PHYSICS REPORTS (Review Section of Physics Letters) 133, No. 6 (1986) 337–401, North-Holland, Amsterdam

A REVIEW OF EXTENDED PROBABILITIES

W. MÜCKENHEIM

Südring 10, D-3400 Göttingen, Fed. Rep. Germany

Including notes by:

G. LUDWIG

Fachbereich Physik der Philipps-Universität, D-3550 Marburg, Renthof 7, Fed. Rep. Germany

C. DEWDNEY, P.R. HOLLAND, A. KYPRIANIDIS and J.P. VIGIER

Institut Henri Poincaré, Laboratoire de Physique Théorique, 11, Rue P. et M. Curie, 75231 Paris Cedex 05, France

N. CUFARO PETRONI

Dipartimento di Fisica dell'Università and Istituto Nazionale di Fisica Nucleare, Via G. Amendola 173, 70126 Bari, Italy

M.S. BARTLETT

Priory Orchard, Priory Avenue, Totnes, Devon TQ9 5HR, U.K.

and

E.T. JAYNES

Department of Physics, Washington University, St. Louis MO 63130, U.S.A.

Received 1 August 1985

Contents:

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 133, No. 6 (1986) 337–401.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 43.00, postage included.

Abstract:
Some results emerging from the formalism of quantum theory seem to indicate probabilities outside of the conventional range between 0 and 1. This article draws attention to arguments in favour of extended probabilities, reviews some approaches to a formal account and interpretation, and presents a collection of statements of distinguished scientists about this strange topic.

1. Introduction

Reading of old papers sometimes opens unexpected views. Although it is well known, for instance, that P.A.M. Dirac was the first to introduce explicitly the concept of negative energy, the number of those who know his remarks about negative probability – closely related to negative energy and invented simultaneously – seems to be very restricted. In one of his papers on this subject [1] for example, "probability" appears in connexion with the adjective "negative" not less than 20 times, not considering the appearance in numerous equations. This concept is used with reservations but, as it seems, not without a certain kind of sympathy. Said paper – not the only one on this topic, as will become clear from the following – meanwhile has been forgotten, at least as far as negative probability is concerned. It shares the fate of many ideas which have been rejected by the majority of their contemporaries. Nevertheless, as history shows, in some cases revision of a verdict happened to prove useful.

Of course, there is nothing simpler than to show by means of the usual axioms that negative probabilities are impossible and self-contradictory, but quite frequently the shortest way fails to reach a hidden aim, and axioms are essentially a matter of convention. Thus an investigation along these lines is perhaps not completely unreasonable as will be pointed out briefly in chapter 2. In chapter 3 some indications in favour of extended probabilities will be presented. It starts with those emerging from non-relativistic quantum mechanical phase-space distributions, invented by Wigner and elaborated by Moyal, Margenau and Hill, Cohen, Wigner and many others. Subsequently the large set of indications appearing in relativistic quantum theory is briefly reviewed, referring to work of Dirac, Gupta and Bleuler (the latter being utilized by Heisenberg) and Feynman. The Einstein–Podolsky–Rosen paradox and few non-quantum examples conclude this chapter. Some attempts to a formal treatment of negative probabilities, first formulated by Bartlett, are reported in chapter 4. Subsequently a collection of opinions about the existence of extended probabilities is presented, and chapter 6 contains short articles by G. Ludwig; C. Dewdney, P.R. Holland, A. Kyprianidis and J.P. Vigier; N. Cufaro Petroni; M.S. Bartlett; E.T. Jaynes, written independently of this review and discussing the topic from different points of view.

According to Szent–Györgyi discovery consists of seeing what everybody has seen and thinking what nobody has thought. The present work – to the author's knowledge the first review of this strange concept – is not intended to meet the second part of Szent–Györgyi's remark, and those who expect to find in it any explicit solution will be disappointed bitterly. All thoughts mentioned here were thought before; its aim is merely to concentrate some ideas dispersed over literature and to repolish what got dusty, in order to encourage those who are interested to "see". May the passed sentence be confirmed or revised.

2. Vindication

Sometimes in the development of scientific theories it happens that fundamental concepts which – up to then – were believed to represent indisputable truth have to be sacrificed in order to resolve contradictions and to facilitate progress. Well-known examples are the extension of the set of natural numbers by negative, non-rational and imaginary numbers, the introduction of non-Euclidian geometry, the elimination of absolute space and time, the admission of entirely negative energies, and, most recently, the discovery of negative temperatures [2].

In the course of development of quantum theory – and, in particular, of its formulation with respect to Lorentz invariance – new problems arose which seemingly require one or the other modification of fundamental concepts which, on the first glance, seem to represent basic axioms of each physical theory. The present state of physics indicates that at least one of the following three axioms cannot be upheld: (1) *The principle of cause and result, or the definite direction of the time-arrow*, (2) *the principle of separability, i.e., the limited velocity of any interaction*, (3) *the concept of an observer-independent, precisely defined reality* [3]. On the other hand, there are quite a lot of plausible arguments in favour of the validity of these physical axioms. To name only few: Instantaneous interactions have to be excluded because there is no objective definition of simultaneity of spatially separated events, unless Newton's absolute space and time were reestablished. (There seems to exist a preferred reference frame [4–6], namely that which makes our universe appear isotropic with respect to the averaged Doppler red-shift or to the blackbody background radiation of temperature 2.7 K. But it would mean an overbiased conclusion to doubt relativity theory on grounds of this accidental natural reference frame.) Further it is meaningless to distinguish between interactions which can be used for signalling and those which can not, as long as the axiom of an observer-independent reality is maintained, because in that case it does not play a role whether an interaction is in principle observable or not. On the other hand, the elimination of an objective reality is hard to swallow since its existence always can be proved "post festum".

In order to circumvent the elimination of the physical axioms mentioned above, an alternative idea has been considered which consists of modifying classical logic. Birkhoff and von Neumann already replaced in the framework on non-relativistic quantum mechanics the distributive law of classical logic by the modular identity [7]:

$$\text{if } a \subseteq c, \quad \text{then} \quad a \cup (b \cap c) = (a \cup b) \cap c .$$

Mittelstaedt excluded the principle of "unrestricted availability" denying expressions like $a \to (b \to a)$ in case a and b are measurements corresponding to non-commuting operators. The "tertium non datur" becomes inapplicable in three-valued logics of Lukasiewicz, Destouches-Février and Reichenbach as well as in the complex-valued logic of von Weizsäcker. Finkelstein and Putnam stressed the analogy between the epistemological situations in logic and geometry. All these attempts, however, are disputable because physics, and in particular quantum theory, is based upon classical logic. Hence, a fundamental modification in this field would entail unforeseeable alterations of our whole set of concepts. Moreover "*we can formulate a system of geometry without the use of geometrical principles, but we cannot formulate a system of logic without the use of logical principles*" [8].

Therefore a less radical change might be considered which consists of modifying probability theory. According to Strauss [9] a different probability theory has to be used in quantum mechanics, and according to Watanabe [10] conditional probability precedes logic. Hence, this step is already implied by above considerations. Nevertheless, just because of the close relationship between conditional probability and logic, it might be preferable to take the plunge at a slightly different, perhaps less sensitive point, namely by eliminating Kolmogorov's axiom, which restricts probability numbers p to the interval $0 \leq p \leq 1$. It is, of course, not necessary to mention that the objections to this idea are of the same strength and plausibility as those objecting to any change of the physical axioms mentioned above; and the principle of cause and result – at least in its strong deterministic meaning – is not maintainable, even when introducing extended probabilities.

Therefore, the attempt of struggling with plausibilities – often enough condemned to fail – will be discontinued in order to present some examples which – truely or falsely – appear to indicate the existence of extended probabilities.

3. Indications

At first sight quantum mechanics looks like a statistical theory, comparable to, say, statistical thermodynamics, and numerous attempts have been made to formulate quantum mechanics in a purely statistical way. But this will be shown to be impossible as long as the physical axioms (1)–(3) mentioned in chapter 2 are maintained and the predictions of quantum theory continue to be verified experimentally. Since the latter cannot *reasonably* be doubted too, we have the choice between equally undesirable alternatives, and the decision seems to be merely a matter of taste or uneducated guess. The reader will be aware that for the present work it has been settled by following Feynman's remark: "The only difference between a probabilistic classical world and the equations of the quantum world is that somehow or other it appears as if the probabilities would have to go negative, and that we do not know, as far as I know, how to simulate" [11].

This chapter is subdivided into four sections, due to the independent importance of four realms for the present subject. These realms are: (1) non-relativistic phase-space distributions or Wigner functions, (2) some approaches to generalize quantum mechanics, (3) the EPR paradox, and (4) some non-quantum indications.

3.1. Phase-space distributions

Even in non-relativistic quantum mechanics negative probabilities creep into the picture. To formulate a conventional (Maxwellian) probability distribution of the coordinates x and momenta p, similarly to statistical mechanics, is plainly excluded by the corresponding uncertainty relation which prevents *at least* the simultaneous *knowledge* of these quantities. Wigner and Szilard, however, found a distribution function which for the first time was applied by Wigner in order to calculate the quantum correction to the gas pressure formula [12]. If a wave function $\psi(x_1, \ldots, x_n)$, abbreviated by $\psi(x)$, is given, the corresponding Wigner function reads

$$P(x, p) = (\pi\hbar)^{-n} \int_{-\infty}^{\infty} \mathrm{d}^n y\, \psi^*(x+y)\psi(x-y)\exp(2\mathrm{i}p\cdot y/\hbar)\,, \tag{3.1}$$

with x, y and p vectors having as many components as has the configuration space of the ψ, namely n. $p\cdot y$ denotes the scalar product. In order to demonstrate the fundamental features of the Wigner function, relevant for the present purpose, it is sufficient to consider a single particle in linear motion. Thus $n = 1$ and the vector symbols will be dropped henceforth.

The Wigner function exhibits remarkable similarities to a probability distribution in that it leads to the correct probabilities for the coordinates when integrated with respect to the momenta (the ration range is always understood to be $(-\infty, \infty)$ unless indicated otherwise),

$$\int P(x, p)\, dp = |\psi(x)|^2 , \tag{3.2}$$

and, vice versa, it gives the proper probabilities for the momenta when integrated over the coordinates,

$$\int P(x, p)\, dx = (2\pi\hbar)^{-1} \left| \int dx\, \psi(x) \exp(-ipx/\hbar) \right|^2 . \tag{3.3}$$

Although Wigner calls it the probability function of the simultaneous values for the coordinates and momenta (in more recent papers the notation "quasi-probability" is adopted [13]) he stresses in the same context, that it cannot really be interpreted in this way "as is clear from the fact, that it may take negative values. But of course this must not hinder the use of it in calculations as an auxiliary function which obeys many relations we would expect from such a probability" [12].

The indefiniteness of P is already suggested by the following consideration [13]. If two state vectors ψ and ϕ are chosen so that they are orthogonal, the equality

$$\int dx \int dp\, P_\psi(x, p) P_\phi(x, p) = 0 \tag{3.4}$$

must hold. Thus it is clear that P cannot be everywhere positive, and it must exhibit a very pathological shape in order to avoid going negative in spite of (3.4).

But the existence of Wigner functions taking negative values is firmly proved by imposing two very general conditions on P which, among others given in [13], can be said to define this type of probability distributions, namely [14]:

(i) $P(x, p)$ should be a Hermitian form of the state vector $\psi(x)$, i.e., with $\hat{M}(x, p)$ a self-adjoint operator,

$$P(x, p) = (\psi(x), \hat{M}(x, p)\psi(x)) . \tag{3.5}$$

This condition makes $P(x, p)$ a real number.

(ii) $P(x, p)$ should give the proper expectation values for all operators which are sums of a function of p and a function of x,

$$\int\int P(x, p)[f(p) + g(x)]\, dp\, dx = \left(\psi, \left[f\left(\frac{\hbar}{i}\frac{\partial}{\partial x}\right) + g(x)\right]\psi\right) . \tag{3.6}$$

This condition is a somewhat milder form of (3.2) and (3.3) which properly have to be understood as axioms of the Wigner function and, in any case, must be satisfied.

Further, it suffices to consider such ψ which are linear combinations $a\psi_1 + b\psi_2$ of any two fixed functions, vanishing in certain intervals of x. Now, by requiring

$$P(x, p) \geq 0 \quad \text{for all } x \text{ and } p \tag{3.7}$$

for *every* such ψ, Wigner obtains a contradiction [14] which in short runs as follows:

Consider an interval I, inside of which $\psi(x) = 0$ and $g(x) \geq 0$, while $g(x) = 0$ outside and $f(p) \equiv 0$ everywhere. Then (3.6) leads to

$$\int P(x, p) g(x) \, \mathrm{d}x = 0 \tag{3.8}$$

for all p (except a set of measure zero) and

$$\iint P(x, p) g(x) \, \mathrm{d}p \, \mathrm{d}x = 0 \,. \tag{3.9}$$

From (3.7) and the condition imposed on $g(x)$ we obtain (Wigner's lemma 2): *If $\psi(x)$ vanishes in an interval I, the corresponding $P(x, p)$ vanishes (except for a set of measure zero) for all values of x in that interval.*

Now, consider two functions $\psi_1(x)$ and $\psi_2(x)$ which vanish outside of two non-overlapping intervals I_1 and I_2, respectively. Because of (3.5) $P(x, p)$ corresponding to $\psi = a\psi_1 + b\psi_2$ will have the form

$$P = |a|^2 P_1 + a^* b P_{12} + a b^* P_{21} + |b|^2 P_2 . \tag{3.10}$$

By setting $b = 0$ it is obvious that P_1 is the Wigner function belonging to ψ_1 (and P_2 that belonging to ψ_2). The meaning of P_{12} and P_{21} is less obvious, but we need not bother, because both must be identically zero. This can be seen by considering any interval I' outside I_1. Since, according to the above lemma, P_1 vanishes almost everywhere in interval I', (3.10) cannot be positive for *every* choice of a and b unless $P_{12} = P_{21} = 0$ outside I_1. The same proof applies to I_2. Thus, instead of (3.10) we have

$$P = |a|^2 P_1 + |b|^2 P_2 \tag{3.11}$$

almost everywhere.

In order to complete the contradiction, let us denote the Fourier transforms of ψ_1 and ψ_2 by $\varphi_1(p)$ and $\varphi_2(p)$, respectively. Equation (3.3) then reads

$$|a|^2 \int P_1(x, p) \, \mathrm{d}x + |b|^2 \int P_2(x, p) \, \mathrm{d}x = |a|^2 |\varphi_1(p)|^2 + |b|^2 |\varphi_2(p)|^2 + 2 \,\mathrm{Re}[a b^* \varphi_1(p) \varphi_2^*(p)] \,. \tag{3.12}$$

Since this must be valid for *all* a and b, we must have identically in p:

$$\varphi_1(p) \varphi_2^*(p) = 0 \,. \tag{3.13}$$

"This is, however, impossible since φ_1 and φ_2, being Fourier transforms of functions restricted to finite intervals, are analytic functions of their arguments and cannot vanish over *any* finite interval" [14].

In order to illustrate this result, the Wigner function formalism may be applied to the paradigm of quantum theory, the linear harmonic oscillator.

From its Hamiltonian

$$H(x, p) = p^2/2m + m\omega^2 x^2/2 \tag{3.14}$$

and the time-independent Schrödinger equation

$$\hat{H}\left(x, \frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x}\right)\psi(x) = E\psi(x) \tag{3.15}$$

it is easy to find the wave function of the ground state

$$\psi_0(x) = (m\omega/\pi\hbar)^{1/4}\exp(-x^2 m\omega/2\hbar) \tag{3.16}$$

corresponding to the energy $E_0 = \hbar\omega/2$. Inserting (3.16) in (3.1) and integrating out in y leads to

$$P_0(x, p) = (\pi\hbar)^{-1}\exp(-x^2 m\omega/\hbar - p^2/m\omega\hbar), \tag{3.17}$$

which does not exhibit any anomaly in that it is non-negative and, when integrated with respect to x, supplies the proper distribution of the momentum

$$\int P_0(x, p)\,\mathrm{d}x = (m\omega\pi\hbar)^{-1/2}\exp(-p^2/m\omega\hbar), \tag{3.18}$$

which is a Gaussian distribution with expectation zero and standard deviation $(\Delta p)^2 = m\omega\hbar/2$. Integrating with respect to p yields, as expected, the square of (3.16),

$$\int P_0(x, p)\,\mathrm{d}p = (m\omega/\pi\hbar)^{1/2}\exp(-x^2 m\omega/\hbar), \tag{3.16'}$$

also a Gaussian distribution with expectation zero and standard deviation $(\Delta x)^2 = \hbar/2m\omega$.

It is well known that Gaussian distributions satisfy Heisenberg's uncertainty relation in its marginal form, i.e., as an equality [15]. From (3.18) and (3.16′) we obtain $(\Delta x)(\Delta p) = \hbar/2$. "It may also be noted that the distributions of momentum and position are statistically independent, because $\int P_0\,\mathrm{d}p \int P_0\,\mathrm{d}x = P_0$. *A fortiori*, the covariance and correlation coefficient are zero" [16].

Clearly, this example does not contradict Wigner's "negativity proof" because the latter only says that *there are* state functions for which the corresponding $P(x, p)$ cannot be everywhere non-negative. One of those is the first excited state of the harmonic oscillator. Using the state function of the first excited level, the same formalism as described above will lead to the corresponding Wigner function. As, however, the calculation is somewhat lengthy, we adopt the elegant formula given in [13]. With H the Hamiltonian of eq. (3.14) and L_n the nth Laguerre polynomial, the Wigner function corresponding to the nth excited state can be expressed by

$$P_n(x, p) = (\pi\hbar)^{-1}(-1)^n\exp(-2H/\hbar\omega)L_n(4H/\hbar\omega) \tag{3.19}$$

or, using (3.17),

$$P_n(x, p) = (-1)^n P_0(x, p)L_n(4H/\hbar\omega). \tag{3.19'}$$

As, P_0 was found to be non-negative everywhere, we have to examine the remaining expression

$$P_n/P_0 = (-1)^n L_n(4H/\hbar\omega)\,. \tag{3.20}$$

The first-order Laguerre polynomial is [17]

$$L_1(u) = 1 - u\,. \tag{3.21}$$

Hence, $P_1(x, p)$ goes negative for

$$H \equiv p^2/2m + m\omega^2 x^2/2 < \hbar\omega/4\,. \tag{3.22}$$

Since the energy of the first excited level of the linear harmonic oscillator is known to be $E_1 = \frac{3}{2}\hbar\omega$, this should never happen. On the other hand, energy conservation can be violated during short time intervals, corresponding to the uncertainty principle. Therefore, in this special case, negative probabilities will only occur in connexion with rather drastic violations of the conservation law. The inversion of this proposition, however, is not true.

This is demonstrated by the Wigner function of the second excited state. The second-order Laguerre polynomial is [17]

$$L_2(u) = 2 - 4u + u^2\,. \tag{3.23}$$

(As we referred the reader extensively to [13], it should be mentioned, that L_2 is incorrectly given there.) Using (3.20) we obtain that $P_2(x, t)$ goes negative for

$$\tfrac{1}{2}\hbar\omega(1 - 2^{-1/2}) < H < \tfrac{1}{2}\hbar\omega(1 + 2^{-1/2})\,, \tag{3.24}$$

distinctly smaller than $E_2 = \frac{5}{2}\hbar\omega$. For $H = 0$ and $H \to \infty$, however, $P_2(x, t)$ is non-negative in spite of the drastic violation of the law of energy conservation, corresponding to these energy values.

We will not leave this illustrative example without noting some general features of Wigner functions of the linear harmonic oscillator.

From

$$\lim_{u\to\infty} L_n(u) = (-1)^n u^n \tag{3.25}$$

and (3.19) we find P_n being positive and asymptotically approaching zero for H going to infinity. In the special case of $H = 0$,

$$L_n(0) = n! \tag{3.26}$$

together with (3.19) makes even-order P_n being positive and odd-order P_n being negative at $H = 0$. Most interesting in the present context is, however, that all these Wigner functions of non-zero order unavoidably will take positive as well as negative values. This can easily be seen from the orthogonality relation [17]

$$\frac{1}{n!}\frac{1}{m!}\int_0^\infty e^{-u}L_n(u)L_m(u)\,du = \delta_{nm} \tag{3.27}$$

by inserting $L_m(u) = L_0(u) \equiv 1$.

The Wigner function in a more general form was also derived and thoroughly analyzed by Moyal [18], who shows in few steps, how it can be obtained from the characteristic function of a basic set of dynamical variables. Further, he shows that the Wigner function can be expanded in terms of "phase-space eigenfunctions", which form a complete orthogonal set in the Hilbert space of the phase-space functions or Wigner functions. Corresponding to our example (3.27), we have in general: "Phase-space eigenfunctions must generally take negative as well as positive values, since they are orthogonal. Only one eigenfunction (generally the ground state one) may possibly be non-negative for all values of the dynamical variables, except for singular eigenfunctions involving delta functions, such as the momenta eigenfunctions" [18]. Moyal considers the partial being negative of the Wigner function as an indication that "where the distribution [. . .] can take negative values, it is not an observable quantity" and concludes "that in applications of the theory, we need not be concerned whether the phase-space distributions are true probabilities, provided that the final results [. . .] are necessarily true, non-negative probabilities" [18].

A closely related function P_M was analyzed by Margenau and Hill [19]. They consider the general case of a wave function ψ having the alternate expansions

$$\psi = \sum_i a_i u_i(\tau) \quad \text{and} \quad \psi = \sum_j b_j v_j(\tau) \tag{3.28}$$

on orthonormal sets u and v, which are, respectively, eigenstates of operators $\hat{X}$ and $\hat{Y}$. Whence, given ψ, the joint probability of the corresponding variables taking values x_i and y_i is expressed by

$$P_M(x_i, y_i) = \mathrm{Re}\left[a_i^* b_j \int u_i^* v_j \, d\tau\right]. \tag{3.29}$$

"Most unfortunately", they observe however, "the defining relation (3.29) does not prevent the joint probabilities from being negative" [19], demonstrating this by means of some examples.

Treating a system with angular momentum $J = 1(=\hbar)$, represented by a state vector

$$\psi = \begin{bmatrix} a \\ b \\ c \end{bmatrix},$$

they obtain a negative joint probability for, e.g., two eigenstates belonging to eigenvalues $J_x = -1$ and $J_y = -1$, provided a, b, and c are suitably chosen.

The probability for simultaneous measurement of position x and linear momentum p of a particle, as derived from (3.29) by using a Gaussian wave packet with mean momentum p_0, results in a distribution

$$P_M(x, p) = f(x, p)\cos(p - p_0)x/\hbar\,,$$

where $f(x, p) \geq 0$, and thus, P_M being indefinite.

Finally, Margenau and Hill discuss the joint probability, again in the simple case of a moving free particle, for position measurements at different times. While they arrive at a satisfactory result for large time differences, they obtain an indefinite result for certain positions x and small time differences.

After briefly considering a different alternative for calculating a joint probability $P(x, p)$, which leads to indefinite results too, the authors consider this feature of joint probabilities reason enough to reject *in toto* their attempt, by means of which arguing they arrive at the desired conclusion, namely, a denial of von Neumann's projection postulate and the reduction of wave packets on measurement. (See however, H. Margenau in chapter 5.)

Cohen [20] could show that a wide class of probability distribution functions is supplied by the rather general expression

$$P(x, p) = (2\pi)^{-2} \int\int\int f(\theta, \tau) \exp(-\mathrm{i}\theta x - \mathrm{i}\tau p + \mathrm{i}\theta u)\psi^*(u - \tau\hbar/2)\psi(u + \tau\hbar/2)\, \mathrm{d}\theta\, \mathrm{d}\tau\, \mathrm{d}u\,. \tag{3.30}$$

Herein f is simply a smearing function. By setting $f \equiv 1$, substituting τ by $-2y/\hbar$ and integrating over θ and u, we obtain the original Wigner function (3.1). The choice of $f(\theta, \tau) = \cos \theta\tau\hbar/2$, using the latter in its exponential form and again integrating over θ and u, leads to a reformulation of (3.29) in terms of x and p.

$$P_{\mathrm{M}}(x, p) = (2\pi)^{-1}\, \mathrm{Re}\left[\psi(x) \int \exp(-\mathrm{i}\tau p)\psi^*(x - \tau\hbar)\, \mathrm{d}\tau\right]. \tag{3.29'}$$

Other distribution functions may be built with different functions f, if only f satisfies the condition

$$f(0, \tau) = f(\theta, 0) = 1 \tag{3.31}$$

in order to yield the correct quantum mechanical marginal distributions.

Cohen imposed the following conditions on a general distribution function $P(x, p)$ [21]: (i) those given by (3.2) and (3.3); (ii) if the quantum mechanical mean value of the operator $\hat{G}$ is $\langle \hat{G} \rangle$, then there should exist a function $g(x, p)$ such that

$$\langle \hat{G} \rangle = \int\int g(x, p) P(x, p)\, \mathrm{d}p\, \mathrm{d}x\,; \tag{3.32}$$

and, for any function K,

$$\langle K(\hat{G}) \rangle = \int\int K(g(x, p)) P(x, p)\, \mathrm{d}p\, \mathrm{d}x\,. \tag{3.33}$$

And he found, that, irrespective of whether P is positive semidefinite or not, condition (ii) can never be satisfied [20–22]. The Wigner function P_1 of the harmonic oscillator, e.g., yields the correct expectation value for the mean energy, but fails to supply the zero-standard deviation, which one should expect from a quantum mechanical energy eigenstate. He concludes: "Of course, it can be argued that the classical formalism does go through as long as we do not insist that the function which must be used to obtain the mean value of a function, K, of g is not identical to $K(g)$. But this would carry us even further from the conceptual basis of classical probability theory than does quantum mechanics itself!" [21].

In spite of this concern the Wigner function ("the rather trivial connections among the various distributions [. . .] leads one to wonder why more than the Wigner distribution need be considered for any calculations" [23]) being essentially nothing else than a convenient version of the density matrix, has found numerous applications. References are given in review articles [24, 25] and [13], the authors of the latter note only few further references, but are preparing a forthcoming article on this subject.

Bopp [26], well aware of the fact that "it is not possible to construct non-negative distribution functions", tried to circumvent the problem by introducing a constant l which "is a certain length, describing the accuracy of measuring space positions." The corresponding distribution function can also be derived from Cohen's general expression (3.30) by suited choice of f, namely [21]:

$$f = \exp(-\tau^2\hbar^2/2l^2 - \tfrac{1}{8}\theta^2 l^2) . \tag{3.34}$$

But, "since f does not satisfy (3.31), we immediately know that it cannot yield correct quantum mechanical marginal distributions. [. . .] The advantage of Bopp's distribution is that of being positive definite for all values of the variables. However, the statistical predictions obtained using [it] will, of course, contain l and are therefore at variance with the usual predictions of quantum theory" [21].

The use of a smearing function f, not fulfilling (3.31) but leaving $P(x, p)$ positive semidefinite, was first proposed by Husimi [27] and has since been employed several times [13]. But we do not wish to elaborate on this topic, because of Wigner's general proof, which essentially means, that a statistical foundation of quantum mechanics can be excluded, unless the currently accepted axioms are modified. The existence of "positive" smearing functions suggests, however, that extended probabilities do appear real on the microscopical level of reality, thus, the Wigner function "once smoothed out over any phase-space region of dimensions larger than or equal to unity [$=\hbar$], is always positive and smaller than one" [28].

3.2. Approaches to quantum theory

The necessity of extended probabilities becomes most distinct if a Lorentz-invariant formulation of quantum theory is attempted. The special role that time plays in non-relativistic theory can, e.g., in the most simple case of particles with no charge and spin, be removed by means of the Klein–Gordon equation which for a single free particle of rest mass m is given by

$$(\Box + m^2)\psi = 0 \tag{3.35}$$

($\hbar = c = 1$, and the usual terminology [29] is adopted). Born's notion, however, according to which the square of the wave function has to be interpreted as probability density [30], necessarily must fail in this context, because $|\psi|^2$ as a scalar violates conservation of total probability. On the other hand, the density proposed by Gordon [31] and Klein [32],

$$\frac{1}{2im}\left(\frac{\partial\psi^*}{\partial t}\psi - \psi^*\frac{\partial\psi}{\partial t}\right), \tag{3.36}$$

satisfies as time component of a four-vector the conservation law, and "thus (3.36) is evidently the correct mathematical form to use" [1], but, clearly, it can go negative.

This is not the only difficulty. If the wave function of a plane wave

$$\psi = \exp[-i(p_0x_0 - p_1x_1 - p_2x_2 - p_3x_3)], \quad x_0 \equiv t, \quad p_0 \equiv E, \tag{3.37}$$

is transformed to the momentum and energy variables, the Gordon–Klein expression (3.36) goes over into

$$|\psi(p_0p_1p_2p_3)|^2 p_0^{-1}\, dp_1\, dp_2\, dp_3, \tag{3.38}$$

"as the probability of the momentum having a value within the small domain $dp_1\, dp_2\, dp_3$ about a value p_1, p_2, p_3, with the energy having the value p_0, which must be connected with p_1, p_2, p_3 by

$$p_0^2 - p_1^2 - p_2^2 - p_3^2 - m^2 = 0. \tag{3.39}$$

The weight factor p_0^{-1} appears in (3.38) and makes it Lorentz invariant, since $\psi(p)$ is a scalar – it is defined in terms of $\psi(x)$ to make it so – and the differential element $p_0^{-1}\, dp_1\, dp_2\, dp_3$ is Lorentz invariant. This weight factor may be positive or negative, and makes the probability positive or negative accordingly. Thus the two undesirable things, negative energy and negative probability, always occur together" [1].

In more recent work [33, 34] Dirac tried to avoid negative energy solutions (and hence, negative probabilities) by referring to an approach of Majorana's [35]. The result, however, entailed some strange consequences like "Zitterbewegung" and an inverse mass–spin dependence. Furthermore it does not allow for a minimal coupling to an external electromagnetic field. (The latter difficulty has been overcome most recently by Sudarshan et al. [36], at least in the special case of particles with fixed mass, zero spin and positive energy.) Dirac seemingly was not convinced of his approach, because some years later he confessed: "I have spent many years looking for a good Hamiltonian which I haven't yet found." [37].

In some older papers [1, 38] Dirac formulates an alternative approach to quantum electrodynamics which allows for a conventional treatment of particles with half-odd integral spin, but unavoidably entails negative probabilities when applied to particles with integral spin, in special cases even demanding probabilities of plus or minus 2, distinctly outside the usual range. On the other hand, this relativistic theory has great advantages over the usual method in that it avoids the most artificial process of renormalization. With respect to the latter, Dirac never changed his mind, qualifying it as a "*working rule*" and considering its results, in spite of their accuracy, as "*not reliable*" [37].

Indeed, as we know that semi-classical treatment of the radiation field as, for example, investigated by Jaynes et al. [39, 40] can be excluded [41–43] (see also section 3.3) and, on the other hand, the commonly applied method of renormalization is a thing between artificial and unphysical, we are left between Scylla and Charybdis, in that our equations contain either probabilities as large as plus or minus 2 or electron masses exceeding that of the whole universe.

Obviously, also Dirac was very sceptical about those "undesirable things, negative energy and negative probability", but he asserts: "Negative energies and probabilities should not be considered as nonsense. They are well-defined concepts mathematically, like a negative sum of money, since the equations which express the important properties of energies and probabilities can still be used when they are negative. Thus negative energies and probabilities should be considered simply as things which do not appear in experimental results. The physical interpretation of relativistic quantum mechanics that one gets by a natural development of the non-relativistic theory involves these things and is thus in contradiction with experiment. We therefore have to consider ways of modifying or supplementing this interpretation" [1].

The essential difference between Dirac's theory and the usual method, going back to Heisenberg and Pauli [44, 45], is, in the case of photons, to extend the range of integration from only positive frequencies $\nu \in (0, \infty)$, for odd photon states n unavoidably implying divergencies, to $(-\infty, \infty)$. Since $f(\nu)$ in the expression

$$\int_{-\infty}^{\infty} f(\nu)\,d\nu, \quad f(\nu) \propto \nu^n \text{ for large } \nu, \tag{3.40}$$

is always a rational algebraic function, the divergencies with odd n values all cancel out. But as there is no simple physical explanation of the traded-in extended probabilities, Dirac bans his considerations to a hypothetical world, which however, shows some similarities to the real world. In his subsequent paper [38] he proposes the following interpretation: "The various parts of the wave function which referred to the existence of positive- and negative-energy photons in the old interpretation now refer to the emissions and absorptions of photons. [...] The probabilities (3.41), equal to 2 and −2, are not physically understandable, but one can use them mathematically in accordance with the rules for working with a Gibbs ensemble. One can suppose a hypothetical mathematical world with the initial probability distribution" for n photons

$$P_n = 2(\varepsilon - 1)^n, \tag{3.41}$$

where ε is a small positive quantity tending to zero, in order that we may have $\Sigma_0^\infty P_n = 1$. "One can work out the probabilities of radiative transition processes occurring in this world. One can deduce the corresponding probability coefficients, i.e., the probabilities per unit intensity of each beam of incident radiation concerned, by using Einstein's laws of radiation. For example, for a process involving the absorption of a photon, if the probability coefficient is B, the probability of the process is

$$\sum_{n=0}^{\infty} nP_nB = -B/2, \tag{3.42}$$

and for a process involving the emission of a photon, if the probability coefficient is A, the probability of the process is

$$\sum_{n=0}^{\infty} (n+1)P_nA = A/2. \tag{3.43}$$

Now the probability of an absorption process, as calculated from the theory, is negative, and that for an emission process is positive, so that, equating these calculated probabilities to (3.42) and (3.43) respectively, one obtains positive values for both B and A. Generally, it is easily verified that any radiative transition probability coefficient obtained by this method is positive.

If now becomes reasonable to assume that *these probability coefficients obtained for a hypothetical world are the same as those of the actual world*" [38].

In a review article "On Dirac's New Method of Field Quantization", Pauli [46] interpreted: "the term 'negative probability' means essentially that observables with only positive eigenvalues can get negative expectation values", a point of view which seems not exhaustive (compare Gupta's formalism of indefinite metrics [47], outlined in chapter 4). Pauli concluded: "The arbitrariness of the rules for the translation of results concerning the hypothetical world into results concerning the actual world and the lack of uniqueness of these rules seems to indicate that new ideas and more radical changes of the present formalism will be necessary in order to get a really satisfactory quantum theory of the electromagnetic field" [46]. Now, 40 years later, the problem has not yet been settled.

Another early example involving extended probabilities in the formalism of quantum theoretical description of radiation is given by extending the work of Weisskopf and Wigner [48] who calculated the natural linewidth of radiative decay of an excited atom. The corresponding transition amplitude (equation 15b of [48]) may be rewritten

$$A(E, t) \propto \frac{e^{-t/2} - e^{iEt}}{i/2 - E} \tag{3.44}$$

with E denoting the difference between actual photon energy and mean state energy E_0 in units of the natural width of the excited state, and t denoting the time interval between excitation and decay in units of the mean lifetime of the state. In case $E_0 \gg 1$, (3.44) is valid for practically the total duration of detectable emission, because it holds up to times t of order E_0.

In finishing their calculation, the authors obtain the well-known Breit–Wigner lineshape in the limit $t \to \infty$.

It is now very interesting to consider the spectral distribution of photons emitted in finite time intervals. For the time interval $(0, t)$ we have

$$|A(E, t)|^2 = \frac{1}{2\pi} \frac{1 - 2\,e^{-t/2}\cos(Et) + e^{-t}}{E^2 + \frac{1}{4}}, \tag{3.45}$$

which undoubtedly is non-negative for every E and t. The spectral distribution emitted *at* time t, however,

$$I(E, t) = d|A(E, t)|^2/dt, \tag{3.46}$$

entails negative values, as easily can be seen from

$$I(E, t) = \frac{1}{2\pi} \frac{(2E\sin(Et) + \cos(Et))\,e^{-t/2} - e^{-t}}{E^2 + \frac{1}{4}}. \tag{3.46'}$$

Further, if the Breit–Wigner lineshape $|A(E, t \to \infty)|^2$ in connection with the law of exponential decay,

$$I_{\text{norm}}(E, t) = \frac{1}{2\pi} \frac{e^{-t}}{E^2 + \frac{1}{4}}, \tag{3.47}$$

is used to normalize $I(E, t)$, we obtain the normalized decay probability density

$$\rho(E, t) = I(E, t)/I_{\text{norm}}(E, t)\,, \tag{3.48}$$

which can take on negative values as well as values exceeding unity, and, if integrated over suited domains $\Delta E \cdot \Delta t$, small compared to unity ($=\hbar$), the normalized probability $\rho(E, t)\,\Delta E\,\Delta t$, which is an observable quantity, may violate both the lower and the upper limit of Kolmogorov's axiom.

These results have been verified by experiments [49–51]. A Mössbauer absorber, covering mainly a negative part of $I(E, t)$ in the 14.4 keV decay of ^{57}Fe, exhibits a significantly enhanced transmission, compared to the non-Mössbauer case which is expected (and usually proved) to absorb least photons at all.

One reason for the difficulties with quantum electrodynamics is the general Lorentz condition, according to which the four-divergence of the electromagnetic potential A must vanish

$$\frac{\partial A^0}{\partial x_0} + \frac{\partial A^1}{\partial x_1} + \frac{\partial A^2}{\partial x_2} + \frac{\partial A^3}{\partial x_3} \equiv \partial_\mu A^\mu = 0\,. \tag{3.49}$$

A photon density obtained from this continuity equation suffers from the same problems as the Gordon–Klein conserved density (3.36) in that it is not positive semidefinite, or, according to the opinion of the respective referee, it does not exist. "This is related to the fact that photons cannot be sharply localized. If they could, we could define the photon density as the number of photons per unit volume in some arbitrary small volume. However, in a relativistic field described by E, A, and B, we cannot define such a density. If this density ($=j^0$) would exist, it would have to form a four-vector j^μ together with [. . .] the corresponding current density, satisfying with j^0 a continuity equation $\partial_\mu j^\mu = 0$. Moreover, j^0 then should be positive definite. Such density four-vector j^μ does not exist for the Maxwell field.

If A is generalized to a four-vector A^μ satisfying the Lorentz condition (3.49), it is possible to form a four-vector $j^\mu = (4\pi)^{-1}F^{\mu\lambda}A_\lambda$, but then $\partial_\mu j^\mu = (8\pi)^{-1}F^{\mu\lambda}F_{\mu\lambda} - j^\lambda A_\lambda \neq 0$, and j^0 is not positive definite" [52]. ($F^{\mu\lambda} = \partial_\lambda A^\mu - \partial_\mu A^\lambda$.)

The same holds for distribution functions, applied in quantum optics in order to describe modes of the electromagnetic field. These functions which have been invented by Glauber [53] and Sudarshan [54], analogously to the Wigner functions, allow to replace the operator description (in this case: the creation and annihilation operators) by distribution functions in terms of c-numbers. (See also H. Paul in chapter 5. A detailed review containing a lot of recent contributions to this topic is given in [13].) However, in some cases it can be shown that such a Glauber–Sudarshan P function $P(\alpha)$ does not exist, or, in other words, that it is not positive semidefinite. Most recently [55] the theoretical investigation of an optical field produced by intracavity four-wave mixing, exhibiting photon antibunching, squeezing, and violation of Cauchy–Schwarz and Bell's inequalities led to the following conclusion (which is also interesting in the context of section 3.3, EPR paradox and Bell's inequality): It may be shown that if $P(\alpha)$ is positive and interpretable as a probability distribution, the above mentioned classical inequalities must hold. "The quantum theory of radiation, however, allows $P(\alpha)$ to be singular and take on negative values for certain fields. [. . .] Thus we deduce that violation of classical inequalities is associated with the non-existence of a positive normalizable representation of the electromagnetic field in terms of the Glauber–Sudarshan P function" [55].

Negative measures (or probabilities) appear also in Feynman's "*Space-time approach to non-relativistic quantum mechanics*" [56], describing the famous path integral method. "Finite results can be obtained under unexpected circumstances because the measure is not positive everywhere, but the contributions from most of the paths largely cancel out. These curious mathematical problems are sidestepped by the subdivision process" [56].

In the same year, 1948, Feynman proposed a symmetric treatment of the longitudinal and the transverse parts of the electromagnetic field [57], the former of which was usually separated and replaced by the Coulomb interaction, because real photons are known to possess only transverse components. This method, also proposed by Majumdar and Gupta [58] and others, was completed by Gupta [59] to a symmetric treatment of the four components of the electromagnetic potential by introducing four types of photons: two transverse, one longitudinal and one scalar. The scalar photons require the use of an indefinite metric. In order to account for the experimental fact that only transverse photons are observable, the usual supplementary condition (3.49) which, as we have seen above, is too restrictive to allow any state of the radiation field to satisfy it is replaced by

$$[\partial_\mu A^\mu]^+ \psi = 0 \quad \text{and} \quad \psi^\dagger [\partial_\mu A^\mu]^- = 0\,, \tag{3.50}$$

where $[\partial_\mu A^\mu]^+$ and $[\partial_\mu A^\mu]^-$ are, respectively, the positive and the negative frequency parts of $\partial_\mu A^\mu$.

In most cases of practical interest, the initial and the final states may be taken to be interaction-free. In such cases longitudinal or scalar photons do not exist in the initial state, and though they may occur in intermediate states, they will again disappear in the final state. "The probability of the emission of a real longitudinal photon is cancelled by the 'negative probability' of the emission of a corresponding scalar photon" [59].

In applying this formalism to the interaction of two electrons, the contribution of the longitudinal and the scalar photons alone is shown to be precisely the Coulomb interaction. A calculation of the electron's self-energy, however, leads to the old difficulties in that the result is divergent.

Bleuler [60] could show that Gupta's symmetrical theory was equivalent with the reduced theory where the longitudinal field is replaced by the static Coulomb interaction. In the course of his paper, he remarks that gauge-invariant quantities are uniquely defined in Gupta's theory. Difficulties in interpreting them do not arise, for negative probabilities here are prevented by the new Lorentz condition (3.50).

As the theory is operating in a Hilbert space with indefinite metric, entailing negative probabilities, the characteristics of indefinite metric will be outlined briefly in chapter 4.

Although in particular Heisenberg is known to be a vehement proponent of the opinion that a physical explanation is not absolutely necessary if a sufficient mathematical formalism is available, he took the opposite point of view with respect to negative probability. In the course of discussing renormalization theory, he wrote in 1959 [61]: "A more detailed investigation of this mathematical scheme, however, made it most probable that those quantities which in common quantum theory have to be interpreted as probabilities, here occasionally can get negative, after the renormalization has been applied. Of course, this would prevent a consistent interpretation of the formalism with respect to a description of matter, because a negative probability is a senseless notion." Some years later he was forced by mathematical difficulties arising from his unified field theory of elementary particles to change his mind [62]. He adopted the indefinite metric, referring to the method of Gupta and Bleuler and the Lee model. Observable quantities are represented by self-adjoint operators in a Hilbert space with indefinite metric. The eigenvalues of these operators need not be real, but imaginary eigenvalues do not conflict with experimental observation, because the norm of the respective "*ghost states*" must vanish. After describing his rather difficult theory, Heisenberg concludes: "The indefinite metric of the Hilbert space implies an essential extension of the usual mathematical scheme of quantum mechanics. From the physical viewpoint this extension can be interpreted by the assumption, that the notion of probability cannot be applied to local events, i.e., to space domains of order 10^{-14} cm or less. The universal length which enters the theory by the constant l or by the proton mass, can be defined as that distance, below which the quantum theoretical representation of probability fails.

These answers to the critical questions may not be the only ones possible. But they seem to be the most simple ones, considering the fact that it has not yet been possible, to formulate a relativistic theory including interactions within the tight frame of Wightman's axioms" [62, p. 136].

The Lee model [63] was introduced in order to study some characteristics of quantum electrodynamics by means of a simplified analogon. Even a brief description of its formalism is beyond the scope of the present work. Therefore we would only mention, that, as Källén and Pauli [64] have pointed out, the basic Schrödinger equations maintain their usual form if states with negative norm are taken into account. Eigenstates with non-zero norm are always real. Similar to original quantum electrodynamics, "in this model divergencies enforce the introduction of an indefinite metric in the Hilbert space. If usual quantum mechanics is extended this way, then the mathematical scheme is consistent and does not cause any unexpected divergence. Indeed, the scheme permits one to construct a unitary S-matrix and the respective physical interpretation. Therefore it seems to be a suited model for the mathematical structure of a non-linear spinor-theory" [62, p. 156].

We will not leave this topic without mentioning a paraphrase of Heisenberg's, which is a rather trivial one but supports our standpoint, not to consider deviations from probability theory (even in its extended form) other than enlarging the range of probability numbers. It reads "... and the norm of a product state is equal to the product of norms of its separate factors" [62, p. 154].

That the problem is still pending can be seen from a paper of Cufaro Petroni and Vigier [65], who confess: "We are confronted here with an old problem characteristic of relativistic quantum mechanics, namely the existence of negative probabilities. [...] Of course, we have no final answer to the question 'What is a negative probability?': We can only quote a proposition for its interpretation [66] in which the negative sign of the probability distribution is interpreted as reflecting the opposite 'charge' values of antiparticles in a particle–antiparticle distribution. We further remark that this is a problem which arises every time we are dealing with particles and antiparticles, and hence that it would be very strange not to meet it here where the possibility of trajectories running backward in time on our lattice are interpreted with the presence of pair creation and annihilation [67]. On the other hand, it is clear that if we had not assumed the possibility of the trajectories running backward in time (i.e., the antiparticle behaviour) all our statistics would be different since it is possible to show that one obtains in this case a classical diffusion equation that one cannot reduce to the quantum Klein–Gordon equation."

In this chapter we have not explicitly considered the Dirac equation

$$(\mathrm{i}\not{\partial} - e\not{A} - m)\psi = 0\,, \tag{3.51}$$

with e and m denoting charge and mass, respectively, and $\not{V} = \gamma_\mu V^\mu$ symbolizing the product of Dirac-matrices γ_μ with a four-vector V [29]. This has been omitted because it has been shown (and can be obtained from textbooks [68] already) that it is impossible to simultaneously maintain positive semidefinite probability densities and a physical meaning of the basic expression (3.39). And it is self-evident that also modifications of the Klein–Gordon equation (3.35) which permit a treatment of charged particles, i.e., second-order equations like the Feynman and Gell–Mann equation [69],

$$[(\mathrm{i}\not{\partial} - e\not{A})(\mathrm{i}\not{\partial} - e\not{A}) - m^2]\psi = 0\,, \tag{3.52}$$

do not permit the conserved density to be positive semidefinite [70]. For a discussion of this topic see the note by N. Cufaro Petroni in section 6.3.

In a series of very recent papers [71–75] the Vigier group, however, proposed how to avoid negative probability solutions under most circumstances. "We now suppose that all physical states correspond to those parts of the Klein–Gordon solutions which, following Einstein's basic principles, have positive energy and probability for particles and antiparticles" [71]. "The negative energy and probability solutions may be associated with positive energy and probability antiparticles by means of the charge conjugation operation" [72].

This formalism and the corresponding stochastic interpretation of quantum theory will not be elaborated here, because it is briefly outlined by its authors in section 6.2. The theory is a causal one, but necessarily suffers from being non-local, and thus it is in contradiction with axiom (2) of chapter 2. On the other hand, local causality (in its strong form) yields Maxwell–Boltzmann statistics and is thus in contradiction with quantum statistics [73].

Concluding, we should keep in mind that mathematics is capable of giving an inconceivably precise picture of reality if the correct equations for a certain realm of reality are applied. Considering how frequently renormalization has to be applied in modern quantum electrodynamics [29] and, furthermore, considering the unsatisfying state of present elementary particles theories (this opinion is explicitly stated in the preface of [29] and [76]), one is forced to conclude that either the correct equations or the correct interpretations (or both) have not yet been found. This, although weakening some of our indications for extended probabilities obtained from perhaps incorrect examples, suggests at least a close reexamination of theories entailing negative probabilities but being satisfying otherwise.

3.3. The Einstein–Podolsky–Rosen paradox

The opinions, whether the EPR paradox is really a paradox, are differing. Bohr [77] and the Copenhagen school refuse to see anything like a paradox in this case (cf. T.W. Marshall and H.J. Treder in chapter 5). However, we will point out briefly, that there is a paradox as long as the axioms (1)–(3) mentioned in chapter 2 are valid *and* the Copenhagen-school prohibition is not obeyed, which forbids to put questions which cannot be answered by the Copenhagen-school interpretation of quantum theory.

The EPR paradox [78] in Bohm's most popular version [79] deals with a spatially separated pair of electrons in the singlet state. That means, if the spin s of one electron (labelled by 1) is measured in direction a and yields the result $s_1(a)$, the same measurement performed on the other electron (2) will yield the result

$$s_2(a) = -s_1(a)\,. \tag{3.53}$$

As (3.53) is valid for any direction a whatsoever, it must be concluded that the result in every direction is determined, or, to put it in other words, that the total spin $S = \sqrt{s(s+1)}$ has a uniquely defined direction, unless one measurement influences the other one, i.e., unless the axiom of separability is dropped. According to the Copenhagen-school interpretation of quantum theory, the simultaneous existence of more than one spin component s, however, is excluded, because the corresponding operators represented by the Pauli matrices do not commute. According to this interpretation it is meaningless (and, hence, forbidden) to ask for more than one spin component, because on one electron only one component can be measured simultaneously, thereafter the original state is destroyed and another component which *could have been measured but has not* may have changed its value (if ever it had a value at all).

The paradox emerging from this interpretation with respect to this and other situations has been distinctly articulated by Schrödinger [80], who puts the only possible logical conclusion in the plain picture: "From so many experiments performed in advance I know that the pupil answers my first question always correctly. Hence it follows, that in any case he *knows* the answers to *both* questions", implying that the Copenhagen-school interpretation was wrong. As, so far, the paradox emerges only from this interpretation, it has been attempted to avoid it by so-called "hidden variables", namely physical quantities which are not observable, but influence the result of a measurement (in the present case: a definite direction of the total spin). The reader interested in this topic is referred to [52].

The argument of Einstein, Podolsky and Rosen, intended only to demonstrate the incompleteness of quantum theory, became really a paradox, when Bell [81] could prove that hidden variables existing in a separable reality (both notions together being usually termed "local hidden variables") cannot yield the quantum theoretical expectation values in some special cases. Bell considers the product of measurements performed in directions a and b on the first and the second electron, respectively, of a correlated pair in the singlet state. The single result $s_1(a, \lambda)$ depends on the direction a of the due measuring device and a hidden variable, usually denoted by λ. But it does not depend on the result $s_2(b, \lambda)$ or the direction b, of the measurement of the second electron. If now the expectation value of the product of the correlated measurements is denoted by $E(a, b)$, and corresponding expressions involving a third direction c are formed, Bell's inequality reads, setting $\hbar = 1$ [81],

$$|E(a, b) - E(a, c)| \leq \tfrac{1}{4} + E(b, c)\,, \tag{3.54}$$

which obviously is violated by the quantum theoretical expectation value $\langle s(a)s(b)\rangle = -\frac{1}{4}\cos(a, b)$, if e.g., b bisects the angle $2\pi/3$, formed by a and c.

In the present context we have to mention, that Bell in deriving (3.54) introduced a hidden variables density $\rho(\lambda)$, which must be positive semidefinite in order to arrive at (3.54). Meanwhile various inequalities of this type have been derived (some are reviewed in [82, 83]), but all necessarily make use of positive semidefinite probabilities (or, what essentially is the same: densities or measures). The most elegant derivation was given by Wigner [84] who, like Bell, considers pairs of correlated electrons in the singlet state, and subdivides the whole set into subsets, the members of which *would* yield identical results s^i,

$$s^i = \{s_1^i(a), s_1^i(b), s_1^i(c), s_2^i(a), s_2^i(b), s_2^i(c)\}\,, \tag{3.55}$$

if all six individual results s_j^i could be determined. If the results in all three directions are determined, then each of the sets s^i must be considered an entity of reality. Wigner, however, shows that the measure of at least one of the sets s^i must be negative, if the directions are not chosen parallel or anti-parallel and the quantum mechanical expectations are correct.

Analogous derivations were given for the experimentally most relevant case of correlated photon pairs [85–87] and, in the most general case of arbitrary spin, Mermin obtains the result that the range of angles, for which a contradiction to quantum theory arises, vanishes as $1/S$ as the classical limit $(S \to \infty)$ is approached [88]. Most recently, even a spin-S Bell inequality was derived that is violated for *any* set of three coplanar detector orientations for *any* value whatever of the spin [89].

As the overwhelming number of experiments performed in this realm [90] verified quantum theory and violated the inequalities based on the existence of local hidden variables, the paradox is completed.

This implies a full justification of the orthodox Copenhagen-school interpretation of quantum theory, asserting that physical quantities which cannot be measured simultaneously cannot exist simultaneously. The usual interpretation, however, (which, like all experimental descriptions, has to be expressed in terms of classical physics) ascribing the uncertainty principle to mechanical disturbance of the considered system [91], necessarily must fail in the EPR case. Hence it is open to question whether it is the correct one at all. It is not a far step to do, to arrive at the opinion that all physical quantities come into being only by measurements, and from that point it is an even smaller step to accept solipsism which, however, will be banned explicitly from the present context. But the results presented above prevent us from applying some familiar concepts.

It is, for example, wrong to ascribe any spin component to an electron, unless it has been measured or prepared, because, if one component came into being accidentally and the electron had been prepared by a Bohm-like process, another component could be measured on the correlated electron, and the former one was in possession of two definite spin components.

Further, it is no longer justified to assume the square of the wave function as describing the probability of an electron to *exist* within a certain domain of space, but we must properly think of it as indicating the probability of an electron to be *contracted* to this domain. Because, if the electron had an exactly defined moment before the coordinate was measured, it could not simultaneously possess an exactly defined coordinate. Thus we have to accept that, after an electron has been localized within, say, 10^{-5} cm, *the only entity corresponding in reality to our concept of the electron's position, has a diameter of* 10^{+5} cm *one second later.* The failure of the concept of an exactly defined, observer-independent reality is precisely the reason for the inapplicability of semi-classical theory of radiation, mentioned in section 3.2, because not even a definite polarization can be ascribed to a photon unless it has been measured or prepared. On the other hand, all the physical quantities which do not exist at present, can be determined retrospectively, because the past is not subject to the uncertainty principle. Hence, time acts like a filter, transforming a presence which is undefined and blurred (except those quantities, which are accidentally measured) to a past which is exactly defined and clear.

There is a manifold of unsatisfying answers, but a satisfying answer to the problem is still lacking. As mentioned above, the answer of the Copenhagen school which is widely accepted reads plainly and concisely: *Don't bother.* But, with respect to the EPR paradox in particular, we then have to accept the following statement too: "The idea which is found to be untenable may, roughly, be said to be that of the independent existence of two entities, the state of system II and one's knowledge of its state, only the latter being affected by measurements made on system I. Quantum theory shows that this is not an adequate concept of the relation between subject and object" [92]. This is not far from solipsism, not to be considered in this article, and the answer it is suggesting to Feynman's famous question "*if a tree falls in a forest and there is nobody there to hear it, does it make a noise*?" [93], may be "*does it fall*?".

On the other hand, nobody can give a nearly plausible solution. Most of those, who dislike above non-solution, tend to drop the axiom of separability by introducing non-local interactions [73, 94], others consider retroactive influences [95], which implicitly are contained in the assumption of non-local interactions too. Some physicists doubt the validity of the crucial experiments [96] and propose sophisticated local hidden variables, giving predictions which come extremely close to those of quantum mechanics [97, 98], some express their despair: "Thus the theorem of Bell proves, in effect, the profound truth that the world is either fundamentally lawless or fundamentally inseparable" [99]. Further, there is Everett's "*Relative state formulation of quantum mechanics*" [100], usually termed "*Many worlds theory*", according to which the whole universe is split into distinct branches by every quantum jump, all existing simultaneously but being not observable from that one occupied by us (cf.

[100–102]). This theory is not verifiable neither is it falsifiable but, at least, it is hard to swallow. After all, we have to mention some attempts to get around Bell's inequality by altering some of the fundamental laws of probability theory, mainly the calculus for joint probabilities, and introducing a generalized probability space, nevertheless keeping probabilities restricted to the usual interval between 0 and 1 [103–108].

As the reader will already be expecting, Bell's inequality can also be circumvented by means of extended probabilities. This claim is valid for all inequalities of this type which have been derived yet, or will ever be derived, as can be demonstrated by a simple example reproducing exactly the quantum mechanical averages for measurements on beams of correlated electrons. The model is not a deterministic one and, hence, Wigner's contradiction (3.55) does not apply. The subsets considered, simply do not exist, because there are not 6 individually determined results, since determinism is not involved. One can speak of the outcome of a measurement but never be sure (except in special cases which belong to a probability value of 0 or 1) what the result of an experiment would have been, if it was not actually performed. Therefore, we are not in trouble with subsets which had to be considered elements of reality. The big disadvantage of this model is, however, that it does not satisfy (3.53) strictly in the single case, but only in the average case. This feature is due to any model of this kind, and at present there is no evidence whether and how this drawback can be removed. Nevertheless, with respect to the other "solutions" of the EPR paradox mentioned above, it seems appropriate at least to have a brief look on that one fitting our general topic.

Clearly there is an infinite number of "solutions" involving extended probabilities. We take the most simple one [109].

The probability for the result "spin up", denoted by $w_+(a, S)$, is assumed to be a linear function of the projection of the hidden variable (which in this case is the total spin S) on the direction a of measurement:

$$w_+(a, S) = C_1 + C_2 a \cdot S . \tag{3.56}$$

As the result of a measurement must be either "spin up" or "spin down", the respective probabilities must satisfy the condition

$$w_+(a, S) + w_-(a, S) = 1 . \tag{3.57}$$

From the symmetry of the problem we obtain

$$w_+(a, S) = 0.5 \quad \text{for} \quad a \cdot S = 0 \tag{3.58}$$

and, as repeated measurements in the same direction with certainty will reproduce the initial result, the last condition necessary is

$$w_+(a, S) = 1 \quad \text{for} \quad a \cdot S = \tfrac{1}{2}(\hbar) . \tag{3.59}$$

Consequently the probability functions read

$$w_+(a, S) = 0.5 + a \cdot S \quad \text{and} \quad w_-(a, S) = 0.5 - a \cdot S . \tag{3.60}$$

These probability functions ranging from $(1 - \sqrt{3})/2$ to $(1 + \sqrt{3})/2$ admit a correct calculation of averages of correlation measurements as described above. In addition they reproduce correctly the quantum mechanical averages concerning repeated measurements on polarized electron beams [110]. A similar treatment for the experimentally most interesting case of correlated photon pairs requires an extension of the usual interval of probability numbers to $[(1-\sqrt{2})/2, (1+\sqrt{2})/2]$ [110].

Further it has been shown [89] that quantum theoretic correlation results for particles with arbitrary spin can be obtained from classical distribution functions, which for large spin approximate δ-functions, but "they cannot, for any value of S, share with ordinary δ-functions the property of being everywhere non-negative. [. . .] they perfectly reproduce the quantum theoretic spin-S EPR correlations. But there is no way to give meaning to the phrase 'spinning about the direction n', and the interpretation as conditional probabilities suffers from the appearance of negative 'probabilities'. The functions are thus reminiscent of Wigner's phase-space 'distribution functions'.

The generalized Bell's theorem establishes that no matter how large S, the representation of the quantum theoretic spin S distribution [. . .] lacks the essential non-negativity of the analogous representation [. . .] of the classical distribution, that makes possible a simple interpretation in terms of conditional distributions" [89].

Of course, there are many physicists, who prefer to sacrifice locality if reality can be kept and negative probabilities are avoided, but it has been proved explicitly [111] that the concept of the total spin S to have a uniquely defined direction excludes that the quantum mechanical expectations can be obtained from any theory making use of this assumption, even if the theory is a *non-local and/or statistical* one (unless something else than the spin influences a spin measurement). Therefore reality cannot exist without extended probabilities. But it is open to question whether it can exist with, because of some strange consequences emerging from this special example. These, however will be left for the discussion.

3.4. Non-quantum indications

There are some examples even in classical physics, which either entail or at least can be interpreted in terms of negative probabilities, being however so simple, that negative probabilities involved are usually overlooked. We need not ambiguously intermingle negative processes with negative probabilities for these processes by, for example, identifying an absorption with an emission appearing with negative probability, in order to demonstrate this.

One interesting example (among others) is cited by Feynman [112], who considers the process of diffusion (like many physicists have been doing before him, but without noticing the following).

"A particle diffusing in one dimension in a rod has a probability of being at x at time t of $P(x, t)$ satisfying $\partial P(x, t)/\partial t = -\partial^2 P(x, t)/\partial x^2$. Suppose at $x = 0$ and $x = \pi$ the rod has absorbers at both ends so that $P(x, t) = 0$ there. Let the probability of being at x at $t = 0$ be given as $P(x, 0) = f(x)$. What is $P(x, t)$ thereafter? It is

$$P(x, t) = \sum_{n=1}^{\infty} P_n \sin nx \exp(-n^2 t) , \qquad (3.61)$$

where P_n is given by

$$f(x) = \sum_{n=1}^{\infty} P_n \sin nx , \qquad (3.62)$$

or

$$P_n = \frac{2}{\pi} \int f(x) \sin nx \, dx . \tag{3.63}$$

The easiest way of analyzing this (and the way used if $P(x, t)$ is a temperature, for example) is to say that there are certain distributions that behave in an especially simple way. If $f(x)$ starts as $\sin nx$ it will remain that shape simply decreasing with time, as $\exp(-n^2 t)$. Any distribution $f(x)$ can be thought of as a superposition of such sine waves. But $f(x)$ cannot be $\sin nx$ if $f(x)$ is a probability and probabilities must always be positive. Yet the analysis is so simple this way that no one has really objected for long" [112].

This example may be claimed to resemble quantum mechanics because of the close relationship between diffusion equation and Schrödinger equation, but the Planck constant clearly is not involved. It seems to be easy now, to find many similar indications, but we will restrict ourselves to one additional example.

The simplest solution describing the behaviour of the classical damped oscillator

$$d^2x/dt^2 + dx/dt + \omega_0^2 x = 0 \tag{3.64}$$

(the damping constant is set equal to unity) is given by

$$x(t) = x_0 \exp(-t/2) \cos \bar{\omega} t , \qquad \bar{\omega} = \sqrt{\omega_0^2 - \tfrac{1}{4}} . \tag{3.65}$$

In case of an electromagnetic oscillator, (3.65) may also be interpreted as the time-dependent amplitude of the field, observed (with the respective time delay not interesting in this context) at a certain point of space. In order to obtain the amplitude in terms of frequency, the usual Fourier transformation has to be applied. If, however, the considered point of space is shielded from the oscillator by a shutter, which opens at time $t = 0$, when the radiation starts, and closes at time t, we obtain for the observable part of the amplitude

$$A(\omega, t) \propto \int_0^t \exp(-\tau/2) \cos(\bar{\omega}\tau) \exp(i\omega\tau) \, d\tau , \tag{3.66}$$

or

$$A(\omega, t) \propto \frac{\exp[(i(\omega - \bar{\omega}) - \frac{1}{2})t] - 1}{i(\omega - \bar{\omega}) - \frac{1}{2}} + \frac{\exp[(i(\omega + \bar{\omega}) - \frac{1}{2})t] - 1}{i(\omega + \bar{\omega}) - \frac{1}{2}} . \tag{3.67}$$

If the damping is very low, i.e., $\bar{\omega} \gg 1$, the second term of (3.67) can be omitted for $\omega \gg 1$. The energy observed in the time interval $(0, t)$ then is given by

$$|A(\omega, t)|^2 \propto \frac{1 - 2\,e^{-t/2} \cos(\omega - \bar{\omega})t + e^{-t}}{(\omega - \bar{\omega})^2 + \frac{1}{4}} . \tag{3.68}$$

Formula (3.68) is the classical analogon of (3.45) and, hence, according to (3.46) and (3.46′) its time derivate will yield extended probabilities for small frequency intervals $\Delta\omega$ if the shutter is opened for short time intervals Δt only. It may be difficult even with modern opto-electronic devices to set up an appropriate experiment. But the validity of this classically derived relation has been demonstrated by a quantum effect, namely the Mössbauer effect (cf. section 3.2).

This section should not be finished without briefly referring to work of D.R. Cox [113] who found use even for complex probabilities in a theory of stochastic processes having discrete states in continuous time. A method invented by Erlang [114], who supposed that life has k stages, the time spent in each following independently an exponential distribution, is generalized by Cox to non-exponential distributions involving complex transition probabilities between the different stages. "Erlang's idea is particularly useful in fairly practical investigations because it gives a routine method of dealing with non-exponential distributions" [113]. The method can be applied to biological processes like division times of bacteria or economical problems like service times of customers in a queue. However, the subdivision of life into fictuous stages is merely a mathematical device and need not have physical significance. The same must hold for all macroscopic examples involving extended probability. In this realm extended probabilities emerge from a mathematical subdivision of an entirely positive function like (3.62) into a sum of functions, some of which can go negative without altering the entirely positive character of the original one. But it is as justified to deal with these subdivision functions as it is justified to represent the number 2 by $5+(-3)$.

On the quantum level of matter extended probabilities may be credited with a higher degree of reality. This is an open question. Nevertheless, as has been shown in the first three sections of this chapter, extended probabilities cannot a priori be banned from quantum theory, at least not without defining exactly what has to be understood by *the* quantum theory. This fact cannot be considered surprising, because one of quantum theory's most fundamental concepts generally is an inherently complex quantity and as such essentially different from, e.g., the complex quantities used in classical electrodynamical calculations. The latter are applied for convenience but can easily be avoided. Nevertheless the popularity of this computing tool may have assisted to the end that the complex wave function has been accepted without too much concern. The complex character of the wave function, although with respect to the description of reality eliminated by a mathematical trick, raises the question, whether (and if yes, which) properties of reality unavoidably cause it. Basically the wave function as fundamental means for the description of reality is scarcely more plausible than the formal use of extended probabilities. Its being defined not in real space and time but rather in complex Hilbert space does not too much improve this situation.

These arguments are far from attempting to doubt the extremely successful formalism of quantum theory. But it should be realized that – as soon as a relation to reality is concerned – the accepted foundations of this formalism are as obscure as its possible consequence discussed in this review.

4. Formal accounts

The commonly accepted interpretation of probability [115] is based on the relative frequency of events exhibiting a certain property (so-called successful trials, henceforth denoted by A) in a number N_0 of events (trials). Let

$$N_0 = N(\mathrm{A}) + N(\neg\mathrm{A}) \qquad (4.1)$$

with $-$A denoting the absence of A, then the probability of observing A is given by

$$p(\mathrm{A}) = \lim_{N_0\to\infty} N(\mathrm{A})/N_0 . \qquad (4.2)$$

This event-related definition has proved well in practical applications, it suffers however from serious philosophical objections (cf. H. Margenau in chapter 5), because the number of events observed or trials performed never can reach infinity and thus the experimental result can take each value between 0 and 1 in practical measurements. Of course there is a probability for the experiment to yield more or less exactly the true probability, but as long as finite sequences are considered, this and any further probability or level of confidence suffer from the same problem. (On the other hand most concepts applied in science can be verified only approximately (cf. M.S. Bartlett in chapter 5).)

In order to avoid the *event-related* character of the frequency interpretation the *fact-related* measure (or density) definition can be applied, if the facts are known. Given a box with 10 balls, 3 of them being red and the other 7 being not red, the relative measure (or density) of the property "red ball" is 0.3 and thus, the probability of finding a red ball by ideal "blind choice" is 0.3 too. The problem with this definition in physics is however, that in most cases the numbers of balls are unknown, and experiments based upon the measurement of relative frequencies are performed just in order to determine the measure or density of a property. In quantum mechanics, moreover, we cannot be sure that definite numbers of balls do exist at all (cf. section 3.3). If, however, the conditions are known and the ideal character of blind choice is guaranteed, the probability obtained from relative measure or density can be understood as a degree of confidence to find the desired result.

If reality was wholly deterministic, probabilities could be applied due to our lack of knowledge only, because every event would have the inherent probability 0 or 1. Referring to the difficulties with quantum theory (especially the EPR paradox, cf. section 3.3) we cannot take this viewpoint. Therefore it seems reasonable, to attribute a probability number to each individual case, reflecting the objective probability for a certain result (the time interval during which a radioactive atom will decay and the direction in which it will emit its decay product, for instance). Of course subjective probabilities (due to unknown parameters which, however, in principle could be determined – whether there is a radioactive atom, for instance) have also to be taken into account in real experiments, but they will not be subject to the following considerations.

Obviously the probability of a certain result will depend not alone on the intrinsic objective state of a physical entity being analyzed, but also on the conditions under which this is done. Hence the whole system has to be taken into account. Herewith we arrive at Popper's "propensity definition". This definition appears to be a very reasonable one (Popper illustrates it by a loaded die, the probability of which to show up the biased number will depend on the strength of the gravitational field present [116]), but the existence of an objective intrinsic state unavoidably implies extended probabilities (a nice example hereof is discussed in [117]). Nevertheless those propensity-like probabilities may be handled by means of the frequency interpretation as follows.

Consider a number N_0 of particles prepared in a pure quantum state, but, deviating from quantum theory, possessing hidden variables λ_i, $i = 1, \ldots, n \le N_0$. For the sake of brevity assume, that the different values λ_i are taken with equal positive probability p_i. The generalization to an inhomogeneously occupied space of hidden variables follows by a straight forward treatment. By an interaction the nature of which need not be specified the values of these hidden variables change from λ_i to λ'_j, $j = 1, \ldots, m \le N_0$, the transition probability being denoted by p_{ij}. By this interaction the pure state may split into $k \leq m$ experimentally distinguishable states one of which is called A. Those values of λ'_j which form this state are denoted by $\lambda'_{j(A)}$. The result of a measurement exhibits $N(A)$ particles in the state A and $N(-A)$ particles in other states $(-A)$. Conservation of particle number requires $N(A) + N(-A) = N_0$. The combined transition probability for state A follows from the experiment in case $N_0 \to \infty$ by

$$p(A) = \frac{1}{n} \sum_i \sum_{j(A)} p_{ij} = \lim_{N_0 \to \infty} N(A)/N_0 . \tag{4.3}$$

If the single transitions are admitted to interfere with each other, it is easy to extend the defining equation of probabilities to that end that not only "normal events" are covered, but also events by which normal ones are eliminated or quenched. By this means – in close analogy to theory of matter – two kinds of events are introduced, namely normal and quenching events. As the individual p_{ij} – due to the uncertainty principle – may be not observable, some of them may exceed the usual range between 0 and 1. A negative probability of an event is interpreted in this formalism as a positive probability for the appearance of a quenching event; a probability exceeding 1 indicates that sometimes more than one transition may happen from a substate populated by only one particle.

While in this picture the usual, macroscopically observable probabilities are restricted to $0 \leq p(A) \leq 1$, the individual transition probabilities only have to obey the normalization

$$\frac{1}{n} \sum_i \sum_j p_{ij} = 1 \tag{4.4}$$

and the left side of (4.3) for distinguishable results. At first sight, one would say that individual transition probabilities are not observable, because of the uncertainty principle. But – as discussed in connection with the time-resolved photon decay – sophisticated devices employing the Mössbauer effect may be capable of detecting something like that.

While this interpretation of extended probabilities by means of quenching events assumes a concrete physical mechanism, a more abstract and, thus, more general approach to introduce extended probabilities makes use of moment-generating and characteristic functions. This access was analysed in a short paper by Bartlett [118]. Before coming to this work, we would like to remind the reader of some definitions and technical expressions.

The function

$$F_r(z) = \sum_{r=0}^{\infty} z^r P(r) , \quad z \in \mathbb{C} \quad \text{and} \quad |z| \leq 1 \tag{4.5}$$

is called *probability-generating function* of the random variable r or of its probability distribution $P(r)$, for the probability distribution is uniquely defined by

$$P(r) = \frac{1}{r!} \left[\frac{\mathrm{d}^r}{\mathrm{d}z^r} F_r(z) \right]_{z=0} . \tag{4.6}$$

Further, for fixed z, the probability-generating function is equal to the expectation value of the random variable z^r:

$$F_r(z) = E\{z^r\}\,. \tag{4.7}$$

For p the probability for a successful single trial, $q = 1 - p$ the complementary probability and n the number of (independent) trials making up the random variable r, another expression for $F_r(z)$ is

$$(pz + q)^n\,. \tag{4.8}$$

The equivalent *moment-generating function* $M_r(t)$ is obtained by substituting e^t for z in $F_r(z)$; and, as suggested by its name, the variance and higher moments of the distribution can be calculated from it. For example, as

$$M_r(t) = \sum_{r=0}^{\infty} e^{tr} P(r) \tag{4.9}$$

for a Bernoullian scheme, the expectation is

$$E\{r\} = \left[\frac{d}{dt} M_r(t)\right]_{t=0}\,,$$

and similarly for further moments.

A further advantage of the moment-generating function is its usual availability for continuous random variables; in fact it is always available (even for non-finite moments) if the exponent is imaginary [119]. The function

$$M_r(t) = E\{\exp(irt)\} \quad \text{with} \quad -\infty < t < \infty \tag{4.10}$$

is then called the *characteristic function* of r.

The short paper by Bartlett [118] entitled "*Negative Probability*" now will be given in some detail. After mentioning the work of Dirac [1] and Moyal ([18], unpublished at that time), Bartlett concludes that it may be "in the highly abstract symbolism of physics that this extension is most likely to prove useful. Moreover, it is in the same field that the relation between the probability theory and its physical interpretation, while tending to agree with the natural viewpoint [1] that the extended theory merely corresponds to a more elaborate accounting system involving debit as well as credit entries, is also a warning not to ascribe individual reality to negative entries in this theoretical book-keeping." Then he considers the probability equation $p = p_1 + p_2$ and applies it to the case of a biased penny with $p_1 = \frac{1}{2}$ and p_2 a correcting term. "We are naturally led to ask whether, if $p < \frac{1}{2}$, and $p_2 < 0$, we can still in any consistent sense regard $p = p_1 + p_2$ as a probability equation, provided we admit negative numbers into the domain of admissible probability numbers." [...] "Since the formal structure for the distributions is unaltered, it is trivial to show that the probabilities in the extended theory must for consistency obey the same rules as before, that is, the addition and multiplication laws. In particular, the familiar method of deriving the distributional moments for binomial frequencies r_1 and r_2 corresponding to n independent trials for the event represented (by above probability equation), by forming the expectation

$$E\{\exp(r_1 t_1 + r_2 t_2)\} = (1 - p + p_1 \exp t_1 + p_2 \exp t_2)^n \tag{4.11}$$

whence are obtained the well-known results

$$\bar{r}_2 \equiv E\{r_2\} = np_2\,, \qquad \sigma^2(r_2) \equiv E\{(r_2 - \bar{r}_2)^2\} = np_2(1 - p_2)\,,$$

$$\operatorname{cov}(r_1, r_2) \equiv E\{(r_1 - \bar{r}_1)(r_2 - \bar{r}_2)\} = -np_1p_2\,,$$

etc., is also valid for p_2 negative.

Having set up a consistent extended theory which still retains the same rules as the orthodox theory, we may next ask what physical interpretation these more general probability numbers can have. In the orthodox theory mathematical limit theorems provide the appropriate theoretical relations between probability numbers and limiting frequency ratios (see, for example, [120]). It is, however, not clear how r_2/n can tend to p_2 in any sense in the extended theory when p_2 is negative since the formal distribution of r_2 only covers the positive range 0 to n. The usual proof (e.g., on p. 9 of my previous paper [120]) in fact breaks down for negative probabilities. Notice, moreover, that the partition or probability-generating function for the distribution of r_2 alone,

$$(1 - p_2 + p_2z_2)^n \tag{4.8'}$$

consists of a number of terms of alternating sign for negative p_2, the sum of the absolute values of the coefficients being $(1 + 2|p_2|)^n$ and tending to infinity with n.

Thus probabilities in the original range 0 to 1, as we might reasonably expect, still retain their special significance. It is only these probabilities which we can immediately relate with actual frequencies; it is only these probabilities, for example, for which the theoretical frequency ratio r/n tends to p with probability one, as n tends to infinity."

The next chapter is entitled "Extraordinary random variables and their characteristic functions", in which is reasoned that the use of negative probabilities "is still possible through their moment-generating or characteristic functions, which still have proper limits. We have, for example, the cumulant function of $r_2/\sqrt{n}$ defined by

$$\log E\{\exp r_2t/\sqrt{n}\} \to \sqrt{n}\cdot p_2t + \tfrac{1}{2}p_2q_2t^2\,, \tag{4.12}$$

where $q_2 = 1 - p_2$." (When expanding the logarithm and the exponent, all higher order terms vanish with $n^{-1/2}$ or faster.) "We may consistently think of $r_2/\sqrt{n}$ as a random normal or Gaussian variable with mean $\sqrt{n}\cdot p_2$ and standard deviation $\sqrt{(p_2q_2)}$. For p_2 negative, this random variable has an imaginary standard deviation; it will be termed, using Eddington's convenient adjective [121], an *extraordinary* random variable." Considering the joint distribution of r_1 and r_2, "the variable $r/\sqrt{n}$ becomes a normal random variable with two correlated components $r_1/\sqrt{n}$ and $r_2/\sqrt{n}$, of which $r_2/\sqrt{n}$ has an imaginary standard deviation, and imaginary correlation with $r_1/\sqrt{n}$.

One further result will be noted. The characteristic function equation

$$M(t) = M_1(t)M_2(t)\,, \tag{4.13}$$

where $M(t) \equiv E\{\exp \mathrm{i}tx\}$, etc., implies that the random variable x is the sum of two independent components x_1 and x_2. Given $M(t)$ and $M_1(t)$, this equation has the solution

$$M_2(t) = M(t)M_1^{-1}(t)\,, \tag{4.13'}$$

[??] if $M_1^{-1}(t)$ can be interpreted as the characteristic function of some random variable. Since $M_1(t)M_1^{-1}(t) = 1$, $M_1^{-1}(t)$ will be said to define the *reciprocal* distribution to $M_1(t)$. This distribution is only an admissible one in our extended theory, since it corresponds to an extraordinary independent random variable y, such that the random variable $x_1 + y$ is certainly zero. A binomial distribution with negative probability p_2 gives rise to a well-behaved reciprocal distribution $(q_2 + p_2 z)^{-n}$, which has the form of the so-called 'negative binomial distribution'. Even on the present theory, however, $M_1^{-1}(t)$ does not necessarily define a distribution function obtainable from it by Fourier inversion."

Bartlett summarizes: "Negative probabilities must always be combined with positive ones to give an ordinary probability before a physical interpretation is admissible. This suggests that where negative probabilities have appeared spontaneously in quantum theory it is due to the mathematical segregation of systems or states which physically only exist in combination." (See also M.S. Bartlett in chapter 6.) This rather general treatment gives an impression of how extended probabilities could appear and be handled in the framework of the frequency interpretation generally, although no physical interpretation is intended.

Turning to the special field of quantum electrodynamics, an account of solving the problems referred to in section 3.2 may be given, following an example of Feynman [112].

"In the quantum theory of electrodynamics, the free photon moving in the z direction is supposed to have only two directions of polarization transverse to its motion x, y. When this field is quantized, an additional interaction, the instantaneous Coulomb interaction, must be added to the virtual transverse photon exchange to produce the usual simple

$$(j_x j'_x + j_y j'_y + j_z j'_z - j_t j'_t)\, e^2/q^2 \tag{4.14}$$

virtual interaction between two currents, j and j'. It is obviously relativistically invariant with the usual symmetry of the space j_x, j_y, j_z and time j_t components of the current (in units where the velocity of light, is $c = 1$). The original starting Hamiltonian with only transverse components does not look invariant. Innumerable papers have discussed this point from various points of view but perhaps the simplest is this. Let the photon have *four* directions of polarization of a vector x, y, z, t no matter which way it is going. Couple the time component with ie instead of e so that the virtual contribution for it will be negative as required by relativity in eq. (4.14). For real photons, then, the probability of a photon emission is negative, proportional to $-|\langle \mathrm{f}|j_t|\mathrm{i}\rangle|^2$ the square of the matrix element of j_t between initial and final states, just as the probability to emit an x photon is $+|\langle \mathrm{f}|j_x|\mathrm{i}\rangle|^2$. The total probability of emitting any sort of photon is the algebraic sum of the probabilities for the four possibilities,

$$|\langle \mathrm{f}|j_x|\mathrm{i}\rangle|^2 + |\langle \mathrm{f}|j_y|\mathrm{i}\rangle|^2 + |\langle \mathrm{f}|j_z|\mathrm{i}\rangle|^2 - |\langle \mathrm{f}|j_t|\mathrm{i}\rangle|^2 . \tag{4.15}$$

It is always positive, for by the conservation of current there is a relation of j_t and the space components of j, $k_\mu j_\mu = 0$ if k_μ is the four-vector of the photon. For example, if k is in the z direction, $k_z = \omega$, and $k_x = k_y = 0$ so $j_t = j_z$ and we see eq. (4.15) is equal to the usual result where we add only the transverse emissions. The probability to emit a photon of definite polarization e_μ is (assume e_μ is not a null vector)

$$-|\langle \mathrm{f}|j_\mu e_\mu|\mathrm{i}\rangle|^2/(e_\mu e_\mu) . \tag{4.16}$$

This has the danger of producing negative probabilities. The rule to avoid them is that only photons whose polarization vector satisfies $k_\mu e_\mu = 0$ and $e_\mu e_\mu = -1$ can be observed asymptotically in the final or initial states. But this restriction is not to be applied to virtual photons, intermediary negative probabilities are not to be avoided. Only in this way is the Coulomb interaction truly understandable as the interchange of virtual photons, photons with time-like polarization which are radiated as real photons with a negative probability.

This example illustrates a small point. If one t photon is emitted with a negative probability $-\alpha$ ($\alpha > 0$), and another t photon is emitted say independently with probability $-\beta$ ($\beta > 0$), the chance of emitting both is positive $(-\alpha)(-\beta) = \alpha\beta > 0$. Should we not expect then to see physical emission of two such photons? Yes, but (if these photons are moving in the z direction) there is a probability to emit z photons α and β also, and there are four emission states: two t photons with probability $+\alpha\beta$; two z photons with probability $+\alpha\beta$; the first z and second t probability $(+\alpha)(-\beta) = -\alpha\beta$ and the first t second z with probabilities $-\alpha\beta$ so again, for total emission rate only transverse photons contribute" [112].

This physical interpretation of extended probabilities in photon emission processes comes very close to that one based on quenching events described in the first part of chapter 4, sharing with it the problem of how to explain that probabilities always interfere precisely enough to prevent negative events from being identified directly.

While this crucial question must be left to further study, we shall now describe the formalism of indefinite metric, which is a useful tool to handle extended probabilities appearing in quantum electrodynamics. This account was invented by Gupta [122] and has been repeated in his recent textbook [47] from which the following paragraphs are excerpted.

After stating that the usual formalism of quantum mechanics is not adequate to deal with all the fields in nature, Gupta develops his more general formalism, considering a space of infinite dimensions in which the components of vectors are in general complex numbers. "[...] let us denote any vector in this space by ψ or Φ. With this space let us associate a dual space such that corresponding to any vector ψ there exists a vector ψ^* in the dual space. Let us assume that the relationship between the two spaces is such that

$$(\psi + \Phi)^* = \psi^* + \Phi^* , \qquad (c\psi)^* = c^*\psi^* , \tag{4.17}$$

$$\Phi^*\psi = \text{number} = (\psi^*\Phi)^* , \tag{4.18}$$

where c is a number. It follows from (4.18) that

$$\psi^*\psi = \text{real number}. \tag{4.19}$$

The only difference between the above assumptions and those of [usual quantum mechanics] is that now the quantity $\psi^*\psi$ is not positive definite, but, according to (4.19), it can be positive, negative, or zero. We shall, therefore, refer to the space under consideration as an indefinite-metric space" [47].

Vectors ψ and ψ^* satisfying the relations (4.17) to (4.19) are called Hermitian conjugates of each other, and the real number $\psi^*\psi$ is called the norm of the vector ψ. Any vector satisfying the condition

$$\psi^*\psi = \pm 1 \quad \text{or} \quad 0 \tag{4.20}$$

is said to be normalized, because we cannot change the sign of the norm of a vector by multiplying it by any number.

If an operator $\hat{A}$ is such that

$$\hat{A}^* = \hat{A}\,, \tag{4.21}$$

it is called Hermitian. The expectation value of a dynamical variable is given by

$$\langle \hat{A} \rangle = \psi^* \hat{A} \psi\,, \tag{4.22}$$

where the operator $\hat{A}$ represents the dynamical variable, and ψ is the normalized state vector. It follows from (4.22) that all quantities of physical interest must be represented by Hermitian operators in order that they have real expectation values. If a is an eigenvalue of $\hat{A}$ belonging to the eigenstate ψ, in usual quantum theory

$$(a - a^*)\psi^* \psi = 0 \tag{4.23}$$

accounts for the reality of eigenvalues of Hermitian operators. Since according to (4.19) the norm $\psi^*\psi$ may vanish, this conclusion is not generally valid here. But we can still infer from (4.23) that all eigenvalues of an Hermitian operator corresponding to eigenvectors of nonvanishing norms are real. "Similarly, we cannot now establish in general the orthogonality of all eigenvectors of an Hermitian operator. But, by following the [usual] treatment, we can still show that those eigenvectors of an Hermitian operator that have non-vanishing norms form an orthogonal set, provided that the eigenvectors corresponding to the same eigenvalue are chosen in a suitable way.

We shall assume that if all the eigenvectors of an Hermitian operator have non-vanishing norms, the orthogonal set of eigenvectors formed by them is a complete set. Then, if the eigenvalues of an Hermitian operator $\hat{A}$ are discrete and all its eigenvectors have non-vanishing norms, we can express any normalized vector Φ as

$$\Phi = \sum_n c_n \psi_n\,, \tag{4.24}$$

where the c_n are numbers, and the ψ_n are orthonormal eigenvectors of $\hat{A}$ corresponding to the eigenvalues a_n, so that

$$\psi_m^* \psi_n = \pm\delta_{mn}, \tag{4.25}$$

where the plus and minus signs correspond to the eigenvectors of positive and negative norms, respectively." Be $f(\hat{A})$ an arbitrary function. Instead of the usual expansion

$$\langle f(\hat{A}) \rangle = \sum_n c_n^* c_n f(a_n) \tag{4.26}$$

we have in the present case

$$\langle f(\hat{A}) \rangle = \sum_n \pm\, c_n^* c_n f(a_n)\,, \tag{4.27}$$

which gives

$$P_n = \pm c_n^* c_n, \tag{4.28}$$

"where P_n is the probability of obtaining a_n as the result of measurement of $\hat{A}$. The result (4.28) shows that vectors with a negative norm represent states occurring with a negative probability. It should be noted that if the vector ψ_r has a negative norm, the expectation value of $\hat{A}$ in this state is $-a_r$. But we cannot regard ψ_r as representing a state of positive probability corresponding to the observable value $-a_r$, because in this state

$$\langle f(\hat{A})\rangle = -f(a_r) \neq f(-a_r). \tag{4.29}$$

[...] Finally, it must be observed that because of some difficulties of physical interpretation, the formalism of quantum mechanics with an indefinite metric can be used only under special circumstances. The main difficulty of physical interpretation arises from the fact that a vector with a negative norm represents a state occurring with a negative probability, which is of course meaningless. [...] Since the above difficulties of physical interpretation are connected with vectors of negative or vanishing norms, an indefinite metric can be used for a quantum mechanical system only when the states represented by vectors of negative or vanishing norms are physically unobservable" [47].

Summarizing: we have (i) proposed a general interpretation of the appearance of extended probabilities on the microscopical level of reality, (ii) presented an account of extended probabilities based upon general probability theory, (iii) reviewed an example for the field of quantum electrodynamics, and (iv) described a formalism of handling extended probabilities in the quantum theoretical framework. All these approaches are obtained from the frequency- or measure-based definitions of probability. These definitions, however, are not the only ones possible, and in some cases they even fail to yield exact results, although probabilities can be defined uniquely. As this fact may turn out to be useful when considering extended probabilities, a different and more general approach, although not dealing with extended probabilities, should be mentioned in the present context. It is based on the invariance principle and has served Jaynes' in solving Bertrand's paradox. The following paragraphs are cited from Jaynes' paper: "*The well-posed problem*" [123], in order to illuminate this principle by means of one of its most striking applications.

"Many statistical problems, including some of the most important for physical applications, have long been regarded as undetermined from the standpoint of a strict frequency definition of probability; yet they may appear well posed or even overdetermined by the principles of maximum entropy and transformation groups. [...] On the viewpoint advocated here, Bertrand's problem turns out to be well posed after all, and the unique solution has been verified experimentally." (This verification, of course has made use of the frequency interpretation.) "We conclude that probability theory has a wider range of useful applications than would be supposed from the standpoint of the usual frequency definitions."

Bertrand's problem, originally stated in terms of drawing a straight line "at random" intersecting a circle, may be formulated as follows in a more concrete way.

"A long straw is tossed at random onto a circle; given that if falls so that it intersects the circle, what is the probability that the chord thus defined is longer than a side of the inscribed equilateral triangle? Since Bertrand proposed it in 1889 this problem has been cited to generations of students to demonstrate that Laplace's 'principle of indifference' contains logical inconsistencies. For, there appear to be many ways of defining 'equally possible' situations, and they lead to different results. Three of these are: Assign uniform probability density to (A) the linear distance between centers of chord and circle, (B) angles of intersections of the chord on the circumference, (C) the center of the chord over the interior area of the circle. These assignments lead to the results $p_A = \frac{1}{2}$, $p_B = \frac{1}{3}$, and $p_C = \frac{1}{4}$, respectively. Which solution is correct?" While most mathematicians cited in [123] believe that the problem is ill posed and, hence, has no definite solution, Jaynes demonstrates that it is possible to obtain the resolution by "*pure thought*".

"Bertrand's problem has an obvious element of rotational symmetry, recognized in all the proposed solutions; however, this symmetry is irrelevant to the distribution of chord lengths. There are two other 'symmetries' which are highly relevant: Neither Bertrand's original statement nor our restatement in terms of straws specifies the exact size of the circle, or its exact location. If, therefore, the problem is to have any definite solution at all, it must be 'indifferent' to these circumstances; i.e., it must be unchanged by a small change in the size or position of the circle. This seemingly trivial statement, as we will see, fully determines the solution" [123].

Observing the mentioned invariances, for a circle of radius R, Jaynes obtains the unique random distribution of chord lengths. "A chord whose midpoint is at (r, θ) has a length $L = 2(R^2 - r^2)^{1/2}$. In terms of the reduced chord lengths, $x \equiv L/2R$, we obtain the universal distribution law"

$$p(x)\,dx = x\,dx/\sqrt{1-x^2}\,, \qquad 0 \le x < 1\,, \tag{4.30}$$

corresponding to the solution A.

"Any rain of straws which does *not* produce a frequency distribution agreeing with (4.30) will necessarily produce different distributions on different circles. But this is all we need in order to predict with confidence that the distribution (4.30) *will* be observed in any experiment where the 'region of uncertainty' is large compared to the circle. [...]

These conclusions seem to be in direct contradiction to those of von Mises [115, 124], who denied that such problems belong to the field of probability theory at all. It appears to us that if we were to adopt von Mises' philosophy of probability theory strictly and consistently, the range of legitimate physical applications of probability theory would be reduced almost to the vanishing point. Since we have made a definite, unequivocal prediction, this issue has now been removed from the realm of philosophy into that of verifiable fact. The predictive power of the transformation group method can be put to the test quite easily in this and other problems by performing the experiments.

The Bertrand experiment has, in fact, been performed by Professor E.T. Jaynes and Dr. Charles E. Tyler, tossing broom straws from a standing position onto a 5 inch diameter circle drawn on the floor. Grouping the range of chord lengths into ten categories, 128 successful tosses confirmed eq. (4.30) with an embarrasingly low value of chi-squared. However, experimental results will no doubt be more convincing if reported by others" [123].

The reason for extensively citing this example is, besides its beauty, the fact that there are other physical problems which are not verifiable by means of the frequency interpretation, nevertheless being subject to defining probability densities. To quote a last time from Jaynes' paper: "Bertrand's paradox has greater importance than appears at first glance, because it is a simple crystallization of a deeper paradox which has permeated much of probability theory from its beginning. In 'real' physical applications when we try to formulate the problem of interest in probability terms we find almost always that a statement emerges which, like Bertrand's, appears too vague to determine any definite solution, because apparently essential things are left unspecified" [123].

If we reject the principle of indifference, e.g., in kinetic gas theory, and insist that the only valid basis for assigning probabilities is frequency in some random experiment, it would again appear that the only way of determining quantities like velocity distribution, pressure fluctuations, etc., is to perform the experiments. With regard to the uncertainty principle and, even more significantly, with regard to extended probabilities, it appears however, that some probabilities assigned to physical quantities on grounds of invariances or other general reasoning, are not verifiable by means of experiments determining relative frequencies. As Jaynes' treatment of Bertrand's problem suggests, such probabilities need not a priori be meaningless.

Definite answers to these questions cannot be given at present but may be obtained by further physical and philosophical research. Those who insist, that probabilities are meaningless without a frequency definition, may be referred to Dirac's very pragmatic interpretation at least of small negative or even complex probabilities, which does not attempt to look behind the curtain, but may serve as a good guide. When developing a theory of functions of non-commuting variables, Dirac obtained a formal probability for these observables to have numerical values: "*This probability turns out to be in general a complex number, but all the same it has some physical meaning, since when it is close to zero one can say that the numerical values are unlikely*" [125].

5. Statements

The first chapters of this review may arouse the (justified) impression of being a pleading for extended probabilities. This is because to plead against this concept would resemble carrying coals to Newcastle. Mainly examples, facts, or considerations in one or the other way suggesting or permitting the use of extended probabilities have been described; and it is superfluous to mention that the overwhelming domain of reality gives no reason for considering this concept.

In order to restore a certain degree of balance, some paraphrases of distinguished scientists about the present topic are quoted below. Also some statements dealing not directly with extended probabilities, but illuminating some points touched in this review are included. Ideas devoted to extensions of probability theory other than by dropping Kolmogorov's axiom, however, have been omitted. The reader interested in this wide field is referred to Gudder's text book [104].

The statements are given in alphabetical order, and, of course, no kind of completeness whatsoever can be claimed.

A. Aspect: "I know that several smart physicists (among them Richard Feynman) have considered the negative probabilities as an issue to the EPR problem. As a simple-minded experimentalist, I can hardly accept such a solution. As a matter of fact, in the experiments that we have performed, a negative probability means that an event that has been detected, and *stored* in the memory of the computer, will be *erased* later, when the event with a negative probability will occur! On the other hand, any attempt to 'understand' with a simple picture the EPR situation raises hard problems" [126].

M.S. Bartlett: "Statistical probabilities [...] may be termed objective in the sense that while we cannot say they exist exactly in the real world they exist outside ourselves *in the theory*, and can ideally be measured. The criticism has sometimes been made that this ideal measurement does not exactly correspond to anything possible in practice, since it involves an infinite series of trials, but this kind of criticism is not peculiar to statistics. Theoretical methods of measuring other quantitative scientific concepts will be found to be only approximately realisable, owing to the inevitable idealisation involved in the theory" [127].

A.O. Barut: "I first got involved in negative metric in connection with the quantization of higher-order spinor equations in one of my first papers [128, 129]. The problem of indefinite metric always arises if one tries to make a field theory of several fermions. I believed we have made recently definite progress by showing that negative norm states are separated from those of positive norm by a superselection rule [130]. Thus negative metric is now a virtue instead of a disease in explaining for example, the absence of $\mu \to e + \gamma$ decay.

As to the EPR problem, [... in] recent calculations I made with P. Meystre [131] we show here that a classical model exists giving exactly the same correlations as quantum mechanics, thus the *program* set by Bell and others in the form $p(a, b) = \int d\lambda\, p(a, \lambda) p(b, \lambda)$ is not a realizable program: the abstract theoretical assumptions are inadequate, not the experiment, nor the classical model or the quantum theory!" [132].

The probabilities used in [131] are the same as given in [109]; cf. eq. (3.60). For a more general treatment see [89].

J.S. Bell: "Unfortunately I cannot think of anything intelligent to say about negative probability, or indeed about the square circle. [There is a] very recent paper by Feynman in which he finds a use for negative probabilities at intermediate steps provided they cancel out at the end. I have great difficulty dividing the world sharply into intermediate steps on the one hand and ends on the other!" [133].

F. Bopp: "As is well known there exists a transformation invented by Wigner, which converts each von Neumannian state-matrix $P = \langle r|P|r'\rangle$ uniquely into a phase-space state-function $f(p, r)$ and vice versa. In case of the harmonic oscillator, the equation defining $f(p, r)$ is identical to Liouville's equation. Though the domains for $\langle u^{+}|P|u\rangle \geq 0$ and $f(p, r) \geq 0$ are different, hence negative averaging weights can appear. This is closely related to the fact, that the second inequality in quantum-physics has to be replaced by

$$\int f(p, r)^2\, d^3p\, d^3r \leq \text{const}/\hbar^3\,.$$

This is a new formulation of the uncertainty relation, independent of the choice of coordinates [134].

Stimulated by works of Heisenberg I once investigated how theories involving negative probabilities can be formulated in Hilbert spaces with finite dimensionality. I remember I have shown but not published that negative probabilities can be eliminated by the requirement (due to Heisenberg) that they do not appear experimentally" [135].

R.F. Bordley: "Most physicists currently feel that quantum behavior such as the n-slit interference experiment requires an abandonment of standard probability theory. This seems rather peculiar because other fields [...] have found standard probability theory perfectly adequate for their use. [...] quantum physics does not, in fact, require an abandonment of standard probability theory" [136].

L. de Broglie: "[...] von Neumann's demonstration implies a hypothesis that is not absolutely unavoidable, and that is not substantiated in the causal theory in question. This hypothesis is that when a system is in a state ψ the distributions of probabilities defined by Wave Mechanics are valid *prior to* any act of measurement. Now, if these laws of probability are stated in the way they should be (for example, $|c_k|^2$ is the probability that an exact measurement of the quantity A will furnish the value a_k), it is immediately clear that they are valid only after the measurement has been performed and prior to the knowledge of the result" [137].

A.A. Broyles: "We shall see here that it is not necessary to *postulate* that $\psi^*(\bar{x})\psi(\bar{x})$ is the coordinate probability distribution. This fact will be *derived* from the mathematics of quantum mechanics with the aid of a coordinate-measuring thought experiment" [138].

S. Bugajski: "Joint probability distributions for uncompatible observables should be non-classical, hence the seemingly most natural property (i) should be rejected" [139]. [(i) expresses that the joint probability distribution for observables A_1, A_2 in a pure state is a probability measure on the real plane.]

W. Döring: "The notion 'probability of a certain measurement result' is in my opinion explained according to quantum mechanics as follows. The same measurement has to be repeated very frequently on the same system in the same state. Then it is simply counted how often the result in question has happened. The ratio of this number and the total number of measurements in the case of a very high number of the latter is called probability. As I do not know, how to visualize a negative number of measuring results, a negative probability has no sense for me. Therefore, in my opinion, that is 'nonsense'. Clearly, this does not exclude the appearance of quantities in quantum mechanics, which can become negative and are called probabilities. I think, who wishes to use this nomenclature, has to explain what is to be understood by this probability, because the explanation known to me then does not make sense" [140].

B. d'Espagnat (in the framework of a proof of the inseparability of quantum theory): "On the other hand, a standard rule of probability theory (the validity of which is here beyond doubt since the probabilities we consider are defined as limits of frequencies on ensembles of events that, being observational results, necessarily obey classical logic) gives [...]" [141].

R.P. Feynman: "If a physical theory for calculating probabilities yields a negative probability for a given situation under certain assumed conditions, we need not conclude the theory is incorrect. Two other possibilities of interpretation exist. One is that the conditions (for example, initial conditions) may not be capable of being realized in the physical world. The other possibility is that the situation for which the probability appears to be negative is not one that can be verified directly. A combination of these two, limitation of verifiability and freedom in initial conditions, may also be a solution to the apparent difficulty. [...] It is not our intention to claim that quantum mechanics is best understood by going back to classical mechanical concepts and allowing negative probabilities (for the equations for the development of [a phase-space distribution] in time are more complicated and inconvenient than those of ψ). Rather we should like to emphasize the idea that negative probabilities in a physical theory does not exclude that theory, providing special conditions are put on what is known or verified. But how are we to find and state these special conditions if we have a new theory of this kind? It is that a situation for which a negative probability is calculated is impossible, not in the sense that the chance for it happening is zero, but rather in the sense that the assumed conditions of preparation or verification are experimentally unattainable."

With regard to a classical treatment of the double-slit experiment Feynman concludes: "With such formulas all the results of quantum statistics can be described in classical probability language, with states replaced by 'conditions' defined by a pair of states (or other variables), provided we accept negative values for these probabilities. This is interesting, but whether it is useful is problematical, for the equations with amplitudes are simpler and one can get used to thinking with them just as well" [112].

N. Grossman: "The quantity $C_n^* C_n$ is clearly a real positive number less than or equal to one, and is interpreted as the probability that a measurement made on a system whose initial state is ψ will yield the eigenvalue corresponding to the eigenfunction U_n" [117].

S.P. Gudder: "Bell concludes that locality, reality or hidden variables must be rejected. [...] There is another alternative which has been missed in the above arguments. The above arguments not only rely upon the concepts of reality, locality, and quantum theory, but also upon probability theory. We contend that reality, locality and the predictions of quantum mechanics can be retained if we alter our concept of probability theory. We argue that the framework of conventional probability theory is too restrictive to describe subatomic systems and must be extended. This extension does not contradict traditional probability theory, but merely enlarges its scope" [108].

J.M. Jauch, after celebrating Heisenberg's important contributions to modern science, he writes in a review of Heisenberg's book [142]: "Since in a later stage of the development of the theory Heisenberg runs into mathematical difficulties, he postulates that the metric of the Hilbert space in which the basic field operates is not positive definite. While it is not at all certain that the mathematical difficulties can be overcome by this artifact, it is interesting to examine the reasons which Heisenberg adduces for this assumption. [...]

As an example of a similar situation in quantum electrodynamics, the so called Gupta–Bleuler method is mentioned, which permits a formal construction of normalizable states satisfying the subsidiary condition and a representation of gauge transformation. Another example is encountered in the Lee model where the appearance of 'ghost' states for sufficiently large coupling constants seem to require a similar hypothesis in a formal treatment of the problem.

In the opinion of this reviewer, such reasoning for the indefinite metric of the Hilbert space is not very convincing. There seems to be a confusion between the definiteness of a quadratic form which is left invariant under a group and the metric which characterizes the physical Hilbert space. The two are not necessarily related. On the other hand, the definiteness of the metric in Hilbert space seems to be required for a physical interpretation of the theory. [...]

What all this shows is that Heisenberg's contention that the indefinite metric is essential cannot be correct. It may at best be a mathematical convenience, but certainly not essential. This is very well demonstrated by precisely the prime example of Heisenberg's, the Gupta–Bleuler method. It is well known that the electromagnetic field can also be quantized with the subsidiary condition in a space with definite metric. Thus, just in this example, the indefinite metric is not at all necessary; it gives at best a slight formal advantage.

As to the example of the Lee model, the situation here is one of mathematical ambiguity in the definition of the Hamiltonian as a self-adjoint operator. The occurrence of 'ghosts' in this theory is, as modern treatments have shown, merely an indication that the formal expression of this Hamiltonian cannot be interpreted as a self-adjoint operator. To attribute any deep physical significance to 'ghosts' and the associated indefinite metric seems to this reviewer to attach too much importance to a mathematically inadequate model theory" [143].

E.T. Jaynes: "In all applications of probability theory, basically the same controversy rages over whether our probabilities represent real situations, or incomplete human knowledge.

If the wave function of an electron is an 'objective' thing, representing a real physical situation, then it would be mystical – indeed, it would require a belief in psychokinesis – to suppose that the wave function can change, in violation of the equations of motion, merely because information has been perceived by a human mind.

If the wave function is only 'subjective' [...], then a new difficulty appears; the relative phases of the wave function at different points have not been determined by our information; yet they determine how the electron moves. [...]

As many have pointed out [...], the Copenhagen interpretation of quantum theory not only denies the existence of causal mechanisms for physical phenomena; it denies the existence of an 'objectively real' world.

But surely, the existence of that world is the primary experimental fact of all, without which there would be no point to physics or any other science; and for which we all receive new evidence every waking minute of our lives. This direct evidence of our senses is vastly more cogent than are any of the deviously indirect experiments that are cited as evidence for the Copenhagen interpretation. [...]

At this point, it is clear that theoretical physics has gone berserk. In quantum theory we have got ourselves into a situation where the objects have become unreal, but the probabilities have become real!" [144].

"Finally, I note that in conventional general probability theory having nothing to do with quantum mechanics, whenever we write a 'probability mixture of conditional probabilities': $p(x) = \int p(x \mid \theta) w(\theta)\, d\theta$, the distribution function $w(\theta)$ may become negative while $p(x)$ remains nonnegative. An old historical example is Laplace's derivation of the Rule of Succession from a probability mixture of binomial distributions. Today this is often called the de Finetti representation theorem because de Finetti proved, in a sense, the converse of Laplace's result. We find [145] that if Laplace's probability density $w(\theta)$ becomes negative we can represent not negative probabilities, but negative correlations, in an exchangeable sequence. This was not heretofore possible with the Laplace–de Finetti choices of $w(\theta)$. So I now call $w(\theta)$ a 'generating function' and do not interpret it as a probability" [146].

H. Jeffreys: "But observable quantities are mathematically real, and probabilities real and non-negative. It is therefore somewhat puzzling to find that the fundamental statements of the [quantum] theories involve complex numbers, non-commutative multiplication, or matrix roots of -1. Real combinations of these are identified with observables, and the probabilities calculated are positive, so that the conditions required are satisfied, but the logical status of the complex quantities remains mysterious. [...] it does not appear that [...] the probability density is correctly given by $\psi\bar{\psi}$" [147].

A. Kolmogorov: "Axiom III: A non-negative real number $P(\mathrm{A})$ is attached to each set A of F. This number $P(\mathrm{A})$ is called the probability of the event A" [148].

G. Lüders: "When discussing negative probabilities I would not emphasize the issue 'to show by means of the usual axioms that negative probabilities are impossible'. Rather, I would like to debate the relation between calculated and 'observed' probabilities. The only physical – and, therefore, at least in principle testable – meaning of quantum mechanical probability predictions known to me, however, is that of relative frequencies of observed results in an ensemble of identical systems being in identical states (certainly a difficult concept when a detailed theory is lacking). The mentioned relative frequency as ratio of non-negative numbers (number of 'successful' trials divided by number of all trials) is non-negative itself. Under this aspect negative probabilities apparently cannot occur in the final result; this being fairly independent of the 'usual axioms'.

Perhaps something else was meant in in the introduction to this article. In any case, however, the question concerning the physical – and, hence, at least in principle testable – and not purely formal meaning of negative probabilities cannot be neglected" [149].

G. Ludwig: "The problem of probability (and in this context possibly negative probability) is basically due to the use of the word 'probability' for completely different facts. Hence one can reasonably assert, that negative probabilities are senseless, and the other one can as reasonably invent 'negative probabilities'; for both do not understand even approximately the same by this word" [150].

H. Margenau (in a comment concerning the introduction to this article only): "[This] account of negative probabilities appears to me perfectly reasonable, especially in view of the fact that there is to my knowledge no entirely acceptable definition even of positive probability. As I have set forth in my earlier writings we do not even have a completely rigorous definition of positive probability, since von Mises definition in terms of the limit of relative frequency must be rejected because the limit does not exist. Hence I find [the introductory] comments wholly acceptable: The subject of probability is not yet closed" [151].

T.W. Marshall: "There is no 'EPR paradox'. In 1935, Einstein, Podolsky and Rosen devised the 'EPR argument' to show that quantum mechanics is incomplete. In 1964 Bell strengthened this argument to the 'EPRB theorem', namely that quantum mechanics cannot be reconciled with any model based on the axioms of local realism. The realist axioms are precisely those which say that, the quantities $\rho(\lambda)$, $P_1(a, \lambda)$ and $P_2(b, \lambda)$ are all positive. Of course it is possible to avoid the consequences of the EPRB theorem by relaxing the requirement that probabilities be non-negative, but this amounts to abandoning, or at least redefining, 'realism'. This is in no sense 'a resolution of the EPR paradox'.

Einstein has already pointed out [152] that, if the world really were nonlocal in the way quantum theory depicts it, then we would not be able to do any science. In my opinion this judgement remains as valid as ever, and I and my colleagues have proved, in a series of publications [96, 97, 153], that claims to have observed this nonlocality experimentally are without foundation. I would say the same applies to any nonrealist features of quantum theory, such as negative probabilities. No meaningful system of hypothesis testing is possible once we admit such entities. For an early statement of mine on negative probabilities see my paper on Random Electrodynamics [154]. That negative probabilities seem to be an inescapable feature of all quantum theories is, for me, precisely the point at which the inadequacy of those theories becomes manifest" [155].

N.D. Mermin: "My feelings about negative probability are rather negative. I agree that they provide a formal 'solution' to the EPR paradox [...but] an explanation in terms of negative probabilities doesn't illuminate anything for me" [156].

P. Meystre: "I would like to suggest that, maybe, *none* of these 'self-evident truths' [locality, reality, and the use of inductive inference] is wrong. In the derivation of Bell's theorem, there was a fourth hypothesis, which one mostly doesn't pay much attention to, namely the fact that probabilities are positive and bounded by 1" [157].

J.L. Park: "In any physical theory which assigns probabilities to possible measurement results, use of the construct *ensemble* is unavoidable, simply because probability in physics means relative frequency" [158].

H. Paul: "I do not believe that negative probabilities actually offer a solution to the EPR problem. In my opinion, only the fundamentally non-classical character of quantum theory becomes obvious by this issue, as one learns from the Wigner distribution. Another nice example is provided by Glauber's P-representation of a radiation-field. There exists a complete equivalence between classical and quantum mechanical description, if the radiation field has a P-representation, i.e. if its density operator can be written as $\rho = \int P(\alpha)|\alpha\rangle\langle\alpha|\, d^2\alpha$, where α denotes a Glauber state and the integration extends over the whole complex plane. (For the sake of simplicity only the single-mode case is considered.) The mentioned equivalence reads:

$$\langle F(q^+, q)\rangle \equiv \mathrm{tr}\{F(q^+, q)\rho\} = \int P(\alpha)F(\alpha^*, \alpha)\, d^2\alpha\,.$$

Here, F is an operator of the radiation-field, more precisely speaking, a normally ordered function of the photon creation and photon annihilation operators q^+ and q, respectively. $F(\alpha^*, \alpha)$ is the corresponding classical function, in which q^+ is substituted by α^* and q by α. In most cases $P(\alpha)$ is positive definite and, hence, can be interpreted as a classical distribution function. But there are cases in which $P(\alpha)$ becomes negative in small α domains. Then the above interpretation can no longer be maintained. However, one might say, also in this case classical physics succeeds to reproduce the quantum mechanical result, but only by means of an unallowed trick, namely by allowing for negative probabilities" [159].

I. Pitowsky: "We can conceive of mathematical situations where a natural concept of probability emerges which is not captured by the usual (Kolmogorov) axioms of probability theory. What I have in mind is not a radical extension of probability (such as introducing negative or complex probability values) but rather a conservative extension [...]" [106].

K.R. Popper: "We thus arrive at the propensity interpretation of probability. It differs from the purely statistical or frequency interpretation only in this – that it considers the probability as a characteristic property of the experimental arrangement rather than as a property of a sequence.

The main point of this change is that we now take as fundamental *the probability of the result of a single experiment*, with respect to its *conditions*, rather than the frequency of results in a sequence of experiments" [116].

G. Richter: "Negative probabilities have been well-known since the thirties. They occur if, in the course of a statistical interpretation of quantum mechanics, e.g., physical quantities of a system in a state with sharp energy are treated as ensemble properties of a set of individual systems, which populate states with less sharp energy. It seems hardly possible to attach an exact physical meaning to such a description. After creation of quantum mechanics so much has been written about this question, that it seems appropriate, first of all to review the existing literature; it will be seen then, that a lot already has been reflected – as is stated in the introduction" [160].

A. Salam plainly and concisely answered (the remark: "Although I expect that your attitude towards negative probabilities is very negative ..."): "Correct, Abdus Salam" [161].

E. Santos: "Probabilities must be identified with ratios between number of successes and number of trials, and this ratio is always non-negative.

There are, however, two instances where negative pseudoprobabilities may be useful. The first one is in formal manipulations, where one uses mathematical techniques similar to those used in probability theory which may be valid even for negative functions. A typical example is the use of the Wigner function (but without attributing to it the meaning of a probability density).

Another instance is the use of expectation values incorrectly called probabilities. For example, it is usually stated that the density of particles is the number of particles per unit volume, but this is incorrect from a mathematical point of view because the limit

$$\lim_{\Delta V\to 0} \Delta n/\Delta V$$

does not exist as Δn is an integer. The correct definition is the expectation value of the number of particles per unit volume. Now, if one is not interested in the number of particles but, say, in the electric charge per unit volume, this can be negative if we have two kinds of particles (e.g., electrons and positrons). In this way, it might be possible to give a physical meaning to the Wigner function as the expectation number of electrons minus positrons per unit volume" [162].

F. Selleri: "I have made a point of honor for me to follow good sense in research, since I saw crazy ideas going around in elementary particle physics: from this point of view I should conclude that negative probabilities do not make sense. *However*, I happen to be very interested in the EPR paradox and I am convinced that all possible solutions of it are necessarily far fetched. One is that quantum mechanics is not correct. Another one is that there is not a separable reality in the world, in spite of the fact that all interactions are known to decrease with distance. A third one is [that] with negative probabilities. A fourth one is the propagation of signals towards the past. And so on. This means that physics is reaching a hot point from which great changes could take place. [...] There is a point which I would like to make. If one has the state vector $|\eta\rangle$ for a pair of correlated particles S and T and if $\{|\psi_i\rangle\}$ and $\{|\varphi_j\rangle\}$ are two orthonormal and complete sets of vectors for S and T, respectively, one can write

$$|\eta\rangle = \sum_{i,j} c_{ij}|\psi_i\rangle|\varphi_j\rangle\,.$$

This is obvious. However von Neumann could prove that $\{|\psi_i\rangle\}$ and $\{|\varphi_j\rangle\}$ can *always* be chosen in such a way that

$$|\eta\rangle = \sum_l \sqrt{\omega_l}|\psi_l\rangle|\varphi_l\rangle$$

with $\sqrt{\omega_l} \geq 0$. This is less obvious, but true. Given then the operators $\hat{A}(a)$, $\hat{B}(b)$ corresponding to two dichotomic observables (values ± 1), one for S, the other for T, one has from the definition of the correlation function

$$P(a, b) = \langle\eta|\hat{A}(a)\otimes\hat{B}(b)|\eta\rangle$$

$$= \sum_{l, l'} \sqrt{\omega_l \omega_{l'}} \langle\psi_l|\hat{A}(a)|\psi_{l'}\rangle\langle\varphi_l|\hat{B}(b)|\varphi_{l'}\rangle$$

$$= \sum_{l, l'} \rho(l, l')\hat{A}(a, l, l')\hat{B}(b, l, l') ; \qquad (*)$$

where $\rho(l, l') = \sqrt{\omega_l \omega_{l'}}$ and

$$-1 \le \hat{A}(a, l, l') \equiv \langle\psi_l|\hat{A}(a)|\psi_{l'}\rangle \le 1 , \qquad -1 \le \hat{B}(b, l, l') \equiv \langle\varphi_1|\hat{B}(b)|\varphi_{l'}\rangle \le 1 .$$

Expression (*) is just like a two-hidden-variable expression

$$P(a, b) = \int \mathrm{d}\lambda\ \mathrm{d}\lambda'\, \rho_0(\lambda, \lambda')\hat{A}(a, \lambda, \lambda')\hat{B}(b, \lambda, \lambda')$$

with an important difference, that

$$\int \mathrm{d}\lambda\ \mathrm{d}\lambda'\, \rho_0(\lambda, \lambda') = 1 ,$$

while

$$\sum_{l, l'} \rho(l, l') = \sum_{l, l'} \sqrt{\omega_l}\sqrt{\omega_{l'}} = \sum_l \omega_l + \sum_{l \ne l'} \sqrt{\omega_l}\sqrt{\omega_{l'}} = 1 + \sum_{l \ne l'} \sqrt{\omega_l}\sqrt{\omega_{l'}} > 1 .$$

Thus, one could say that quantum mechanics differs from hidden variables theory simply because it implies the existence of probabilities larger than unity; but this is very near to negative probabilities as also [109] shows. This seems to be the formal origin of the violations of Bell's inequality, but I am not sure if it is physically meaningful" [163].

A. Shimony: "I am negative about negative probability. There are two generic concepts of probability. One is epistemic, and includes personal probability and logical probability. In both cases it is possible to give a fundamental proof for the non-negativeness of probability, once reasonable assumptions are made. The other is ontic, which includes the frequency and the propensity concepts – the latter postulating that probability makes sense in an individual case, but that the evidence for a propensity comes from an ensemble. Because of the role of frequencies for both versions of ontic probability, non-negativeness is essential.

I do believe that quantum mechanics makes us change our fundamental concepts. It makes us introduce objective indefiniteness and entanglement. My main worry about negative probabilities is that it uses a rather formal device to shield us from facing the radical metaphysical consequences of quantum mechanics" [164].

A. Siegel: "I have no *a priori* objection to negative probabilities. As with negative energy solutions of the Dirac equation, or the concept of neutrinos, they will have to wait for their acceptance until they are proved useful or convenient, or turn out to be 'real'. 'Realness' in science is of course a relative concept. A purely mental construct can gradually acquire 'reality' if it explains an experimental result not easily explained otherwise.

One circumstance under which I would accept negative probabilities would be if their use led to some important simplification in probability theory or calculation, or to correct results not otherwise readily obtainable. I do not know of any such achievements.

The formulation of probability theory now dominant – of events and their frequencies of occurrence – of course contains no room for negative probabilities. It is easy to see, for example, that if negative probabilities were permitted, the expectation values of essentially positive quantities (the squares of numbers, for example) could have negative values; to accept this we would have to redefine our concept of expectation value.

Until some good reason turns up to compensate us adequately for such violations of intuition, any formulation of negative probabilities will be useful only in highly restricted circumstances.

In this connection the 'Wigner distribution' [12] comes to mind. With this formalism, one avoids the awkwardness of probability amplitudes in quantum mechanical calculations and gains an illusion of having an orthodox probability distribution in p and q. This probability distribution, of course, takes on negative values for certain sets in p, q space. The use of the Wigner distribution is hemmed in by a set of rules and regulations designed to guarantee agreement with orthodox quantum mechanics. These rules prevent any unphysical consequences. In particular, the above-mentioned negative expectation values for positive observables cannot occur under them. In the case of the Wigner distribution, one is simply trading one theoretical fiction – the probability amplitude – for another, the negative probability. Since the negative probability is not permitted to have any physical consequences as such, there is no reason to object to its presence, and I see no reason to condemn the use of the Wigner distribution if it makes life simpler in any situation. The equivalence of the Wigner distribution to the wave function makes it clear that the properly regulated use of 'ensembles' containing negative probabilities is no more objectionable, in principle, than the use of probability amplitudes. But this is, of course, not the same as an endorsement of negative probabilities *per se*" [165].

P. Suppes: "The joint 'distribution' derived by the methods of Wigner [12] and Moyal [18] is not a genuine probability distribution at all. If this 'distribution' is accepted as the most reasonable one possible, then we may infer a stronger result than the Heisenberg uncertainty principle, namely, not only are position and momentum not precisely measureable simultaneously, they are not simultaneously measurable at all. [...] Koopman [166] is concerned to argue that quantum mechanics provides no real evidence for changing the foundations of probability. And with this I am in complete agreement, for the simple reason that there is, so far as I know, no substantial argument for making any such changes. The use of functions like the improper joint densities [...] is for purposes of calculation. In no sense do they help make a case for changing the basic laws of probability. Far from it, rather it may be argued that we can use their very improperness as a key to inferring what is not possible to measure experimentally" [16].

H.J. Treder: "In particular Heisenberg has dealt with the meaning of 'negative probabilities' in Dirac's sense. At the time he conceived the idea that measurements beyond the elementary length become senseless, because they would imply negative probabilities for the measurement results. Certainly negative probabilities are not involved in the Einstein–Podolsky–Rosen paradox. As I recently [167–169] stated for another time, Bohr's arguments are absolutely conclusive in this context;

also experimentally situations have been realized meanwhile, which claim the EPR paradox to be impossible (..., die das EPR-Paradoxon für unmöglich erklären)" [170].

V.S. Varadarajan: "The Kolmogorov axioms are precise, concise and lead to an extensive theory; the pure mathematician asks for nothing more. However there remains the problem of understanding the phenomenological background to these axioms. [...] it is found that when thus formulated [as a Boolean σ-algebra], the theory of probability does not include the situations that arise in quantum physics" [171].

C.F. von Weizsäcker: "An indefinite measure on the Hilbert space has been considered frequently and is used concretely, e.g., in quantum electrodynamics by the method of Bleuler and Gupta; also by Heisenberg's spinor-field theory. But since it has turned out impossible, to conceive negative probability in a visualizable sense, the indefinite measure was no longer understood as probability. I do not think that fundamental difficulties have appeared in this context. If probability is defined, e.g., as prediction of a relative frequency, clearly it has to be non-negative" [172].

E.P. Wigner: "I fully agree that the concept of a negative probability is in contradiction to the usual definition of the probability concept. However, other quantities from which an actual probability can be calculated, are often called 'probabilities' " [173].

H.D. Zeh: "I think that negative probabilities do not *explain* anything. On the contrary the necessary negativity shows that the intended explanation fails. In my opinion probabilities can only make sense if interpreted as relative frequencies. Negative frequencies, however, can have formal meaning only (as numbers in calculations), and, therefore, do prevent an interpretation" [174].

I.S. Zlatev: "Here we leave open the question about the possible physical picture which leads to quasidistributions. Let us adduce just a few arguments which prove eloquently that quasidistributions are closely connected with a classical probability treatment. This connection lies in the fact that, if they are formally considered as distributions, one can follow a standard procedure and get expressions that not only allow but also require an interpretation as probability distributions. [...] a clear physical picture should be sought behind the quasidistributions which themselves may be devoid of a physical meaning" [175].

6. Notes

This chapter continues the previous one in that its contents was written independently of (but destined to appear in) this review. It differs from chapter 5 in that the contributions give a more detailed and elaborate account of some special aspects connected with the subject under discussion.

The first note, written by G. Ludwig, examines the physical meaning of negative probabilities from the conceptual point of view, discussing the concept of axioms in physics in a more subtle way than it was applied in chapter 2.

The second note, written by C. Dewdney et al., contains a concise account of the problem of negative probabilities in connection with quantum electrodynamics – touched in section 3.2 – with special respect to the causal stochastic interpretation invented by its authors and capable of avoiding negative probabilities under most circumstances. This topic is also discussed by the following note by N. Cufaro Petroni treating the consequences implied by the use of the Feynman and Gell–Mann equation and presenting an alternative interpretation of the resulting non-positive conserved densities.

This chapter is concluded with notes by M.S. Bartlett and E.T. Jaynes, presenting some examples and considering the topic from an open-minded pragmatic point of view.

6.1. *Have negative probabilities a physical meaning*? (G. Ludwig)

The discussions concerning physical concepts often suffer from using the same words for very different concepts. Not to distinguish between theoretical auxiliary concepts and basic concepts is a second source of misunderstandings. Here basic concepts mean such concepts which are used to compare a theory with experiments. Therefore, it is necessary first to classify the various concepts.

6.1.1. *Classification of physical concepts*

There must be a possibility of comparing a physical theory with experiments; otherwise the theoretical framework cannot claim to be a "physical" theory, since it will be invariant against any experience. Physical concepts are necessary in order to compare a theory with experiments, concepts by means of which the experimental results can be described; we say, by means of which the "observational report" can be formulated (see [176, 177] and [178, chapter XIII]). We will call these concepts the basic concepts of the theory.

The meaning of these concepts is not explained by the theory under consideration. It must be explained before formulating the theory; it may be explained by other yet known theories, which we will call "pretheories" (relative to the theory under consideration). For the concept of pretheory see also [176–178].

We say that those facts which can be described by the basic concepts, can be "directly" measured. That many other things can also be measured with the help of the theory is very well known. What, however, is the meaning of this paraphrase: with the help of the theory?

A physical theory does not merely consist of a description of the observational report applying the basic concepts; in addition, it contains as an essential part a formulation of "physical laws". These laws *and* the basic concepts allow for "a definition of new concepts" named "theoretic concepts". The physical meaning of these concepts is given by the meaning of the basic concepts *and* the "definitions". These definitions also furnish methods how to measure "indirectly" facts which can be described by these new theoretic concepts (see [176] and [178, chapter XIII]).

In addition to the basic and theoretic concepts described above, many theories include a third kind of concepts, namely those which will be called "theoretic auxiliary concepts". The following example may serve to illuminate this point. With the help of Boltzmann's collision equation we may deduce the Navier–Stokes equations. For this theory we can use the velocity field, the mass density field, and the temperature field as the basic concepts of the theory. A lot of different theoretic concepts can be deduced from these basic concepts, e.g., the heat transfer. The auxiliary concept of atoms, however, cannot be derived. Exactly because of this fact, the positivists objected to the "atomic hypothesis".

Nowadays nobody will forbid the introduction of theoretic auxiliary concepts. On the contrary, if a new theory is lacking, all kinds of ideas, formulations, concepts, even crazy ones and contradictions, are permitted as an intermediate phase of the development towards a sufficient theory. Nevertheless, if a theory has taken a valid form, then there remains the task to decide to which extent the theoretic auxiliary concepts can be replaced by theoretic concepts, i.e., to which extent the auxiliary concepts are "real."

Concerning the "atomic hypothesis" we know that the development of physics has offered *new* experimental results by investigating the interactions of macrosystems, which made it possible to develop a theory of these interactions. This theory entails atoms as theoretic concepts and, in this sense, as real facts (see [178]).

This classification, briefly outlined above, can be replaced by a rigorous treatment introducing a form of theory which we called an "axiomatic basis". As these chains of reasoning would exceed the space of this note, the interested reader may be referred to [176, 179] and [178, chapter XIII]. It should be mentioned, however, that the theoretic auxiliary concepts appear in the mathematical part of an axiomatic basis in axioms of the form: "there are such ..., that ..."; e.g., "there are such entities, called atoms which emit light, and the light can be measured directly by spectral devices".

In the philosophy of science we sometimes find the opinion that modern theories cannot be formulated without theoretic auxiliary concepts. On the contrary I have formulated the task of obtaining every physical theory in a form free of theoretic auxiliary concepts. I have shown that this is possible for quantum mechanics [178]. Such a form of theory is best suited to investigate all possible new theoretic concepts. Concerning quantum mechanics it is possible in this way to recover the microsystems and their structure as theoretic and, therefore, real concepts [178].

In the development of physics it is a very essential step to realize that some concepts, originally introduced as theoretic auxiliary concepts, can be deduced as theoretic and, hence, real concepts.

In this modern classification of physical concepts the old controversy between Galilei and the church was nothing else than the question whether the motion of the planets around the sun is only a theoretic auxiliary concept or a theoretic, i.e., a real concept. While the notions of Copernicus rather had the character of auxiliary concepts, Galilei, due to the advanced state of experimental results (i.e. a more comprehensive domain of applications) was right in declaring these notions to be realistic concepts, a point of view which later on was confirmed by the theories of Newton and Einstein.

6.1.2. Probability as a basic concept

The most important concept of probability in physics is a basic concept, i.e., a concept by which a theory can be compared with experiments: Probability as a mathematical picture is to be compared with reproducible frequencies occurring in experiments.

A general mathematical structure covering all cases in physics which involve experimental frequencies in comparison with a mathematical formalism is given in [176, 178, 180]. The main concepts are experimental "selection procedures", which can be treated formally by a set S of subsets of a set M. This set S fulfills the following axioms:

$$(a, b \in S) \wedge (a \subset b) \Rrightarrow (b \backslash a \in S)\,, \tag{6.1}$$

$$(a, b \in S) \Rrightarrow (a \cap b \in S)\,. \tag{6.2}$$

Here $\wedge$ is the logical "and", $\Rrightarrow$ is the logical implication, and $b \backslash a$ is the relative complement of a in b. A set of statistical selection procedures is determined by the additional axiom: *There is a real function*

$$p\colon T \overset{\text{def}}{=} \{(a, b) \mid (a, b \in S) \wedge (a \supset b) \wedge (a \neq \emptyset)\} \xrightarrow{p} [0, 1]$$

satisfying the following conditions:

$$(a_1, a_2 \in S) \wedge (a_1 \cap a_2 = \emptyset) \wedge (a_1 \cup a_2 \in S) \Rrightarrow p(a_1 \cup a_2, a_1) + p(a_1 \cup a_2, a_2) = 1\,, \tag{6.3}$$

$$(a_1, a_2, a_3 \in S) \wedge (a_1 \supset a_2 \supset a_3) \wedge (a_2 \neq \emptyset) \Rightarrow p(a_1, a_3) = p(a_1, a_2)p(a_2, a_3)\,, \tag{6.4}$$

$$(a_1, a_2 \in S) \wedge (a_1 \supset a_2 \neq \emptyset) \Rightarrow p(a_1, a_2) \neq 0\,. \tag{6.5}$$

$p(a, b)$ is called the probability of the selection procedure b relative to a. The real number p shall be compared with reproducible frequencies. That means, if among a number N of trials selected according to a there is a number N_+ of trials satisfying also the finer selection procedure b, then the frequency N_+/N has to be approximately equal to p; the precision of the approximation being determined by the degree of reproducibility of the frequencies.

It is obvious that for this basic concept the axiom $0 \leq p \leq 1$ is trivial since $0 \leq N_+/N \leq 1$. Also the other axioms (6.1)–(6.5) are not refutable by experience since they are prescriptions for the "correct" way of performing experiments; they are "norms". If an experiment is in contradiction to one of these axioms, then it has to be eliminated from the observational report; it does not belong to the domain of application (the "fundamental domain") of the theory. The same holds for such experiments where the frequencies are not approximately reproducible. If, for instance, we repeat an experiment only ten times, then the reproducibility of the frequencies is not ascertained and, therefore, the experiment is not useful for a comparison with this theory.

A very well-known objection to this basic concept of probability is that probability itself can be compared with frequencies only with a certain probability. This objection is wrong. If the number N of repetitions of an experiment is large enough, it is possible to obtain reproducibility of frequencies as precisely as reproducibility of trajectories in point mechanics, for instance. If a physicist would get a frequency strongly deviating from all other reproducible frequencies in some experiment, he would not ascribe this to a very exceptional case but would look for "disturbances" assuming that the experiment had not been performed "correctly" (with respect to the issue under investigation).

This basic concept of probability is not in contradiction with the introduction of a subjective weight for special irreproducible experiments as tests of a theory. If, for instance, only insufficiently small series of experiments are available, they may be attributed with weights in order to compensate for the uncertainty resulting from these series of experiments performed as tests for the (objective) basic concept of probability. These subjective weights, also called subjective probabilities, however, will not be discussed further, since I have never seen anyone introducing negative weights.

The basic concept of probability, characterized by the axioms (6.1)–(6.5), is the same for all physical theories, classical theories as well as quantum mechanics (see [178]). This concept defines the method of comparing theory and experiment and will not be changed in the future. Only the idealization due to which p is a mapping $T \xrightarrow{p} [0, 1]$ can be changed by introducing finite imprecision sets (see [176, 181]). The only empirical contents of the axioms (6.1)–(6.5) is that experiments fulfilling these axioms as norms are *possible*.

In the same way the basic concept of a (pseudo-) Euclidian geometry in the laboratories has never been changed. On the contrary this basic concept is *necessary* to interpret the non-(pseudo-) Euclidian geometry of Einstein's gravitation theory.

6.1.3. Probability as a theoretic real concept

In many physical theories the basic concept of probability is extended to other concepts than "selection procedures", i.e., there are new concepts deduced from the basic concepts. The interpretation of these theoretic concepts is given by the due deduction from the basic, already interpreted concepts (see section 6.1.1).

An example for such a deduced concept of probability is delivered by quantum mechanics. The deduction is presented in [178] and [180]. We get tr(WE) as the probability to register the "decision effect" E on systems in an "ensemble" (also denoted "state") which is represented by the self-adjoint operator W in a Hilbert space with $0 \leq W$ and $\mathrm{tr}(W) = 1$; E is a projection operator. "Decision effect" and "ensemble" ("state") are new deduced concepts.

In this form of quantum mechanics (i.e., in the axiomatic basis given in [178] and [180]) also the Wigner distribution function on the phase space Γ can be deduced. This measure on Γ can take negative values. The deduction shows, however, that the values of this measure are not the values p of the basic probabilities. Therefore, I propose not to use the same notation "probability" for this measure but to maintain the notation "measure". The physical interpretation of this measure is (at least as long as we take quantum mechanics on said axiomatic basis) supplied by the deduction. Signed measures as the Wigner measure are very well known in mathematics and represent nothing extraordinary. Only a new interpretation deviating from that deduced from the axiomatic basis would be extraordinary. Such a new interpretation would be possible only by introducing additional theoretic auxiliary concepts in quantum mechanics.

6.1.4. Probability as a theoretic auxiliary concept

A completely different concept of probability is the theoretic auxiliary one. Auxiliary concepts are primarily imagined concepts involving no direct relation to experiments. Since one can imagine whatever one likes, no constraints are to be imposed on the imagined concepts, except the following two: the imagined concepts

(1) must not be self-contradictory,
(2) must not show up any contradiction with experience.

We may, for instance, add theoretic auxiliary concepts to quantum mechanics (as founded in [178, 180]). Then condition (2) requires that the auxiliary concepts are not in contradiction with quantum mechanics (at least with very high precision [176, 177, 181]) since quantum mechanics is a very precise theory in its fundamental domain, i.e., the domain of application. Such a theoretic auxiliary concept could consist of a stochastic model with an imagined "probability structure". Since this probability structure is only an imagined notion, one may introduce also negative probabilities if one wants to do so. Imagined concepts are free concepts. It is merely a matter of taste, what concepts are preferred.

But have these imagined concepts any physical meaning? That is the decisive question.

It is by no means decisive that someone believes in the reality of additional theoretic auxiliary concepts (e.g., in probability as something like a propensity in the real world), since such a believer cannot convince a non-believer as long as *decisive experiments* are lacking, or, to put it in other words, as long as the theoretic auxiliary concepts cannot be replaced by theoretic (i.e., physically real) concepts.

It is alright that theories involving additional theoretic auxiliary concepts are called *hidden variables* theories. In this sense also these auxiliary probability concepts deal with *hidden* probabilities. To discuss whether these *hidden* concepts have a realistic meaning is fruitless.

The historical development of the atomic hypothesis to a realistic physical theory serves as a good example to demonstrate what is necessary for a realistic interpretation of present hidden variables theories. Aerodynamics alone was not suited to demonstrate the reality of atoms. It was necessary to extend the domain of applications beyond that of aerodynamics. The Brownian motion is an example for such new applications of an atomistic theory.

A hidden variables model for quantum mechanics (perhaps including negative probabilities) has to supply with high precision the same results in the fundamental domain of quantum mechanics and has to go beyond this domain of applications if this theory should be more than only an imagination. It has to lead us to new regions of applications, to experiments which do not belong to the fundamental domain of quantum mechanics. Then we will have a *more comprehensive* theory than quantum mechanics and on an axiomatic basis of this new theory we can decide which of the concepts at first introduced as auxiliary are real.

By an exactly equivalent procedure it was possible to show that the originally introduced auxiliary concepts of atoms are real concepts. An axiomatic basis of quantum mechanics demonstrates this [178]. My objection against most of the imagined models for quantum mechanics is that these are "pure" imaginations, i.e., they give no extension of the fundamental domain of quantum mechanics. Without an extension of the fundamental domain (i.e., the domain of application), however, the introduced auxiliary concepts have no physical meaning.

6.2. *Negative probabilities and second-order wave equations* (C. Dewdney, P.R. Holland, A. Kyprianidis and J.P. Vigier)

The causal Stochastic Interpretation of Quantum Mechanics (SIQM) provides, by introducing the notion of a real trajectory in spacetime for quantum particles acted on by the quantum potential [182–184], an imagery which renders intelligible quantum processes in a way that the Copenhagen interpretation (CIQM) cannot do. In this note we discuss the interpretative problem raised by relativistic quantum mechanics, namely, the mathematical existence of negative probability density and negative energy solutions to second-order wave equations, which, as in all other quantum processes, we argue can only be coherently treated by assuming the real physical existence of paths. Indeed, this is an important issue in the SIQM since the very existence of paths in spacetime implies positive probability distributions and, moreover, in accordance with Einstein's basic principles, all material drift motions should be time-like and propagate positive energy forward in time.

It is sometimes erroneously stated that the only way out of the problem of negative probability solutions to the Klein–Gordon (KG) equation is to reject the first quantized formalism in favour of second quantization. In fact, this is not so and it is possible to show by a Hamiltonian method due to Feshbach and Villars [185, 186] that for certain well-behaved external potentials the KG solutions may be split into positive- and negative energy parts associated, respectively, with positive and negative probability.

To see this, let us start from the charged scalar wave equation

$$(D_\mu D^\mu - m^2)\psi = 0\,, \tag{6.6}$$

where $D_\mu = \partial_\mu - \mathrm{i}eA_\mu$, e and m are the charge and mass, respectively, of the particle moving in the external field A_μ, the metric has the signature $(-+++)$ and the units are chosen so that $\hbar = c = 1$. Equation (6.6) may be expressed in the form

$$\mathrm{i}\dot{\Psi} = H(e)\Psi\,, \tag{6.7}$$

where Ψ is a two-component wave function and H is a 2×2 matrix Hamiltonian. One can show that for the inner product $\langle \Phi, \Psi \rangle = \int \Phi^* \sigma_3 \Psi \, d^3x = \int j^0 \, d^3x$ where the conserved current is

$$j^\mu = \psi^* \left(\frac{1}{i} \overleftrightarrow{\partial}^\mu - ieA^\mu \right) \psi \tag{6.8}$$

($\overleftrightarrow{\partial}^\mu$ means that ∂^μ is to be applied to the left and right wave functions in turn), the mean value of H in any state is positive: $\langle \Psi, H\Psi \rangle > 0$. It follows that, with $H\Psi = E\Psi$, the space of solutions of eq. (6.7) splits into two disjoint subsets: $\{E > 0, \langle \Psi, \Psi \rangle > 0\}$ and $\{E < 0, \langle \Psi, \Psi \rangle < 0\}$. The latter subset of solutions may be mapped onto positive-energy, positive-probability antiparticle solutions by means of the charge conjugation, $\Psi(x) \to \Psi^c(x) = \sigma_1 \Psi^*(x)$, since from eq. (6.7), $H(-e)\Psi^c = -E\Psi^c$ and $j_0^c = -j_0$.

Thus, within the CIQM, one can show formally how, for stationary states, the signs of energy and integrated probability are correlated and that negative probability solutions may be physically interpreted, as antiparticles running backwards in time. Note though that the local values of probability density may become negative and that such motions remain uninterpreted. We shall now show [75] how in the causal interpretation we are able to prove a stronger result than that just given, and in a way which is technically easier and physically clearer. Our approach extends some brief remarks of de Broglie [187] concerning this problem.

Substituting $\psi = \exp(P + iS)$, where P, S are real scalars, in (6.6) yields the Hamilton–Jacobi and conservation equations:

$$(\partial^\mu S - eA^\mu)(\partial_\mu S - eA_\mu) = -M^2 , \tag{6.9}$$

$$\partial_\mu j^\mu = 0 , \tag{6.10}$$

where $M^2 = m^2 - \Box P - \partial^\mu P \, \partial_\mu P$ is de Broglie's variable rest mass and $j^\mu = 2e^{2P}(\partial^\mu S - eA^\mu)$ is the current (6.8).

The assumption of the SIQM is that the KG particle has a drift velocity $u^\mu = dx^\mu/d\tau$ where τ is the proper time along paths parallel to j^μ. In terms of the momentum $P^\mu = \partial^\mu S - eA^\mu$, $u^\mu = M^{-1}P^\mu$ with $u_\mu u^\mu = -1$, from eq. (6.9).

Defining a scalar density $\rho = Me^{2P}$ we may express eq. (6.10) in the form [188]

$$D\rho/D\tau \equiv \partial_\mu(\rho u^\mu) = 0 .$$

(Note that $D/D\tau$ is an average derivative in a volume element [188].) From this it follows that along a line of flow

$$\omega M \, e^{2P} = K \tag{6.11}$$

where ω is a volume element of fluid, and K is a real or pure imaginary constant (which however varies from one drift line to another). If on an initial space-like surface the motion is time-like then from eq. (6.9) M is real and so is K. Now, in the rest frame, $u^0 E = M$ where the particle energy $E = \partial^0 S - eA^0$. It follows that if initially the motion is future pointing with $E > 0$ then $M > 0$ which implies $K > 0$ (since $e^{2P} > 0$ and $\omega > 0$ always) and we see from eq. (6.11) that the time-like and positive-energy character of the motion is preserved all along a trajectory. Moreover, the sign of the probability density $j^0 = 2\,e^{2P}E$ is correlated with the sign of E and will remain positive along a line of flow if the initial motion has $E > 0$.

Identical arguments lead to an association of past pointing negative-energy motions with negative probability densities and this coupling is preserved along a line of flow if initially $E<0$. Such solutions may be mapped onto positive-energy, positive-probability-density antiparticle solutions by the charge conjugation $\psi^c = \psi^*$.

These results, proved in the rest frame, evidently remain valid under orthochronous Lorentz transformations.

We have thus succeeded in separating the solutions to the causal KG equation into two disjoint subsets, $\{E<0,\ j^0<0\}$ and $\{E>0,\ j^0>0\}$, and shown that the causal laws of motion prevent the development of one type of solution into the other. This reasoning holds for all external fields A_μ which maintain the time-likeness of the momentum P^μ. Should the external potential be strong enough pair creation may occur and the separation of the solutions breaks down. In addition, we assume that the initial motion is associated with a wave packet so that the initial total probability is unity. This is an important point since de Broglie [183] has shown how with a plane KG wave incident on a partially reflecting mirror superluminal motions apparently occur in the region of the Wiener fringes. It seems that these unphysical motions are a consequence of the excessive abstraction implied by the use of plane waves.

It is emphasized that we have only been able to overcome the difficulty of negative probabilities by assuming that particles possess well-defined spacetime trajectories, and that they are subject to action by the quantum potential (contained in M). With these assumptions, we can immediately associate the sign of particle energy with the sign of local probability density, an energy, moreover, which is well-defined and continuously variable for all possible particle motions (and not just for stationary states). The initial character of these motions is preserved for all time by the Hamilton–Jacobi and conservation equations.

If one accepts that the quantum mechanical formalism is complete then one must accept Feynman's statement [11] that there is no way of eliminating negative probabilities from the intermediate stages of, for example, an interference calculation. The problem of their physical interpretation then cannot be avoided.

However, if one accepts the introduction of trajectories in the description, then our demonstration above shows how the positive character of probability is preserved at every stage of the calculation. The association of positive probabilities with positive energy is of course in accordance with the principles of relativity theory.

We note finally that although our discussion here has been confined to a single KG particle, our method may be applied to the elimination of negative probabilities from the theory of the many-body KG system [73], the spin-1 Proca equation [72], and the spin-$\frac{1}{2}$ Feynman–Gell–Mann equation [74].

6.3. Non-positive conserved densities in relativistic quantum mechanics (N. Cufaro Petroni)

In a series of recent papers [70, 189, 190] the author proposed a theory of spin $\frac{1}{2}$ particles based on a second-order wave equation: the so-called 4-component Feynman and Gell–Mann equation [69] that, in the presence of an external electromagnetic field A_μ, by rewriting eq. (3.52) takes the following form ($\hbar = c = 1$):

$$[(\mathrm{i}\partial_\mu - eA_\mu)(\mathrm{i}\partial^\mu - eA^\mu) - \tfrac{1}{2}eF_{\mu\nu}\sigma^{\mu\nu} - m^2]\psi = 0\,. \tag{6.12}$$

Beyond a series of well-known remarks [69] about the use of (6.12), the main reasons that induced the authors to adopt it instead of the Dirac equation

$$(\mathrm{i}\not{\partial} - e\not{A} - m)\psi = 0 \tag{6.13}$$

are the following:

(a) if we are looking for a causal interpretation of the quantum equations, a classical analogy can be found only starting from a second-order differential equation;

(b) if we want to interpret the appearance of a Quantum Potential [70, 182, 189–193] in the classical equations as the global effect of a stochastic process induced on the particle by a subquantum medium, namely, the so-called Dirac aether [65, 184, 194–196], we must use only second-order differential equations [197–201];

(c) both the preceding steps are essential in a coherent causal physical interpretation [202–204] of the non-local quantum effects suggested by the experiments [205–207] on the EPR paradox (cf. section 3.3).

Let us start with a discussion of the relations between the 4-spinors ψ solutions of (6.12) and those solutions of (6.13). From this standpoint we can define the following sets:

$$\mathscr{F} = \{\psi \,|\, (I - \not{D}^2)\psi = 0\}\,, \qquad \mathscr{D}_\pm = \{\psi \,|\, (I \pm \not{D})\psi = 0\}\,, \tag{6.14}$$

with

$$D_\mu = \frac{1}{m}(\mathrm{i}\partial_\mu - eA_\mu)\,. \tag{6.15}$$

Of course $\mathscr{F}$ is the set of the solutions of (6.12) and $\mathscr{D}_-$ the set of the solutions of (6.13). Moreover, to show what $\mathscr{D}_+$ is, the following propositions can be proved (note that the meaning of $\ni'$ is "such that"):

$$P_1 \quad \mathscr{D}_+ \cap \mathscr{D}_- = \{\psi = 0\}\,, \qquad P_2 \quad \mathscr{D}_+ \cup \mathscr{D}_- \subset \mathscr{F}\,,$$

$$P_3 \quad \forall \psi_+ \in \mathscr{D}_+ \exists! \psi_- \in \mathscr{D}_- \ni' \psi_+ = \gamma_5 \psi_-\,.$$

By recalling what γ_5 represents for the symmetry operations on spinors [29], we can conclude that $\mathscr{D}_+$ contains the antiparticle wave functions moving backward in space-time and with the sign of energy inverted with respect to the particle wave functions (solutions of the Dirac equation (6.13)) belonging to $\mathscr{D}_-$. In a word we could say that $\mathscr{F}$ contains both particle and antiparticle solutions and we will show now that it contains also all their superpositions. In fact it can be proved that

$$P_4 \quad \forall \psi \in \mathscr{F} \exists \psi_+ \in \mathscr{D}_+ \wedge \psi_- \in \mathscr{D}_- \ni' \psi = \psi_+ + \psi_-\,,$$

since it is very easy to see that, if $\psi \in \mathscr{F}$, the spinors $\psi_\pm = \frac{1}{2}(I \mp \not{D})\psi$, belonging to $\mathscr{D}_\pm$, are such that $\psi_+ + \psi_- = \psi$.

The mixing of particles and antiparticles in the general solutions of (6.12) is particularly evident in the form of the conserved density. It is well known [70, 189, 190] that eq. (6.12) can be deduced from the scalar Lagrangian density

$$\mathscr{L} = \overline{(\mathrm{i}\not{\partial} - e\not{A})\psi}(\mathrm{i}\not{\partial} - e\not{A})\psi - m\bar{\psi}\psi\,, \tag{6.16}$$

so that the conserved current density is

$$j_\mu = \frac{1}{m}\mathrm{Re}[\bar{\psi}\gamma_\mu(\mathrm{i}\partial\!\!\!/ - eA\!\!\!/)\psi], \tag{6.17}$$

that, of course, does not coincide with the Dirac current $\bar{\psi}\gamma_\mu\psi$ unless the spinor ψ is a solution of (6.13). As a consequence the conserved density j_0 will not be positive definite. This feature, that is common to all the relativistic second-order quantum equations, forbids a direct statistical interpretation of (6.12) with the conserved density playing the role of a probability density.

The way out proposed by Dirac [29] was, in some sense, the restriction of the physically acceptable states to the solutions of (6.13), so that the conserved density becomes $\psi^\dagger\psi \geq 0$. We will now analyze another interpretation [12, 66] of the appearance of non-positive densities: j_0 is not a probability density, but an auxiliary function which obeys many relations we would expect from such a probability. In fact it behaves like an average charge density where, for mixtures of particles and antiparticles, "charge" must be understood in the widest sense, i.e., as a certain property which distinguishes between particles and antiparticles which are identical to each other in all other respects (electric charge, baryon number, etc.). In this sense we can calculate averages of physical observables exactly like in the ordinary probability calculus but using a "probability measure" which is not positive.

From this standpoint the connection between the non-positivity of j_0 and the mixing of particles and antiparticles is better elucidated if we consider that in the preceding section it was settled that each $\psi \in \mathscr{F}$ is a superposition of ψ_- and ψ_+, namely, of a particle and an antiparticle state. As a consequence, we get from P_4 that

$$\begin{aligned} j_\mu &= \mathrm{Re}[(\bar{\psi}_+ + \bar{\psi}_-)\gamma_\mu D\!\!\!\!/(\psi_+ + \psi_-)] \\ &= \mathrm{Re}[(\bar{\psi}_+ + \bar{\psi}_-)\gamma_\mu(-\psi_+ + \psi_-)] \\ &= \bar{\psi}_-\gamma_\mu\psi_- - \bar{\psi}_+\gamma_\mu\psi_+ , \end{aligned} \tag{6.18}$$

i.e., the conserved current j_μ is the difference of two Dirac-like conserved currents for particles and antiparticles. Moreover,

$$j_0 = \psi_-^\dagger\psi_- - \psi_+^\dagger\psi_+ , \tag{6.19}$$

that means that the non-positive conserved density j_0 is always a difference of two positive Dirac densities for particles and antiparticles. The result (6.19) is perfectly coherent with the point of view that considers j_0 as a non-positive measure, because a classical result of the measure theory [208] states that each real measure is the difference of two positive measures.

In conclusion we make the following remarks:

(1) All the theories ruled by a second-order relativistic wave equation will lead to the appearance of non-positive conserved densities j_0.

(2) If we want to interpret directly this j_0 as a probability density we have two ways: (a) to restrict ourselves to the solutions with positive definite density (Dirac equation; see also the contribution of Dewdney et al. in section 6.2); (b) to generalize our idea of probability in order to physically interpret its negative (and greater than one) values.

(3) Before cutting off a part of solutions or modifying our conception probability, however, we should explore the possibility, discussed in this note, that j_0 be interpreted not as a probability density but as a difference of probability densities, so that it could be used as a non-positive real measure to calculate averages of random variables in a theory containing both particles and antiparticles.

6.4. Some aspects of negative probabilities (M.S. Bartlett)

I do not think my views on negative probability have changed much since my 1945 note [118]. It may perhaps be helpful to illustrate Eddington's concept of an *extraordinary* random variable by citing the error variance of a sample mean of n observations taken without replacement from a *finite* 'universe' of N individuals. This is $\sigma^2(1-n/N)/n$, where σ^2 is the variance of one observation, or equivalently $\sigma^2/n - \sigma^2/N$, where the first term represents the usual formula when the universe is infinite, and the second term is the *negative* correction term due to measuring the uncertainty from the mean of the finite total universe.

I found formula (3.46) intriguing, as it relates to my comments on the uncertainty principle in my article 'The Paradox of Probability in Physics' [209]. There I noted (p. 39) that any spectral distribution (in the orthodox sense used by physicists, communication engineers and statisticians, as a distribution of 'frequencies', where here is meant by 'frequency' reciprocal periodicity) cannot be defined at a time instant t, only over an extended time.

The connection with a 'spectral distribution' of energy in quantum mechanics arises from the quantum relation $E = h\nu$, where ν is frequency in the above sense.

Feynman's example (section 3.4) of a *probability* in classical physics which is expressed as a sum of terms, some of which are negative, is interesting; but such expressions are very familiar to those working with random processes, and are merely a warning to physicists never to ascribe reality to the *individual* terms of such expansions unless there is other justification. This is underlined with Feynman's example, as there is an *alternative* expansion, also involving negative terms, obtained by the 'method of images'. Thus compare my formula [210]

$$f_t(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{b} \cos\left[\left(r+\tfrac{1}{2}\right)\frac{\pi x}{b}\right] \exp\left[-\tfrac{1}{2}\left(r+\tfrac{1}{2}\right)^2 \pi^2 \sigma^2 t/b^2\right], \tag{6.20}$$

which is equivalent to Feynman's, with my alternative formula [210]

$$f_2(x) = \sum_{s=-\infty}^{\infty} (-1)^s (2\pi\sigma^2 t)^{-1/2} \exp\left[-\tfrac{1}{2}(x-2bs)^2/(\sigma^2 t)\right]. \tag{6.21}$$

The notion of 'deletions' mentioned by some discussants in connection with negative probability seems consistent with my own example of 'bias' (cf. chapter 4) which can be positive or negative. A special but simple case is the negative probability $-q$ associated with *any* (positive) probability p by the equation $p = 1 - q$. Whereas a series of trials for p could be represented by a series of 0's except where the occurrences were denoted by 1's, the *same* series would consist of all 1's (denoting certainty), with deletions to 0's representing the negative component $-q$.

It may be argued that what is meant by 'separate existence' can be a matter of convention even in classical problems (not involving probability) with *positive* terms only.

Consider a communication engineer's wave form such as ...  ... . Common-sense tells us that this is a wave form involving a *single* frequency. But a standard spectral analysis would be equivalent to expanding the wave form $f(t)$, say, in a Fourier series which is

$$f(t)=\frac{4a}{\pi}\sum_{s=1}^{\infty}\frac{(-1)^s}{2s+1}\cos[(2s+1)t]\,, \tag{6.22}$$

if we take the period as 2π, and the height of the wave form $f(t)$ from $-a$ to a.

The spectrum of $f(t)$ will be proportional to the squares of the coefficients of the cosine terms, and thus consist of a series of positive *discrete* lines, the dominant one at the basic frequency ($s=0$), but with higher harmonics at $s=1,2,\ldots$ with amplitudes varying as $1/(2s+1)^2$. Such a spectrum could be readily demonstrated, but no one, as far as I know, would suggest in this example that the further discrete lines in the spectrum imply separate real 'entities'. (Notice in this example that any ambiguity in interpretation *from the spectrum alone* is due to the loss of the phases; in the *linear* Fourier expansion above, the phases are synchronous, this additional information allowing the recovery and recognition of the original simple wave form.)

I have always regarded the *ad hoc* introduction of probability in quantum theory by Max Born, as an interpretation of the wave function, as puzzling [209], so that, when I once asked Born about this, I was surprised to hear that he saw no difficulties! However, other aspects than negative probability may prove more important. For example, the paradox of Maxwell's demon was resolved by Szilard in terms of entropy and information (which is *negative* entropy). Attempts to derive quantum theory in terms of the Brownian motion of particles have revealed both similarities and differences (see, for example my article 'The Paradox of Probability in Physics' [209]; or, more recently, [211]), and particles may have to be regarded as special cases of Schrödinger's *event sequences.*

6.5. *Negative probability* (E.T. Jaynes)

Since probabilities are defined, in our textbooks, to be non-negative, one would think that in a well-run world there would never be any occasion to bring up the notion of negative probability. Yet in our world the topic seems to arise more frequently than it should. We are led to ask whether the seeming appearances of negative probability have only the nature of error or triviality; or whether they are hinting that something useful might be gained by extending the notion of probability beyond the domain of non-negative real numbers.

The possibility of such an extension cannot be dismissed a priori, particularly since there is good historical precedent for such generalizations. Students today may be surprised to learn that the concept of a negative number was a very difficult one, long regarded with suspicion. Even after the notion of negative numbers had been assimilated, the square of a number seemed, by definition, to be an inherently non-negative quantity. One can imagine the derision with which the idea was first received, that there are numbers whose squares are negative. Traces of this still persist in the epithet "imaginary number" which expresses exactly the initial attitude toward them.

With these lessons from history, we might approach the notion of negative probability with some caution. Rather than greeting it with derision, we can raise it as a serious question whether generalizations to negative – or even complex – probabilities would be (1) possible without contradictions; (2) useful in applications.

But that same caution restrains us from uncritical acceptance of such ideas until their consistency and usefulness are shown. Not all such generalizations are valid; we have become, perforce, accustomed to the idea of a Government owning a negative amount of money, but not to a man having a negative number of pennies in his pocket.

Others have noted various cases of seeming appearances of negative probability. As we should expect, most are merely erroneous, trivial, or frivolous; but some, associated with quantum theory, are indeed thought-provoking.

Eugene Wigner was my thesis advisor and I have been teaching Wigner distribution function theory for many years. However, that the Copenhagen interpretation of the mathematics may sometimes suggest negative probabilities does not bother or surprise me in the least; it merely confirms my judgment that the Copenhagen interpretation is wrong. Eventually, the quantity now called "probability" will turn out to have some other meaning.

To state the writer's conclusions as they stand presently, we see no insuperable obstacle to such an extension, but question whether a complete consistent extension could have any new content. By a "complete consistent formalism" for probability theory we mean a set of rules for calculation with two properties: (I) Given a set of propositions (A, B, C, . . .), it must show how to calculate not only their probabilities, but also the probabilities of all the propositions that can be formed from them by the logical operations of conjunction, disjunction, and negation. (II) Two methods of performing a calculation, each permitted by the rules, must lead to the same result.

The theorems of R.T. Cox [212] obtain the conclusion that, if we represent degrees of plausibility by real numbers, then the rules of conventional probability theory are uniquely determined, to within a change of variables that does not affect their content, by these completeness and consistency requirements. Thus we conjecture that, given any probability formalism using negative probabilities, if it is complete and consistent in the above sense, then a change of variables will reduce it to the present non-negative probability system.

For this and other reasons it is not clear how negative probabilities could lead to any useful results that are not already available from conventional probability theory; i.e., if the negative probability formalism led to different conclusions from the same information, then one could exhibit an inconsistency. Negative probability seems to us usable as a calculational device, much as a bookkeeper may pass through negative numbers in calculating a necessarily non-negative sum. Nevertheless, it appears unnecessary for any purposes we can think of.

When such a radical innovation is proposed, the onus is not on you and me to either refute it or accept it; but on the innovator to justify it by demonstrating that it has positive advantages not otherwise attainable. We stand ready – indeed, eager – to see such demonstrations, which if they could be produced would be not only stimulating, but of decisive importance for the future of probability theory (they would also have to be rather subtle, finding some way to sidestep Cox's results). But to date we have not seen any such demonstrations substantive enough to make us alter the way we now teach probability theory.

7. Discussion

A variety of examples presented in this review show that extended probabilities – though living a wretched existence below the surface of orthodox physics – are more widely spread than is commonly perceived. Here and there they rise to the surface and behave like "true" probabilities in all respects – except one. Therefore it seems appropriate to speak of "extended probabilities", suggesting by this wording a close relationship to true probabilities which, after all, cannot be denied completely. But it is certainly the minor task to find a suited nomenclature, and those who dislike this expression may substitute it by quasi-, pseudo-, meta- or even para-probabilities, -densities or -measures.

The fundamental questions to be discussed (and in part already discussed in the foregoing chapters)

may be listed in progressing order as follows: Are extended probabilities
(1) free of being self-contradictory?
(2) useful as a formal device in calculations?
(3) in agreement with experience?
(4) suited to improve the situation of present physics?
(5) enlarging the domain of application (in the sense of section 6.1)?
(6) a necessary concept of a theory which describes a maximum domain of reality by requiring a minimum number of axioms?
(7) to be considered an essential aspect of reality?

As already stated in the introduction, no final answers can be given to most of these questions. To a large extent this is due to the fact that their meaning depends strongly on the interpretation of the notions used to formulate the questions, in particular that of the last one. Further, the answers will depend on the choice of axioms to be maintained or eliminated.

What has to be understood, for example, by "an essential aspect of reality"? The spectrum of answers reaches from "nothing" via "everything which is an unavoidable concept of a well-proved theory" to "those things which can be touched by hands and seen by eyes". But can reality be touched and seen? Certainly it cannot. If understood in its ontologic sense, a final answer to question (7) can never be obtained; and, if arguing physically, we cannot deal with reality itself but with concepts serving to make a picture of reality. Physics is a method to gain more and more-detailed concepts reflecting reality while permanently testing and, if necessary, changing previous ones. The aim is to map the properties of reality as closely as possible – a task which in its last consequence is tantamount to approaching the zero mark of the Kelvin scale in thermodynamics. Thus we are forced to adopt the classical meaning of "an essential aspect of reality", which may be defined by "everything which is a necessary concept of a theory which describes a maximum domain of reality by requiring a minimum number of axioms". This makes questions (6) and (7) becoming synonymous, and it entails the undesirable consequence that reality changes when a new theory creates new concepts. Unfortunately no better definition is available, apart from pure ontological speculation. Hence, we will treat questions (6) and (7) as expressing the same contents.

At present there is no evidence leading to an affirmative answer to these questions, but, on the other hand, also a flat denial does not appear to express the adequate point of view with respect to the non-absolute state of fundamental concepts of scientific theories. Substantial modifications of current notions are to be expected and one may be aware (or afraid) of the possibility of this esoteric concept of extended probability to become "an essential aspect of reality" in future – whatever then might be understood by this phrase.

The Mössbauer experiments referred to in section 3.2 may be considered as giving some indication along this line. Of course, different explanations of the experimental results have been discussed. Commonly the absorber is assumed to act like a resonant filter, reverberating in response to the incident damped oscillation [51], but this picture is debatable on grounds of the high internal conversion ($\alpha = 8.26 \pm 0.19$, [213]) of the 14.4 keV decay of ^{57}Fe investigated. Also rescattering or similar processes can be excluded due to experimental tests performed parallel to the original experiments [50]. After all, even if the experimental evidence is questioned, there remains the fact that the theory derived from quantum mechanics as well as from classical physics, namely eq. (3.46′), contains extended probabilities without involving reverberating absorbers or analogous devices.

The same holds for the phase space distributions discussed in section 3.1. As long as the axiom of an exactly defined reality is upheld, attributing simultaneously coordinate and momentum to each individual, extended probabilities will result. However, as shown in the framework of the harmonic oscillator, they will appear only in connexion with rather drastic violations of the axiom of energy conservation. Maintaining this axiom, energy as a substance of changeable but not destructible character should not be created or annihilated. The violation of energy conservation possible during short time intervals corresponding to the uncertainty relation $(\Delta E)(\Delta t) \approx \hbar/2$, then rather might be thought of as some kind of "energy oscillation" in that sense that a system may transfer energy from one time interval to another one which is not connected to the former; to put it in other words, we may speculate on extended probabilities as accounting for deviations from the macroscopic arrow of time.

Applying the concept of quenching events examined in chapter 4, we have to face the situation that events which have been detected and stored, say in a note book, will be erased later on [126]. This concept exhibits even more unreasonable features, if applied to the EPR–Bohm experiment. An event appearing with positive probability in one of the detectors together with a correlated one appearing with negative probability in the other detector, will result in cancelling a (hopefully) preceding pair of events of same character (identical or opposite results of spin measurements), which occurred with entirely positive (or entirely negative) probability. No idea about the dimensions of the space-time interval allowing for this kind of interference to occur can be given; and if it could, easily experiments exceeding such a limit could be set up. The list of strange consequences can be continued, and the model of quenching events is highly questionable and justified by nothing than the fact that other attempts to explain how nature works in the EPR case suffer from the same doubtfulness. In no case extended probabilities are compatible with the notion of local particles existing in an exactly defined reality. On the other hand, the non-locality entailed is conceptually different from direct non-local interactions sometimes considered as a resolution of the EPR paradox, since the result of a measurement (spin up or down as normal or quenching event) in fact does not depend on any distant circumstances. The non-locality involved rather may be compared to the non-locality of negative amounts of money. The corresponding positive amount must be somewhere else (and the total sum of all money cannot be negative). Thus the non-locality is shifted from the experimental situation to the level on which the results are interpreted and compared.

Whether the problems appearing in quantum electrodynamics result from the "reality" of extended probabilities is open to question. But it should be noted that numerous attempts have been made the results of which were not capable of eliminating the pending problems. Each attempt suffers from unreasonable features; also the most recent theory by the Vigier group (cf. section 6.2) does not avoid extended probabilities in case of potentials which are strong enough to allow for pair creation.

Here we have gathered a collection of problems of present day physics, which may be treated by using extended probabilities, and hence, be reduced to one single problem, namely the interpretation of extended probabilities.

Nevertheless, there remain two essential points which prevent an affirmative answer to questions (6) and (7): The first obstacle is caused by the lack of a reasonable interpretation. The second is due to the fact that at present there is no consistent theory of extended probability, but only some formalisms which allow for a reproduction of established results. On the other hand, theoretic concepts and auxiliary theoretic concepts as defined by Ludwig (cf. section 6.1) cannot be sharply distinguished; the border between these sets is blurred and is at least depending on the development of experimental physics; and it seems worthwhile to recall the paraphrase of Siegel's (see chapter 5) according to which "*a purely mental construct can gradually acquire 'reality' if it explains an experimental result not easily explained otherwise*", which is completely in agreement with the physical definition of reality. Thus, from the general trend mentioned by few examples in chapter 2 (see also section 6.5) and by numerous others in history of science, we find some support to express our denial in the form of a *conditional* "no".

Turning to question (5) it must be confessed that there is no evidence at present for new experimental applications accomplished or facilitated by the concept of negative probability. This, however, may be due to the scarce research up to now invested along these lines.

Considering question (4) we may arrive at different answers. If the Copenhagen interpretation (forming quantum theory together with the current formalism) is accepted, then quantum theory is with no doubt a splendid theory describing the properties of reality with outstanding accuracy (at least in the original domain of non-relativistic quantum *mechanics*) and no improvement seems possible. Then extended probabilities arising from its formalism are merely due to a mathematical segregation of physical entities which is not permitted by the Copenhagen interpretation. This interpretation, however, cannot be considered an interpretation in the sense of describing, explaining or visualizing anything on a stage beyond the Copenhagen prohibition. Although this stage is sometimes called the ontologic or metaphysical one, in order to distinguish it from "true" physics, this declaration is highly debatable because the only offence of this arguing consists of universally applying some concepts which are supported and even required by other well-proved physical theories. From this viewpoint the Copenhagen interpretation can be charged with precisely its characteristic verdict, i.e., to entail an artificial segregation of things which inherently belong together unless "physics" is used synonymously to "engineering". Once having crossed the border created by the Copenhagen prohibition we are immediately faced to a variety of problems; and applying the Copenhagen interpretation to this domain (which it is not destined for) it may be compared rather to astrology than to astronomy. Hence, violating the Copenhagen prohibition but maintaining the fundamental axioms, we arrive at situations which are described by equations enforcing extended probabilities. From this point of view, they might be considered as improving the situation of present day physics in case a consistent theory of extended probabilities was achievable (which, however, has not yet been presented).

In order to treat question (3) we must distinguish between different examples of extended probabilities discussed in the previous chapters. Obviously there is no model making them compatible with experimental results obtained in EPR experiments in any reasonable way. On the other hand, there is direct evidence of experimental observation in case of radiative decay analyzed by Mössbauer spectroscopy (see section 3.2). Certainly, it is impossible to observe negative probabilities, but it is as impossible to observe negative energy states. In the latter case, only the absence of an electron in the Dirac sea or the radiation created during the transition from a positive-energy state to a negative-energy state can directly be detected. The same applies to extended probabilities. If, for example, a density is known to be normalized and the result of an experimental integration, extending over a certain space, exceeds unity, this must be interpreted as direct experimental evidence for negative parts of said density. Thus, it can be concluded that in some cases extended probabilities are in agreement with experience. (Their disagreement with experience in other cases may be either due to as yet unknown circumstances or to erroneous application of the concept.)

In addition, a concept can be useful even when direct experimental evidence is lacking. The concept of atoms led Boltzmann and Maxwell to extremely useful formulas [123] and Jaynes' solution of Bertrand's problem would not loose its beauty if an experimental verification was not possible by presently available experimental means. The same applies to extended probabilities in connexion with phase space distributions. In purely formalistic application the interpretative problems may be side-stepped as long as no experimentally falsifiable consequences result. Thus question (2) deserves an affirmative answer although there remains the serious objection [164, 174] that the use of extended probabilities may possibly veil the more fundamental problems of physics without presenting an acceptable solution to them after all. This is with no doubt an important objection which should never be overlooked, but, on the other hand, also current quantum theory does not supply anything but a successful formalism (admittedly a more universal one than extended probabilities) if observed from what sometimes is declared as the "ontologic level".

There are several ways to treat extended probabilities (cf. chapter 4) starting with Bartlett's biassed penny via Feynman's partition of the diffusion function to Gupta's metric. Although strange issues like complex standard deviations may be involved self-contradictions in the formalism of extended probabilities should be avoidable.

In any case, until detailed formalisms will have been established, one may use extended probabilities and interpret the results by Dirac's very pragmatic concept (see the end of chapter 4), but must be carefully aware of their formal character.

Quite a lot of speculations have been presented and the discussion already has become too lengthy with respect to the fact that – as stated in the introduction – no final answers can be given. But the set of "final answers" in physics is very small. Nevertheless this paper should not be finished without at least one statement which can claim to represent an absolute truth: *Kolmogorov's axiom may hold or not; the probability for the existence of negative probabilities is not negative.*

Acknowledgements

I would like to thank all those scholars who supplied suggestions and statements by private communication or furnished papers prior to publication, and, in particular, Professors H. Paul, F. Selleri, C.F. von Weizsäcker and E.P. Wigner for valuable information not covered by the first draft of this review. I am very indebted to Dr. N. Cufaro Petroni, Professors E.T. Jaynes, G. Ludwig and J.P. Vigier and his group for contributing notes. It is a special pleasure for me to express my warmest thanks to Professor Bartlett (whose paper on negative probability originally gave rise to this review) for his kind help in many respects.

References

[1] Dirac, P.A.M., Proc. Roy. Soc. Lond. A 180 (1942) 1.
[2] Siegman, A.E., An Introduction to Lasers and Masers (McGraw-Hill, New York, 1971) p. 503.
[3] Mückenheim, W., Phys. Bl. 39 (1983) 331.
[4] Bondi, H., Observatory 82 (1962) 133.
[5] Bergmann, P.G., Found. Phys. 1 (1970) 17.
[6] Prokhovnik, S.J., Found. Phys. 3 (1973) 351.
[7] An exhaustive review of the ideas touched in this paragraph including plenty of references is given in: Jammer, M., The Philosophy of Quantum Mechanics (Wiley, New York, 1974) chapter 8.
[8] Loc. cit. p. 416.
[9] Loc. cit. p. 356.
[10] Loc. cit. p. 409.
[11] Feynman, R.P., Int. J. Theor. Phys. 21 (1982) 467.
[12] Wigner, E., Phys. Rev. 40 (1932) 749.
[13] A comprehensive review of Wigner functions and related distribution functions is given in: Hillery, M., R.F. O'Connell, M.O. Scully and E.P. Wigner, Phys. Rep. 106 (1984) 123.
[14] Wigner, E.P., Quantum-mechanical distribution functions revisited, in: Perspectives in Quantum Theory, W. Yourgrau and A. van der Merwe, eds. (MIT Press, Cambridge MA, 1971) p. 25.

[15] Pauli, W., Die allgemeinen Prinzipien der Wellenmechanik, in: Handbuch der Physik, vol. V. part 1, S. Flügge, ed. (Springer, Berlin, 1958) p. 20.
[16] Suppes, P., Phil. Science 28 (1961) 378.
[17] Rottmann, K., Mathematische Formelsammlung (Bibliographisches Institut, Mannheim, 1960) p. 103.
[18] Moyal, J.E., Proc. Cambridge Phil. Soc. 45 (1949) 99.
[19] Margenau, H. and R.N. Hill, Progr. Theor. Phys. 26 (1961) 722.
[20] Cohen, L., J. Math. Phys. 7 (1966) 781.
[21] Cohen, L., Phil. Science 33 (1966) 317.
[22] Cohen, L., Thesis, Yale University, 1966.
[23] Summerfield, G.C. and P.F. Zweifel, J. Math. Phys. 10 (1969) 233.
[24] Carruthers, P. and F. Zachariasen, Rev. Mod. Phys. 55 (1983) 245.
[25] Imre, K.,E. Özizmir, M. Rosenbaum and P.F. Zweifel, J. Math. Phys. 8 (1967) 1097.
[26] Bopp, F., The principles of the statistical equations of motion in quantum-theory, in: Observation and Interpretation, S. Körner, ed. (Butterworths, London, 1957) p. 194.
[27] Husimi, K., Proc. Phys. Math. Soc. Japan 22 (1940) 264.
[28] Iagolnitzer, D., J. Math. Phys. 10 (1969) 1241.
[29] Compare, e.g., Bjorken, J.D. and S.D. Drell, Relativistic Quantum Mechanics (McGraw-Hill, New York, 1964).
[30] Born, M., Z. Phys. 37 (1926) 863, footnote.
[31] Gordon W., Z. Phys. 40 (1926) 117.
[32] Klein, O., Z. Phys. 41 (1927) 407.
[33] Dirac, P.A.M., Proc. Roy. Soc. Lond. A 322 (1971) 435.
[34] Dirac, P.A.M., Proc. Roy. Soc. Lond. A 328 (1972) 1.
[35] Majorana, E., Nuovo Cim. 9 (1932) 335.
[36] Sudarshan, E.C.G., N. Mukunda and C.C. Chiang, Phys. Rev. D 25 (1982) 3237.
[37] Dirac, P.A.M., Talk on the meeting of Nobel-prize laureates, Lindau 1982, Phys. Bl. 38 (1982) 282.
[38] Dirac, P.A.M., Comm. Dublin Inst. Adv. Stud. A, no. 1 (1943).
[39] Jaynes, E.T. and F.W. Cummings, Proc. IEEE 51 (1963) 89.
[40] Jaynes, E.T., Phys. Rev. A 2 (1970) 260.
[41] Clauser, J.F., Phys. Rev. A 6 (1972) 49.
[42] Clauser, J.F., Phys. Rev. D 9 (1974) 853.
[43] Jaynes, E.T. in: Coherence and Quantum Optics, L. Mandel and E. Wolf, eds. (Plenum Press, New York, 1973) p. 35.
[44] Heisenberg, W. and W. Pauli, Z. Phys. 56 (1929) 1.
[45] Heisenberg, W. and W. Pauli, Z. Phys. 59 (1929) 168.
[46] Pauli, W., Rev. Mod. Phys. 15 (1943) 175.
[47] Gupta, S.N., Quantum electrodynamics (Gordon and Breach, London 1977) p. 30 ff.
[48] Weisskopf, V. and E. Wigner, Z. Phys. 63 (1930) 54.
[49] Holland, R.E., F.J. Lynch, G.J. Perlow and S.S. Hanna, Phys. Rev. Lett. 4 (1960) 181.
[50] Lynch, F.J., R.E. Holland and M. Hamermesh, Phys. Rev. 120 (1960) 513.
[51] Wu, C.S., Y.K. Lee, N. Benczer-Koller and P. Simms, Phys. Rev. Lett. 5 (1960) 432.
[52] Belinfante, F.J., A Survey of Hidden-Variables Theories (Pergamon, Oxford 1973) p. 254.
[53] Glauber, R.J., Phys. Rev. Lett. 10 (1963) 84.
[54] Sudarshan, E.C.G., Phys. Rev. Lett. 10 (1963) 277.
[55] Reid, M.D. and D.F. Walls, Phys. Rev. Lett. 53 (1984) 955.
[56] Feynman, R.P., Rev. Mod. Phys. 20 (1948) 367.
[57] Feynman, R.P., Phys. Rev. 74 (1948) 1430.
[58] Majumdar, R.C. and S.N. Gupta, Proc. Nat. Inst. Sci. Ind. 13 (1947) 187.
[59] Gupta, S.N., Proc. Phys. Soc. A 63 (1950) 681.
[60] Bleuler, K., Helv. Phys. Acta 23 (1950) 567.
[61] Heisenberg, W., Physik und Philosophie (Ullstein, Frankfurt, 1968) p. 134.
[62] Heisenberg, W., Einführung in die Einheitliche Feldtheorie der Elementarteilchen (Hirzel, Stuttgart, 1967).
[63] Lee, T.D., Phys. Rev. 95 (1954) 1329.
[64] Källén, G. and W. Pauli, Kong. Dansk Vid. Selskab Mat. Fys. Medd. 30 (1955) No. 7.
[65] Cufaro Petroni, N. and J.P. Vigier, Found. Phys. 13 (1983) 253.
[66] Vigier, J.P. and Ya. P. Terletskij, Sov. Phys. J.E.T.P. 13 (1961) 356.
[67] Feynman, R.P., Phys. Rev. 76 (1949) 749, 769.
[68] See ref. [29], chapter 1.
[69] Feynman, R.P. and M. Gell-Mann, Phys. Rev. 109 (1958) 193.
[70] Cufaro Petroni, N., Ph. Gueret and J.P. Vigier, Nuovo Cim. 81B (1984) 243.

[71] Holland, P.R. and J.P. Vigier, Positive probabilities and the principle of equivalence for spin zero particles in the causal stochastic interpretation of quantum mechanics, Nuovo Cim. B (1985) to be published.
[72] Holland, P.R., A. Kyprianidis and J.P. Vigier, Phys. Lett. 107 A (1985) 376.
[73] Cufaro Petroni, N., C. Dewdney, P. Holland, A. Kyprianidis and J.P. Vigier: Causal space-time paths of individual distinguishable particle motions in n-body quantum systems: elimination of negative probabilities, Lett. Nuovo Cim. (1985) to be published.
[74] Gueret, Ph., P.R. Holland, A. Kyprianidis and J.P. Vigier, Phys. Lett 107 A (1985) 379.
[75] Cufaro Petroni, N., C. Dewdney, P. Holland, A. Kyprianidis and J.P. Vigier, Phys. Lett. 106 A (1984) 368.
[76] Bjorken, J.D. and S.D. Drell, Relativistic Quantum Fields (McGraw-Hill, New York, 1965).
[77] Bohr, N., Phys. Rev. 48 (1935) 696.
[78] Einstein, A., B. Podolsky and N. Rosen, Phys. Rev. 47 (1935) 777.
[79] Bohm, D., Quantum Theory (Prentice-Hall, New York, 1951) p. 611 ff.
[80] Schrödinger, E., Die Naturwissenschaften 23 (1935) 844.
[81] Bell, J.S., Physics 1 (1964) 195.
[82] Paul, H., Fortschr. Phys. 28 (1980) 633.
[83] Clauser, J.F. and A. Shimony, Rep. Progr. Phys. 41 (1978) 1881.
[84] Wigner, E.P., Am. J. Phys. 38 (1970) 1005.
[85] Clauser, J.F., M.A. Horne, A. Shimony and R.A. Holt, Phys. Rev. Lett. 23 (1969) 880.
[86] Clauser, J.F. and M.A. Horne, Phys. Rev. D 10 (1974) 526.
[87] Bell, J.S., in: Proc. Int. School of Physics "Enrico Fermi", Course IL, B. d'Espagnat, ed. (Academic, New York, 1971) p. 171.
[88] Mermin, N.D., Phys. Rev. D 22 (1980) 356.
[89] Mermin, N.D., Generalization of Bell's theorem to higher spins and higher correlations, preprint, to appear in the proceedings of the Symposium on Fundamental Questions in Quantum Mechanics, SUNY Albany, 1984.
[90] The first decisive experiment was reported by Freedman, S.J. and J.F. Clauser, Phys. Rev. Lett. 28 (1972) 938; a review of early experiments is given in [83]; for more recent references cf. [3].
[91] Heisenberg, W., Physikalische Prinzipien der Quantentheorie (Bibliographisches Institut, Mannheim, 1958).
[92] Furry, W.H., Phys. Rev. 49 (1936) 393.
[93] Feynman, R.P., R.B. Leighton and M. Sands: The Feynman Lectures on Physics. Quantum Mechanics (Addison-Wesley, Reading MA, 1966) 2–8.
[95] Popper, K., Found. Phys. 12 (1982) 971.
Rietdijk, C.W., Found. Phys. 8 (1978) 615.
[96] Selleri, F., Lett. Nuovo Cim. 39 (1984) 252.
[97] Marshall, T.W., E. Santos and F. Selleri, Phys. Lett. 98 A (1983) 5.
[98] Garuccio, A. and F. Selleri, Phys. Lett. 103 A (1984) 99.
[99] Stapp, H.P., Phys. Rev. D 3 (1971) 1303.
[100] Everett III, H., Rev. Mod. Phys. 29 (1957) 454.
[101] Wheeler, J.A., Rev. Mod. Phys. 29 (1957) 463.
[102] DeWitt, B., Phys. Today, Sept (1970) 30.
[103] Bub, J., Found. Phys. 3 (1973) 29.
[104] Gudder, S.P., Stochastic Methods in Quantum Mechanics (Elsevier North-Holland, New York, 1979).
[105] Gudder, S.P., Int. J. Theor. Phys. 19 (1980) 163.
[106] Pitowsky, I., Phys. Rev. Lett. 48 (1982) 1299.
[107] Gudder, S. and N. Zanghi, Nuovo Cim. 79 B (1984) 291.
[108] Gudder, S.P., Reality, Locality and Probability, preprint, University of Denver, MS-R-8404.
[109] Mückenheim, W., Lett. Nuovo Cim. 35 (1982) 300.
[110] Mückenheim, W., Extended probabilities and the EPR-paradox, to be published.
[111] Mückenheim, W., Contradiction of a vector-like spin, to be published.
[112] Feynman, R.P., Negative Probability, preprint.
[113] Cox, D.R., Proc. Cambridge Phil. Soc. 51 (1955) 313.
[114] Brockmeyer, E., H.L. Halstrøm and A. Jensen, The Life and Works of A.K. Erlang (Copenhagen, 1948).
[115] Cf., e.g., von Mises, R., Probability, Statistics and Truth (Macmillan, New York, 1957).
[116] Popper, K.R., The propensity interpretation of the calculus of probability and the quantum theory, in: Observation and Interpretation, S. Körner, ed. (Butterworths, London, 1957) p. 65.
[117] Grossman, N., Phil. Science 39 (1972) 451.
[118] Bartlett, M.S., Proc. Cambridge Phil. Soc. 41 (1945) 71.
[119] Bartlett, M.S., private communication.
[120] Bartlett, M.S., J. Roy. Statist. Soc. 103 (1940) 1
[121] Eddington, A.S., Commun. Dublin Inst. Adv. Stud. A, no. (1943) p. 5.
[122] Gupta, S.N., Canad. J. Phys. 35 (1957) 961.

[123] Jaynes, E.T., Found. Phys. 3 (1973) 477.
[124] von Mises, R. in: Mathematical Theory of Probability and Statistics, H. Geiringer, ed. (Academic, New York, 1964) p. 160.
[125] Dirac, P.A.M., Rev. Mod. Phys. 17 (1945) 195.
[126] Aspect, A., private communication.
[127] Bartlett, M.S., Dialectica 3, no. 9/10 (1949) 104.
[128] Barut, A.O. and G.H. Mullen, Ann. of Phys. 20 (1962) 184.
[129] Barut, A.O. and G.H. Mullen, Ann. of Phys. 20 (1962) 203.
[130] Crawford, J.P. and A.O. Barut, Phys. Rev. D 27 (1983) 2493.
[131] Barut, A.O. and P. Meystre, Phys. Lett. 105 A (1984) 458.
[132] Barut, A.O., private communication.
[133] Bell, J.S., private communication.
[134] Bopp, F.: Statistische Mechanik bei Störung des Zustandes eines physikalischen Systems durch die Beobachtung, in: Werner Heisenberg und die Physik in Unserer Zeit, F. Bopp, ed. (Vieweg, Braunschweig, 1961) p. 128.
[135] Bopp, F., private communication.
[136] Bordley, R.F., J. Math. Phys. 24 (1983) 2411.
[137] de Broglie, L., Non-Linear Wave Mechanics (Elsevier, Amsterdam, 1960) p. 70.
[138] Broyles, A.A., Phys. Rev. D 25 (1982) 3230.
[139] Bugajski, S., Z. Naturforsch. 33a (1978) 1383.
[140] Döring, W., private communication.
[141] d'Espagnat, B., Phys. Rep. 110 (1984) 201.
[142] Heisenberg, W., Introduction to the Unified Field Theory of Elementary Particles (Interscience, London, 1966) (English version of [62])
[143] Jauch, J.M., Found. Phys. 1 (1970) 183.
[144] Jaynes, E.T., Predictive Statistical Mechanics, A summary of five talks, preprint to appear in the proceedings of NATO ASI, Santa Fe, New Mexico, 1984.
[145] Jaynes, E.T., Some applications and extensions of the de Finetti representation theorem, preprint to appear in: Bayesian Inference and Decesion Techniques with Application: Essays in Honor of Bruno de Finetti, P.K. Goel and A. Zellner, eds. (North-Holland, Amsterdam, to be published).
[146] Jaynes, E.T, private communication.
[147] Jeffreys, H., Phil. Magazine (7) 33 (1942) 815.
[148] Kolmogorov, A., Grundbegriffe der Wahrscheinlichkeitsrechnung (Springer, Berlin, 1933).
[149] Lüders, G., private communication.
[150] Ludwig, G., private communication.
[151] Margenau, H., private communication.
[152] Einstein, A., Dialectica 2 (1948) 320.
[153] Marshall, T.W., E. Santos and F. Selleri, Lett. Nuovo Cim. 38 (1983) 417.
[154] Marshall, T.W., Proc. Roy. Soc. Lond. A 276 (1963) 475.
[155] Marshall, T.W., private communication.
[156] Mermin, N.D., private communication.
[157] Meystre, P., Is reality really real? - An introduction to Bell's inequalities, preprint to appear in the proceedings of NATO ASI, Santa Fe, New Mexico, 1984.
[158] Park, J.L., Phil. Science 35 (1968) 205.
[159] Paul, H., private communication.
[160] Richter, G., private communication.
[161] Salam, A., private communication.
[162] Santos, E., private communication.
[163] Selleri, F., private communication.
[164] Shimony, A., private communication.
[165] Siegel, A., private communication.
[166] Koopman, O., Proc. Appl. Math. vol. VII, Applied Probability, New York, 1947, p. 97 ff.
[167] Treder, H.-J., Annalen der Phys. (7) 40 (1983) 227.
[168] Treder, H.-J., Epistemol. Lett. 34 (1983) 17.
[169] Treder, H.-J., Epistemol. Lett. 35 (1983) 49.
[170] Treder, H.-J., private communication.
[171] Varadarajan, V.S., Comm. Pure and Appl. Math. 15 (1962) 189.
[172] v. Weizsäcker, C.F., private communication.
[173] Wigner, E.P., private communication.
[174] Zeh, H.D., private communication.
[175] Zlatev, I.S., Compt. rend. de l'Académie bulgare des Sciences 27 (1974) 1489.

[176] Ludwig, G., Grundstrukturen einer physikalischen Theorie (Springer, New York, 1978).
[177] Ludwig, G., Restriction and Embedding, in: Reduction in Science, Structure, Examples, Philosophical Problems, W. Balzer, D. Pearce, H.-J. Schmidt, eds. (Reidel, Dordrecht, 1984).
[178] Ludwig, G., An Axiomatic Basis for Quantum Mechanics (2 volumes) (Springer, New York, 1985).
[179] Ludwig, G., Zeitschrift für allgemeine Wissenschaftstheorie XII/1 (1981) 55.
[180] Ludwig, G., Foundation of Quantum Mechanics (2 volumes) (Springer, New York, 1983 and 1985).
[181] Ludwig, G., Imprecision in Physics, in: Structure and Approximation in Physical Theories, A. Hartkämper and H.-J. Schmidt, eds. (Plenum Press, New York, 1981).
[182] Bohm, D., Phys. Rev. 85 (1952) 166.
[183] See ref. [137] in general.
[184] Vigier, J.P., Astron. Nachr. 303 (1982) 55.
[185] Feshbach, H. and F. Villars, Rev. Mod. Phys. 30 (1958) 24.
[186] Rizov, V.A., H. Sazdjian and I.T. Todorov, On the relativistic quantum mechanics of two interacting spinless particles, Orsay preprint (1984).
[187] de Broglie, L., The Current Interpretation of Wave Mechanics (Elsevier, Amsterdam, 1964).
[188] Halbwachs, F., Théories relativistes des fluides à spin (Gauthier-Villars, Paris, 1960).
[189] Cufaro Petroni, N., Ph. Gueret and J.P. Vigier, Phys. Rev. D 30 (1984) 495.
[190] Cufaro Petroni, N., Ph. Gueret, A. Kyprianidis and J.P. Vigier, An alternative derivation of the spin dependent quantum potential, preprint.
[191] Bohm, D., Phys. Rev. 85 (1952) 180.
[192] Bohm, D., Phys. Rev. 89 (1953) 458.
[193] Bohm, D. and J.P. Vigier, Phys. Rev. 96 (1954) 208.
[194] Dirac, P.A.M., Nature 168 (1951) 906.
[195] Dirac, P.A.M., Nature 169 (1952) 702.
[196] Sinha, K.P., C. Sivaram and E.C.G. Sudarshan, Found. Phys. 6 (1976) 65, 717.
[197] Lehr, W. and J. Park, J. Math. Phys. 18 (1977) 1235.
[198] Guerra, F. and P. Ruggiero, Lett. Nuovo Cim. 23 (1978) 529.
[199] Cufaro Petroni, N. and J.P. Vigier, Phys. Lett. 73 A (1979) 289.
[200] Cufaro Petroni, N. and J.P. Vigier, Phys. Lett. 81 A (1981) 12.
[201] Cufaro Petroni, N. and J.P. Vigier, J. Phys. A 17 (1984) 599.
[202] Cufaro Petroni, N., Ph. Droz-Vincent and J.P. Vigier, Lett. Nuovo Cim. 31 (1981) 415.
[203] Cufaro Petroni, N. and J.P. Vigier, Phys. Lett. 88 A (1982) 272.
[204] Cufaro Petroni, N. and J.P. Vigier, Phys. Lett. 93 A (1983) 383.
[205] Aspect, A., G. Grangier and G. Roger, Phys. Rev. Lett. 47 (1981) 460.
[206] Aspect, A., G. Grangier and G. Roger, Phys. Rev. Lett. 49 (1982) 91.
[207] Aspect, A., J. Dalibard and G. Roger, Phys. Rev. Lett. 49 (1982) 1804.
[208] Bourbaki, N., Elements de Mathématiques; XIII - Integration, sect. 4 (Hermann, Paris, 1965).
[209] Bartlett, M.S., Probability, Statistics and Time (Chapman and Hall, London, 1975).
[210] Bartlett, M.S., An Introduction to Stochastic Processes (Cambridge Univ. Press, Cambridge, 1978) 3rd ed.
[211] Gaveau, B., T. Jacobson, M. Kac and L.S. Schulman, Phys. Rev. Lett. 53 (1984) 419.
[212] Cox, R.T., The Algebra of Probable Inference (Johns Hopkins Univ. Press, Baltimore, 1961) reviewed by E.T. Jaynes, Am. J. Phys. 31 (1963) 66.
[213] Freund, H.U. and J.C. McGeorge, Z. Phys. 238 (1970) 6.