeta_0 C_{00} + \zeta_1 C_{01} - \zeta_1 (C_{01} - C_{01}) \sum_{\eta_i>0} \eta_i.$$

A22. One calculates that the positive eigenvalue, as a function of the prior, is

$$\eta^+(\xi) = \tfrac{1}{2}(1-\beta) - \tfrac{1}{2}\left(1 + \beta^2 - 2\beta \cos(\theta_1 - \theta_2)\right)^{1/2},$$

for

$$\beta = \frac{\xi}{1-\xi}.$$

A23. The classical and quantum Bayes decision rules displayed above for our basic spin problem, together with the earlier discussion of the problem of making joint simultaneous measurements, now permits a detailed demonstration of the nonexistence of a data set which could be fed to competing decision rules for a data-based evaluation.

We show that: *there can be no single data set that can be jointly processed by both the classical and the quantum rules given in the example.* Basically, this is because the two rules in the example are specified by p.o.m.'s whose associated spectral measures do not commute.

As proof, recall a theorem of functional analysis on commuting spectral measures [see, e.g., Holevo (1982) p. 71]. Begin by checking that for any spin operator σ_π, the only other spin operator σ_φ commuting with it must have $\theta = \varphi$, or $= \varphi + \pi$. In our example the classical and quantum rules are neither identical nor separated by a displacement angle of π, hence they cannot be jointly measurable. Thus they cannot both be derived from a single bivariate, physically realizable data set.

Of course, other decision problems, not discussed here, may have optimal classical and quantum rules which do have commuting spectral measures. Examples are problems for which the two density operators commute; the quantum and classical rules are then identical. But cases where the quantum decision operators commute (or where a classical rule and a quantum rule commute) are evidently infrequent. We are not aware of any decision problems which allow jointly measurable data for the optimal quantum and the classical rules.

The argument above assumes that the existence of a joint distribution is synonymous with the possibility of making a joint measurement. This assumption is standard in quantum theory; see Varadarajan (1962), Prugovečki (1981, p. 260) or Bohm (1986, p. 72).

A24. Indeed, even for $n = 1$, the operator τ is not itself obviously realizable. Only fairly recently (Swift and Wright, 1979) was it shown that a general spin operator (an arbitrary linear combination of the Pauli spin matrices) could always be realized by using a *generalized SG device*, in which there is an inhomogeneous *electric* field in addition to the usual inhomogeneous magnetic field.

A25. In both approaches above (Bayes and Neyman–Pearson) sample size was not explicitly an issue; it was fixed at $n = 1$. Consider again density operators ρ_0 and ρ_1, but now suppose that multiple sample outcomes are available.

It follows from one of the axioms of QM, not discussed above, that the density operator corresponding to n samples from identically prepared systems is the n-fold tensor product of the one-sample density operators, and these act on the n-fold tensor product of the underlying Hilbert space.

This is so even though the particles measured in the n-sample experiment are identical (bosons or fermions), because the n-sample experiment distinguishes them in time, space or both. For example, the same particle can be run repeatedly through the experimental apparatus at n different times, or n particles can each be run once through n different (numbered, noninteracting) copies of the apparatus.

With either arrangement, and using either the Bayes or the Neyman–Pearson approach, the operator $\tau^{(n)}$ for the problem with sample size n is formed from

$$\rho_1^{(n)} = \rho_1 \otimes \cdots \otimes \rho_1 \qquad (n \text{ copies})$$

$$\rho_0^{(n)} = \rho_0 \otimes \cdots \otimes \rho_0 \qquad (n \text{ copies})$$

to give

$$\tau^{(n)} = \rho_1^{(n)} - \lambda \rho_0^{(n)}.$$

As can be checked, $\tau^{(n)}$ is generally not of the form $\tau \otimes \cdots \otimes \tau$ (n copies).

Deriving the optimal operator, that is, the positive eigenspace projector for $\tau^{(n)}$, for this decision problem may not be easy. Moreover, whether the approach is Bayes or Neyman–Pearson, and even if the density operators commute or an optimal operator can be written as a tensor *monomial*, it may not be obvious how to design a device to physically realize the optimal operator: the implementation problem persists for sample size greater than 1.

A final comment is in order about n-sample optimal quantum detectors. As we have seen the optimal detector for the M-ary decision problem, based on a sample of size n, requires using an operator on an n-fold tensor product Hilbert space, and this may explicitly *not* call for measuring individual outcomes for each of the n particles. That is, optimal processing for this problem could become unobtainable if information about individual outcomes were insisted upon, for whatever reason. Hence, in a sequential decision problem, statistical power, or Bayes risk, could be in principle seriously impaired if experimental inquiries were made about the outcomes of any of the n individual particles: *looking at any intermediate outcome may significantly degrade overall performance.* The modifications needed here for sequential quantum decision making are an open problem. Potentially significant, nonintuitive changes from procedures for classical optimal sequential detection and estimation may be required for quantum data. Again, this forms part of the currently active study of *quantum stochastic processes*; see the references in Section 5.

A26. The point just made and the earlier example on quantum coin tossing underscore the following: the process of quantum mechanical measurement is a procedure for estimating a parameterized density operator. In particular, various observables can be rated by their efficiencies when considered as detectors or estimators. This greatly clarifying view of *quantum measurement as statistical detection and estimation* was evidently first proposed by Helstrom (1976, p. 6). It places quantum-consistent statistical procedures and the statistician close to the center of both quantum measurement theory and practical experiment.

The value of the decision-theoretic approach to quantum measurement is also illustrated by the following facts. For Gaussian states, the uniform minimum variance unbiased estimator (quantum UMVUE) of the displacement is the outcome of a measurement of the position operator; see Holevo (1982, p. 261). But for non-Gaussian states, the unprocessed outcome of measuring the position operator can be shown to no longer be the optimal estimate. This statistical estimation approach thus suggests a systematic search for optimal displacement estimators for other classes of states.

A27. The alert reader may spot a possible fatal flaw in this proposal: in the larger space a joint distribution is obtained, but this would then appear to suggest that all marginal distributions would also exist. Thus, for position and momentum, say, there is no bivariate marginal for these measurements on the base space. Hence no joint distribution on any larger space could be presumed to exist either. However, in QM no contradiction arises. That is, while we have seen that two marginals can exist, and have no joint distribution, it is also true that a joint distribution can be found (on a larger Hilbert space) such that the univariate marginals reduce to position and momentum say, but no bivariate marginal is required to exist. In still other words, *marginals can exist without the existence of a joint distribution, and the converse in QM holds as well: a joint distribution can exist while a subset of marginals may have no joint distribution.* This converse has apparently not been noted before.

A good example of this converse can be constructed from the extension of the position and momentum

operators, P and Q, on Hilbert space $\mathcal{H}$, say, to operators $P^{\#}$ and $Q^{\#}$ acting on $\mathcal{H}^{\#}$ (containing a copy of $\mathcal{H}$); see Holevo (1982, p. 123) for full details. It is found that $P^{\#}$ and $Q^{\#}$ commute on $\mathcal{H}^{\#}$ and have a common set of eigenvectors. Thus, they also have a joint distribution. On the other hand the restriction of $P^{\#}$ and $Q^{\#}$ to $\mathcal{H}$ yields operators with identical measurement statistics as that of P and Q. There is no contradiction here because, very loosely speaking, $\mathcal{H}$ does not have enough vectors to supply a basis of common eigenvectors whereas $\mathcal{H}^{\#}$ does.

ACKNOWLEDGMENTS

We are grateful to Carl Helstrom, Jeffrey Shapiro, Kevin McCrimmon, Jack Yanovski, Linda Hirsch, Larry Wasserman, Mark Schervish, Carl Morris and Robert Kass for much encouragement, many helpful comments and numerous excellent questions that focused our thinking and writing. Stuart Personick provided very useful (unpublished) literature. We also want to gratefully acknowledge the considerable time and attention paid by the small, but grand army of not-named (because unobserved by us) reviewers of this manuscript who participated in its protracted development.

REFERENCES

ACCARDI, L. and KOROLIUK, D. (1992). Stopping times for quantum Markov chains. *J. Theoret. Probab.* **5** 521–535.

AKHIEZER, N. I. and GLAZMAN, I. M. (1981). *Theory of Linear Operators in Hilbert Space* (2 vol.). Pitman, Boston.

BALLENTINE, L. E. (1970). The statistical interpretation of quantum mechanics. *Rev. Modern Phys.* **42** 358–381.

BELAVKIN, V. P. (1991). Continuous non-demolition observation, quantum filtering, and optimal estimation. In *Quantum Aspects of Optical Communications* (C. Bendjaballah, O. Hirota and S. Reynaud, eds.) 151–163. Springer, Berlin.

BELTRAMETTI, E. G. and CASSINELLI, G. (1981). *The Logic of Quantum Mechanics*. Addison-Wesley, Reading, MA.

BENDJABALLAH, C., HIROTA, O. and REYNAUD, S., eds. (1991). *Quantum Aspects of Optical Communications*. Springer, Berlin.

BLANKENBECLER, R. and PARTOVI, M. H. (1988). Quantum density matrix and entropic uncertainty. In *Maximum Entropy and Bayesian Methods in Science and Engineering* **1** 235–244. Kluwer, Dordrecht.

BOHM, A. (1986). *Quantum Mechanics: Foundations and Applications*. Springer, New York.

BUTKOVSKIY, A. G. and SAMOILENKO, YU. I. (1990). *Control of Quantum-Mechanical Processes and Systems*. Kluwer, Dordrecht. [English translation of the Russian original published by Nauka Publishers, Moscow (1984).]

CHARBIT, M., BENDJABALLAH, C. and HELSTROM, C. W. (1989). Cutoff rate for the M-ary PSK modulation channel with optimal quantum detection. *IEEE Trans. Inform. Theory* **35** 1131–1133.

COHEN-TANNOUDJI, C., DIU, B. and LALOË, F. (1977). *Quantum Mechanics* (2 vol.). Wiley, New York.

DAVIES, E. (1978). Information and quantum measurement. *IEEE Trans. Inform. Theory* **24** 596–599.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

DIRAC, P. A. M. (1958). *Quantum Mechanics*, 4th ed. Oxford Univ. Press.

FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic, San Diego.

FEYNMAN, R. P. (1951). The concept of probability in quantum mechanics. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 533–541. Univ. California Press, Berkeley.

FEYNMAN, R. P., LEIGHTON, R. B., and SANDS, M. (1965). *The Feynman Lectures on Physics* **3**. Addison-Wesley, Reading, MA.

FIVEL, D. I. (1991). Geometry underlying no-hidden-variable theorems. *Phys. Rev. Lett.* **57** 285–289.

GABOR, D. (1950). Communication theory and physics. *Philosophical Magazine* **41** 1161–1197.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

GLEASON, A. M. (1957). Measures on the closed subspaces of a Hilbert space. *J. Math. Mech.* **6** 885–893

GOTTFRIED, K. (1966). *Quantum Mechanics*. Benjamin, New York.

GREENBERGER, D. M., HORNE, M. A., SHIMONY, A. and ZEILINGER, A. (1990). Bell's theorem without inequalities. *Amer. J. Phys.* **58** 1131–1134.

GUDDER, S. P. (1979). *Stochastic Methods in Quantum Mechanics*. North-Holland, Amsterdam.

GUDDER, S. P. (1988). *Quantum Probability*. Academic, San Diego.

HELSTROM, C. W. (1967). Minimum mean-squared error of estimates in quantum statistics. *Phys. Rev. Lett.* **25** 101–102.

HELSTROM, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic, New York.

HELSTROM, C. W., CHARBIT, M. and BENDJABALLAH, C. (1987). Cut-off rate performance for phase shift keying modulated coherent states. *Optics Communications* **64** 253–255.

HOLEVO, A. S. (1973). Statistical decision theory for quantum systems. *J. Multivariate Anal.* **3** 337–394.

HOLEVO, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland, Amsterdam.

HOLEVO, A. S. (1991). Statistical inference for quantum processes. In *Quantum Aspects of Optical Communications* (C. Bendjaballah, O. Hirota and S. Reynaud, eds.) 127–137. Springer, Berlin.

HOSKINS, R. F. (1979). *Generalised Functions*. Wiley, New York.

INGARDEN, R. S. (1976). Quantum information theory. *Rep. Math. Phys.* **10** 43–72.

JAUCH, J. M. (1968). *Foundations of Quantum Mechanics*. Addison-Wesley, Reading, MA.

JAYNES, E. T. (1989). Clearing up mysteries—the original goal. In *Maximum Entropy and Bayesian Methods* 1–27. Kluwer, Dordrecht.

LEVINE, R. D. (1989). The statistics of quantum mechanical wavefunctions. In *Maximum Entropy and Bayesian Methods* 117–122. Kluwer, Dordrecht.

LOUTON, T. (1975). A theorem on simultaneous observability. *Pacific J. Math.* **59** 147–159.

MALLEY, J. D. (1986). *Optimal Unbiased Estimation of Variance Components. Lecture Notes in Statist.* **39**. Springer, New York.

MALLEY, J. and HORNSTEIN, J. (1992). *Quantum Statistical Inference*. Preprint.

MARGENAU, H. (1963). Measurements in quantum mechanics. *Ann. Physics* **23** 469–485.

MERMIN, N. D. (1990a). *Boojums All the Way Through*. Cambridge Univ. Press.

MERMIN, N. D. (1990b). What's wrong with these elements of reality? *Physics Today* **43** 9–11.

MERMIN, N. D. (1990c). Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.* **65** 3373–3376.

MESSIAH, A. (1961). *Quantum Mechanics* (2 vol.). North-Holland, Amsterdam.

NAGAOKA, H. (1990). A new lower bound of Cramér–Rao type for quantum state estimation (abstract). In *IEEE Internat. Symp. Inform. Theory* January 14–19.

OHYA, M. (1983). On compound state and mutual information in quantum information theory. *IEEE Trans. Inform. Theory* **29** 770–774.

OZAWA, M. (1991). Quantum limits of measurements and uncertainty principle. In *Quantum Aspects of Optical Communications* (C. Bendjaballah, O. Hirota and S. Reynaud, eds.) 3–17. Springer, Berlin.

PERES, A. (1990). Incompatible results of quantum measurements. *Phys. Lett. A* **151** 107–108.

PERSONICK, S. D. (1970). Efficient analog communication over quantum channels. Technical Report, 477, Research Laboratory of Electronics, MIT.

PERSONICK, S. D. (1971). Application of quantum estimation theory to analog communication over quantum channels. *IEEE Trans. Inform. Theory* **17** 240–246.

POPPER, K. R. (1968). *The Logic of Scientific Discovery*. Harper & Row, New York.

POPPER, K. R. (1982). *Quantum Theory and the Schism in Physics*. Rowman and Littlefield, Totowa, NJ.

PRUGOVEČKI, E. (1981). *Quantum Mechanics in Hilbert Space*. Academic, New York.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

SCHIFF, L. I. (1968). *Quantum Mechanics*, 3rd ed. McGraw-Hill, New York.

SCHUMACHER, B. (1990). Information from quantum measurements. In *Complexity, Entropy, and the Physics of Information. SFI Studies in the Sciences of Complexity* (W. H. Zurek, ed.) 8 29–37. Addison-Wesley, Reading, MA.

SEELY, J. and ZYSKIND, G. (1971). Linear spaces and minimum variance unbiased estimation. *Ann. Math. Statist.* **42** 691–703.

SEKIGUCHI, M., HIROTA, O. and NAKAGAWA, M. (1990). Application of quantum minimax rule to general ternary quantum state signals. In *Proceedings of the 1990 IEEE International Symposium on Information Theory*, San Diego, CA [IEEE Catalog Number 90 CH 2711-0].

SHAPIRO, J. H. (1980). On the near-optimum binary coherent-state receiver. *IEEE Trans. Inform. Theory* **26** 490–491.

SHAPIRO, J. H. and SHEPARD, S. R. (1991). Quantum phase measurement: A system-theory perspective. *Phys. Rev. A* **43** 3795–3818.

SHAPIRO, J. H., SHEPARD, S. R. and WONG, N. C. (1989). Ultimate quantum limits on phase measurement. *Phys. Rev. Lett.* **62** 2377–2380.

SHAPIRO, J. H. and WAGNER, S. S. (1984). Phase and amplitude uncertainties in heterodyne detection. *IEEE Journal of Quantum Electronics* **20** 803–813.

SHILOV, G. E. (1968). *Generalized Functions and Partial Differential Equations*. Gordon and Breach, New York.

SHIMONY, A. (1989). Conceptual foundations of quantum mechanics. In *The New Physics* (P. Davies, ed.) Chap. 13, 373–395. Cambridge Univ. Press.

SWIFT, A. R. and WRIGHT, R. (1979). Generalized Stern–Gerlach experiments and the observability of arbitrary spin operators. *J. Math. Phys.* **21** 77–82.

VAN FRAASSEN, B. C. (1991). *Quantum Mechanics: An Empiricist View*. Clarendon, Oxford.

VARADARAJAN, V. S. (1962). Probability in physics and a theorem on simultaneous observability. *Comm. Pure Appl. Math.* **15** 189–217.

VON NEUMANN, J. (1932, 1955). *Mathematical Foundations of Quantum Mechanics*. Princeton Univ. Press.

VOURDAS, A. (1990). Information in quantum optical binary communications. *IEEE Trans. Inform. Theory* **36** 208–211.

YAMAMOTO, Y., MACHIDA, S., SAITO, S., IMOTO, N., YANAGAWA, T., KITIGAWA, M. and BJÖRK, G. (1990). In *Progress in Optics* **28** 87–179. North-Holland, Amsterdam.

YOSIDA, K. (1980). *Functional Analysis*. Springer, New York.

YUEN, H. P., KENNEDY, R. S. and LAX, M. (1975). Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Trans. Inform. Theory* **21** 125–134.

YUEN, H. P. and SHAPIRO, J. H. (1980). Optical communication with two-photon coherent states. III. Quantum measurements realizable with photoemissive detectors. *IEEE Trans. Inform. Theory* **26** 78–92.

ZAUDERER, E. (1983). *Partial Differential Equations of Applied Mathematics*. Wiley, New York.

ZYSKIND, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.