Int J Theor Phys (2008) 47: 2–14
DOI 10.1007/s10773-007-9551-7

An Operational Characterization for Optimal Observation of Potential Properties in Quantum Theory and Signal Analysis

Sven Aerts

Received: 17 October 2006 / Accepted: 30 August 2007 / Published online: 28 September 2007

Abstract Quantum logic introduced a paradigm shift in the axiomatization of quantum theory by looking directly at the structural relations between the closed subspaces of the Hilbert space of a system. The one dimensional closed subspaces correspond to testable properties of the system, forging an operational link between theory and experiment. Thus a property is called actual, if the corresponding test yields "yes" with certainty. We argue a truly operational definition should include a quantitative criterion that tells us *when* we ought to be satisfied that the test yields "yes" *with certainty*. This question becomes particularly pressing when we inquire how the usual definition can be extended to cover potential, rather than actual properties. We present a statistically operational candidate for such an extension and show that its representation automatically captures some essential Hilbert space structure. If it is the nature of observation that is responsible for the Hilbert space structure, then we should be able to give examples of theories with scope outside the domain of quantum theory, that employ its basic structure, and that describe the optimal extraction of information. We argue signal analysis is such an example.

Keywords Potential property · Quantum logic · Quantum theory · Signal analysis · Optimal observation

1 What Do We Observe?

Clearly observation is a process between the observer and the observed, but what is that we seek to observe? A starting point for an answer is provided for in the famous 1936 article by John von Neumann and Garret Birkhoff, The Logic of Quantum Mechanics [4]. They realized that much of the essence of quantum theory lies in the structural relations between the closed subspaces of the Hilbert space $\mathcal{H}$. This was a very important step because the one

This work was supported by the Flemish Fund for Scientific Research (FWO) project G.0362.03N.

S. Aerts (✉)
Center Leo Apostel for Interdisciplinary Studies (CLEA) and Foundations of the Exact Sciences (FUND), Department of Mathematics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
e-mail: saerts@vub.ac.be

dimensional closed subspaces of the Hilbert space $\mathcal{H}$ have an interpretation as operational procedures in the laboratory that test propositions with only two possible answers: "yes" or "no". Birkhoff and von Neumann called such propositions "operational propositions", nowadays, we simply call them "properties". In general properties hold only with a certain probability, but in case that probability is 1, the corresponding property is called actual. Since the state of the system can be defined by means of its actual properties, it seems that to observe a system, is to positively identify its actual properties. Of course, such an assumption can be criticized for being too idealistic. Nevertheless, it is a starting point, and we will soon generalize this notion of observation.

2 States and Properties

In a more recent formulation of quantum logic [1], we start with the set Q of yes–no questions that one can meaningfully ask about a system S. We assume that to each yes–no question, there corresponds a physical test (or experiment) α. Tests that deliver the answer "yes" with certainty then lead to the concept of an actual property.

Definition 1 (Actual property) If α yields a "yes" with certainty (if we would decide to perform the test) for a system S, then we say that there is a property a that is *actual* for that system S.

If two tests are equivalent, (that is, whenever one test results in a certain "yes", the other would certainly also have obtained "yes", if the other test results in a certain "yes", the first test necessarily would also have given "yes"), they naturally test the same property. More accurately, a property can be identified with an equivalence class of tests. We see that for an actual property the question of what optimal observation is becomes trivial: the result of test α for an actual property a, should always give "yes". In quantum mechanics, however, a typical proposition such as "the value of a given physical quantity A lies in the range Δ" will only be found to hold with a certain probability. Obviously, the corresponding test does not yield "yes" with certainty and the property is not actual, but potential. In absence of certainty, we will throughout this paper assume it yields "yes" with a fixed probability.[1] What should we do in this case? Let us naively pursue the analogy with the actual property and tentatively propose

Definition 2 (Potential property—tentative version) If test α yields "yes" with probability p for a system S, then property a is called *potential* with degree p for S.

To take this definition seriously, we have to answer what it operationally means to say a test yields "yes" with probability p. Clearly, the relative frequency should converge to p for many repetitions of the test. But how many tests do we need to perform? If the sequence of "yes" and "no" answers is truly stochastic, then isn't it possible, with some small but finite probability, to choose a starting point of the yes–no sequence of answers and cut it

[1] It is quite conceivable that some fluctuating phenomena in nature may fail to converge nicely to a fixed relative frequency. It is however more difficult to conceive how these should be treated at all! Luckily a very large and interesting class of phenomena does enjoy the property that the relative frequency converges. If not, at least our proposal can serve as a means to operationally falsify the label "potential property" to such a phenomena.

off at a certain point to get pretty close to *any* p in the unit interval? Surely such cutting procedure is unworthy of scientific standards, but the point remains that we want an operational definition that excludes this possibility. We want to know *when* we will be satisfied. Is this after 10 measurements, or after 10^9 measurements? Clearly for some phenomena a few measurements are sufficient, whereas other phenomena require delicate and numerous measurements. So the *number* of measurements *in itself* doesn't really tell us what we need, what we need is a measure of uncertainty. This question was cleverly sidestepped in the definition of an actual property because in that case we simply demand that it yields "yes" *with certainty*. But then again, what is certainty? How many measurements reveal certainty? We will attempt to take this question seriously and see that, somewhat miraculously, Hilbert space theories already account for this problem in a really beautiful way. To do so, we will first need a short excursion into the mother of all stochastic processes: the Bernoulli trial.

3 Bernoulli Processes

A Bernoulli trial is an experiment with a random outcome taken from only two possible outcomes: "Success" and "Failure", "1" and "0", "yes" or "no", ... A Bernoulli process consists of a sequence of independent identically distributed Bernoulli trials. Here is a definition: consider a finite or infinite sequence of independent random variables $X_1, X_2, X_3, \ldots$, such that:

1. For each i, the value of X_i is either 0 or 1.
2. For all values of i, the probability that $X_i = 1$ is the same number p.

By sheer combinatorial counting we obtain that, if the experiment is repeated N times, the probability of precisely k successes equals

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

To obtain average and standard deviation, we calculate the moment generating function:

$$\begin{aligned} G(t) &= \sum_{k=0}^{N} \exp(tk) \binom{N}{k} p^k (1-p)^{N-k} \\ &= [p \exp(t) + 1 - p]^N. \end{aligned}$$

It is then easy to show that the average value of successes is

$$\begin{aligned} \mu &= E[X] = G'(0) \\ &= Np. \end{aligned}$$

The variance, or square of the standard deviation, for the number of successes is

$$\begin{aligned} \sigma^2[X] &= E[X^2] - E[X]^2 \\ &= G''(0) - G'(0)^2 = Np(1-p). \end{aligned}$$

With X being the number of successes in N trials, we define the *fraction of successes* as a *new* random variable $\bar{X} = X/N$.

It's expectation value is

$$E[\bar{X}] = E[X/N] = \frac{Np}{N} = p$$

and the variance of the fraction of successes is given by

$$\begin{aligned}\sigma^2[\bar{X}] &= \sigma^2[X/N] \\ &= \frac{1}{N^2} Np(1-p) \\ &= \frac{p(1-p)}{N}.\end{aligned}$$

We have included these well-known facts in our presentation to show just how little we need to get these results, and because they can help in solving our main problem through the next, truly classic result: the Chebyshev inequality.

Theorem *For* any *discrete random variable X with mean* μ *and standard deviation* σ, *and for any positive number c, we have*

$$P(|X - \mu| > c\sigma) < \frac{1}{c^2}.$$

In words: the probability that a random variable will differ more than c standard deviations from its mean value, is never greater than the square of the reciprocal of c. The true strength of the Chebyshev inequality is not that it provides a tight bound (it usually does not), but rather that it is extremely general, in that we only require of the random variable that its mean and standard deviation exist. So in any case, knowing the standard deviation, gives us at least a clue how far we can expect the relative frequency to lie from the probability p. An additional problem here, is that the Chebyshev inequality is a limiting theorem for large N. For a finite ensemble our estimate for the standard deviation *depends* on our estimate for the probability,[2] so we are still not done.

4 Maximum Predictive Power

There exists in the literature a very nice argument due to Summhammer that shows how probability amplitudes arise from a purely statistical argument. This is most useful to us, because the argument simultaneously solves our problem. We will present it briefly in a form and notation that is close to [13]. Summhammer starts with the observation that, in quantum mechanics, the magnitude of physical observable quantities is derived from probabilities. Estimates for these probabilities are experimentally obtained through the relative frequency of detector clicks. Assume then that we have a dichotomic experiment, where the result "1" was obtained n_1 times, and the result "0" was obtained $n_2 = N - n_1$ times. The relative

[2]The author learned from Andrei Khrennikov (private communication) that a similar problem was one of the main objections to von Mises' original frequentist approach. Perhaps this analysis has some bearing on that problem too.

frequency of outcome 1 is then n_1/N, and this is also our best estimate for the probability and associated uncertainty interval

$$p_1 = \frac{n_1}{N}, \qquad \Delta p_1 = \sqrt{\frac{p_1(1-p_1)}{N}},$$

and likewise for p_2. From this estimate we derive the magnitude of a quantity χ by means of an unspecified relation

$$\chi = \chi(p_1).$$

Its associated uncertainty interval is then

$$\Delta\chi = \left|\frac{\partial\chi}{\partial p_1}\right|\Delta p_1 = \left|\frac{\partial\chi}{\partial p_1}\right|\sqrt{\frac{p_1(1-p_1)}{N}}.$$

We want to concentrate on the observation of those random variables χ for which additional data always decreases the uncertainty. To do so, we require that

$$\Delta\chi(N+1) < \Delta\chi(N). \tag{1}$$

Many functions will satisfy this requirement, and to further constrain the set of possible $\chi's$, Summhammer introduces the notion of *maximum predictive power*:

Definition 3 (Maximum predictive power) A random variable χ allows for maximum predictive power, if the uncertainty $\Delta\chi$ depends only on the number of trials N, and not on the value of n_1.

In this way we can know *in advance* how many times the experiment needs to be repeated to reach a certain prescribed uncertainty level. This translates mathematically as

$$\sqrt{N}\Delta\chi = \left|\frac{\partial\chi}{\partial p_1}\right|\sqrt{p_1(1-p_1)} = C.$$

The solution of which is given by

$$\chi = C\arcsin(2p_1 - 1) + D$$

and its inverse is

$$p = \frac{1+\sin(\frac{\chi-D}{C})}{2} = \cos^2(\pi/4 - (\chi - D)/2C).$$

As remarked in [13], any smooth function $\alpha(\chi)$ of our random variable χ for which equal intervals along χ correspond to equal intervals along α, will also be a solution. To give one particular solution, choose $C = 1$, and D equal to $-\pi/2$, to recover

$$p = \cos^2(\chi/2). \tag{2}$$

Let us come back to our initial problem: how many tests do we need to perform for it to be meaningful to say a test yields "yes" with probability p? To put it more generally, if we infer χ from p, when do we know χ sufficiently precisely? The answer we give here is that,

for any confidence interval $\Delta\chi$ *you* require for your estimate of χ, we can tell *in advance* how many measurements you need, *regardless* of the state (or p). And this can be done, if we insist that the relation between the random variable and the probability is of the nature of (2). The only really essential requirement to obtain this result, is that the probability is the same fixed number in [0, 1] for each separate trial. To summarize, we can state a more sensible definition of a potential property as follows:

Definition 4 (Potential property) If, for *any* prescribed uncertainty interval Δp, one can say *in advance* how many measurements are necessary such that test α yields "yes" with probability p *within* an uncertainty interval Δp, then property a is called potential with degree p and uncertainty Δp for S.

What makes this definition operational is that for any desired uncertainty level the theorist requires for the acceptance or refutation of a conjectured potential property, the experimenter knows in advance how many measurements he needs to come to a conclusion. The halting problem for the experimenter is solved in a pragmatic way. Absolute certainty still follows only after an infinite number of measurements, but for any allowable uncertainty in the estimate of the degree of potentiality, we know how often we have to question nature for it to be a reliable judge on our trial. We have explained before that the same problem exists for an actual property: how many times do we have to repeat the measurement to conclude that the property actually holds? The answer again depends on how certain one wants to be. If we set a certain threshold on Δp, we can estimate the uncertainty by means of the variance. However, there is one additional complication here. If we have obtained 9 times the answer "yes", our best estimate for the probability is the relative frequency of "yes" answers, which is also 1. This implies the experimental variance is zero, which does not really tell us anything useful. However we can remedy this problem by artificially adding a single outcome with the opposite answer. This implies an estimate for p as 0.9 and the corresponding estimate for the variance (0.09) provides a lower bound for the estimate of Δp.

What is surprising here, is that the representation of a property in Hilbert space quantum mechanics as a closed one dimensional subspace automatically fulfills this requirement, as the probability is always of the form of (2). On the other hand it means that for the statistical estimation of observable quantities in quantum theory (at least without additional classical errors), we always have that more data decreases one's uncertainty, as required in (1) and that for any desired uncertainty level, we can say in advance how many measurements are required. In this sense quantum measurements for vector states are optimal. Interestingly, there was another argument in the literature which started from quite different considerations, but also mathematically hinges essentially on Bernoulli trials, that hinted at a similar result.

5 Optimally Distinguishing Probability Distributions

Suppose, as before, we have an experiment with two possible outcomes, say 0 and 1. Depending on the state of the system, the experiment can be characterized by the probability p of the occurrence of outcome 1. Suppose then that we have prepared four ensembles, E_1, E_2, E_3 and E_4, of systems such that the first ensemble E_1, is characterized by a value $p_1 = 1$, E_2 with $p_2 = 0.9$, E_3 with $p_3 = 0.55$ and E_4 with $p_4 = 0.45$. By making measurements on members of the ensembles, we are to estimate which of the four values of p pertain to that ensemble. Clearly, it is easier to distinguish (in a finite set of trials for the experiment) E_1 from E_2, than it is to distinguish E_3 from E_4, even though $\triangle p$ is 0.1 in both

cases. This idea can be made qualitative, as Wootters has done [17, 18]. Suppose that we are to distinguish two ensembles, one characterized by p, the other by $p + \triangle p$. For N trials of an experiment with two possible outcomes, the sequence of outcomes constitute once more a Bernoulli sequence, with associated standard deviations

$$\sigma_1 = \sqrt{\frac{p(1-p)}{N}}, \tag{3}$$
$$\sigma_2 = \sqrt{\frac{(p+\triangle p)(1-p-\triangle p)}{N}}.$$

We will say that two such states are *distinguishable* in N trials of the experiment, iff the difference $\triangle p$ is greater than the sum of these two standard deviations

$$\triangle p > \sigma_1 + \sigma_2. \tag{4}$$

Define the statistical distance between the two states as the reciprocal of the square root of the number (N_0) of measurements necessary to make $\triangle p$ equal to the sum of these two standard deviations:

$$d(p, p+\triangle p) = \frac{1}{\sqrt{N_0}} : \triangle p = \sigma_1 + \sigma_2. \tag{5}$$

To the first order in $\triangle p$, using 3, this is equal to

$$d(p, p+\triangle p) = \frac{1}{2}\frac{|\triangle p|}{\sqrt{p(1-p)}}.$$

For two arbitrary biases, $\triangle p$ may be large, and we will denote the respective probabilities as p_1 and p_2. The definition of statistical distance, as given by R.A. Fisher who used this definition to study genetic drift, is then

$$d(p_1, p_2) = \lim_{N\to\infty} \frac{n}{\sqrt{N}},$$

where n is the maximum number of intermediate probabilities that can still be distinguished in N trials according to the former definition (4). Wootters shows this to be

$$d(p_1, p_2) = \frac{1}{2}\int_{p_1}^{p_2} \frac{dp}{\sqrt{p(1-p)}} = \arccos(\sqrt{p_1 p_2} + \sqrt{(1-p_1)(1-p_2)}).$$

The result extends to the case for an experiment with k mutually exclusive outcomes $x_1, x_2, \ldots, x_k$, with respective probabilities $p_1(x_i)$ and $p_2(x_i)$, where it reads

$$d(p_1, p_2) = \arccos\left(\sum_{i=1}^{k} \sqrt{p_1(x_i)\, p_2(x_i)}\right). \tag{6}$$

The probabilities can be summarized as vectors π_1 and π_2 that lie in the standard $(k-1)$-simplex in Euclidean k-space:

$$\pi_i = (p_i(x_1), p_i(x_2), \ldots, p_i(x_k)), \quad i = 1, 2. \tag{7}$$

If we map (7) to the unit sphere,

$$\psi_i = (\sqrt{p_i(x_1)}, \sqrt{p_i(x_2)}, \ldots, \sqrt{p_i(x_k)}), \quad i = 1, 2, \tag{8}$$

we can rewrite the result of (6) as

$$\begin{aligned} d(p_1, p_2) &= \arccos(\langle \psi_1, \psi_2 \rangle) \\ &= \measuredangle(\psi_1, \psi_2). \end{aligned} \tag{9}$$

It is well-known that any unitarily invariant Riemannian metric on the unit sphere is a trivial function of $d(p_1, p_2)$, so (8) leads to a representation of the probabilities in such a way that the only metric between the two unit vectors, is proportional to the statistical distance. At no point have we invoked quantum features; yet we obtain a structure that—at least partially—resonates with the formalism of quantum theory.[3] If this line of reasoning is not flawed, examples of quantum-like theories should exist outside the domain of quantum physics.

6 Signal Analysis in Operator Form

Another famous theory where Hilbert space provides for the basic setting, is signal analysis. Its similarities with quantum theory come out best if we reformulate it, as is very popular in modern accounts of the subject [6, 11], in an operator framework. We will introduce the main concepts of such a formulation very briefly. Let us work in $\mathcal{L}^2(\mathbb{R}) = \mathcal{H}(\mathbb{C})$ and start with a signal with bounded energy:

$$\begin{aligned} x &\in \mathcal{H}(\mathbb{C}), \\ \textit{Energy} &= \int_{-\infty}^{+\infty} |x(t)|^2 dt < +\infty. \end{aligned}$$

So signal-analysis' natural setting is that of $\mathcal{L}^2(\mathbb{R})$, the square integrable functions over the complex numbers, equipped with the inner product

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f^*(t) g(t) d\mu(t).$$

An observable quantity is described by a linear self-adjoint (Hermitian) operator $\hat{A}$, with expectation value

$$\langle \hat{A} \rangle = \langle x, \hat{A}x \rangle.$$

Likewise, the expectation value of a function g of an operator is given by

$$\langle g(\hat{A}) \rangle = \langle x, g(\hat{A})x \rangle.$$

[3] Wootters did extend his argument to the quantum domain. This is necessary and non-trivial because the inferential power of the experimenter in the quantum domain for the distinguishing of two states, depends on whatever observable he chooses to measure. Therefore Wootters defines the statistical distance between two preparations (states) as the *larges*t distance when the preparations are analyzed by the *most discriminating* apparatus. Again, he recovers (9). Because the Hilbert space angle obtains only for the most discriminating apparatus, this is one more indication that Hilbert space representations in some sense already incorporate maximal statistical performance of observation.

There are several motivations for a reformulation of signal analysis in terms of operators. Firstly, many signal analytic operations like filtering, delay, scaling, Doppler-shift, etc., can be regarded as linear (often unitary) operators that act on the signal. As an example, we give the Doppler-shift in radar applications, which can be well approximated as

$$x(t) \to x_{(\tau,\delta)}(t) = x(t-\tau)e^{2\pi i\delta t}.$$

A second motivation comes from the uniformization of various approaches in a single theoretical framework that the operator formalism provides. A third motivation is that different representations of a signal (time, frequency, scale, ...) can readily be obtained. A typical situation occurs when we have a signal $x(t)$ in the time domain and seek its frequency representation. In an operator framework we define two (self-adjoint) operators $\hat{T}$ and $\hat{F}$ by their action on a signal as follows:

$$\begin{aligned}(\hat{T}x)(t) &\equiv t.x(t),\\ (\hat{F}x)(t) &\equiv \frac{-i}{2\pi}\frac{dx(t)}{dt}.\end{aligned}$$

The signal can be expanded in a basis of eigenvectors $y(t)$ of $\hat{F}$:

$$x(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} X(\nu)\exp(2\pi i\nu t)dt.$$

The coefficients $X(\nu)$ are obtained through inversion of this integral transform:

$$\begin{aligned}X(\nu) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp(-2\pi i\nu t)x(t)dt\\ &= [\mathcal{F}(x)](\nu).\end{aligned}$$

Exactly as is the case for quantum theory, the operational signal-analytic quantities are given as the expectation values of the operators

$$\begin{aligned}\langle\hat{T}\rangle &= \langle x, \hat{T}x\rangle\\ &= \int_{-\infty}^{+\infty} t|x(t)|^2dt,\\ \langle\hat{F}\rangle &= \langle x, \hat{F}x\rangle\\ &= \int_{-\infty}^{+\infty} x^*(t)\left(\frac{-i}{2\pi}\frac{dx}{dt}(t)\right)dt\\ &= \int_{-\infty}^{+\infty}\nu|X(\nu)|^2dt.\end{aligned}$$

As fourth motivation, we call attention to the fact that, in some cases, the operator framework offers strong calculational advantages in comparison with the standard formulation. For example, define the average of a general function g of frequency (or time, or scale) as

$$\langle g(\hat{F})\rangle = \int_{-\infty}^{+\infty} x^*(t)g\left(\frac{-i}{2\pi}\frac{d}{dt}\right)x(t)dt.$$

Using Parseval's relation, this can be also rewritten as

$$\langle g(\hat{F})\rangle = \int_{-\infty}^{+\infty} g(\nu)|X(\nu)|^2 dt.$$

So we have two ways of obtaining the average of a function g of frequency in two distinct ways: one employing the Fourier transform, the other directly on the time signal through the operator formalism. That the first formula often represents a much easier way to calculate the average is advocated with examples in, for example, [5].

Note also that, because the operators $\hat{T}$ and $\hat{F}$ do not commute ($[\hat{T}, \hat{F}] = \frac{i}{2\pi}\hat{I}$), an uncertainty relation (reminiscent of the famous Heisenberg inequality) holds:

$$\Delta\hat{T} \cdot \Delta\hat{F} \geq \frac{1}{4\pi}.$$

7 Joint Observable Representations

Let $\psi \in \mathcal{L}^2(\mathbb{R})$ and define $[\hat{Q}](q) = q$ and $[\hat{P}](q) = -i\hbar\frac{d}{dq}\psi(q)$. As is well-known from quantum theory, the expectation value of the position operator $\hat{Q}$ is given by:

$$\langle\hat{Q}\rangle = \int_{-\infty}^{+\infty} q|\psi(q)|^2 dq$$

and likewise for the momentum operator $\hat{P}$. Recall that for a probability density function $f(x)$, the kth moment is defined as

$$\mu_k = \int_{-\infty}^{+\infty} x^k f(x) dx.$$

Interpreting the $|q(t)|^2$ and $|X(\nu)|^2$ as probability densities (positive and finite), this coincides with the usual definition of averages. It is then natural to assume that $|\psi(q)|^2$, $|\psi(p)|^2$ (with an abuse of notation that is customary in physics, the second) are densities in the position and momentum space respectively. It is now reasonable to ask whether there exists a *joint-representation* $\rho(q, p)$ such that the correct marginals are obtained and the distribution is normalized:

$$\int_{-\infty}^{+\infty} \rho(q, p)dx = |\psi(p)|^2, \qquad \int_{-\infty}^{+\infty} \rho(q, p)dp = |\psi(q)|^2,$$
$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho(q, p)dqdp = 1.$$

In fact, Wigner needed such a joint distribution to calculate quantum corrections to the second virial coefficient [16] for quantum gases. He obtained:

$$\rho(q, p) = \frac{1}{\pi\hbar}\int_{-\infty}^{+\infty} \psi^*(q + y)\psi(q - y)\exp\left(\frac{2ipy}{\hbar}\right)dy. \tag{10}$$

This is however not a positive density unless ψ is Gaussian [12], and was hence called a quasi-density function or quasi-distribution. In 1948 Ville [15] proposed to calculate a joint

time-frequency representation of a signal through the formula:

$$WV(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}x^*\left(t+\frac{\tau}{2}\right)x\left(t-\frac{\tau}{2}\right)\exp(i\omega\tau)d\tau. \tag{11}$$

The former function is now commonly called the Wigner–Ville (quasi) distribution. It is obvious why this should be the case, for the joint time-frequency representation of (11)—disregarding the ontological status of the quantities—is only a trivial transformation away from the Wigner distribution (10).

8 Similarities and Differences

Let us briefly sum up the main similarities we have obtained between quantum theory and signal analysis in the following table.

Quantum theory	Signal analysis
State $\psi(q,t)\in\mathcal{L}^2(\mathbb{R})$	Signal $x(t)\in\mathcal{H}(\mathbb{C},\mathbb{R})$
$\lvert\psi\rvert^2$	$\lvert x\rvert^2$
$\psi(p)=\mathcal{F}[\psi]$	$X(\nu)=\mathcal{F}[x]$
$[\hat{Q},\hat{P}]=i\hbar\hat{I}$	$[\hat{T},\hat{F}]=\frac{i}{2\pi}\hat{I}$
$\Delta\hat{P}\cdot\Delta\hat{Q}\geq\frac{h}{4\pi}$	$\Delta\hat{T}\cdot\Delta\hat{F}\geq\frac{1}{4\pi}$
$\rho(q,p)$	$WV(t,\omega)$

We see that indeed most quantum theoretic quantities have a counterpart in signal analysis. More details on the profound relationship between these two subjects can be found in [2, 5–7, 12]. *Yet we have never heard of the measurement problem in signal analysis, or of non-locality. Why is it that no counterpart of these notorious philosophical issues exists in signal analysis when it does in quantum theory?* We believe the main difference lies in the ontological status of the most basic concept in signal analysis (the signal) and in quantum theory (the state of the system). In signal analysis, the signal that we have actually measured consists of the data that we have actually measured and so it is the end point in the physical process of measurement, and the starting point for the data analysis. After one has obtained the signal, one can perform any transformation one wants, without destroying (or even slightly altering) the signal. In particular, signals can be copied freely. In quantum theory, on the other hand, the state of the system is the subject of the measurement; it is that what we want to measure. In general it is unknown and cannot be copied with certainty. Different representations of the state can readily be obtained theoretically through the Dirac transformation theory, but to *actually obtain* measurement results in one representation or another, we have to use an entirely different measurement setup. The measurement changes the state of the system, making it impossible to know the outcome another (possibly non-compatible) measurement would have given. In short, a signal has the status of classical information, whereas the quantum state has the status of quantum information. Of course, the result of measuring an observable for a given quantum system is an outcome, and indeed has again the status of classical information. Alas, the quantum information is no longer of the same nature once this outcome has been obtained. Because of this sharp ontological distinction between a state and a signal, they can exhibit very explicit differences. For example, according to quantum theory, a compound system is to be described in the tensor

product space of the state spaces of its constituents, giving rise to the spectacular probabilistic effects of entanglement, such as the EPR correlations for the singlet state. This feature is entirely absent in signal analysis. In spite of these differences, the striking similarities have often led to a cross-fertilization between the two fields (for examples, see [2, 6]). One is left to wonder what rôle the extremely well-developed formalism of quantum theory can have in the further development of other theories of observation in which the basic setting is formed by the complex Hilbert space, such as the recently founded theories of shapes [10] and information retrieval [14].

9 Concluding Remarks

We started with an inquiry into the probabilistic implications of taking the concept of potential property seriously in an operational sense. We proposed that a potential property can only be inferred to be of degree p for any desired uncertainty level Δp, if one can tell *in advance* (that is without knowing p but with knowledge of the desired uncertainty level Δp) how many measurements are required to reach such a conclusion. This is indeed an operational definition, because for any uncertainty level desired, we know when our process of inquiry will end. Drawing from the work of Summhammer, we showed that this leads to a natural representation of the potential property as an element of the unit sphere. Similarly Wootters has obtained the Hilbert space distance from considerations of optimally distinguishing between probability distributions. If it is not nature, but the requirement of optimal inference that dictates the basic setting should be a Hilbert space, then there should exist other examples of theories of observation outside the quantum domain with similar features. Signal analysis provides for such an example, and we believe it does so because it is concerned with a formalization of observation in the sense of optimal statistical inference. The idea of observation as optimal inference goes back at least to Helmholtz [9], but has many contemporary advocates (see [3, 8] and the references found there). We believe that the fact that both fields employ the same formalism, in spite of the fact that they deal with different parts of reality, makes that point of view even more plausible, for what unites them on a conceptual level, is that both are ultimately concerned with the optimal extraction of information from nature.

References

1. Aerts, D.: The One and the Many: towards a unification of the quantum and the classical description of one and many physical entities. Ph.D. thesis, Vrije Universiteit Brussel (VUB) (1981)
2. Aerts, S., Aerts, D., Schroeck, F.E.: The necessity of combining mutually incompatible perspectives in the construction of a global view: quantum probability and signal analysis. In: Aerts, D., D'Hooghe, B., Note, N. (eds.) Worldviews, Science and Us: Redemarcating Knowledge and Its Social and Ethical Implications. World Scientific, Singapore (2005)
3. Aerts, S.: Quantum and classical probability as Bayes-optimal observation. quant-ph/0601138 (2006)
4. Birkhoff, G., von Neumann, J.: The logic of quantum mechanics. Ann. Math. **37**, 823–843 (1936)
5. Cohen, L.: Time–Frequency Analysis. Prentice Hall, New York (1995)
6. Flandrin, P.: Time–Frequency/Time-Scale Analysis. Wavelet Analysis and Its Applications, vol. 10. Academic Press, London (1999)
7. Gabor, D.: Theory of communications. J. Inst. Elect. Eng. **93**, 429–457 (1943)
8. Geisler, W.S., Kersten, D.: Illusion, perception and Bayes. Curr. Opin. Neurobiol. **13**, 150–158 (2003)
9. Helmholtz, H.: Handbuch der Physiologischen Optik. Voss, Leipzig (1867)
10. Kendall, D.G., Barden, D., Carne, T.K., Le, H.: Shape and Shape Theory. Wiley Series in Probability and Statistics. Wiley, New York (1999)

11. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press, London (1998)
12. Schroeck, F.E.: Quantum Mechanics on Phase Space. Kluwer Academic, Dordrecht (1996)
13. Summhammer, J.: Maximum predictive power and the superposition principle. Int. J. Theor. Phys. **33**, 171–178 (1994)
14. van Rijsbergen, K.: The Geometry of Information Retrieval. Cambridge University Press, Cambridge (2004)
15. Ville, J.: Theorié et applications de la notion de signal analytique. Cables Transm. **2A**(1), 61–74 (1948)
16. Wigner, E.P.: On the quantum mechanical correction for thermodynamic equilibrium. Phys. Rev. **40**, 749–759 (1932)
17. Wootters, W.K.: The acquisition of information from quantum measurements. Ph.D. dissertation, University of Texas at Austin (1980)
18. Wootters, W.K.: Statistical distance and Hilbert space. Phys. Rev. D **23**, 357–362 (1981)