Missing Link Between Probability Theory and Quantum Mechanics: the Riesz-Fejér Theorem

F. H. Fröhner

Forschungszentrum Karlsruhe, Institut für Neutronenphysik und Reaktortechnik,
D-76021 Karlsruhe, Germany

Z. Naturforsch. **53 a**, 637–654 (1998); received January 28, 1998

Quantum mechanics is spectacularly successful on the technical level but the meaning of its rules remains shrouded in mystery even more than seventy years after its inception. Quantum-mechanical probabilities are often considered as fundamentally different from classical probabilities, in disregard of the work of Cox (1946) – and of Schrödinger (1947) – on the foundations of probability theory. One central question concerns the superposition principle, i. e. the need to work with interfering wave functions, the absolute squares of which are probabilities. Other questions concern the relationship between spin and statistics or the collapse of the wave function when new data become available. These questions are reconsidered from the Bayesian point of view. The superposition principle is found to be a consequence of an apparently little-known mathematical theorem for non-negative Fourier polynomials published by Fejér in 1915 that implies wave-mechanical interference already for classical probabilities. Combined with the classical Hamiltonian equations for free and accelerated motion, gauge invariance and particle indistinguishability, it yields all basic quantum features – wave-particle duality, operator calculus, uncertainty relations, Schrödinger equation, CPT invariance and even the spin-statistics relationship – which demystifies quantum mechanics to quite some extent.

Key words: Superposition Principle; Wave Packets; Logical Inference; Wave-Particle Duality; Quantum Mechanics.

1. Introduction

Quantum mechanics is usually introduced either axiomatically, with states of physical systems represented by vectors in Hilbert space, or historically, showing how crucial experiments led theoreticians like Planck, Einstein, Bohr and Sommerfeld to a first form of quantum theory, and how Heisenberg, Schrödinger, Pauli, Dirac, Feynman, and others succeeded, with much trial and error, in establishing the rather general and consistent quantum mechanics as we know it today. In neither approach does quantum theory look completely compelling from a logical point of view. There are nagging questions up to this day, at least for some: why microphysical systems behave sometimes like particles and sometimes like waves, about the exact meaning of the complex wave functions, about their collapse when measurements are made, about the role of the observer or his consciousness, about the unexpected interference and quantisation phenomena, about the different statistics for fermions and bosons, etc. Others insist that such questions were either answered long ago, or that they are meaningless, or that in view of the great practical success of the theory any doubts are out of place. At least there is consensus that superposition and interference of waves plays a key role, with wave intensities that can be interpreted as probability densities. Dirac, for example, wrote [1] "I believe that this concept of the probability amplitude is perhaps the most fundamental concept of quantum theory."

In view of the debate about the interpretation of quantum mechanics, still going on seventy years after its inception, it is quite astonishing that there is a mathematical theorem, proved and published well before quantum mechanics was developed, that raises the status of the superposition principle from puzzling empirical discovery to a definite mathematical property of all inherently positive distribution functions such as beam intensities or probability densities. The theorem was published in 1915 by Fejér [2] in

Reprint requests to Dr. F. H. Fröhner;
E-mail: Fritz.Froehner@inr.fzk.de.

0932-0784 / 98 / 0800-0637 $ 06.00

Hungarian and German – languages in which many founding fathers of quantum mechanics were fully conversant (von Neumann and Wigner spoke both) – yet it went unnoticed. Equally remarkable is the fact that the discoverers of the theorem, Riesz and Fejér, never seemed to realise its importance in science. There is not the slightest hint of its usefulness in electrodynamics, communication theory, and especially quantum mechanics even in a rather late edition of the book by Riesz and Sz.-Nagy [3] containing a proof of the "Fejér-Riesz lemma". In the following sections it will be argued that this theorem can be regarded as a missing link between probability theory and quantum mechanics, permitting derivation of all the main features of quantum mechanics in a rather inevitable logical chain of fairly simple arguments. Quantum mechanics can thus be considered as an especially powerful extension of ordinary probability theory, useful also for other than space-time processes, as has been conjectured e. g. by C. F. von Weizsäcker who wrote "I propose the view that general or abstract quantum theory is a general theory of probabilities and nothing else" [4]. Many features of quantum mechanics may have been unexpected but with the key provided by Riesz and Fejér they are not inexplicable.

2. The Riesz-Fejér Theorem and Probabilities

In 1915 L. Fejér, well known for his work on Fourier series, published a proof [2] due to F. Riesz of the following theorem (see Appendix): Each real, non-negative Fourier polynomial (truncated Fourier series) of order n (maximal wave number n) can be expressed as the absolute square of a complex Fourier polynomial of at most the same order,

$$0 \le \rho(x) \equiv \sum_{l=-n}^{n} c_l\, e^{ilx} = \Big|\sum_{k=0}^{n} a_k\, e^{ikx}\Big|^2 \equiv \big|\psi(x)\big|^2 , \tag{1}$$

where the complex Fourier polynomial $\psi(x)$ is completely unrestricted, in contrast to the Fourier polynomial $\rho(x)$ which is restricted by the requirements of reality ($c_{-l} = c_l^*$) and non-negativity. Our notation anticipates the rather obvious application to quantum-mechanical probability densities, ρ, and probability wave functions, ψ, without excluding application to other inherently positive quantities such as intensities of classical energy-carrying waves. The coefficients a_k are Fourier transforms of the wave function $\psi(x)$. Fourier techniques are most convenient whenever wave or particle propagation constrained by initial or boundary conditions is to be described, and they are especially powerful if they permit free use of Fourier expansions, unhampered by reality and non-negativity conditions. Constraints such as point sources, diaphragms, slits, scatterers, etc. define, together with a wave equation for the Fourier components, eigenvalue problems whose eigenfunctions are all those waves which are possible under given experimental circumstances.

Since the wave function $\psi(x)$ for given $\rho(x)$ is determined only up to a phase factor, we may define it by

$$\sum_{k=0}^{n} a_k\, e^{ikx} \equiv e^{inx/2}\psi(x) \tag{2}$$

and introduce modified coefficients

$$\phi_k \equiv \sqrt{2\pi}\, a_{k+n/2}. \tag{3}$$

The resulting Fourier transform pair

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \sum_{k=-n/2}^{n/2} \phi_k\, e^{ikx}, \tag{4}$$

$$\phi_k = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \mathrm{d}x\; \psi(x)\, e^{-ikx}, \tag{5}$$

has the especially convenient symmetric form commonly used in quantum mechanics. The wave functions thus defined are

$$n = 0, \quad \psi(x) = \frac{1}{\sqrt{2\pi}}\phi_0 , \tag{6}$$

$$n = 1, \quad \psi(x) = \frac{1}{\sqrt{2\pi}}\left(\phi_{-1/2}\, e^{-ix/2} + \phi_{+1/2}\, e^{+ix/2}\right),$$

$$n = 2, \quad \psi(x) = \frac{1}{\sqrt{2\pi}}\left(\phi_{-1}\, e^{-ix} + \phi_0 + \phi_{+1}\, e^{+ix}\right),$$

$$\vdots \qquad\qquad \vdots$$

Although $\rho(x)$ has 2π periodicity and the sum in (2) likewise, the same is true only for wave functions for

even n. Those for odd n have 4π periodicity, so one has

$$\psi(x) = \pm\psi(x+2\pi) = \psi(x+4\pi) \text{ for } n = \begin{cases} 0,2,4,\dots \\ 1,3,5,\dots \end{cases}. \tag{7}$$

The reason is, of course, the phase factor $e^{inx/2}$ pulled out in (2). As will become clearer below, the plus-minus sign appearing after one full period leads to the spin-statistics relationship and thus to the Bose-Einstein and Fermi-Dirac statistics of quantum mechanics.

Integration of the function $\rho(x)$ over one period yields

$$\int_{-\pi}^{\pi} \mathrm{d}x|\psi(x)|^2 = \sum_{k=-n/2}^{n/2} |\phi_k|^2 \tag{8}$$

(Parseval's theorem). If we interpret this as a total probability we find that any continuous, periodic probability density $|\psi(x)|^2$ in x-space is related via Fourier transformation to discrete probabilities $|\phi_k|^2$ in a dual k-space – *periodic x entails quantised k.*

In classical as well as quantum mechanics (infinite) Fourier series occur naturally when spatial rotations are studied with their obvious 2π periodicity. In other cases, like spatial translation, it is customary to introduce Fourier series by the familiar device of the periodicity box. We note that

(a) infinite Fourier series can be approximated by finite Fourier polynomials to any desired accuracy if only the order n is chosen high enough;

(b) a smooth transition to Fourier integrals describing arbitrary nonperiodic processes is achieved if the periodicity box is made bigger and bigger relative to the physical system considered.

In view of these uneventful generalisations one may consider the Riesz-Fejér theorem as equivalent to the wave-mechanical superposition principle: *Probability densities as inherently non-negative quantities can be represented as absolute squares of wave functions that in their turn can be expressed as linear superpositions of orthogonal functions.* In (1) the orthogonal functions describe standing waves in a (one-dimensional) periodicity box. Other possible orthogonal bases – spherical waves, angular momentum eigenfunctions etc. – can be invoked by unitary transformations. An immediate consequence is *quantisation: periodic probability density functions in one space are accompanied by discrete probabilities in a dual space*, both spaces related by Fourier transformation of the wave functions. A further consequence is the appearance of two families of eigenfunctions with 2π and 4π periodicity (for *bosons and fermions*). All this is just part of Fourier theory, *valid for all periodic non-negative functions, in particular for all periodic probability densities,* not only quantum-mechanical ones.

Historically, the superposition principle had been established first as a puzzling empirical feature of the quantum world, before M. Born recognised that the absolute square of the wave function can be interpreted as a probability density. If, on the other hand, one starts with probabilities, the superposition principle, far from puzzling, appears as a theorem, applicable not only in quantum mechanics but also to nonquantal probabilities and signal intensities (cf. e. g. Feller [5] on L^2 theory and especially Cohen [6] on time-frequency analysis). The much discussed role of the phases of the superposed functions seems to be simple: They are needed for a faithful reproduction of the non-negative probability density $\rho(x)$ in (1). We shall see below that they also define possible forces.

3. Free Particle With Uncertain Initial Location

Let us consider a classical particle. Its energy as a function of its position $\boldsymbol{r}$ and (generalised) momentum $\boldsymbol{p}$ is given by the Hamiltonian $H = H(\boldsymbol{r},\boldsymbol{p},t)$; its motion is gouverned by Hamilton's canonical equations

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \frac{\partial H}{\partial \boldsymbol{p}}, \quad \frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -\frac{\partial H}{\partial \boldsymbol{r}}. \tag{9, 10}$$

For given initial phase space coordinates, $\{\boldsymbol{r}(0),\boldsymbol{p}(0)\}$, one obtains the trajectory in phase space, $\{\boldsymbol{r}(t),\boldsymbol{p}(t)\}$, by integration of the canonical equations, for $t < 0$ as well as for $t > 0$. If the initial coordinates are uncertain, lying somewhere in a phase space domain $D(0)$, there is a multitude of possible trajectories. At time t the possible values of $\boldsymbol{r}(t)$ and $\boldsymbol{p}(t)$ lie in a domain $D(t)$ that has the same size as $D(0)$: The canonical equations imply zero divergence in phase space,

$$\sum_{j=1}^{3} \left(\frac{\partial}{\partial r_j}\frac{\mathrm{d}r_j}{\mathrm{d}t} + \frac{\partial}{\partial p_j}\frac{\mathrm{d}p_j}{\mathrm{d}t} \right) = 0 \tag{11}$$

(Liouville's theorem, valid already separately for each pair r_j, p_j).

More generally, the initial uncertainty may be described by a continuous probability density. In thermodynamics one is accustomed to assign joint probabilities for position and momentum of the particle but if one considers probability distributions ("ensembles") of particle trajectories in ordinary space some care is required. If the physically allowed trajectories $\boldsymbol{r}(t)$ are specified, at least in principle, it is enough to assign position probabilities at some particular time. Those for other times can then be deduced from the functions $\boldsymbol{r}(t)$, and the velocity or momentum distributions, too.

Let us consider a time-dependent spatial probability density $\rho(\boldsymbol{r}, t) = |\psi(\boldsymbol{r}, t)|^2$ in a periodicity box so large that the Fourier polynomials of the Riesz-Fejér theorem can be replaced by Fourier integrals. The resulting wave function and its Fourier transform,

$$\psi(\boldsymbol{r}, t) = (2\pi)^{-3/2} \int \mathrm{d}^3k\, \phi(\boldsymbol{k}, t)\, e^{+i\boldsymbol{k}\cdot\boldsymbol{r}}\,, \qquad (12)$$

$$\phi(\boldsymbol{k}, t) = (2\pi)^{-3/2} \int \mathrm{d}^3r\, \psi(\boldsymbol{r}, t)\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\,, \qquad (13)$$

are both normalised to unity, $\int \mathrm{d}^3r|\psi(\boldsymbol{r}, t)|^2 = \int \mathrm{d}^3k \cdot|\phi(\boldsymbol{k}, t)|^2 = 1$. The best estimates (under quadratic loss) of positions and wave vectors are then given by the expectation values

$$\begin{aligned}\langle \boldsymbol{r}(t)\rangle &= \int \mathrm{d}^3r\, |\psi(\boldsymbol{r}, t)|^2\, \boldsymbol{r} \\ &= \int \mathrm{d}^3k\, \phi(\boldsymbol{k}, t)^* \frac{i\partial}{\partial \boldsymbol{k}} \phi(\boldsymbol{k}, t),\end{aligned} \qquad (14)$$

$$\begin{aligned}\langle \boldsymbol{k}(t)\rangle &= \int \mathrm{d}^3k\, |\phi(\boldsymbol{k}, t)|^2\, \boldsymbol{k} \\ &= \int \mathrm{d}^3r\, \psi(\boldsymbol{r}, t)^* \frac{\partial}{i\partial \boldsymbol{r}} \psi(\boldsymbol{r}, t)\end{aligned} \qquad (15)$$

since $|\phi|^2$ is the probability density in $\boldsymbol{k}$-space corresponding to the probability density $|\psi|^2$ in the dual $\boldsymbol{r}$-space. We note that the factor $\boldsymbol{k}$ in $\boldsymbol{k}$-space is to be replaced by the operator $-i\partial/\partial \boldsymbol{r}$ in $\boldsymbol{r}$-space, and the factor $\boldsymbol{r}$ in $\boldsymbol{r}$-space by the operator $i\partial/\partial \boldsymbol{k}$ in $\boldsymbol{k}$-space, since application of the operators to the Fourier components $e^{\pm i\boldsymbol{k}\cdot\boldsymbol{r}}$ produces the required integrands. (The well known proof involves Fourier transformation of ψ and ϕ, integration by parts under the assumption of vanishing probability densities for $k \to \infty$ and $r \to \infty$, and use of the completeness relations $(2\pi)^{-3} \int \mathrm{d}^3r\, e^{i\boldsymbol{r}\cdot(\boldsymbol{k}-\boldsymbol{k}')} = \delta(\boldsymbol{k} - \boldsymbol{k}')$ and $(2\pi)^{-3} \int \mathrm{d}^3k\, e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} = \delta(\boldsymbol{r} - \boldsymbol{r}')$.) An immediate consequence are the commutation relations in $\boldsymbol{r}$ or $\boldsymbol{k}$ space, $r_jk_{j'} - k_{j'}r_j = i\delta_{jj'}$ $(j, j' = 1, 2, 3)$. The familiar (classical) wave-mechanical uncertainty relations for Fourier transforms that follow from the Cauchy-Schwarz inequality are

$$\Delta r_j\, \Delta k_{j'} \geq \frac{1}{2}\delta_{jj'}\,, \quad (j, j' = 1, 2, 3)\,, \qquad (16)$$

where Δr_j and $\Delta k_{j'}$ are standard (root-mean-square) errors (see e. g. Cohen [6]).

The specific time dependence $\phi(\boldsymbol{k}, t) = \phi(\boldsymbol{k}, 0) \cdot e^{-i\omega t}$ describes a superposition of plane waves propagating with phase velocities ω/k in directions $\boldsymbol{k}/k$,

$$\psi(\boldsymbol{r}, t) = (2\pi)^{-3/2} \int \mathrm{d}^3k\, \phi(\boldsymbol{k}, 0)\, e^{+i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}\,, \qquad (17)$$

$$\phi(\boldsymbol{k}, t) = (2\pi)^{-3/2} \int \mathrm{d}^3r\, \psi(\boldsymbol{r}, 0)\, e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}+\omega t)}\,. \qquad (18)$$

In general, the frequencies for different wave lengths will differ, $\omega = \omega(k)$. The expectation value is

$$\begin{aligned}\langle \omega\rangle &= \int \mathrm{d}^3r\, \psi(\boldsymbol{r}, t)^* \frac{i\partial}{\partial t} \psi(\boldsymbol{r}, t) \\ &= \int \mathrm{d}^3k\, \phi(\boldsymbol{k}, t)^* \frac{i\partial}{\partial t} \phi(\boldsymbol{k}, t),\end{aligned} \qquad (19)$$

which shows that for averaging purposes ω is equivalent to the operator $i\partial/\partial t$ in both representations.

From (17) and (18) one finds in $\boldsymbol{k}$ representation the expectation values (best estimates under quadratic loss in the language of decision theory [7]) for the position and the wave vector

$$\begin{aligned}\langle \boldsymbol{r}(t)\rangle &= \int \mathrm{d}^3k\, \phi(\boldsymbol{k}, 0)^* \Big(\frac{i\partial}{\partial \boldsymbol{k}} \phi(\boldsymbol{k}, 0) + \phi(\boldsymbol{k}, 0) \frac{\partial \omega}{\partial \boldsymbol{k}}\, t\Big) \\ &= \langle \boldsymbol{r}(0)\rangle + \Big\langle \frac{\partial \omega}{\partial \boldsymbol{k}} \Big\rangle_{t=0} t\,,\end{aligned} \qquad (20)$$

$$\langle \boldsymbol{k}(t)\rangle = \int \mathrm{d}^3k\, \phi(\boldsymbol{k}, 0)^* \boldsymbol{k} \phi(\boldsymbol{k}, 0) = \langle \boldsymbol{k}(0)\rangle\,, \qquad (21)$$

which obviously describes linear translation with constant group velocity $\langle \partial\omega/\partial \boldsymbol{k}\rangle$. As was to be expected,

wave packets constructed of undistorted plane waves can move only along straight lines, like physical free particles not influenced by forces.

So far we used only (Fourier-Riesz-Fejér) wave theory but now we can establish contact with Hamiltonian particle kinematics. We identify the motion of the wave packet with the expected motion of the particle and compare the wave-mechanical (classical) time derivatives

$$\frac{\mathrm{d}\langle \boldsymbol{r}\rangle}{\mathrm{d}t} = \left\langle \frac{\partial \omega}{\partial \boldsymbol{k}} \right\rangle , \quad \frac{\mathrm{d}\langle \boldsymbol{k}\rangle}{\mathrm{d}t} = 0 , \tag{22, 23}$$

and the expectation values that follow from Hamilton's canonical equations for a classical free particle,

$$\frac{\mathrm{d}\langle \boldsymbol{r}\rangle}{\mathrm{d}t} = \left\langle \frac{\partial H}{\partial \boldsymbol{p}} \right\rangle , \quad \frac{\mathrm{d}\langle \boldsymbol{p}\rangle}{\mathrm{d}t} = 0 , \tag{24, 25}$$

with the same weighting by the spatial probability distribution (17) (hence for the same ensemble of possible trajectories). Evidently one can take $\boldsymbol{k} \propto \boldsymbol{p}$ and $\omega \propto H$ (disregarding a possible additive constant for H – for $\boldsymbol{p}$ such a constant is excluded by isotropy). If we denote the common proportionality constant by $\hbar$ we get de Broglie's particle-wave transcription,

$$H = \hbar\omega , \quad \boldsymbol{p} = \hbar \boldsymbol{k} , \tag{26, 27}$$

and from (16) Heisenberg's quantum-mechanical uncertainty relations,

$$\Delta r_j \, \Delta p_{j'} \geq \frac{\hbar}{2} \delta_{jj'} , \quad (j, j' = 1, 2, 3) , \tag{28}$$

replacing Liouville's theorem (11). The equality sign applies if both the spatial and the momentum probability density functions are three-dimensional Gaussians. A sharply peaked spatial density implying a diffuse momentum density means particle-like behavior of the wave packet. On the other hand the behavior is wave-like if the momentum is well defined but not the location. Whether a particle or a wave description is more appropriate depends on the state of information about the particle. *It is not the particle but the wave packet encoding this information that exhibits wave-particle duality.*

This is the crucial point where Planck's quantum of action, $\hbar$, enters the scene, tying together two classical formalisms, Hamiltonian particle mechanics and the Fourier-Riesz-Fejér wave mechanical extension of probability theory, whereby quantum mechanics is ushered in. The decisive new feature is that in the probabilistic description of a particle with uncertain coordinates the momentum distribution is fully determined by the position distribution and vice versa, via unitary (Fourier) transformation of the wave function. As a consequence of this *rather special entanglement – the wave function is Fourier transformed rather than the probability density* – and of the finite value of $\hbar$, the accuracy with which the momentum can be specified is limited by the accuracy of the location. Empirically, $\hbar$ is found to be a natural constant, not merely a formal proportionality factor that can be made arbitrarily small. This clashes with the use of joint probability distributions for particle positions and momenta in statistical mechanics, and limits the classical phase space concept to situations where $\hbar$ can be treated as negligibly small.

Expectation values of physical quantities that depend on both $\boldsymbol{r}$ and $\boldsymbol{p}$ can be calculated from ψ or ϕ with the appropriate operators. For example, the best estimate of the orbital angular momentum with respect to the origin, $\boldsymbol{r} = 0$, is

$$\begin{aligned} \langle \boldsymbol{r} \times \boldsymbol{p} \rangle &= \int \mathrm{d}^3 r \, \psi^* \Big(\boldsymbol{r} \times \frac{\hbar}{i} \frac{\partial}{\partial \boldsymbol{r}} \Big) \psi \\ &= \int \mathrm{d}^3 k \, \phi^* \Big(\frac{i\partial}{\partial \boldsymbol{k}} \times \hbar \boldsymbol{k} \Big) \phi . \end{aligned} \tag{29}$$

Caution is required, however. The employed operators must be Hermitean (self-conjugate) in order to produce *real* expectation values as is required for physical observables. For the orbital angular momentum there is no problem but other products of noncommuting operators are not Hermitean. For example, if one wants to calculate the expectation value $\langle \boldsymbol{r} \cdot \boldsymbol{p} \rangle$ one must use the operator $(\boldsymbol{r} \cdot \boldsymbol{p} + \boldsymbol{p} \cdot \boldsymbol{r})/2$ with $\boldsymbol{p} = -i\hbar\partial/\partial \boldsymbol{r} \equiv -i\hbar\nabla$.

As mentioned already, quantisation is a conspicuous consequence of the introduction of probability waves and operator calculus by means of the Riesz-Fejér theorem. It is characterised by the existence of eigenvalue equations with discrete eigenvalues. The most important eigenvalue problems are, of course, those defined by the Schrödinger equation,

$$H\psi = i\hbar\dot{\psi}, \tag{30}$$

together with initial or boundary conditions. This central equation of quantum mechanics follows from (19) with $H = \hbar\omega$ (and the notation $\dot{\psi} \equiv \partial\psi/\partial t$). With $H(\boldsymbol{r}, \boldsymbol{p}, t)$ in operator form ($\boldsymbol{p} = -i\hbar\nabla = -i\hbar\partial/\partial \boldsymbol{r}$)

it is the wave equation determining the time evolution of $\psi(\boldsymbol{r}, t)$ and of its Fourier transform $\phi(\boldsymbol{k}, t)$. In this sense it is the wave-mechanical analog of Hamilton's particle equations (9) and (10). If ψ is an eigenfunction with eigenvalue E one has $\langle H\rangle = E$ and $(\Delta H)^2 = \langle H^2\rangle - \langle H\rangle^2 = 0$: The energy predicted under quadratic loss is E without uncertainty.

Similarly, if ψ tends towards an eigenfunction of the momentum operator $-i\hbar\boldsymbol{\nabla}$ with eigenvector $\boldsymbol{p}$ the momentum uncertainty goes to zero while the position uncertainty becomes infinite. The limit, with $\psi(\boldsymbol{r}) \propto \exp(i\boldsymbol{p}\cdot\boldsymbol{r}/\hbar)$, is a useful idealisation for a particle in a beam, but in practice the momentum cannot be quite sharp since the beam dimensions and thus the position uncertainty may be huge compared to the particle dimensions but not really infinite (which perhaps could be taken as a hint that physical particles are not really mathematical points, and that $\hbar$, Planck's quantum of action, may have something to do with their finite size). Quite generally, whenever ψ is one of the eigenfunctions of some operator, the variance of the corresponding physical quantity vanishes.

4. Introduction of Forces via Local Gauge Transformation

So far we have considered superpositions of plane waves and found that they correspond to spinless free particles whose Hamiltonian is $H_0 = \boldsymbol{p}^2/2m$ in the nonrelativistic case. We can generalise to accelerated motion, i. e. to forces. Let us assume that the spatial probability density for accelerated motion, $\rho = |\psi|^2$, coincides with that for free motion, $\rho_0 = |\psi_0|^2$, at time $t = 0$. At this time the wave functions can only differ by a phase factor. We must therefore have

$$\psi_0 = \psi \, \exp(-i\Lambda e/\hbar c)\,, \tag{31}$$

where the phase is written in a form that is convenient for our purposes. The real "gauge function" Λ cannot be a mere constant but must depend on $\boldsymbol{r}$ and t if ρ is ever to differ from ρ_0. Inserting (31) into the nonrelativistic Schrödinger equation for free motion,

$$H_0\psi_0 = \frac{1}{2m}(-i\hbar\boldsymbol{\nabla})^2\psi_0 = i\hbar\dot{\psi}_0\,, \tag{32}$$

one finds for the accelerated wave packet

$$\frac{1}{2m}\Big(-i\hbar\boldsymbol{\nabla} - \frac{e}{c}\boldsymbol{\nabla}\Lambda\Big)^2\psi = \Big(i\hbar\frac{\partial}{\partial t} + \frac{e}{c}\dot{\Lambda}\Big)\psi\,. \tag{33}$$

We conclude that the nonrelativistic Schrödinger equation for a particle influenced by forces must have the general form

$$H\psi = \Big[\frac{1}{2m}(-i\hbar\boldsymbol{\nabla} - \frac{e}{c}\boldsymbol{A})^2 + e\Phi\Big]\psi = i\hbar\dot{\psi}\,, \tag{34}$$

where the real scalar Φ and the real vector $\boldsymbol{A}$ are subject to modification by $-\dot{\Lambda}/c$ and $\boldsymbol{\nabla}\Lambda$. We recognise them as the scalar and vector potentials of the electromagnetic force field, and e as the particle charge that specifies how strongly the particle responds to the field. The form of the Schrödinger equation is invariant under "local gauge transformations of the first kind" of the wave function,

$$\psi \;\rightarrow\; \psi' = \psi \, \exp(i\Lambda e/\hbar c)\,, \tag{35}$$

in combination with "gauge transformations of the second kind" of the electromagnetic potentials,

$$\boldsymbol{A} \;\rightarrow\; \boldsymbol{A}' = \boldsymbol{A} + \boldsymbol{\nabla}\Lambda\,, \tag{36}$$

$$\Phi \;\rightarrow\; \Phi' = \Phi - \frac{1}{c}\dot{\Lambda}\,. \tag{37}$$

Neither probabilities nor observables can depend on the arbitrary gauge function Λ. This means that ψ, $\boldsymbol{A}$, and Φ are merely auxiliary formal quantities, in contrast to the given probability density ρ and the measurable electric and magnetic field strengths

$$\boldsymbol{E} = -\boldsymbol{\nabla}\Phi - \frac{1}{c}\dot{\boldsymbol{A}}\,, \tag{38}$$

$$\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}\,. \tag{39}$$

Operators for observable quantities may, for the same reason, contain spatial and temporal derivatives only in the gauge invariant combinations

$$-i\hbar\boldsymbol{\nabla} - \frac{e}{c}\boldsymbol{A} = \boldsymbol{p} - \frac{e}{c}\boldsymbol{A}\,, \tag{40}$$

$$+i\hbar\frac{\partial}{\partial t} - e\Phi = H - e\Phi\,. \tag{41}$$

Canonical quantities like $\boldsymbol{p}$ or $\boldsymbol{p}^2/2m$ are observables only if no forces act, Newtonian quantities like $m\boldsymbol{v} \equiv (\boldsymbol{p} - e\boldsymbol{A}/c)$ or $m\boldsymbol{v}^2/2 + V(\boldsymbol{r})$ always. Time derivatives of expectation values for observables can be

calculated after the pattern

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle O\rangle = \langle\dot{\psi}|O|\psi\rangle + \langle\psi|\dot{O}|\psi\rangle + \langle\psi|O|\dot{\psi}\rangle = \langle\dot{O} + \frac{1}{i\hbar}(OH - HO)\rangle, \tag{42}$$

where we employed Dirac's bra-ket notation $\langle O\rangle = \langle\psi|O|\psi\rangle = \int \psi^* O\psi \mathrm{d}^3 r$ for expectation values. With the Hamiltonian of (34) one finds readily for the most important observables [8]

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \boldsymbol{r}\rangle \equiv \langle \boldsymbol{v}\rangle = \frac{1}{m}\langle \boldsymbol{p} - \frac{e}{c}\boldsymbol{A}\rangle , \tag{43}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle m\boldsymbol{v}\rangle = e\langle \boldsymbol{E} + \frac{1}{2c}(\boldsymbol{v}\times\boldsymbol{B} - \boldsymbol{B}\times\boldsymbol{v})\rangle, \tag{44}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \frac{m\boldsymbol{v}^2}{2}\right\rangle = \frac{e}{2}\langle \boldsymbol{v}\cdot\boldsymbol{E} + \boldsymbol{E}\cdot\boldsymbol{v}\rangle , \tag{45}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \frac{m\boldsymbol{v}^2}{2} + V\right\rangle = \frac{e}{2}\langle \boldsymbol{v}\cdot\boldsymbol{E}_t + \boldsymbol{E}_t\cdot\boldsymbol{v}\rangle . \tag{46}$$

Equation (43) gives the relationship between velocity and generalised momentum while the following equations show that any acceleration is due to the Lorentz force, that the kinetic energy is changed by the electric but not by the magnetic field, and that the total energy is changed only by the time-dependent part $\boldsymbol{E}_t$ of the electric field, while the static part, $\boldsymbol{E}_0 \equiv \boldsymbol{E} - \boldsymbol{E}_t$, gives rise to the potential energy

$$V \equiv -e\int_{\infty}^{\boldsymbol{r}} \mathrm{d}\boldsymbol{r}'\cdot\boldsymbol{E}_0(\boldsymbol{r}') . \tag{47}$$

(Only static forces can define a potential energy.)

We note with interest that the Schrödinger equation (34) exhibits one more invariance: Upon CPT transformation, i. e. combined time reversal T, spatial reflexion P, and charge conjugation C,

$$t \to -t,\ \boldsymbol{r} \to -\boldsymbol{r},\ e \to -e, \tag{48}$$

one finds the original Schrödinger equation (34) again, but now for ψ^* instead of ψ. This is true if under PT not only $\partial/\partial\boldsymbol{r}$ and $\partial/\partial t$ change sign but the complete gauge-invariant combinations (40) and (41) – which leaves the electromagnetic field (38), (39) unchanged. Now C (together with $\psi \to \psi^*$) means particle-antiparticle exchange, while PT implies full reversal of all motions, temporal as well as spatial, as in a movie running backwards. One concludes that backward motion of particles (in the space-time PT sense) can also be treated formally as forward motion of antiparticles, in the same electromagnetic field, as recognised in the relativistic case by Stueckelberg [9] and further elaborated by Costa de Beauregard [10].

It is remarkable that the mere existence of *an arbitrary phase* of the wave function *leads unambiguously to the electromagnetic interaction* and thus to the Lorentz force as the only possible influence on the motion of a spinless charged particle. Although we demonstrated this for nonrelativistic particles only, it is easy to see that it must be true for relativistic particles, too: Hamilton's equations of motion hold for relativistic as well as nonrelativistic Hamiltonians, and the concept of position uncertainty at a given time remains viable, with all consequences. Any Hamiltonian depends on the momentum by definition, therefore the corresponding wave equation, whether relativistic or not, must contain spatial and temporal derivatives only in the gauge invariant combinations (40) and (41) [11]. The Klein-Gordon and Dirac equations are examples [12]. In any case, gauge invariance and electromagnetic fields show their basic simplicity and inevitability most clearly in the relativistic formalism. The strategy of deriving the form of an interaction from phase arbitrariness is due to Weyl [13] as well as the term gauge invariance. More recently it has been the decisive tool for the construction of the electro-weak theory and of quantum chromodynamics (cf. e. g. [14, 15]).

5. Angular Momenta and Spinors

All spatial probability wave packets possess angular periodicity, $\rho(2\pi + \alpha) = \rho(\alpha)$, around any fixed axis in ordinary space. This implies discrete angular momentum eigenvalues and the possibility to expand the wave function in terms of the corresponding eigenfunctions, viz. spherical harmonics, $\psi(\boldsymbol{r}, t) = \sum_{\ell,m} a_{\ell m}(r, t)Y_\ell^m(\Omega)$. In order to explore the consequences, consider a physical system whose total angular momentum is due to the orbital motion of a bound spinless particle, for instance a hydrogen-like atom (without spin). Let $\boldsymbol{R}$ be the center-of-mass position of the whole system and $\boldsymbol{r}$ the position of the bound particle relative to $\boldsymbol{R}$. If the system has expected spin 1 (in units of $\hbar$), all expansion terms with

$\ell \neq 1$ vanish and the eigenfunction expansion is

$$\begin{aligned}\Psi(\boldsymbol{R},\boldsymbol{r},t) &= \psi_+(\boldsymbol{R},t)\chi_+(\boldsymbol{r}) + \psi_0(\boldsymbol{R},t)\chi_0(\boldsymbol{r}) \\ &\quad + \psi_-(\boldsymbol{R},t)\chi_-(\boldsymbol{r}) \\ &= u_x(\boldsymbol{R},t)\xi(\boldsymbol{r}) + u_y(\boldsymbol{R},t)\eta(\boldsymbol{r}) \\ &\quad + u_z(\boldsymbol{R},t)\zeta(\boldsymbol{r})\end{aligned} \tag{49}$$

with

$$u_x = \frac{1}{\sqrt{2}}(\psi_+ + \psi_-),\ u_y = \frac{i}{\sqrt{2}}(\psi_+ - \psi_-),\ u_z = \psi_0, \tag{50}$$

and

$$\begin{aligned}\chi_\pm(\boldsymbol{r}) &= f(r)Y_1^{\pm1}(\theta,\phi) \\ &= \sqrt{\frac{3}{4\pi}}\frac{f(r)}{r}\frac{x \pm iy}{\sqrt{2}} \equiv \frac{\xi \pm i\eta}{\sqrt{2}}, \\ \chi_0(\boldsymbol{r}) &= f(r)Y_1^0(\theta,\phi) \\ &= \sqrt{\frac{3}{4\pi}}\frac{f(r)}{r}z \equiv \zeta,\end{aligned} \tag{51}$$

The vector (ξ,η,ζ) is seen to behave like (x,y,z) under coordinate transformations, i. e. as a polar vector, and the vector $\boldsymbol{u} = (u_x,u_y,u_z)$ likewise, since the wave function must remain unchanged. For a beam of randomly oriented spin-1 atoms, left and right handed circular polarisation must be equivalent,

$$\langle|\psi_+|^2\rangle = \langle|\psi_-|^2\rangle\,, \tag{52}$$

and also all Cartesian coordinates of $\boldsymbol{u}$, which implies $\langle\boldsymbol{u}\rangle = 0$ for the mean vector and

$$\langle|u_x|^2\rangle = \langle|u_y|^2\rangle = \langle|u_z|^2\rangle = \langle\boldsymbol{u}^2\rangle/3 \tag{53}$$

for the variances. With the linear relations (50) one obtains equal probability densities for all three orientation eigenstates,

$$\langle|\psi_+|^2\rangle = \langle|\psi_0|^2\rangle = \langle|\psi_-|^2\rangle = \langle\boldsymbol{u}^2\rangle/3. \tag{54}$$

As a consequence, a beam of randomly oriented particles is split into three equally intense components in any field that acts differently on the three angular momentum eigenstates, as does for example the inhomogeneous magnetic field, $\boldsymbol{B}$, of a Stern-Gerlach magnet. Its force, $\boldsymbol{F} = \nabla(\boldsymbol{m}\cdot\boldsymbol{B})$, is proportional (a) to the field gradient and (b) to the component of the magnetic moment, $\boldsymbol{m}$, and therefore also of the orbital angular momentum, $\boldsymbol{r}\times\boldsymbol{p}$, along the direction of the field gradient. For any orientation of the magnet the wave function can be expanded in terms of three spherical harmonics which cause the wave packet to exhibit three discrete observable eigenvalues of the angular momentum orientation. Contrary to common belief, this quantisation is not restricted to quantum systems. Simply because of the 2π periodicity of rotations in ordinary space it must be true for all spatial probability distributions with finite extension. No matter what experimental method one employs to measure the angular momentum, and no matter how one orients the applied force field, the possible (probability-weighted) internal motions always seem to conspire in such a way that *the wave packet as a whole behaves as if the angular momentum were quantised.* This is true although we have not constrained them except by demanding rotational periodicity and a bound system. It is important to realise that as a consequence an experimenter's selective capabilities, for example to prepare beams of particles with specified *true* spins, are fundamentally limited.

The averages over the internal coordinates x,y,z that we encounter here are more easily calculated with 3-component "spinors". We introduce a complete basis of spinors,

$$\chi_+ = \begin{pmatrix}1\\0\\0\end{pmatrix},\ \chi_0 = \begin{pmatrix}0\\1\\0\end{pmatrix},\ \chi_- = \begin{pmatrix}0\\0\\1\end{pmatrix}, \tag{55}$$

whose orthonormality conditions (in obvious notation) are

$$\begin{aligned}\chi_m^\dagger\chi_{m'} &= \int \mathrm{d}^3r\ \chi_m(\boldsymbol{r})^*\chi_{m'}(\boldsymbol{r}) \\ &= \int_0^\infty |f(r)|^2\mathrm{d}r \int_{4\pi} Y_1^m(\Omega)^*Y_1^{m'}(\Omega)\mathrm{d}\Omega = \delta_{mm'},\end{aligned} \tag{56}$$

where the dagger denotes the Hermitean conjugate. (We assume the radial function $f(r)$ to be normalised appropriately.) In spinor notation the wave function is a three-component spinor, too,

$$\begin{aligned}\boldsymbol{\Psi}(\boldsymbol{R},t) &= \psi_+(\boldsymbol{R},t)\boldsymbol{\chi}_+ + \psi_0(\boldsymbol{R},t)\boldsymbol{\chi}_0 + \psi_-(\boldsymbol{R},t)\boldsymbol{\chi}_- \\ &= \begin{pmatrix}\psi_+(\boldsymbol{R},t)\\ \psi_0(\boldsymbol{R},t)\\ \psi_-(\boldsymbol{R},t)\end{pmatrix},\end{aligned} \tag{57}$$

normalised by

$$\int \mathrm{d}^3R\, \Psi(\boldsymbol{R},t)^\dagger \Psi(\boldsymbol{R},t) = \int \mathrm{d}^3R \int d^3r\, |\Psi(\boldsymbol{R},\boldsymbol{r},t)|^2 = 1. \tag{58}$$

In essence the spinor notation replaces the wave function $\Psi(\boldsymbol{R},\boldsymbol{r},t)$ by three wave functions $\psi_m(\boldsymbol{R},t)$ that are originally the coefficients of an eigenfunction expansion in the space of the intrinsic coordinates. Otherwise the intrinsic coordinates themselves are no longer visible. In a similar way one can describe systems with expected integer spins 2, 3, ... (cf. e. g. [16]). For all of these the wave function has 2π angular periodicity which ensures the same for the probability density.

We have, however, not exhausted all possibilities yet. As we saw above, the condition $\rho(2\pi+\alpha) = \rho(\alpha)$ is not only fulfilled if $\psi(2\pi+\alpha) = +\psi(\alpha)$, with 2π periodicity, but also if $\psi(2\pi+\alpha) = -\psi(\alpha)$, with 4π periodicity (cf. (7)). Therefore another family of possible spins (or representations of the rotation group) exists, with half-integer eigenvalues, 1/2, 3/2, ..., as discovered by Cartan long before the advent of quantum mechanics (see [17]). The familiar spinor formalism for particles with spin 1/2, with two-component spinors

$$\chi_+ = \begin{pmatrix}1\\0\end{pmatrix}, \quad \chi_- = \begin{pmatrix}0\\1\end{pmatrix}, \tag{59}$$

and

$$\Psi(\boldsymbol{R},t) = \psi_+(\boldsymbol{R},t)\chi_+ + \psi_-(\boldsymbol{R},t)\chi_- = \begin{pmatrix}\psi_+(\boldsymbol{R},t)\\ \psi_-(\boldsymbol{R},t)\end{pmatrix}, \tag{60}$$

is the exact analog of the three-component formalism introduced here for systems with spin 1. We took spin 1 as an example because it allows to demonstrate explicitly, without need to go beyond the concept of trajectories or orbits with *orbital* angular momentum, how the spinor formalism accounts for "curled-up" internal degrees of freedom.

In more general probabilistic problems, base spinors like those in (59) and (55) can be employed to enumerate alternative possibilities. Their orthonormality, $\chi_m^\dagger \chi'_m = \delta_{mm'}$, indicates that the alternatives are mutually exclusive, while the completeness relation, $\sum_m \chi_m \chi_m^\dagger = \mathbb{1}$, means that there are no other ones. The two mutually exclusive possible outcomes of Bernoulli trials (success or failure, head or tail, spin up or down, etc.) can, for example, be represented by the two-component base spinors of (59).

6. Indistinguishable Particles: the Spin-statistics Relationship

So far we have considered probability distributions for single particles only. The generalisation to several particles looks straightforward but if the particles are indistinguishable there are nontrivial consequences. If two equal particles collide, for instance in proton-proton scattering, one must allow for two alternatives: Any registered outgoing particle may either be the incoming beam particle or, with equal probability, the target particle. Let us assume that there are two particles, labeled by 1 and 2, at center-of-mass positions $\boldsymbol{R}_a$ and $\boldsymbol{R}_b$, with intrinsic coordinates $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$, respectively. Without information about correlations between them the maximum entropy principle [18, 19] directs us to assign independent probability densities,

$$\rho(\boldsymbol{R}_a,\boldsymbol{r}_a,\boldsymbol{R}_b,\boldsymbol{r}_b) = \rho_1(\boldsymbol{R}_a,\boldsymbol{r}_a)\,\rho_2(\boldsymbol{R}_b,\boldsymbol{r}_b) = |\psi_1(\boldsymbol{R}_a,\boldsymbol{r}_a)\,\psi_2(\boldsymbol{R}_b,\boldsymbol{r}_b)|^2. \tag{61}$$

If the particles are not distinguishable, ψ_1 and ψ_2 must be the same function of the intrinsic polar coordinates but the orientations of the intrinsic coordinate frames may differ. Let the polar angles differ by θ and the azimuths by ϕ, as indicated in Figure 1 (a). We can therefore write the wave function as

$$\Psi(\boldsymbol{R}_a,\boldsymbol{r}_a,\boldsymbol{R}_b,\boldsymbol{r}_b) = \psi(\boldsymbol{R}_a,\boldsymbol{r}_a)\psi(\boldsymbol{R}_b,\boldsymbol{r}_b). \tag{62}$$

Interchange of the particle positions but not of the orientations results in

$$\Psi(\boldsymbol{R}_b,\boldsymbol{r}_a,\boldsymbol{R}_a,\boldsymbol{r}_b) = \psi(\boldsymbol{R}_b,\boldsymbol{r}_a)\psi(\boldsymbol{R}_a,\boldsymbol{r}_b), \tag{63}$$

with mixed-up orientations as indicated in Figure 1 (b). The orientations can be restored by the rotations indicated by arrows in Figs. 1 (b), (c), and (d): First make the polar axes parallel by letting $\theta_b \to \theta_b - \theta = \theta_a$, so that the wave function becomes, in obvious notation,

$$\Psi(\boldsymbol{R}_b,\vartheta_a,\varphi_a,\boldsymbol{R}_a,\vartheta_a,\varphi_b) = \psi(\boldsymbol{R}_b,\vartheta_a,\varphi_a)\psi(\boldsymbol{R}_a,\vartheta_a,\varphi_b). \tag{64}$$

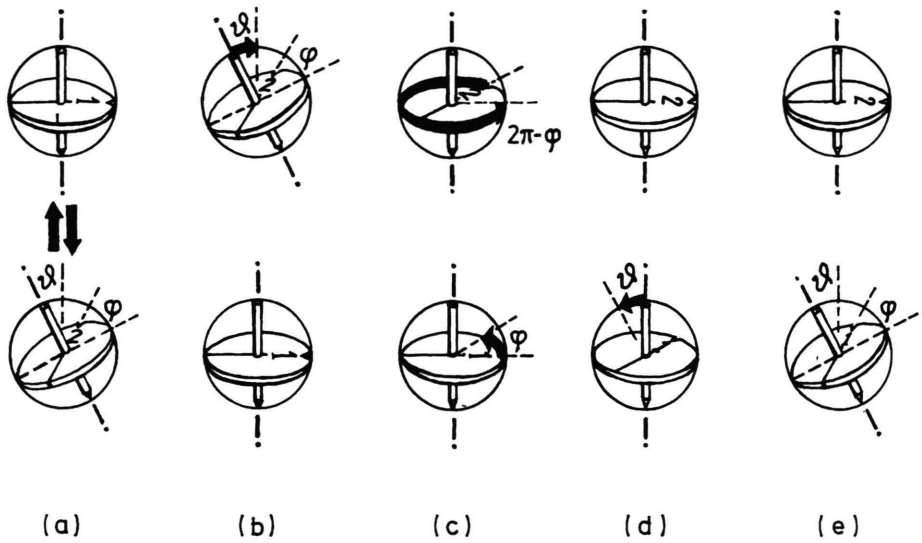

Fig. 1. Probability density distributions for two indistinguishable particles (schematic): (a) initial state, (b) state after interchange without rotations, (c) state with parallel polar axes, (d) state with correct azimuths, (e) final state, indistinguishable from (a).

Next rotate both wave packets about the polar axes to get the correct azimuths. With $\varphi_b \to \varphi_b + 2\pi - \varphi = \varphi_a + 2\pi$ and $\varphi_a \to \varphi_a + \varphi = \varphi_b$ one finds

$$\Psi(\boldsymbol{R}_b, \vartheta_a, \varphi_b, \boldsymbol{R}_a, \vartheta_a, \varphi_a + 2\pi) \tag{65}$$
$$= \psi(\boldsymbol{R}_b, \vartheta_a, \varphi_b)\psi(\boldsymbol{R}_a, \vartheta_a, \varphi_a + 2\pi).$$

Finally rotate the polar axis of the particle at $\boldsymbol{R}_b$ through the angle ϑ, which gives

$$\Psi(\boldsymbol{R}_b, \vartheta_b, \varphi_b, \boldsymbol{R}_a, \vartheta_a, \varphi_a + 2\pi) \tag{66}$$
$$= \psi(\boldsymbol{R}_b, \vartheta_b, \varphi_b)\psi(\boldsymbol{R}_a, \vartheta_a, \varphi_a + 2\pi)$$
$$= \pm\psi(\boldsymbol{R}_b, \boldsymbol{r}_b)\psi(\boldsymbol{R}_a, \boldsymbol{r}_a).$$

The absolute square is now again the same as initially, (61), and the configurations shown in Figs. 1 (a) and 1 (e) are indistinguishable. Evidently exchange of two indistinguishable particles and restoration of the orientations involves a full rotation of one of the particles around its intrinsic polar axis which changes the sign of the wave function if the spin is half-integer ((7), see also [20]). Adding the wave functions, or more correctly, the state vectors for the two alternatives "no exchange" and "exchange" one obtains the total state vector for two indistinguishable particles. In abbreviated (Dirac) notation one has, properly normalised,

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\Big[\psi(1,2)|1,2\rangle \pm \psi(2,1)|2,1\rangle\Big] \tag{67}$$
$$\text{for } \begin{cases} \text{bosons,} \\ \text{fermions.} \end{cases}$$

The base vectors $|1,2\rangle$ and $|2,1\rangle$ are a complete, orthonormal set, representing the two alternatives "no exchange" and "exchange", and could also be written as two-component base spinors (cf. (59), (60) and text following thereafter). The probability is just what one expects, $\langle\Psi|\Psi\rangle = (|\psi(1,2)|^2 + |\psi(2,1)|^2)/2$. We have thus obtained the spin-statistics relationship in its simplest form: Wave functions for systems of indistinguishable particles with *integer spin* must be *symmetric* in all particle coordinates, including spin coordinates, which entails Bose-Einstein statistics, whereas wave functions for particles with *half-integer spin* must be *antisymmetric*, which entails Fermi-Dirac statistics. The spin-statistics relationship is widely believed to be inexplicable without relativity and quantum field theory. Here it appears, however, as a nonrelativistic consequence of the two angular periodicities allowed by the Riesz-Fejér theorem for wave functions in ordinary space.

7. EPR Entanglement and Bell Inequalities

In the spin version of the famous Einstein-Podolsky-Rosen thought experiment [21] one considers a particle with spin zero that decays into two equal particles flying apart in opposite directions, each with spin 1/2. Because angular momentum is conserved, the spins of the two particles must be antiparallel, $\boldsymbol{\sigma}_1 = -\boldsymbol{\sigma}_2$. If one of the spin components of particle 1 is measured with a Stern-Gerlach magnet as pointing up, the same spin component of particle 2 is immediately known to be pointing down (which can be confirmed experimentally). This is a logical inference and has nothing to do with spooky superluminal action at a distance. More generally, one finds that the covariance of arbitrary spin coordinates $(\boldsymbol{a}\cdot\boldsymbol{\sigma}_1)$ and $(\boldsymbol{b}\cdot\boldsymbol{\sigma}_2)$ is given by

$$\langle(\boldsymbol{a}\cdot\boldsymbol{\sigma}_1)(\boldsymbol{\sigma}_2\cdot\boldsymbol{b})\rangle = -\boldsymbol{a}\cdot\boldsymbol{b} = -\cos(\boldsymbol{a},\boldsymbol{b}), \tag{68}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are unit vectors along two arbitrary analyzer directions. This result is obtained quantum-mechanically with the singlet state (total spin zero) described by the antisymmetric fermion wave function for the two discrete possibilities "spin up" and "spin down" of the two particles,

$$\boldsymbol{\Psi} = \frac{1}{\sqrt{2}}\left[\begin{pmatrix}1\\0\end{pmatrix}_1\begin{pmatrix}0\\1\end{pmatrix}_2 - \begin{pmatrix}0\\1\end{pmatrix}_1\begin{pmatrix}1\\0\end{pmatrix}_2\right], \tag{69}$$

and the spin coordinates by Pauli matrices (see [12]),

$$\boldsymbol{\sigma}_j = \{\sigma_{xj},\sigma_{yj},\sigma_{zj}\} \qquad (j=1,2) \tag{70}$$
$$= \left\{\begin{pmatrix}0&1\\1&0\end{pmatrix}_j, \begin{pmatrix}0&-i\\i&0\end{pmatrix}_j, \begin{pmatrix}1&0\\0&-1\end{pmatrix}_j\right\}.$$

The subscripts 1 and 2 refer to particles 1 and 2, and the subscripted matrix operators act only on column vectors (spinors) with the same subscript. The description is the same for any orientation of the coordinate frame. Expectation values are to be calculated as $\langle\ldots\rangle \equiv \boldsymbol{\Psi}^\dagger\ldots\boldsymbol{\Psi}$ which yields

$$\langle\boldsymbol{a}\cdot\boldsymbol{\sigma}_1\rangle = 0\,, \tag{71}$$

$$\langle(\boldsymbol{a}\cdot\boldsymbol{\sigma}_1)^2\rangle = 1\,, \tag{72}$$

$$\langle\sigma_{x1}^2\rangle = \langle\sigma_{y1}^2\rangle = \langle\sigma_{z1}^2\rangle = \langle\boldsymbol{\sigma}_1^2\rangle/3 = 1 \tag{73}$$

(similarly for $\boldsymbol{b}$ and $\boldsymbol{\sigma}_2$). So the spin projection on any unit vector $\boldsymbol{a}$ has expectation value zero with unit variance, while the expectation value of the squared spin angular momentum is $\langle\boldsymbol{\sigma}_1^2\rangle(\frac{\hbar}{2})^2 = \frac{1}{2}(\frac{1}{2}+1)\hbar^2$. The covariance of the two spin projections considered is found to be given by (68) as stated above.

It is often claimed that the covariance (68) cannot be obtained classically, or at least not together with spin quantisation. Its confirmation by experiment is then taken as evidence that the spin coordinates cannot exist simultaneously before a measurement reveals one of them, in accordance with N. Bohr's epistemological (Copenhagen) interpretation [22] of quantum mechanics but at variance with the ontological view of Einstein, Podolsky and Rosen [21]. Since, however, the quantum-mechanical result (68) does not contain Planck's constant one expects a classical derivation to be feasible. Let us therefore consider the spin $\boldsymbol{\sigma}_1 = -\boldsymbol{\sigma}_2$ as an ordinary vector for which all orientations are equally probable. Expectation values are then to be calculated classically as

$$\langle\ldots\rangle \equiv \int_0^\infty d\sigma_1\,\rho(\sigma_1)\int_{-1}^{+1}\frac{d(\cos\theta)}{2}\int_0^{2\pi}\frac{d\phi}{2\pi}\cdots\,, \tag{74}$$

where $\rho(\sigma_1)$ is the probability density of the length $\sigma_1 = |\boldsymbol{\sigma}_1|$ of both spin vectors and θ, ϕ are polar angle and azimuth of $\boldsymbol{\sigma}_1$. Without any difficulty one finds

$$\langle(\boldsymbol{a}\cdot\boldsymbol{\sigma}_1)(\boldsymbol{\sigma}_2\cdot\boldsymbol{b})\rangle = -\langle(\boldsymbol{a}\cdot\boldsymbol{\sigma}_1)(\boldsymbol{\sigma}_1\cdot\boldsymbol{b})\rangle = -\frac{\langle\boldsymbol{\sigma}_1^2\rangle}{3}(\boldsymbol{a}\cdot\boldsymbol{b}) \tag{75}$$

which, with $\langle\boldsymbol{\sigma}_1^2\rangle/3 = 1$ (cf. (73)), is equal to the quantum-mechanical result. Hence the correlation measurements alone do not rule out the ontological viewpoint, i. e. reality of unobserved spin components. This conclusion is not changed if we also take spin quantisation into account. It, too, follows already classically from the 2π periodicity of rotations as we have seen. Therefore the reality of unobserved spin coordinates need not be questioned. A temptation to introduce *hidden variables* exists only if one tries to treat the spin eigenvalues, $+1/2$ or $-1/2$, measured along $\boldsymbol{a}$ and $\boldsymbol{b}$, as if they were true particle spins rather than global properties of the wave packet encoding incomplete information about them, and if one considers $\boldsymbol{\sigma}_1 = -\boldsymbol{\sigma}_2$ as true not for each particle pair but only on average. The inequalities derived by Bell [23] from these premises are, in fact, contradicted by

experiment. Hidden variables are not needed for an understanding of the correlations if one distinguishes clearly between the particles themselves and information about them – the natural variables of the problem, spin coordinates, are enough.

The common misunderstanding to interpret measurable expectation values as true values is fostered mainly by the unfortunate and misleading use of the word *"state"* in quantum mechanics for what is actually *"information about the state"* of a physical system, in particular about its preparation, and also by futile attempts to endow probability amplitudes with physical properties while paying lip service to Born's probability interpretation. The ingenious and elaborate experimental checks on increasingly complicated correlations between particle spins or photon polarisations (cf. e. g. [24]) look often more like attempts to check the Riesz-Fejér theorem than like investigations of the physics. One is reminded of experiments proposed earlier in this century to check other mathematical consequences of Fourier theory, for example the existence of side bands in amplitude-modulated beams of optical or radio waves (see [25]). For further clarification let us see how the basic probabilistic concepts apply in quantum mechanics.

8. The Basic Rules of Probability Theory and Quantum Mechanics

Quantum mechanical "operator-valued" probabilities are often said to differ fundamentally from "ordinary" ones, in disregard of the work of Cox [26] who proved that any scheme of inductive inference, i. e. of reasoning in the face of uncertainty, must be either equivalent to ordinary probability theory or inconsistent, with probabilities understood as encoding incomplete information on a numerical scale of plausibility or rational expectation, in the tradition of Bernoulli and Laplace (and Heisenberg [27]). Cox proved this by demonstrating that for any formal system of inference obeying the rules of ordinary Aristotelian logic the most general consistency conditions can be cast in the form of two functional equations whose solutions are the basic sum and product rules

$$P(A|C) + P(\bar{A}|C) = 1, \tag{76}$$

$$\begin{aligned} P(AB|C) &= P(A|BC)P(B|C), \\ &= P(B|AC)P(A|C), \end{aligned} \tag{77}$$

from which probability theory unfolds. (A and B represent propositions such as "the coin shows head" or "the neutrino rest energy is between 0 and 25 eV", AB means "both A and B are true", $\bar{A}$ means "A is false", and $P(A|C)$ denotes the probability of A given C. Our notation indicates that all probability assignments are conditional, based either on empirical or theoretical information or on assumptions C. The two forms of the product rule reflect the symmetry $AB = BA$.) It is interesting that Schrödinger [28] arrived independently at similar conclusions. Criticism that Cox had assumed differentiability of his probability functions was met by Rényi [29] who gave a proof without this assumption. It appears that any claim of an essential difference between ordinary and quantum probabilities must overcome the obstacle of Cox's proof.

An immediate consequence of the two forms of the product rule (77) is Bayes' theorem in its simplest form,

$$P(A|BC) = \frac{P(B|AC)P(A|C)}{P(B|C)}. \tag{78}$$

This theorem is fundamental to scientific reasoning. It provides nothing less than a formal model for updating of information with new evidence, or *learning from observations*. Suppose we are interested in some hypothesis A (for instance about the value of a half-life) to which one can assign an initial probability ("prior") $P(A|C)$ (from nuclear systematics or previous half-life measurements, with C specifying isotope and decay type). Suppose further that we receive new data B (counts), and that we also have a theoretical model of the experiment (involving the exponential decay law, counting statistics, and experimental details such as geometry, source specifications and counter efficiency) from which we can calculate, for arbitrary half-life A, the "likelihood" $P(B|AC)$ of observing the data B. The updated probability ("posterior") is essentially proportional to the product of likelihood and prior, $P(B|AC)P(A|C)$, the denominator in (78) acting merely as a normalisation constant. Updating can be repeated whenever new data become available, the old posterior becoming the new prior in each step. It should be understood that the historical terms "prior" and "posterior" have a logical rather than a temporal meaning. They simply mean "without" and "with" the new data taken into account. It should also be understood that *probabilities are not relative frequencies* although frequency estimates can be derived from them (see e. g. [7, 18]).

In order to see how these rules apply in quantum mechanics, let us look at a system described by spatial wave amplitudes $\psi_{\boldsymbol{r}}$ or, alternatively, by momentum wave amplitudes $\phi_{\boldsymbol{k}}$, so that the state vector is (in Dirac notation)

$$|\Psi\rangle = \sum_{\boldsymbol{r}} \psi_{\boldsymbol{r}} |\boldsymbol{r}\rangle = \sum_{\boldsymbol{k}} \phi_{\boldsymbol{k}} |\boldsymbol{k}\rangle, \tag{79}$$

where $|\boldsymbol{r}\rangle$ and $|\boldsymbol{k}\rangle$ are base vectors in position and momentum space, representing possible alternatives that are mutually exclusive ($\langle \boldsymbol{r}|\boldsymbol{r}'\rangle = \delta_{\boldsymbol{r}\boldsymbol{r}'}$, $\langle \boldsymbol{k}|\boldsymbol{k}'\rangle = \delta_{\boldsymbol{k}\boldsymbol{k}'}$) and complete ($\sum_{\boldsymbol{r}} |\boldsymbol{r}\rangle\langle \boldsymbol{r}| = \mathbb{1}$). The sum over all probabilities,

$$\langle \Psi|\Psi\rangle = \sum_{\boldsymbol{r}} |\psi_{\boldsymbol{r}}|^2 = \sum_{\boldsymbol{k}} |\phi_{\boldsymbol{k}}|^2 = 1, \tag{80}$$

is consistent with the sum rule (76) as long as we work either with the $\boldsymbol{r}$ or the $\boldsymbol{k}$ description. Intercalation of the completeness relations, $\sum_{\boldsymbol{r}} |\boldsymbol{r}\rangle\langle \boldsymbol{r}| = 1$ and $\sum_{\boldsymbol{k}} |\boldsymbol{k}\rangle\langle \boldsymbol{k}| = 1$, produces a mixed form of the normalisation condition,

$$\begin{aligned} \sum_{\boldsymbol{r}} \sum_{\boldsymbol{k}} \langle \Psi|\boldsymbol{r}\rangle\langle \boldsymbol{r}|\boldsymbol{k}\rangle\langle \boldsymbol{k}|\Psi\rangle &= \sum_{\boldsymbol{r}} \sum_{\boldsymbol{k}} \mathrm{Re}\big(\psi_{\boldsymbol{r}}^* \langle \boldsymbol{r}|\boldsymbol{k}\rangle \phi_{\boldsymbol{k}}\big) \\ &\equiv \sum_{\boldsymbol{r}} \sum_{\boldsymbol{k}} \tilde{P}(\boldsymbol{r}, \boldsymbol{k}|\Psi) = 1. \end{aligned} \tag{81}$$

The real summand $\tilde{P}(\boldsymbol{r}, \boldsymbol{k}|\Psi)$ looks like the joint probability distribution of $\boldsymbol{r}$ and $\boldsymbol{k}$ given Ψ in so far as it readily yields the correct marginal distributions $P(\boldsymbol{r}|\Psi) = |\psi_{\boldsymbol{r}}|^2$ and $P(\boldsymbol{k}|\Psi) = |\phi_{\boldsymbol{k}}|^2$ if summed over $\boldsymbol{k}$ or $\boldsymbol{r}$, and also correct expectation values for observables. It is not a true probability distribution, however, as it can assume negative values (Margenau and Hill [30]), like other expressions that have been proposed as joint distributions (e. g. by Wigner [31]).

In fact, there is no room for a joint distribution. The wave functions $\psi_{\boldsymbol{r}}$ and $\phi_{\boldsymbol{k}}$ can be considered as coordinates of the same unit vector $|\Psi\rangle$ in two coordinate frames that differ by a "rotation" described by a unitary matrix having elements $\langle \boldsymbol{r}|\boldsymbol{k}\rangle$,

$$\psi_{\boldsymbol{r}} = \langle \boldsymbol{r}|\Psi\rangle = \sum_{\boldsymbol{k}} \langle \boldsymbol{r}|\boldsymbol{k}\rangle \phi_{\boldsymbol{k}}, \tag{82}$$

$$\phi_{\boldsymbol{k}} = \langle \boldsymbol{k}|\Psi\rangle = \sum_{\boldsymbol{r}} \langle \boldsymbol{k}|\boldsymbol{r}\rangle \psi_{\boldsymbol{r}}. \tag{83}$$

(These are, of course, the Fourier transforms (12) and (13) in Dirac notation). The wave amplitudes for $\boldsymbol{k}$ are therefore completely determined by those for $\boldsymbol{r}$, given the elements $\langle \boldsymbol{r}|\boldsymbol{k}\rangle$ of the unitary matrix of Fourier transformations. Taking the absolute square of the spatial probability amplitude (82) one finds with (83)

$$|\psi_{\boldsymbol{r}}|^2 = \sum_{\boldsymbol{k},\boldsymbol{k}'} \langle \Psi|\boldsymbol{k}\rangle\langle \boldsymbol{k}|\boldsymbol{r}\rangle\langle \boldsymbol{r}|\boldsymbol{k}'\rangle\langle \boldsymbol{k}'|\Psi\rangle \tag{84}$$

$$= \sum_{\boldsymbol{k}} |\langle \boldsymbol{r}|\boldsymbol{k}\rangle|^2 |\phi_{\boldsymbol{k}}|^2 + 2\mathrm{Re} \sum_{\boldsymbol{k}} \sum_{\boldsymbol{k}'<\boldsymbol{k}} \phi_{\boldsymbol{k}}^* \langle \boldsymbol{k}|\boldsymbol{r}\rangle\langle \boldsymbol{r}|\boldsymbol{k}'\rangle \phi_{\boldsymbol{k}'}.$$

Summation over all alternatives $\boldsymbol{r}$ produces $\sum_{\boldsymbol{r}} |\psi_{\boldsymbol{r}}|^2 = \sum_{\boldsymbol{k}} |\phi_{\boldsymbol{k}}|^2$, the cross terms with $\boldsymbol{k} \neq \boldsymbol{k}'$ cancelling each other. These interference terms are commonly considered as a hallmark of quantum mechanics, not present in classical probability theory. Our equations are, however, purely classical as $\boldsymbol{k}$ has not been replaced yet by $\boldsymbol{p} = \hbar\boldsymbol{k}$. Furthermore, (84) is completely analogous to the expression $\boldsymbol{x}'^{\dagger}\boldsymbol{x}' = \boldsymbol{x}^{\dagger}\boldsymbol{O}^{\dagger}\boldsymbol{O}\boldsymbol{x}$ that is valid for a real vector and a rotation described by an orthogonal matrix $\boldsymbol{O}$, and contains analogous cross terms.

The paradigmatic example for interfering probability waves is the particle version of Young's famous double slit experiment, with a parallel beam of monoenergetic particles which can be described by a state vector $|\Psi\rangle = |\boldsymbol{k}\rangle$ or the corresponding wave function in $\boldsymbol{r}$-representation

$$\psi_{\boldsymbol{r}} = \langle \boldsymbol{r}|\boldsymbol{k}\rangle \propto e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{85}$$

Consider a plane S normal to the beam. The probability for a beam particle to reach a final position $\boldsymbol{r}'$ from any one of the many possible positions $\boldsymbol{r}$ in the plane assumes the form

$$\big|\langle \boldsymbol{k}|\boldsymbol{r}'\rangle\big|^2 = \Big|\sum_{\boldsymbol{r}\in S} \langle \boldsymbol{k}|\boldsymbol{r}\rangle\langle \boldsymbol{r}|\boldsymbol{r}'\rangle\Big|^2 \tag{86}$$

if we intercalate the completeness relation for the $\boldsymbol{r}$. One recognises (86) as a formal expression of Huygens' principle: Each point on a wave front can be considered as the centre of an outgoing spherical wave, with

$$\langle \boldsymbol{r}|\boldsymbol{r}'\rangle \propto \frac{e^{ik|\boldsymbol{r}'-\boldsymbol{r}|}}{|\boldsymbol{r}'-\boldsymbol{r}|} \tag{87}$$

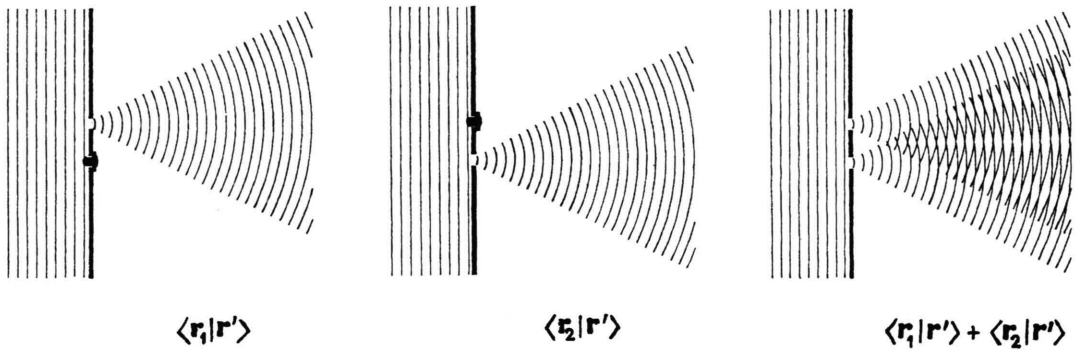

Fig. 2. Young double-slit interference experiment: Single- and double-slit wave functions (schematical).

as the transition amplitude for the transfer from $\boldsymbol{r}$ to $\boldsymbol{r}'$, and all these elementary waves are superposed.

Next let us assume that the mathematical plane S is the place of an absorbing screen with slits at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, so that there are only two possibilities for particles to pass the screen, either through slit 1 or slit 2. These two *mutually exclusive* alternatives are formally represented by a complete set of orthonormal base kets $|\boldsymbol{r}_1\rangle$ and $|\boldsymbol{r}_2\rangle$. Instead of (86) one has then

$$\left|\langle \boldsymbol{k}|\boldsymbol{r}'\rangle\right|^2 = \Big|\sum_{s=1}^{2}\langle \boldsymbol{k}|\boldsymbol{r}_s\rangle\langle \boldsymbol{r}_s|\boldsymbol{r}'\rangle\Big|^2, \tag{88}$$

with elementary waves emerging only at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, and interfering in the overlap zone (see Figure 2). We stress again that there is no question of particles passing both slits simultaneously. We are definitely dealing with mutually exclusive possibilities but in the framework of the Fourier-Riesz-Fejér theory of probability wave amplitudes.

If only the slit at $\boldsymbol{r}_1$ is open the sum in (86) reduces to a single term and we get the expected result, without interference,

$$\left|\langle \boldsymbol{k}|\boldsymbol{r}'\rangle\right|^2 = \left|\langle \boldsymbol{k}|\boldsymbol{r}_1\rangle\right|^2\left|\langle \boldsymbol{r}_1|\boldsymbol{r}'\rangle\right|^2, \tag{89}$$

and analogously if the other slit is open instead.

There could be yet another state of information: It might be known that only one slit is open but not which one. According to the Principle of Insufficient Reason one must then assign equal probabilities of 1/2 to both alternatives and average the two one-slit probabilities accordingly,

$$\left|\langle \boldsymbol{k}|\boldsymbol{r}'\rangle\right|^2 = \frac{1}{2}\sum_{s=1}^{2}\left|\langle \boldsymbol{k}|\boldsymbol{r}_s\rangle\right|^2\left|\langle \boldsymbol{r}_s|\boldsymbol{r}'\rangle\right|^2. \tag{90}$$

This result, a so-called statistical mixture, is also valid if one knows that the two slits are opened and closed alternately, either periodically or at random (as in so-called delayed-choice experiments) but in such a way that the total opening times of both are equal.

Although it should be no surprise that different states of knowledge lead to different probability assignments, as always in probability theory, long discussions have evolved about the presence of interference if both slits are open, and their absence if one of them is closed, or about the fact that, if both slits are open, one must sum over probability amplitudes after the pattern

$$\psi_{ac} = \sum_b \psi_{ab}\psi_{bc}, \tag{91}$$

whereas in the case of one unknown slit one must sum over their absolute squares, i. e. probabilities,

$$P_{ac} = \sum_b P_{ab}P_{bc}. \tag{92}$$

The differences were related to the mysterious "collapse of the wave function" or "reduction of the wave packet" that seems to occur whenever a measurement reveals which of several possibilities is the true one. The one-slit situation was considered as a kind of measurement furnishing "which-way" information, in contrast to the two-slit situation. Bayes' theorem is rarely mentioned in these discussions, although its importance can hardly be overstated. With proper understanding of Bayesian updating there is nothing strange about a change of the wave function when additional data or new information are included in the formal process of logical inference. Their incorporation by means of Bayes' theorem inevitably changes all prior probabilities to posterior ones, as common

sense and probability theory demand. As soon as we learn, for instance, that in the spin version of the EPR experiment one spin is observed as pointing up, we can have no doubt any more that the other spin is pointing down. Similarly with Young's double-slit experiment: As soon as we learn that only one slit is open, the (interfering) double-slit wave function must be replaced by the appropriate (noninterfering) one-slit function or, if it is not known which slit is open, by a "statistical mixture" of the two possible one-slit wave functions. Since this is not a physical but a *logical* change, questions about its sudden (superluminal) occurrence throughout physical space do not arise. It was Heisenberg [32] himself who wrote that the reduction of the wave packet is caused not by "a physical, but rather, so to say, a mathematical process. With the sudden change of our knowledge also the mathematical presentation of our knowledge undergoes of course a sudden change." It is quite possible to reason even backwards in time if the new evidence is relevant to the past. From this standpoint there does not seem to be much need for a special measurement theory as expounded for instance by Omnès [33].

We conclude that there is no fundamental difference between classical and quantum-mechanical probabilities. Therefore the basic rules of classical probability theory are valid also for probability waves and thus for quantum mechanics. The sum rule (76) assumes the form

$$P(A|C) + P(\bar{A}|C) = \left|\psi(A|C)|A\rangle + \psi(\bar{A}|C)|\bar{A}\rangle\right|^2 = 1, \tag{93}$$

where $|A\rangle$ and $|\bar{A}\rangle$ constitute a complete orthonormal state vector base representing the facts that A and $\bar{A}$ are mutually exclusive alternatives and that they are the only ones. Whether one sums probabilities or state vectors, the final result is the same. This justifies what we did when we added state vectors for the alternatives "no exchange" and "exchange" of two indistinguishable particles to find the total wave function (67). Although superposition of A and $\bar{A}$ – of Schrödinger's cat [34] simultaneously alive and dead – does occur at the level of the auxiliary probability wave amplitudes (Fourier components) introduced by way of the Riesz-Fejér theorem, the final result contains only the wave intensities, i. e. conventional probabilities.

What about the product rule? Suppose we open the box enclosing the cat and Schrödinger's deadly contraption, measure the cat's heart beat and find it normal. Knowing the conditional probabilities $P(B|A) = 1$ and $P(B|\bar{A}) = 0$, where A stands for "alive" and B for "beat noticeable", we can apply Bayes' theorem and infer that the cat is alive: $P(\bar{A}|BC) \propto P(B|\bar{A})P(\bar{A}|C) = 0$, hence $P(A|BC) = 1$ (where C stands for "contraption"). Bayes' theorem and thus the basic multiplication rules (77) are found to work in quantum mechanics as well, with

$$\begin{aligned} P(AB|C) &= P(B|AC)P(A|C) \\ &= |\psi(B|AC)\chi(A|C)|^2. \end{aligned} \tag{94}$$

The special case of independent propositions A and B, where $P(A|BC) = P(A|C)$ and $\psi(A|BC) = \psi(A|C)$, was already utilised when we dealt with independent internal and centre-of-mass coordinates in (49), or with independent coordinates of two particles in (62). Often $P(A|C)$ is a conventional prior while $P(B|AC)$ is a wave-mechanical transition probability, as in "statistical mixtures" (see (90)).

The final conclusion is

(1) traditional probability theory can be extended by means of the Riesz-Fejér superposition theorem, without change of the basic sum and product rules from which it unfolds, hence *without violation of Cox's consistency conditions*;

(2) the resulting *probability wave theory* turns out to be essentially the formalism of quantum mechanics inferred by physicists with great effort from the observation of atomic phenomena.

From the Bayesian point of view the nature and interpretation of probabilities in traditional probability theory, probability wave theory and quantum mechanics need not be considered as different. The mathematics of probability waves is more powerful, however, comprising superposition and interference with all algebraic consequences, in particular operator calculus and eigenvalue equations.

9. Summary

The formalism of quantum mechanics, in the traditional axiomatic or historical presentation, looks mysterious. It emerges rather naturally, however, if one treats position and momentum uncertainties for classical point particles wave-mechanically, by means of the Riesz-Fejér superposition theorem, which by the way dispels any doubts about the *linearity* of

the theory. The theorem permits unrestricted use of Fourier series – the proper tool for dealing with temporal and spatial constraints – in a way that guarantees non-negativity of all probabilities. All the basic features of quantum mechanics are obtained readily – wave-particle duality, operator calculus and commutation rules, uncertainty relations, Schrödinger equation, periodicity-related quantisation of angular momenta and other physical quantities, etc. Moreover, elementary requirements for possible forms of the arbitrary phases of probability amplitudes lead unequivocally to the correct form of the electromagnetic interaction. The spin-statistics relationship for indistinguishable particles is a logical consequence of the natural periodicity of spatial rotations. Planck's quantum of action appears automatically, as a "blurring" parameter, as soon as the kinematics of probability wave packets is equated with Hamiltonian particle kinematics. *Nonlocality* (instantaneous collapse of the wave function throughout physical space if new information is taken into account) follows from strict adherence to Born's interpretation of $|\psi|^2$ as a probability density in combination with the Bayesian scheme for the updating of knowledge. There is no reason to doubt that physical quantities, such as the spin coordinates in the spin version of the Einstein-Podolsky-Rosen experiment, have a *reality independent of the observer*, in obvious contrast to eigenfunction expansions and eigenvalues that reflect *his* choice of measurement and thus of *his* preferred reference frame. From this viewpoint *quantum mechanics looks much like an error propagation (or rather information transmittal) formalism for uncertainty-afflicted physical systems that obey the classical equations of motion.* Difficulties already present in Hamiltonian mechanics, for instance with the infinite electromagnetic self-energy of charged point particles, must then also be expected in quantum theory.

The formalism is *holistic*, taking into account all probability waves fitting into a given experimental setup, and thereby the ensemble of all *possible* trajectories, as becomes especially clear in R. Feynman's path integral formulation [35]. For a given path, $\boldsymbol{r} = \boldsymbol{r}(t)$, the momentum $\boldsymbol{p}$ is not independent of $\boldsymbol{r}$ but related by *differentiation* along the path, in contrast to statistical mechanics where joint distributions are postulated for $\boldsymbol{r}$ and $\boldsymbol{p}$ without questioning mutual compatibility, and then extrapolated in time by *integration*. The two interfering cylindrical waves emerging with equal phases from the two slits in the particle version of *Young's double-slit experiment* do not indicate that the electron can pass both slits simultaneously but only that two slits are open for it and we do not know through which one it will go. Comparison with experiment requires that observables such as relative frequencies are estimated from probabilities [7, 18]. The statistical samples, for instance the number of registered particles, are so large in typical diffraction experiments that relative frequencies and probabilities hardly differ numerically. In this sense an observed Young diffraction pattern shows the probabilities rather directly (see the results of Möllenstedt and Jönsson [36] for an early realisation). In other cases, especially in high-energy physics, the number of observed events may be quite small (as small as one) but probability theory and thus quantum mechanics remain applicable, the only difference being that estimated root-mean-square errors become larger, hence predictions less certain. *Quantum mechanics can thus be understood as a powerful extension of ordinary probability theory*, particularly well suited for dealing with ensembles of particle trajectories or chains of space-time events fitting into given experimental configurations.

Quantum mechanics treats positions and momenta in symmetric fashion. Our exposition emphasised probabilities for particles and deduced wave-like behavior for spatially extended wave packets (states of information). One could, it seems, equally well consider light waves and deduce the particle-like behavior of photons. The formal symmetry may be misleading, however. The location aspect appears to be more natural for massive particles for which the generalised momentum (wave length) is not gauge invariant as we saw. For the massless photons it is just the particle aspect (well defined position) that is problematic while for radio waves or radar pulses the wave aspect seems natural. A related question is whether it is necessary to quantise also the electromagnetic field. The electromagnetic field intensities are positive definite quantities like the probability densities for electrons, so the Riesz-Fejér theorem is applicable. Actually, Fourier techniques and the superposition principle were applied routinely to electromagnetic field strengths and potentials long before quantum mechanics appeared on the scene. Such problems belong to quantum electrodynamics and will not be further discussed here. As far as ordinary quantum mechanics is concerned it should have become clear that it can be demystified to quite some extent with the Riesz-Fejér theorem.

The question seems not so much " How can it be like that?" but rather "Could it be otherwise?"

What remains mysterious is the *irreducible uncertainty and lack of control* caused by the empirical finite and universal value of Planck's quantum of action. That its value is the same for electrons, nucleons, photons etc. is not surprising since their mutual interactions must conserve energy and momentum. In fact, Jaynes [37] found that action is conserved, too, as a consequence of probability conservation, i. e. that there is an integral of motion which can be identified with $\hbar$, if a spinless hydrogen-like atom is coupled to a (classical) electromagnetic field in a cavity. The role of $\hbar$ as a limit to the attainable information and control in microphysics has been clear ever since Heisenberg [27] discussed his uncertainty relations: Particle trajectories and orbits are always affected by a non-removable minimum blur. As finite particle size would produce a similar blur, one is tempted to ask if quantum mechanics can perhaps be viewed as a kind of minimum information (maximum entropy) generalisation of probabilistic Hamiltonian mechanics *from mass points to particles with finite extension* (spatial distribution) *and internal motion* (momentum distribution, spin). How this conjecture fits in with others, such as solitons, zitterbewegung, or superstrings, remains to be seen.

Appendix: Proof of the Riesz-Fejér Theorem

The proof presented by Fejér [2] as due to Riesz, and also contained in the book by Riesz and Sz.-Nagy [3], is given here in slightly different notation. Consider the real Fourier polynomial

$$\rho(x) = \rho(x)^* = \sum_{l=-n}^{n} c_l\, e^{ilx}, \quad (c_l = c_{-l}^*). \tag{A1}$$

Defining the polynomial $g(z)$ as

$$\begin{aligned} g(z) \equiv\ & c_n^* + \ldots + c_1^* z^{n-1} + c_0 z^n \\ & + c_1 z^{n+1} + \ldots + c_n z^{2n}, \end{aligned} \tag{A2}$$

one can write $\rho(x) = e^{-nix} g(e^{ix})$ or, as the polynomial $\rho(x)$ is non-negative,

$$\rho(x) = |g(e^{ix})|. \tag{A3}$$

The polynomial $g(z)$ is of degree $2n$ if $c_n \neq 0$, so that $g(0) \neq 0$. If z_k is a solution of $g(z) = 0$,

$$g(z_k) = c_n^* + \ldots + c_n z_k^{2n} = 0\,, \tag{A4}$$

then $1/z_k^*$ is another solution,

$$g(\frac{1}{z_k^*}) = [(c_n^* + \ldots + c_n z_k^{2n}) z_k^{-2n}]^* = 0\,. \tag{A5}$$

One concludes that each root z_k inside the unit circle is accompanied by another root $1/z_k^*$ outside, with equal multiplicities of the roots inside and the accompanying ones outside. (Remember that $z_k = 0$ can be excluded.) Equation (A3) shows that there are no solutions on the unit circle if the polynomial is definitely positive – which we may assume without loss of generality, as we can always add a small positive quantity ϵ and let it vanish eventually. Thus one has

$$g(z) = c_n \prod_{k=1}^{n} (z - z_k)(z - \frac{1}{z_k^*}) \tag{A6}$$

(where not all the z_k are different if there are multiple roots). For $z = e^{ix}$ one obtains

$$\rho(x) = |g(e^{ix})| = \left| \sqrt{c_n} \prod_{k=1}^{n} \frac{e^{ix} - z_k}{\sqrt{z_k}} \right|^2, \tag{A7}$$

which is the absolute square of a Fourier polynomial of the same order as $\rho(x)$, so that one can write

$$\rho(x) = \sum_{l=-n}^{n} c_l e^{ilx} = |\psi(x)|^2\,, \quad -\pi < x \leq \pi, \tag{A8}$$

$$\begin{aligned} \psi(x) &= e^{i\alpha} \sqrt{\left| \frac{c_n}{z_1 \ldots z_n} \right|} \prod_{k=1}^{n} (e^{ix} - z_k) \\ &= e^{i\alpha} \sum_{k=0}^{n} a_k\, e^{ikx} \ (\alpha \text{ arbitrary}). \end{aligned} \tag{A9}$$

This completes the (constructive) proof that each non-negative real Fourier polynomial can be written as the absolute square of a complex one of (at most) the same order (same highest harmonic). The complex Fourier polynomial is mathematically more convenient and more flexible because it is not subject to the non-negativity requirement and contains an arbitrary phase.

Acknowledgements

It is a pleasure to thank R. Balian, R. E. Collins, O. Costa de Beauregard, B. Roy Frieden, and R. N. Silver for discussions and challenging criticism, and G. Kessler and the Karlsruhe Nuclear Fusion Project for support.

[1] P. A. M. Dirac, in: The Past Decade in Particle Theory, (E. C. G. Sudarshan and Y. Ne'eman, Eds.), Gordon & Breach, London 1973, p. 747.
[2] L. Fejér, J. reine u. angew. Math. **146**, 53 (1915); reprinted in Gesammelte Werke, Vol. I, Birkhäuser Verlag, Basel 1970, p. 842.
[3] F. Riesz and B. Sz.-Nagy, Vorlesungen über Funktionalanalysis, VEB Verlag der Wissenschaften, Berlin 1973, pp. 108-109.
[4] C. F. von Weizsäcker, in Quantum Theory and the Structures of Time and Space (L. Castell et al., Eds.), Carl Hanser, München 1975, p. 220.
[5] W. Feller, An Introduction to Probability Theory and its Applications, Vol. II, Wiley & Sons, New York 1966.
[6] L. Cohen, Time-Frequency Analysis, Prentice Hall, Eaglewood Cliffs 1995.
[7] F. H. Fröhner, Nucl. Sci. Eng. **126**, 1 (1997).
[8] Kuo-Ho Yang, Ann. Phys. New York **101**, 62 (1976).
[9] E. C. G. Stueckelberg, Helv. Phys. Acta **14**, 588 (1941), **15**, 23 (1942).
[10] O. Costa de Beauregard, Time, the Physical Magnitude, Reidel, Dordrecht, 1983; cf. also Phys. Rev. Lett. **50**, 867 (1983); Proc. Int. Symp. Foundations of Quantum Mechanics, Tokyo, 1983, pp. 233-241.
[11] D. H. Kobe, Amer. J. Phys. **46**, 342 (1978).
[12] P. A. M. Dirac, Principles of Quantum Mechanics, 3rd ed., Clarendon Press, Oxford 1947.
[13] H. Weyl, Ann. Phys. **59**, 101 (1919).
[14] D. J. Griffiths, Introduction to Elementary Particles, J. Wiley & Sons, New York 1987.
[15] T. Kugo, Eichtheorie, Springer-Verlag, Berlin 1997.
[16] F. Hund, Materie als Feld, Chap. 7, Springer-Verlag, Berlin 1954.
[17] É. Cartan, Leçons sur la théorie des spineurs, Herman, Paris 1937; English translation: The Theory of Spinors, Dover, New York 1981.
[18] E. T. Jaynes, Papers on Probability, Statistics and Statistical Physics (R. D. Rosenkrantz, Ed.), Reidel, Dordrecht 1983; 2nd paperback edition: Kluwer, Dordrecht 1989.
[19] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957); reprinted in [18].
[20] R. P. Feynman, in Elementary Particles and the Laws of Physics: the 1986 Dirac Memorial Lectures, (R. MacKenzie and P. Doust, Eds.), pp. 1-59, Cambridge U. Press, Cambridge 1987.
[21] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[22] N. Bohr, Phys. Rev. **48**, 696 (1935).
[23] J. S. Bell, Physics **1**, 195 (1964).
[24] A. Aspect and P. Grangier, in: Advances in Quantum Phenomena, (E.G. Beltrametti and J.-M. Lévy-Leblond, Eds.) pp. 201-213, Plenum Press, New York 1995.
[25] E. T. Jaynes, in: Maximum Entropy and Bayesian Methods, (W.T. Grandy and L.H. Schick, Eds.), pp. 1-13, Kluwer, Dordrecht 1991.
[26] R. T. Cox, Amer. J. Physics **14**, 1 (1946).
[27] W. Heisenberg, Die physikalischen Prinzipien der Quantentheorie, Hirzel, Leipzig 1930; English: The Physical Principles of Quantum Theory, Univ. Press, Chicago 1930.
[28] E. Schrödinger, Proc. Roy. Irish Acad. **51 A**, 51 (1947); **51 A**, 141 (1947); reprinted in Collected Papers, vol. 1, pp. 463-478, 479-484, Austrian Acad. Sci., Vienna 1984.
[29] L. Rényi, Valószinüségszámítás (= Probability Calculus), Budapest 1954. For a slightly generalised presentation see J. Aczél, Vorlesungen über Funktionalgleichungen und ihre Anwendungen, Birkhäuser Verlag, Basel 1961, Section 7.1.4.; English translation: Lectures on Functional Equations and Their Applications, Academic Press, New York 1964.
[30] H. Margenau and R. N. Hill, Prog. Theor. Phys. **26**, 722 (1961).
[31] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[32] W. Heisenberg (translated from a letter to Renninger, dated February 2, 1960, in: M. Jammer, The Philosophy of Quantum Mechanics, Wiley, New York 1974)
[33] R. Omnès, The Interpretation of Quantum Mechanics, Princeton U. Press, Princeton 1994.
[34] E. Schrödinger, Naturwiss. **23**, 807 (1935); reprinted in Collected Papers, vol. 3, pp. 484-489, Austrian Acad. Sci., Vienna 1984.
[35] R. P. Feynman and A. R. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill, New York 1965.
[36] G. Möllenstedt and C. Jönsson, Z. Physik **155**, 472 (1959).
[37] E. T. Jaynes, in Coherence and Quantum Optics, (L. Mandel and E. Wolf, Eds.), Plenum Press, New York 1973, pp. 35-81.