Quantum Mechanics in View of Information *

Dénes Petz **

The inception of quantum theory was forced by observation of subtle discrepancies between some experimantal facts and laws of classical physics. Today, without quantum physics we could not explain the properties of superfluids, the action of lasers or the color of stars.

At the Royal Institution of Great Britain, London, on April 27, 1900, Lord Kelvin presented a talk entitled "*Nineteenth Century Clouds over the Dynamical Theory of Heat and Light*", in which he addressed himself to some difficulties of the otherwise well-established theory [MR]. Atoms of a heated material emit light (electromagnetic waves) with very specific discrete frequencies which are characteristic of the atom. In October, 1900, as a first step towards quantum theory, the German physicist, Max Planck proposed the idea that electromagnetic oscillations occur only in "quanta". Describing the energy distribution in the blackbody radiation, he made use of the fundamental constant h. It is a very tiny quantity by everyday standard and nowadays h is known as Planck's constant. The same constant played a crucial role in the light-quantum hypothesis of Albert Einstein which explained several peculiar phenomena connected with the emission and absorption of light. He was awarded by the Nobel Prize for this discovery. His succesful application of the quantum concept was followed by others. The rule of Niels Bohr required that the angular momentum of an electron orbitting about the nucleus can occur in integer multiples of $h/2\pi$. Quantum theory turned into one of the most revolutionary fields in physics in the early twentieth century.

In the present form, quantum mechanics arose of two independent schemes initiated by Werner Heisenberg and Erwin Schrödinger in the "turbulent" years 1925–1927. Heisenberg's "matrix mechanics" and Schrödinger's "wave mechanics" seemed rather different at the first sight but the equivalence of the two schemes was soon established as a result of the work of Paul Dirac, Pascual Jordan and several other excellent physicists. Subsequently, the Hungarian/American mathematician, John von Neumann developed an axiomatic approach based on linear operators of abstract Hilbert spaces.

Quantum mechanics is a statistical theory in great deal. As Max Born pointed out in 1926: "*The motion of particles conforms the law of probability but the probability itself is propagated in accordance with causality.*" Saying that the probability of a certain outcome of an experiment is p, we mean that if the experiment is repeated many times, the fraction of those which give this outcome is roughly p. For example, assume that a beam of electrons is fired to a screen for observation. The sentence "The probability that an electron hits a part of the screen is 10%" is understood in a statistical way: 10% of all electrons hitting the screen are detected on the chosen area. In order to show the probabilistic aspects of quantum theory more deeply, an idealized experiment is descibed which helps to understand the "interference of probabilities".

---

* The Japanese translation is to be published in the journal "KAGAKU".

** Research Institute for Mathematical Sciences, Kyoto University, on leave from the Technical University of Budapest.

A monochromatic light source S emits light (that is, photons) in all directions. Many of the photons pass through the two slits of the screen A and they are received on the parallel screen B, which can be a photographic plate. What is observed when one slit is open? The light bulb emits a huge number of photons, a portion of which passes through the slit and yield a seemingly rather uniform illumination of screen B. To have something more interesting, the light intensity should be reduced and the slits should be small in width, 0.001 mm would do. The smooth distribution of detected light intensity was a statistical effect, at smaller intensity of light emission the illumination on the screen is made up of individual spots (in accordance with our particle picture of photons)[Pe]. When both slits are open and they are cca. 0.15 mm apart, the illumination appears to be wavy, an iterference pattern consisting of bands is observed. The pattern of illumination seems to be completely different from what it was with a single open slit. It is not at all true that the illumination doubles when the second slit is opened. At the bright places the intensity can be three-four times what it was before and on the other hand, the intensity may go down drastically at other places (see Figure 1). The photons behave like waves and not like particles. Reinforcing and destructive interference are features of ordinary wave propagation.

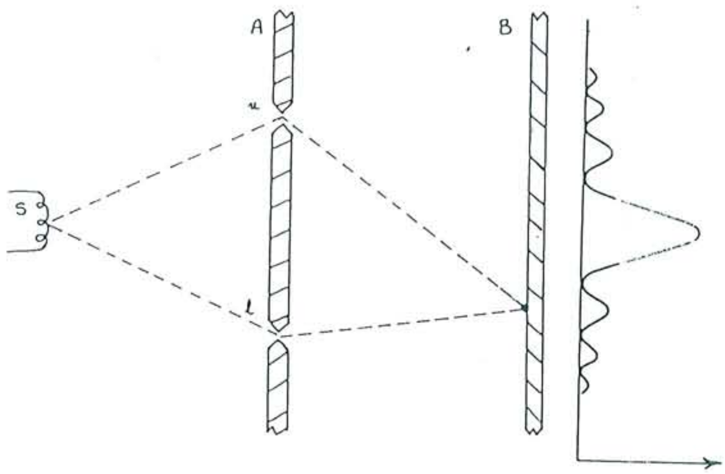

Figure 1. The two-slit experiment: Wavy pattern of intensity is observed on screen B when both slits are open.

Let H be a small part of screen B, and let r_u, r_l, r be the relative frequency of photons hitting H when the upper, the lower and both slits are open. Let us try to apply probability theory naively and let E_u be the event that a photon issued by S goes through the upper slit and it is detected afterwards on the chosen small area H. Define the event E_l symmetrically: the photon passes through the lower slit and is received on H. Finally, let E be the event that the photon hits H when both slits are open. In this case we do not care which slit it goes through. Let us interpret the relative frequencies r_u, r_l and r as probabilities of the corresponding events. Since any photon passes through only one of the two slits, the events E_l and E_u should be considered as exclusive alternatives for a photon hitting H. Hence $r_l + r_u = r$ should hold, which, however, is in contradiction with the experimental result. The way out of this contradiction is to reject the very classical particle picture that the photon goes through one of the slits only and to use Hilbert space probability theory. It is not suggested here that a single photon goes through both slits. In quantum mechanics, to ask "Did the photon pass through the upper slit?" is just not a right question and in the logical structure of quantum mechanics we are not permitted to assign a probability to the "histories" E_u and E_l [Om].

By the interference of the probabilities r_l and r_u, we mean that both $r_l + r_u < r$ and $r_l + r_u > r$ may happen (depending on the position of the area H). In quantum mechanics probabilities are computed as the squared modulus of a state vector which is typically in an infinite dimensional complex Hilbert space. Nevertheless, the 2 dimensional Euclidean space (or rather the plain of complex numbers) is suitable for explanation. When w, z and $w + z$ are vectors, the familiar cosine rule tells us that

$$|w + z|^2 = |w|^2 + |z|^2 + 2|w|\,|z| \cos\theta \,,$$

where θ is the angle between w and z (see Figure 2).

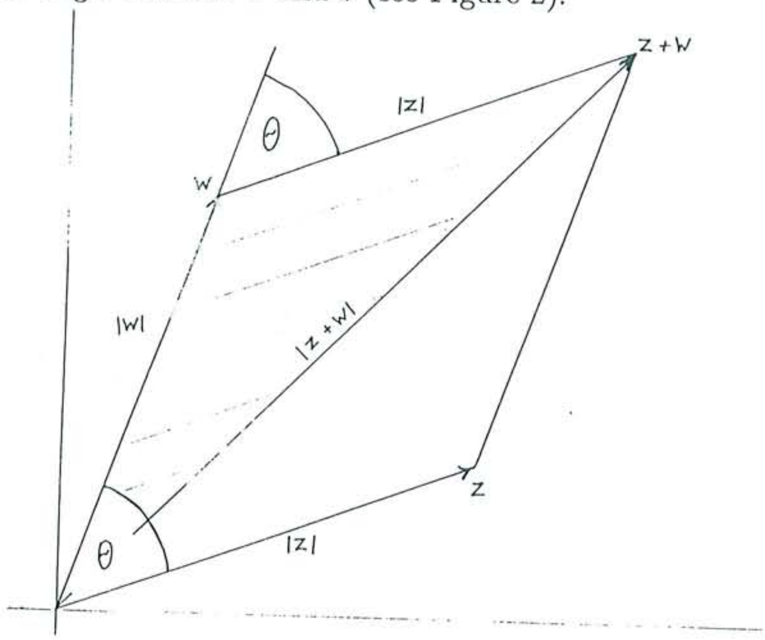

Figure 2. The cosine rule is responsible for the quantum interference of probabilities. The interference depends on the sign of the correction term $2|w|\,|z| \cos\theta$.

If the probabilities r_l, r_u and r are thought of as $|w|^2$, $|z|^2$ and $|w + z|^2$, then we have

$$r = r_l + r_u + 2\sqrt{r_l r_u} \cos\theta \,.$$

Here is the interference of probabilities. If $w = -z$ or correspondingly $\cos\theta = -1$, then $r = 0$, and this is the case of destructive interference. At the brightest point in the screen of the two-slit experiment (both slits are open), we have $w = z$, that is $\cos\theta = 1$ and $r = 4r_u$. This is in accordance with the observation that the intensity of the illumination can be four times the intensity observed when just the upper slit is open.

State vectors have two representations in the Hilbert space of a quantum particle. One is associated with position and the other with monentum, for the sake of simplicity these representations may be viewed as two different coordinate systems.

The quantum mechanical particle is regarded as being spread out spatially, rather than always concentrated at a definite point. When the state vector ψ is viewed in the position representation, then the probability density $p_Q = |\psi|^2$ desrcibes the position distribution in the three dimensional space. (Actually, the state vector is a function with complex values and it is often called state function, wave function or wavepacket.) For graphical convenience, we constrain the particle to one degree of freedom, so it moves along the x-axis and its position distribution $p_Q(x)$ is a function of a single variable (see Figure 3). If $p_Q(x)$ is sharply peaked, the particle is well-localized, it stays with high probability in a small interval. The sharpness of a probability distribution is expressed by the Boltzmann-Gibbs differential entropy

$$H(p) = -\int_{-\infty}^{\infty} p(x) \log p(x)\, dx\,,$$

which is close to 0 for a very peaked distribution [OP]. If ψ is the state vector in the position representation then its Fourier transform describes the state in the momentum representation. It is the peculiarity of the Fourier transform that it produces a strongly non-localized state function in the momentum space from a sharp one in the position space (Figure 3).

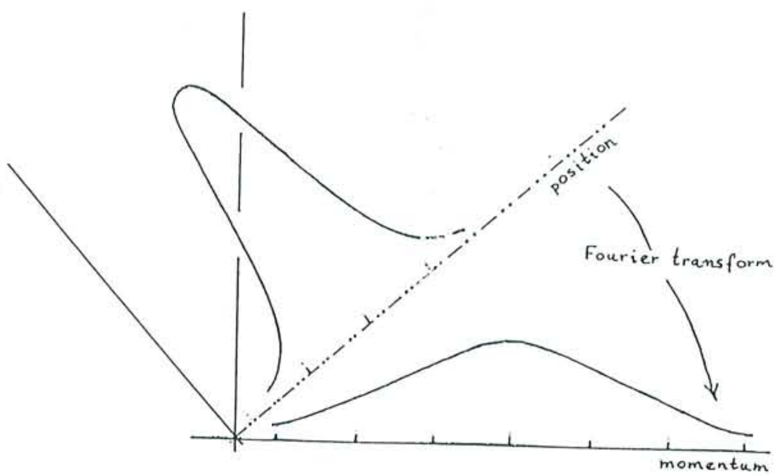

Figure 3. The Fourier transform switches from the position representation to the momentum representation and it increases uncertainty of a localized position wavefunction.

Most readers guess that the uncertainty relation will follow soon. Yes, it will, but in an unconventional form. The momentum-position uncertainty

$$H(p_P) + H(p_Q) \geq 1 + \log \pi$$

tells us, in terms of differential entropy, that there exists a positive lower bound for the sum of uncertainties (called also entropies) which prevent them from being very small at the same time [OP]. The above entropic uncertainty relation was conjectured in 1957 by Hirschmann but the exact proof came only in 1975 after a thorough analysis of the Young inequality. Paul Dirac would be pleased: Fundamental physics goes together with deep mathematics.

In the course of the two-slit experiment, a two-peaked wavefunction may develop, the superpostion of two, spatially separated Gaussian wavepackets has this form (Figure 4).

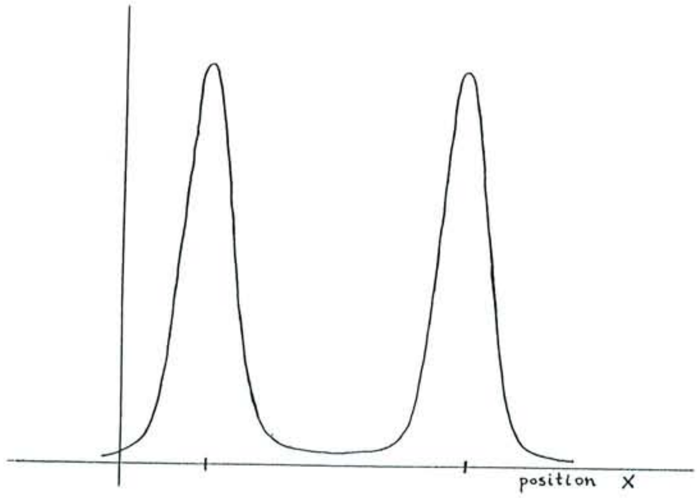

Figure 4. A two-peaked wave function (the superposition of Gaussians without phase difference). A quantum particle with a doubly peaked position wavefunction is apparently at two places.

A photon with a doubly peaked position wavefunction is apparently at two different places, and, while time is passing, it is apparently able to travel along different routes. There are recent investigations (or speculations?) that the ability of a quantum system to be at many places at the same time, may be used in a computer, to branch out computations into many paths, at least in principle,. The difficulty is getting the paths to interfere in a useful way at the end in order to bring out something with reasonable probability. A theoretical quantum computer might be applied effectively to problems whose solution depends on many computation paths [Be].

The probabilistic formulation of quantum mechanics follows the point of view of experimentalists. Many basic concepts are the same as in statistical mechanics or in ordinary mathematical statistics. The quantum system under study should be regarded as a particular sample drawn from a statistical ensemble of identically prepared systems subject to possible individual variations in their measured properties. A state specifies our information at a given time on the statistical ensemble. It does not describe the properties of an individual system. Experimentalists are rarely able to prepare pure states in their laboratories. A realistic state is nothing else but the collection of expectation values of all relevant physical observables. As early as 1927 von Neumann introduced the concept of density matrix for the description of statistical mixtures of pure states, the latter ones are given by state vectors in a Hilbert space [vN]. In general a quantum measurement is incomplete in the sense that it fails to provide an exhaustive determination of the state of the system. It is worthwile to make a short stop here and to have a glance at the historical development of sciences. Quantum mechanics had to cope with the problem of incomplete information at the end of the 1920's. On the other hand, Kolmogorov's fundamental work in probability theory appeared years later. The break-through of rigorous statistical mechanics followed only decades later.

Density matrices do not have direct physical meaning, they come from the statistical viewpoint. Manipulation with them is sophisticated mathematics but density matrices are really needed to deal with spins, quantum fields and macroscopic systems. Von Neumann attachediated an information quantity, called today "von Neumman entropy", to a density matrix. When p_i's are the eigenvalues of the density matrix, the von Neumann entropy is

$$-\sum_i p_i \log p_i \,.$$

This formula might be familiar as Shannon's information measure. To tell the truth, we have to say that von Neumann arrived at his formula by a thermodynamical consideration. It was Shannon who initiated the interpretation as "uncertainty measure" or "information measure". The American electric engineer/scientist Claude Shannon created communication theory in 1948. Many years later he told [TM]: "*My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.'*"

Both the von Neumann and the Shannon entropies are informational/statistical quantities of the same kind. The von Neumann entropy concerns an ensemble of quantum systems while the other is the statistical uncertainty of the signal ensemble.

$$D = \begin{pmatrix} \frac{1}{2} + x_1 & x_2 + \mathrm{i}x_3 \\ x_2 - \mathrm{i}x_3 & \frac{1}{2} - x_1 \end{pmatrix}$$

$$r = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

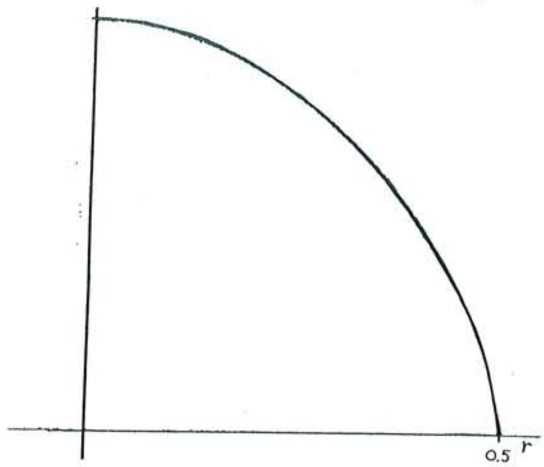

Figure 5. The von Neumann entropy of a 2-by-2 density matrix D is a decreasing function of the parameter r. The value $r = 1/2$ corresponds to pure states, they have vanishing entropy.

It might be in order to note that the use of the notion of entropy does not assume always an explicit or virtual statistical ensemble. There are entropy quantities attached to an individual object. Entropy is the amount of randomness present in a system (independently if it is quantum or not). But why should information be measured by randomness? The statement that a message has high information content is the same that it is extracted from a large class of alternatives, or it is very random. (The information content is the number of digits (bits) of the number of alternatives.) Since quantum mechanics is a statistical theory at a very basic level, entropy, a key idea in dealing with incomplete information and randomness, must play a central role in the theory[Pa]. Does fundamental physics rely upon information theory in this way? Or more strongly: Is the Planck constant h another name for the basic information unit, bit? I do not think so. After all, quantization is not exactly and not only discretization (as it might be suggested by the begining of quantum theory).

Information theory has had remarkable success in searching the performance limits of communication channels. It has developed concepts and methods for the study of randomness. The informational viewpoint may enter (but should not invade) quantum physics. The entropic uncertainty relation of position and momentum is a simple example of the utility of entropy but entropic methods have been applied to a variety of problems. The collapse of the wavepacket, which is far from so dramatic as it sounds, is the prototype of an irreversible process. In the reduced density matrix, only the diagonal survives and the decay of the off-diagonal terms increases the entropy. The reduction of the density matrix (or in particular, that of the wave packet) amounts to keeping the whole information about the observables compatible with the measured one and getting rid of all the remaining information. This is an example of maximization of uncertainty. Geometric interpretation of the evolution of a system as the motion of a point in the space of states has been a fruitful idea in several contexts. It is tempting to endow the state space with the geometric structure deduced from entropy and watch the evolution of the system on a curve finding its direction by a continuous minimization of the information [Ba].

While information theory searches for the ultimate limits of information transfer, we have in mind the capacity of a communication channel, mathematical statistics is interested in the ultimate bounds of estimation. Analysing population statistic, Ronald Aylmer Fisher made the observation that when one has two distinguish between two populations, each of them given by a finite probability distribution, not exactly the probabilities play the principal role but rather the square roots of these probabilities [Ka]. On the other hand, the probabilities are arising in quantum mechanics as the squared modulus of the state vector. Is this an odd coincidence, or, something more? We believe in the latter possibility but a fully satisfactory explanation is difficult to give today. Here are some hints. Fisher found that the natural parametrization of the probability simplex $\{(p_1, p_2, p_3) : p_i > 0 \text{ and } p_1+p_2+p_3 = 1\}$ is given by $z_i = \sqrt{p_i}$. In this way we can visualize the positive orthant portion of the sphere of radius 1. When this parametrization is adequate, the correct (more precisely, geodesic) distance between the probability distributions (p_1, p_2, p_3) and (p_1', p_2', p_3') is the angle between the vectors $(\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3})$ and $(\sqrt{p_1'}, \sqrt{p_2'}, \sqrt{p_3'})$. There is something similar with density matrices. A 2-by-2 density matrix may be parametrized by (x_1, x_2, x_3) as in Figure 5. The state space (that is, the set of density matrices) is wiewed as the part of a sphere (in the 4 dimensional space). The matrix D in Figure 5 is just the point

$$(x_1, x_2, x_3, x_4), \qquad \text{where } x_4 = \sqrt{\det D}\,.$$

This is not all! The geodesic distance between two points is well-known distance, called after Bures in the context of density matrices [Uh]. Quantum estimation theory will benefit from the understanding of the geometric analogy between the quantum state space and Fisher's old discovery in statistics.

Let me close this report from the vivid boundary of quantum mechanics, information theory, statistics, and geometry with a very concise message. The curved "arena of density matrices" [Uh] will be busy place in the near future. It will be occupied by researchers, rather than fighters, making interdisciplinary efforts to understand this new area.

REFERENCES

[Ba] R. Balian, On the principles of quantum mechanics and the reduction of the wave packet, Am. J. Phys. **57**(1989), pp. 1019–1027

[Be] C. H. Bennett, Certainty from uncertainty, Nature **362** (22 April 1993), pp. 694–695

[Ka] R.E. Kass, The geometry of asymptotic inference, Stat. Sci. 4(1989), pp. 184–234

[MR] J. Mehra, H. Rechenberg, *The Historical Development of Quantum Theory*, Volumes 1–5 (Springer, 1982, 1987)

[vN] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932)

[OP] M. Ohya, D. Petz, *The Use of Quantum Entropy* (Springer, 1993)

[Om] R. Omnès, Consistent interpretations of quantum mechanics, Rev. Mod. Phys., **64**(1992), pp. 339–382

[Pa] M.H. Partovi, "Entropy and quantum mechanics", in *Complexity, Entropy and the Physics of Information*, ed. W.H. Zurek (Addison-Wesley, 1990)

[Pe] R. Penrose, *The Emperors's New Mind* (Oxford University Press, 1989)

[TM] M. Tribus, E.C. McIrvine, Energy and information, Scientific American, **224** (September 1971), pp. 178–184

[Uh] A. Uhlmann, Density operators as an arena for differential geometry, Conference Lecture, Torun, December 1992