

Mathematische Begründung der Quantenmechanik.

Mathematical Foundation of Quantum Mechanics

By J. v. Neumann, Göttingen.

Presented by M. Born at the meeting of 20 May 1927.

Introduction

I. The formulations of “quantum mechanics” given by Heisenberg, Dirac, Born, Schrödinger and Jordan¹ have raised many entirely new concepts and questions, of which we should like to single out the following:

α . It has become apparent that the behaviour of an atomic system is somehow connected with a certain eigenvalue problem—how this is to be formulated will be touched upon later, in § XII—and, in particular, that the values of the characteristic quantities describing the system are the eigenvalues themselves.

β . In this way the long-sought fusion of the continuous (classical-mechanical) and the discontinuous (quantized) in the world of atoms has been satisfactorily achieved: an eigenvalue spectrum can, after all, have both continuous and discontinuous parts.

γ . Furthermore, indications have appeared in the new quantum mechanics that the laws of nature (or at least the quantum laws known to us) do not determine atomic events uniquely and causally; rather, the elementary laws merely specify probability distributions, which degenerate into causal precision only in exceptional cases.

δ . The eigenvalue problem appears in different guises: as the eigenvalue problem of an infinite matrix (i.e., its transformation to diagonal form), and as that of a differential equation. Nevertheless, the two formulations are equivalent: for the matrix (regarded as a linear transformation) arises from the differential operator (which, applied to the “wave function”, gives the left-hand side of the

¹Several papers by the authors named in the 1926/27 volumes of *Ann. der Physik*, *Zeitschr. für Physik*, and *Proc. of the Royal Soc.*

differential equation²), when one passes from the “wave function” to its expansion coefficients with respect to a complete orthogonal system.³ (The matrix then effects the corresponding transformation of these expansion coefficients.)

ε. Both methods of treatment have their difficulties. With the matrix method one is, in fact, almost always faced with an insoluble problem: transforming the energy matrix to diagonal form. This is possible only when there is no continuous spectrum,⁴ i.e., this treatment is one-sided (although in the opposite sense to classical mechanics): only the discontinuous (quantized) appears in it. (The hydrogen atom—which also has a continuous spectrum⁵—therefore cannot be treated correctly in this way.) One can, to be sure, make do by using “continuous matrices”,⁶ but this procedure (actually a simultaneous operation with matrices and integral-equation kernels) can probably be carried out with mathematical rigour only with great difficulty: one must introduce such concepts as infinitely large matrix elements or infinitely closely neighbouring diagonals.

ζ. In the treatment by the differential-equation method, the probability assumptions of the matrix method were initially absent. (They will be discussed in detail further below.) This omission was remedied by Born, and later by Pauli and Jordan; nevertheless, the complete procedure, as developed by Jordan into a closed system,⁷ is also open to serious mathematical objections. For one cannot avoid admitting so-called improper eigenfunctions as well (cf. § IX); for example, the function $\delta(x)$ first used by Dirac, which is supposed to possess the following (absurd) properties:

$$\delta(x) = 0 \quad \text{for } x \neq 0,$$

²The typical (Schrödinger) differential equation of quantum mechanics has the form

$$H\psi = \lambda\psi,$$

where H is a differential operator, λ the eigenvalue parameter, and ψ the “wave function”, upon which certain regularity and boundary-value conditions are imposed, giving rise to the eigenvalue problem.

³Schrödinger, *Ann. der Physik*, vol. 79/8, p. 734 (1926).

⁴Hellinger, inaugural dissertation, § 4 (Göttingen, 1907). The necessity of this condition is immediately evident: the spectrum remains invariant under a transformation, and a diagonal matrix has only a discontinuous spectrum (with the diagonal elements as eigenvalues).

⁵Schrödinger, *Ann. der Physik*, vol. 79/4, p. 367 (1926).

⁶See principally Dirac, *Proc. of the Royal Soc.*, vol. 113, p. 621 (1927).

⁷*Zeitschr. für Physik*, vol. 40, nos. 11/12, p. 809 (1927). See also a paper by Hilbert, Nordheim and the author, shortly to appear in the *Math. Ann.*

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

A particular difficulty with Jordan is that one must calculate not only his transforming operators (whose integral kernels are the “probability amplitudes”), but also the range of variables into which the transformation is made (i.e., the eigenvalue spectrum).

∅. A defect common to all these methods, however, is that they introduce into the calculation elements which are in principle unobservable and physically meaningless: the eigenfunctions must be calculated, although, owing to their normalization, they remain undetermined up to a constant of absolute value 1 (the “phase” $e^{i\varphi}$), and indeed, in the case of κ -fold degeneracy (i.e., a κ -fold eigenvalue), up to a κ -dimensional orthogonal⁸ transformation among themselves. The probabilities appearing as final results are, admittedly, invariant; but it is unsatisfactory and unclear why the detour through the unobservable and non-invariant should be necessary.

The present paper attempts to give a method which remedies these defects and, as we believe, brings together the present statistical standpoint in quantum mechanics in a unified and consistent manner.

Wherever calculation did not belong to the essence of the matter, we have avoided entering into purely mathematical questions; nevertheless, our arguments are to be regarded as essentially mathematically rigorous. It was unavoidable that the greater part of the paper should be devoted to explaining and establishing the formal concepts subsequently applied.

Accordingly, § II–§ XI are preparatory in character; the actual subject is treated only in § XI–§ XIV.

⁸What is meant is the complex orthogonality of the transformation matrix $\{\alpha_{\mu\nu}\}$:

$$\sum_{\rho=1}^{\kappa} \alpha_{\mu\rho} \overline{\alpha_{\nu\rho}} = \begin{cases} 1, & \mu = \nu, \\ 0, & \mu \neq \nu, \end{cases}$$

which leaves the “Hermitian unit form”

$$\sum_{\rho=1}^{\kappa} x_{\rho} \overline{y_{\rho}}$$

invariant. The name “unitary” is customary for this in the mathematical literature.

Hilbert Space

II. As we saw in I.δ, the eigenvalue problems of quantum mechanics occur in two principal forms: as eigenvalue problems of infinite matrices (or, what is the same thing, bilinear forms), and as eigenvalue problems of differential equations.

We shall consider the two forms separately and emphasize their common features. In doing so, we shall deal entirely with facts which have long been known mathematically and offer nothing new physically either, since they occur essentially in the paper by Schrödinger cited in note 3 and in several papers by Dirac. Nevertheless, it may be appropriate to develop everything in context, and it is also useful to put these discussions first, in order to motivate the abstract concepts of the following sections.

Let us first consider the matrix formulation. Here an infinite matrix is given (representing the energy; how it is obtained will be discussed later), and the task is to transform it to diagonal form (for the diagonal terms are then the energy levels⁹). Let us assume that this can be done without difficulty (i.e., that there is only a point spectrum; cf. I.ε and note 4).

We denote the energy matrix by

$$H = \{h_{\mu\nu}\} \quad (\mu, \nu = 1, 2, \dots);$$

it is to be assumed Hermitian, i.e.,

$$h_{\mu\nu} = \overline{h_{\nu\mu}}.$$

We seek a transformation matrix

$$S = \{s_{\mu\nu}\} \quad (\mu, \nu = 1, 2, \dots)$$

which accomplishes the following: it is orthogonal, i.e.,

$$\sum_{\rho=1}^{\infty} s_{\mu\rho} \overline{s_{\nu\rho}} = \begin{cases} 1, & \mu = \nu, \\ 0, & \mu \neq \nu, \end{cases}$$

and $S^{-1}HS$ has diagonal form.

We denote the matrix $S^{-1}HS$ by W ; let the diagonal terms of this (diagonal) matrix be $w_1, w_2, \dots$. The requirement is then:

$$HS = SW,$$

⁹See, for example, Born, *Probleme der Atomdynamik*, p. 86 (Berlin, 1926).

$$\sum_{\rho=1}^{\infty} h_{\mu\rho} s_{\rho\nu} = s_{\mu\nu} w_{\nu}.$$

That is, the ν th column of S , $s_{1\nu}, s_{2\nu}, \dots$, is transformed by H into w_{ν} times itself. Thus every column of S is a solution of the eigenvalue problem: to find those sequences $x_1, x_2, \dots$ which are transformed by H into a multiple—say w times—of themselves. ($x_1, x_2, \dots$ is then an eigensequence, and the proportionality factor w an eigenvalue; the trivial solution $0, 0, \dots$ is, of course, excluded. Thus the eigenvalue belonging to $s_{1\nu}, s_{2\nu}, \dots$ is w_{ν} .)

One can now show that these are essentially the only solutions; more precisely: there are no eigenvalues distinct from $w_1, w_2, \dots$, and if w is an eigenvalue, its eigensequences are the linear combinations of all columns $s_{1\nu}, s_{2\nu}, \dots$ for which $w_{\nu} = w$. (Cf. [Appendix 1](#).)

Consequently, the determination of the matrix s is essentially accomplished as soon as the eigenvalue problem, formulated as above, has been completely solved.

In the differential-equation formulation the situation is even clearer: an eigenvalue problem is given from the outset. A differential operator H is given (thus, for example, for the rotator, oscillator and hydrogen atom, respectively,^{T1}

$$\begin{aligned} & \frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \dots \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \dots, \\ & \frac{d^2}{dq^2} \dots - \frac{16\pi^4}{h^2} \nu_0^2 q^2 \dots, \\ & \frac{\partial^2}{\partial x^2} \dots + \frac{\partial^2}{\partial y^2} \dots + \frac{\partial^2}{\partial z^2} \dots + \frac{8\pi^2 m e^2}{h^2} \frac{1}{\sqrt{x^2 + y^2 + z^2}} \dots \end{aligned}$$

and one seeks a function ψ (in our examples ψ is regarded as a function of $\vartheta, \varphi; q; x, y, z$, respectively) for which

$$H\psi = w\psi,$$

i.e., which is transformed by H into a multiple of itself (the eigenvalues w are again the energy levels, apart from a factor $8\pi^2 m/h^2$). Of course, ψ must satisfy certain regularity requirements and must not vanish identically.¹⁰

What is the common underlying feature of all these cases? Evidently this: in each case a manifold of certain quantities is given (namely, that of all number

^{T1}The original omits the exponent 2 from the derivative symbols in the rotator's φ -term and in the oscillator term. The intended second derivatives are written explicitly here.

¹⁰Schrödinger, *Ann. der Physik*, vol. 79/4, p. 361, and no. 6, p. 489 (1926).

sequences $x_1, x_2, \dots$, or of all functions ψ of two angles ϑ, φ , or of one coordinate q , or of three coordinates x, y, z , together with a linear operator H in this manifold. In each case all solutions of the eigenvalue problem belonging to H are sought, i.e., all (real) numbers w for which there is a non-vanishing element f of this manifold such that

$$Hf = wf.$$

These eigenvalues w then represent the energy levels.

Our task is now to pass from this unified formulation to a unified problem. We shall do this by showing that all the manifolds just mentioned (as well as, generally, all those to which the questions customary in quantum mechanics today can lead) are essentially identical with one another; i.e., that they can all be obtained from a single manifold (to be described in the following sections) by a mere change of names.

For this purpose, however, we must specify more precisely which number sequences or functions are to be admitted into the above manifolds; i.e., we must state the regularity and boundary conditions, which are of decisive importance in eigenvalue problems.

III. We begin with the manifold of number sequences $x_1, x_2, \dots$. Here it is natural to require the finiteness of the square sum $\sum_{n=1}^{\infty} |x_n|^2$. This is indeed satisfied by the solutions of the eigenvalue problem (as long as there is only a point spectrum), which are the columns $s_{1\nu}, s_{2\nu}, \dots$ of the matrix S (cf. § II); for, as was required there,

$$\sum_{\rho=1}^{\infty} s_{\mu\rho} \overline{s_{\mu\rho}} = \sum_{\rho=1}^{\infty} |s_{\mu\rho}|^2 = 1.$$

(When a continuous spectrum is present, the eigenvalue problem can no longer be solved within it by eigensequences $x_1, x_2, \dots$ with finite $\sum_{n=1}^{\infty} |x_n|^2$. It will become apparent below that our approach nevertheless permits a complete treatment of continuous spectra as well.)

Moreover, even the simplest linear transformations H (i.e., matrices $\{h_{\mu\nu}\}$, $\mu, \nu = 1, 2, \dots$) fail when applied to sequences $x_1, x_2, \dots$ with arbitrary $\sum_{n=1}^{\infty} |x_n|^2$: the series $\sum_{\nu=1}^{\infty} h_{\mu\nu} x_{\nu}$ need not converge. (As will become apparent, all the matrices of quantum mechanics are such that the row square sums $\sum_{\nu=1}^{\infty} |h_{\mu\nu}|^2$ are all finite; and for these, the finiteness of $\sum_{\nu=1}^{\infty} |x_{\nu}|^2$ also implies the convergence of $\sum_{\nu=1}^{\infty} h_{\mu\nu} x_{\nu}$.¹¹)

¹¹By the inequality

$$|h_{\mu\nu} x_{\nu}| \leq \frac{1}{2} |h_{\mu\nu}|^2 + \frac{1}{2} |x_{\nu}|^2,$$

Finally, it is precisely this restriction of the range of variation of $x_1, x_2, \dots$ which has proved most successful in the theory of infinite matrices (on which, or on a suitable generalization of which, the mathematical construction of quantum mechanics must rest).¹²

It is thus well motivated to delimit, in this case, the following manifold: all sequences $x_1, x_2, \dots$ of complex numbers with finite $\sum_{n=1}^{\infty} |x_n|^2$. The name (complex) Hilbert space is customary for it.

Let us now pass to the manifolds of functions mentioned in § II. Here the situation regarding the subsidiary conditions (i.e., regularity and boundary conditions) is more involved. It is generally customary to require twice differentiability, as well as single-valuedness, vanishing at infinity or at the boundary of the domain of definition, and the like. How do we arrive at a unified standpoint here?

Let the domain of definition of the functions in question (in our examples: the ϑ, φ space, i.e., the sphere; the q space, i.e., the straight line; the x, y, z space, i.e., ordinary space) be called Ω . Naturally, only functions ψ defined single-valuedly in Ω , to which the operator H is applicable, come into consideration. Thus, if H is, for example, a second-order differential operator (as is generally the case in quantum-mechanical problems), twice differentiability of ψ is required.

To be sure, only at those points of Ω where this actually makes the application of H possible; where, for instance, the coefficients of H have singularities (such as $x = y = z = 0$ in the hydrogen atom), this is not necessary.¹³ Admittedly, as is well known, it is precisely the requirement of regular behaviour at these

the series $\sum_{\nu=1}^{\infty} h_{\mu\nu} x_{\nu}$ even converges absolutely.

¹²The theory of infinite matrices, or bilinear forms, was essentially founded by Hilbert through a complete clarification of the situation (in particular, the solution of the eigenvalue problem) in the important classes of so-called completely continuous and (the more inclusive class of) bounded bilinear forms. (Cf. *Gött. Nachrichten*, Math.-Phys. Klasse, 1906, pp. 159–227.)

A large part of the mathematical obscurities and difficulties of quantum mechanics arises from the fact that even the simplest of its underlying operators (= matrices–bilinear forms) do not belong to the class of bounded ones dealt with by Hilbert.

How the eigenvalue problem of arbitrary, hence also unbounded, operators can be uniquely solved has recently been shown by the author (to appear shortly in the *Math. Annalen*). Naturally, in this paper too we shall have to investigate carefully the formulation of the eigenvalue problem for unbounded operators.

¹³For example, the fundamental solution for the hydrogen atom (in Schrödinger's terminology $n = 1, l = 0$) is

$$\psi(x, y, z) = e^{-\sqrt{x^2+y^2+z^2}},$$

and therefore has a conical point at $x = y = z = 0$.

points which causes an eigenvalue problem to arise at all; but one can (and must) enforce this by substantially weaker conditions.

We shall therefore require $\int_{\Omega} |\psi|^2 dv$ (where dv denotes the element of length, area or volume in Ω , respectively) to remain finite: this excludes too strong a divergence at the singularities of H , and also guarantees part of the boundary conditions, namely vanishing at infinity. The outcome will show that our choice was the right one.

Certain boundary conditions (vanishing at the boundary of a finite domain of definition), however, have not yet been included. To clarify their role, let us recall the following:

For an eigenvalue problem to be possible at all, we had to require Hermitian symmetry for matrices H ,

$$H = \{h_{\mu\nu}\}, \quad h_{\mu\nu} = \overline{h_{\nu\mu}};$$

and for differential equations the condition of self-adjointness is required: that

$$\int_{\Omega} \{\psi_1 \cdot \overline{H\psi_2} - H\psi_1 \cdot \overline{\psi_2}\} dv$$

vanish for all functions ψ_1, ψ_2 which vanish (sufficiently strongly) at the boundary of the domain of integration. (I.e., $\{\psi_1 \cdot \overline{H\psi_2} - H\psi_1 \cdot \overline{\psi_2}\} dv$ is to be the differential of an expression in $\psi_1, \overline{\psi_2}$ and their derivatives.) Here the boundary conditions may be quite important: let, for example, Ω be the interval $0, 1$ and H the operator $i d/dx$. Then

$$\begin{aligned} & \int_{\Omega} \{\psi_1 \cdot \overline{H\psi_2} - H\psi_1 \cdot \overline{\psi_2}\} dv \\ &= \int_0^1 \{\psi_1(x) \cdot [-i\overline{\psi_2'(x)}] - i\psi_1'(x) \cdot \overline{\psi_2(x)}\} dv \\ &= -i \int_0^1 \{\psi_1(x)\overline{\psi_2'(x)} + \psi_1'(x)\overline{\psi_2(x)}\} dv \\ &= -i[\psi_1(x)\overline{\psi_2(x)}]_0^1. \end{aligned}$$

This certainly vanishes if ψ_1, ψ_2 vanish at the endpoints of Ω ; but without any boundary conditions on ψ_1, ψ_2 it may very well be $\neq 0$.

We therefore require further (and, as will become apparent, this requirement absorbs the last essential remainder of the boundary conditions) that ψ be such that the self-adjointness of H is preserved with respect to it.

In all, therefore, we required: ψ is to be defined single-valuedly in Ω ; it is to be such that H is applicable to it and is of self-adjoint character, and

$$\int_{\Omega} |\psi|^2 dv$$

is to be finite. As one sees, the only non-trivial condition (the “regularity requirement” which causes the eigenvalue problem to arise) is the last.

(The objection readily suggests itself that this requirement cuts off access to the continuous spectrum—where the eigenfunctions are never square-integrable. We arranged matters analogously for matrices as well. It will become apparent, however, that precisely this restriction permits a particularly favourable—and mathematically completely rigorous—conception of the continuous spectrum, fully doing justice to its physical meaning, without admitting its eigenfunctions otherwise than as improper objects. Cf. §X.)

Let us summarize:

We have a space $\mathfrak{H}$ (in the matrix interpretation, that of all sequences $x_1, x_2, \dots$ with finite $\sum_{n=1}^{\infty} |x_n|^2$; in the differential-equation interpretation, that of all functions ψ defined in Ω with finite $\int_{\Omega} |\psi|^2 dv$, where Ω may moreover be a surface of arbitrarily many dimensions—bounded or unbounded, finite or infinite), whose elements we shall denote by $f, g, \dots$ for the sake of uniformity. Furthermore, a linear operator H is given (i.e., one which assigns to certain elements f of $\mathfrak{H}$ other elements Hf of $\mathfrak{H}$, and for which

$$\begin{aligned} H(af) &= aHf & (a \text{ a complex constant}), \\ H(f+g) &= Hf + Hg \end{aligned}$$

hold), which is also symmetric or self-adjoint, respectively. We can formulate this last condition uniformly too. For if by $Q(f, g)$ we understand

$$\sum_{n=1}^{\infty} x_n \overline{y_n} \quad \text{or} \quad \int_{\Omega} \varphi \overline{\psi} dv,$$

respectively (thus $f = x_1, x_2, \dots$, $g = y_1, y_2, \dots$ in the matrix case, and $f = \varphi$, $g = \psi$ in the differential-equation case), it simply reads: for the admissible f, g , let

$$Q(f, Hg) = Q(Hf, g).$$

For this operator H we now seek the solutions of the eigenvalue problem:

$$Hf = wf \quad (f \neq 0, w \text{ a real constant}).$$

This gives us, admittedly, only the point spectrum to begin with; but the general conception to be reached in this way will enable us subsequently to encompass the continuous spectrum as well.

As one sees, these formulations now differ only in the underlying space $\mathfrak{H}$. Is it not possible to interpret this too in some unified way?

IV. A natural course would be to exploit the analogous structure of the expressions $\sum_{n=1}^{\infty} |x_n|^2$ and $\int_{\Omega} |\psi|^2 dv$, $\sum_{n=1}^{\infty} x_n \overline{y_n}$ and $\int_{\Omega} \varphi \overline{\psi} dv$, and say: a space R is given which may be discontinuous or continuous (the “space” 1, 2, $\dots$, or an Ω); we consider all functions in R (all sequences $x_1, x_2, \dots$, or all functions ψ in Ω , respectively). There is an “integration over R ” ($\sum_{n=1}^{\infty} x_n$ or $\int_{\Omega} \psi dv$, respectively), and we admit as “proper” functions only those whose squared absolute value has a finite “integral over R ” ($\sum_{n=1}^{\infty} |x_n|^2$ or $\int_{\Omega} |\psi|^2 dv$ finite, respectively). (Of course, R could, where appropriate, be of mixed type, containing, for example, discrete points and intervals, and so on.)

Essentially this course has been followed by Dirac with great consistency and undoubted success in several papers fundamental to quantum mechanics.¹⁴ If we nevertheless seek another solution, the reason is that the analogy sketched above remains a rather superficial one, so long as we adhere to the customary standard of mathematical rigor.

For example, when R is the “space” 1, 2, $\dots$, so that $\mathfrak{H}$ comprises the sequences $x_1, x_2, \dots$, all linear operators H can be represented by matrices $\{h_{\mu\nu}\}$.^{T2}

$$H(x_1, x_2, \dots) = (y_1, y_2, \dots),$$

$$y_{\mu} = \sum_{\nu=1}^{\infty} h_{\mu\nu} x_{\nu}.$$

If we wish to carry over this analogy to the case in which Ω is the interval 0, 1 (so that $\mathfrak{H}$ contains all functions defined there), we must require every linear operator H to be represented by an integral kernel $\Phi(x, y)$:

$$H\varphi(x) = \int_0^1 \Phi(x, y)\varphi(y) dv.$$

But this is not the case even for the very simplest operators (for example, the “identity,” which takes every φ into itself); if one nevertheless wishes to pretend

¹⁴Several papers in the 1926/27 volumes of the *Proceedings of the Royal Society*.

^{T2}The original writes y_n on the left of the matrix-component formula but uses μ as the free row index on the right. The free index has been made consistent here.

that this false statement is correct, one must, as Dirac does, consider “improper” integral kernels, and so on.

We take a different course, which in its essentials already underlies Schrödinger’s “proof of equivalence” (namely, the equivalence of the matrix and differential-equation interpretations; cf. [note 3](#)), and which follows from a mathematical fact long known. To describe it, some preliminaries are necessary.

Let Ω again be any surface (1-, 2-, ...-dimensional, bounded or unbounded, finite or infinite), and dv the differential element of integration on Ω . We consider all (complex-valued) functions $f, g, \dots$ defined on Ω with finite $\int_{\Omega} |f|^2 dv$; these are to form the “space” $\mathfrak{H}$. We write $Q(f, g)$ for $\int_{\Omega} f \bar{g} dv$ (the only combination in which the integral over Ω will be used below). We write $Q(f)$ for $Q(f, f)$ (the finiteness of $Q(f)$ describes $\mathfrak{H}$!).

We call two elements f, g of $\mathfrak{H}$ **orthogonal** if

$$Q(f, g) = 0,$$

and a system $f_1, f_2, \dots$ a **normalized orthogonal system** if

$$Q(f_{\mu}, f_{\nu}) = \begin{cases} 1, & \text{for } \mu = \nu, \\ 0, & \text{for } \mu \neq \nu. \end{cases}$$

Finally, a **complete normalized orthogonal system** is one to which no further f can be added while preserving normalized orthogonality. (This plainly amounts to the same thing as saying that, apart from 0, there is no f orthogonal to all the f_n .)

It can be shown that there are complete orthogonal systems in $\mathfrak{H}$ (for this and all the following assertions stated without proof, see [§ VI^{T3}](#), where they will be discussed more precisely); let $\varphi_1, \varphi_2, \dots$ be one such system.

If f is a function in $\mathfrak{H}$, we call

$$c_{\mu} = Q(f, \varphi_{\mu}) \quad (\mu = 1, 2, \dots)$$

its expansion coefficients (with respect to $\varphi_1, \varphi_2, \dots$). The sum $\sum_{\mu=1}^{\infty} |c_{\mu}|^2$ is always finite (and equals $Q(f)$; this is the so-called Parseval formula), and the series $\sum_{\mu=1}^{\infty} c_{\mu} \varphi_{\mu}$ converges to f in a certain sense. It need not converge at a single point of Ω ; but if we follow the deviation of its N th partial sum from f (that is,

^{T3}The original refers to § V; the orthogonal-system results cited here are established in [§ VI](#).

$\sum_{\mu=1}^N c_\mu \varphi_\mu - f$) over the whole of Ω , and characterize this total behavior by the integral of its squared absolute value,

$$Q\left(\sum_{\mu=1}^N c_\mu \varphi_\mu - f\right),$$

then this integral converges to 0 as $N \rightarrow \infty$ (the smaller it is, the more closely $\sum_{\mu=1}^\infty c_\mu \varphi_\mu$ fits f throughout Ω , rather than at any particular point of Ω). That is,^{T4}

$$\lim_{N \rightarrow \infty} Q\left(\sum_{\mu=1}^N c_\mu \varphi_\mu - f\right) = 0.$$

(This kind of convergence is called *convergence in the mean*—*convergence en moyenne*; it asserts less than “pointwise” convergence, but is the appropriate concept for orthogonal expansions.)

The converse also holds (the theorem of Fischer and F. Riesz¹⁵):

If $c_1, c_2, \dots$ is any sequence of complex numbers with finite $\sum_{\mu=1}^\infty |c_\mu|^2$, then the (function) sum $\sum_{\mu=1}^\infty c_\mu \varphi_\mu$ converges in the mean (that is, there is an f in $\mathfrak{H}$ such that $Q(f - \sum_{\mu=1}^N c_\mu \varphi_\mu)$ tends to 0 as $N \rightarrow \infty$), and its sum f has the expansion coefficients $c_1, c_2, \dots$

We thus have a one-to-one correspondence between the f of $\mathfrak{H}$ and the sequences $c_1, c_2, \dots$ with finite $\sum_{\mu=1}^\infty |c_\mu|^2$. This correspondence is plainly linear: if f goes into $c_1, c_2, \dots$, then af (a a complex constant) goes into $ac_1, ac_2, \dots$; and if f, g go into $c_1, c_2, \dots$ and $d_1, d_2, \dots$, respectively, then $f + g$ goes into $c_1 + d_1, c_2 + d_2, \dots$.^{16, T5} Furthermore, $Q(f, g)$ goes into $\sum_{\mu=1}^\infty c_\mu \bar{d}_\mu$ (this is the more general Par-

^{T4}In the following limit the original prints the lower summation index as $\mu = 0$. It has been corrected to $\mu = 1$, consistently with the system and partial sums defined immediately above. In the converse statement below, the original also uses $n = 1$ under a sum with summand $|c_\mu|^2$; that lower index has likewise been made $\mu = 1$.

¹⁵Göttinger Nachrichten, Math.-Phys. Klasse, 1907, pp. 116–122.

¹⁶That the finiteness of $\sum_{\mu=1}^\infty |c_\mu|^2$, $\sum_{\mu=1}^\infty |d_\mu|^2$, or of $\int_\Omega |f|^2 dv$, $\int_\Omega |g|^2 dv$, implies that of $\sum_{\mu=1}^\infty |c_\mu + d_\mu|^2$, or of $\int_\Omega |f + g|^2 dv$, respectively (and thus the finiteness of $Q(f), Q(g)$ implies that of $Q(f + g)$), follows at once from the relation

$$|u + v|^2 \leq |u + v|^2 + |u - v|^2 = 2|u|^2 + 2|v|^2.$$

^{T5}In original note 16, the exponent in $\int_\Omega |g|^3 dv$ is printed as 3. It is corrected to 2, consistently with $Q(g)$ and the adjoining square-integrability conditions.

seval formula:

$$c_\mu = Q(f, \varphi_\mu), \quad d_\mu = Q(g, \varphi_\mu) \quad (\mu = 1, 2, \dots),$$

$$Q(f, g) = \sum_{\mu=1}^{\infty} c_\mu \overline{d_\mu},$$

which will also be discussed in the next section).

But since in the space of $x_1, x_2, \dots$ we denoted precisely $\sum_{\mu=1}^{\infty} x_\mu \overline{y_\mu}$ by $Q(x, y)$ (x for $x_1, x_2, \dots$, y for $y_1, y_2, \dots$), this means that $\mathfrak{H}$ can be mapped one-to-one onto the space of all sequences $x_1, x_2, \dots$ with finite $\sum_{\mu=1}^{\infty} |x_\mu|^2$, in such a way that the operations af (a a complex constant), $f + g$, $Q(f, g)$ —that is, all the operations we have used so far in describing quantum mechanics—go into themselves (that is, into their analogues in the space just mentioned). Thus all the spaces $\mathfrak{H}$ (belonging to different Ω) differ from the space of sequences $x_1, x_2, \dots$ with finite $\sum_{\mu=1}^{\infty} |x_\mu|^2$ (and hence from one another) only in the naming of the elements and operations; in all their properties, however, they must agree completely.^{T6}

That is: even without introducing “continuous matrices” and “improper entities,” the different sequence and function spaces underlying quantum mechanics are essentially identical, even with absolute mathematical rigor.

V. The space of all sequences of complex numbers $x_1, x_2, \dots$ with finite $\sum_{\mu=1}^{\infty} |x_\mu|^2$ is called complex Euclidean space of infinitely many dimensions, or complex Hilbert space; we shall denote it by $\mathfrak{H}_0$.

As we saw in the last few sections, all function spaces $\mathfrak{H}$ have formal properties agreeing with those of $\mathfrak{H}_0$ and with one another; only the notation or interpretation of their elements differs. It is therefore natural to characterize all these spaces by their common properties, and to call a space possessing all the characteristic properties to be specified an **abstract Hilbert space**. One could then conceive of all the special sequence and function spaces as arising by giving names to its elements (interpreting them as sequences $x_1, x_2, \dots$ or functions ψ on a surface Ω), rather as in relativity theory one obtains special space-time interpretations from the metric homogeneous four-dimensional “world” by introducing special coordinate systems (thus naming the world points by quadruples of numbers).

Thus our task is to describe abstract Hilbert space on the basis of its “intrinsic” properties, that is, properties that can be formulated without reference to

^{T6}In the preceding comparison, the original’s second occurrence of the finiteness condition is printed as $\sum x_\mu^2$; the absolute-value signs required for complex sequences have been restored.

the interpretation of its elements as sequences or functions. We shall solve it by specifying five properties from which, as will be shown, everything else does indeed follow. These five properties concern an “abstract Hilbert space $\overline{\mathfrak{H}}$,” with elements $f, g, \dots$, in which the operations af (a a complex constant), $f + g$, $Q(f, g)$ are defined ($af, f + g$ are again elements of $\overline{\mathfrak{H}}$, whereas $Q(f, g)$ is a complex number); these operations (especially $Q(f, g)$) are now to be regarded independently of their definition in § III: we assume of them only the five axioms to be stated.

Of course, they are satisfied if we replace $\overline{\mathfrak{H}}$ by its “discontinuous realization” $\mathfrak{H}_0$ (the space of all sequences $x_1, x_2, \dots$ with finite $\sum_{n=1}^{\infty} |x_n|^2$, $Q(x, y) = \sum_{n=1}^{\infty} x_n \overline{y_n}$), or by any of the “continuous realizations” $\mathfrak{H}$ (the space of all functions φ defined on Ω with finite $\int_{\Omega} |\varphi|^2 dv$ ^{T7}, $Q(\varphi, \psi) = \int_{\Omega} \varphi \overline{\psi} dv$); cf. Appendix 2. Thus, when one wishes to make the concrete meaning of our conditions clear, one may replace the abstract $\overline{\mathfrak{H}}$ by one of the special realizations just mentioned.

We now give the five characteristic properties A–E of the abstract Hilbert space $\overline{\mathfrak{H}}$. Immediately after each we append its simplest consequences; at the end of the whole comes the proof that they do in fact determine $\overline{\mathfrak{H}}$ completely, by showing that $\overline{\mathfrak{H}}$ can be mapped one-to-one onto $\mathfrak{H}_0$, with invariance of the operations $af, f + g, Q(f, g)$ (the idea of the proof was already sketched for function spaces $\mathfrak{H}$ in § IV). That the spaces $\mathfrak{H}_0, \mathfrak{H}$ really are to be regarded as instances of $\overline{\mathfrak{H}}$ (that is, possess properties A–E) is discussed in Appendix 2.

Our five axioms A–E for the abstract (complex) Hilbert space $\overline{\mathfrak{H}}$ read as follows:

A. $\overline{\mathfrak{H}}$ is a linear space.

That is: in $\overline{\mathfrak{H}}$ there are an addition $f + g$ and a multiplication af (f, g in $\overline{\mathfrak{H}}$, a a complex number, $f + g, af$ in $\overline{\mathfrak{H}}$); these satisfy the familiar laws of the analogous operations on vectors (namely, existence of 0, commutativity and associativity of addition, distributivity and associativity of multiplication).

A fundamental concept that can be formulated on the basis of A alone is that of **linear independence**: elements $f_1, f_2, \dots, f_k$ of $\overline{\mathfrak{H}}$ are called linearly independent if

$$a_1 f_1 + a_2 f_2 + \dots + a_k f_k = 0$$

implies $a_1 = a_2 = \dots = a_k = 0$. Another is that of the **linear manifold spanned** by a subset $\mathfrak{M}$ of $\overline{\mathfrak{H}}$: it is the set of all $a_1 f_1 + a_2 f_2 + \dots + a_k f_k$, where $a_1, a_2, \dots, a_k$ are arbitrary complex numbers and $f_1, f_2, \dots, f_k$ are arbitrary elements of $\mathfrak{M}$.^{T8}

^{T7}The original uses ψ in this finiteness condition, although the functions just specified are denoted by φ . The symbol has been made consistent.

^{T8}The original ends this definition with “elements of $\overline{\mathfrak{H}}$.” Since it defines the manifold spanned

B. $\overline{\mathfrak{H}}$ is a metric space; its metric arises from a Hermitian-symmetric bilinear form $Q(f, g)$.

That is: there is a function $Q(f, g)$ (defined for all f, g of $\overline{\mathfrak{H}}$ and taking complex numbers as values) with the following properties:

1. $Q(af, g) = aQ(f, g)$ (a a complex constant).
2. $Q(f_1 + f_2, g) = Q(f_1, g) + Q(f_2, g)$.
3. $Q(f, g) = \overline{Q(g, f)}$.

(Because of 3, properties 1 and 2 imply

$$\begin{aligned} 1'. \quad Q(f, ag) &= \bar{a}Q(f, g), \\ 2'. \quad Q(f, g_1 + g_2) &= Q(f, g_1) + Q(f, g_2). \end{aligned}$$

Properties 1, 2, 1', 2' express the Hermitian bilinearity of Q , and 3 its symmetry.) By 3, $Q(f, f)$ is always real; we further require (again writing $Q(f)$ for $Q(f, f)$):

4. $Q(f) \geq 0$, with equality only when $f = 0$.

From 1–4 it follows without difficulty that we always have^{17,T9}

$$|Q(f, g)| \leq \sqrt{Q(f) \cdot Q(g)},$$

by the specified subset $\mathfrak{M}$, $\mathfrak{M}$ is required here.

¹⁷One readily calculates (a, b real constants):

$$\begin{aligned} Q(af + bg) &= a^2Q(f) + ab[Q(f, g) + Q(g, f)] + b^2Q(g) \\ &= a^2Q(f) + 2ab\Re Q(f, g) + b^2Q(g). \end{aligned}$$

(We denote the real and imaginary parts of the complex number z by $\Re z, \Im z$.) The left side is always ≥ 0 , and the right side is a quadratic form in a, b ; hence its discriminant is ≤ 0 :

$$\begin{aligned} [\Re Q(f, g)]^2 - Q(f) \cdot Q(g) &\leq 0, \\ |\Re Q(f, g)| &\leq \sqrt{Q(f) \cdot Q(g)}. \end{aligned}$$

If we replace f here by $e^{i\varphi}f$ (φ a real constant), the right side remains unchanged, while the left side becomes

$$|\Re\{e^{i\varphi}Q(f, g)\}| = |\cos \varphi \Re Q(f, g) - \sin \varphi \Im Q(f, g)|.$$

Since its maximum is

$$\sqrt{[\Re Q(f, g)]^2 + [\Im Q(f, g)]^2} = |Q(f, g)|,$$

we do indeed have

$$|Q(f, g)| \leq \sqrt{Q(f) \cdot Q(g)}.$$

and furthermore

$$\begin{aligned}\sqrt{Q(af)} &= |a|\sqrt{Q(f)}, \\ \sqrt{Q(f+g)} &\leq \sqrt{Q(f)} + \sqrt{Q(g)}.\end{aligned}$$

These last two relations justify regarding $\sqrt{Q(f)}$ as the absolute magnitude of f , and $\sqrt{Q(f-g)}$ as the distance between f and g .¹⁸

Thus Q does indeed impose a metric, a concept of distance, on the space $\overline{\mathfrak{H}}$. Expressions such as “continuous,” “bounded,” “arbitrarily close,” and so forth thereby acquire a meaning in $\overline{\mathfrak{H}}$ as well.

C. $\overline{\mathfrak{H}}$ has infinitely many dimensions.

That is: there are arbitrarily large (finite) numbers of linearly independent elements in $\overline{\mathfrak{H}}$.

D. There is a sequence everywhere dense in $\overline{\mathfrak{H}}$.

Furthermore:

$$\begin{aligned}Q(af) &= a\bar{a}Q(f) = |a|^2Q(f), \\ \sqrt{Q(af)} &= |a|\sqrt{Q(f)},\end{aligned}$$

and

$$\begin{aligned}Q(f+g) &= Q(f) + Q(f,g) + Q(g,f) + Q(g) \\ &= Q(f) + 2\Re Q(f,g) + Q(g) \\ &\leq Q(f) + 2\sqrt{Q(f) \cdot Q(g)} + Q(g) \\ &= (\sqrt{Q(f)} + \sqrt{Q(g)})^2, \\ \sqrt{Q(f+g)} &\leq \sqrt{Q(f)} + \sqrt{Q(g)}.\end{aligned}$$

¹⁹In the second line of the first calculation in original note 17, the original prints $b^2Q(a)$ in place of $b^2Q(g)$. The latter, consistent with the preceding line, is used here.

¹⁸The last relation gives the fundamental postulate for any distance:

$$\text{distance}(f, h) \leq \text{distance}(f, g) + \text{distance}(g, h).$$

Our distance $\sqrt{Q(f-g)}$ is, in $\mathfrak{H}_0$ and $\mathfrak{H}$ (where we know Q),

$$\sqrt{\sum_{n=1}^{\infty} |x_n - y_n|^2} \quad \text{and} \quad \sqrt{\int_{\Omega} |f - g|^2 dv},$$

respectively, and thus the natural generalization of distance in ordinary Euclidean spaces.

That is: there is a sequence $f_1, f_2, \dots$ such that arbitrarily close to every element f of $\bar{\mathfrak{H}}$ there are terms of this sequence.

E. The Cauchy convergence condition holds in $\bar{\mathfrak{H}}$.

That is: every sequence $f_1, f_2, \dots$ in $\bar{\mathfrak{H}}$ satisfying the Cauchy convergence condition (for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that $N \leq m \leq n$ implies $\sqrt{Q(f_m - f_n)} \leq \varepsilon$) is convergent (there is an f in $\bar{\mathfrak{H}}$ such that for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ for which $N \leq m$ implies $\sqrt{Q(f_m - f)} \leq \varepsilon$).

Postponing to [Appendix 2](#) the question whether $\mathfrak{H}_0, \mathfrak{H}$ satisfy these conditions, we proceed to draw the most immediate consequences of A–E. They culminate in the repeatedly mentioned possibility of mapping $\bar{\mathfrak{H}}$ onto $\mathfrak{H}_0$; in the course of deriving them, however, we gain an insight into the structure of the orthogonal systems in $\bar{\mathfrak{H}}$ that is fundamental to our later considerations.

For completeness we shall carry out these considerations, based on familiar mathematical arguments, without gaps in the next section; they may be skipped, provided the principal results are noted.

VI. First we must give a few simple definitions of a general nature.

f is an **accumulation point** of a subset $\mathfrak{M}$ of $\bar{\mathfrak{H}}$ if there are points of $\mathfrak{M}$ arbitrarily close to f (cf. the remarks following B).^{T10} A subset $\mathfrak{M}$ of $\bar{\mathfrak{H}}$ is **closed** if all its accumulation points belong to it; it is **everywhere dense** if every f of $\bar{\mathfrak{H}}$ is an accumulation point of it; it is **dense in** $\mathfrak{N}$ if it is a subset of $\mathfrak{N}$ and every point of $\mathfrak{N}$ is an accumulation point of it.

$\mathfrak{M}$ is a **linear manifold** if it spans itself as a linear manifold; that is, together with $f_1, f_2, \dots, f_k$ it contains $a_1 f_1 + a_2 f_2 + \dots + a_k f_k$.

Two elements f, g are **orthogonal** if $Q(f, g) = 0$. $\mathfrak{M}$ is a **normalized orthogonal system** if, for all f, g of $\mathfrak{M}$,

$$Q(f, g) = \begin{cases} 1, & \text{for } f = g, \\ 0, & \text{for } f \neq g. \end{cases}$$

$\mathfrak{M}$ is a **complete normalized orthogonal system** if no further f can be added while preserving its normalized orthogonal character; or, what is plainly the same, if no f (apart from 0) is orthogonal to all g of $\mathfrak{M}$.

We can now proceed to derive the announced theorems.

Theorem 1. Every normalized orthogonal system $\mathfrak{M}$ is finite or a sequence; every complete normalized orthogonal system is a sequence.

^{T10}The original prints “points of $\bar{\mathfrak{H}}$ ” at this place; the subset $\mathfrak{M}$ is intended in this definition.

Proof. Let $\mathfrak{M}$ be a normalized orthogonal system, and $f_1, f_2, \dots$ the sequence everywhere dense in $\overline{\mathfrak{H}}$. For any two distinct elements f, g of $\mathfrak{M}$,

$$Q(f - g) = Q(f) - 2\Re Q(f, g) + Q(g) = 1 - 0 + 1 = 2,$$

so their distance is $\sqrt{2}$. For each f of $\mathfrak{M}$ there is an element of the sequence $f_1, f_2, \dots$ whose distance from it is less than $\frac{1}{2}\sqrt{2}$; by the above, two different f of $\mathfrak{M}$ correspond to different elements of the sequence. Thus $\mathfrak{M}$ has at most as many elements as the sequence, as was to be proved. If, on the other hand, $\mathfrak{M}$ is complete, we must show that $\mathfrak{M}$ cannot be finite. That is: for finitely many $\varphi_1, \dots, \varphi_k$ there is an orthogonal $f \neq 0$. The linear manifold spanned by $\varphi_1, \dots, \varphi_k$ does not contain $k + 1$ linearly independent elements; hence $\overline{\mathfrak{H}}$ has elements f lying outside it. Consequently,

$$f - c_1\varphi_1 - \dots - c_k\varphi_k$$

is never 0, but for $c_n = Q(f, \varphi_n)$ ($n = 1, 2, \dots, k$) it is orthogonal to all $\varphi_1, \dots, \varphi_k$, as was to be proved.

Theorem 2. Let $\varphi_1, \varphi_2, \dots$ be a normalized orthogonal system. Then every series

$$\sum_{n=1}^{\infty} Q(f, \varphi_n) \overline{Q(g, \varphi_n)}$$

is absolutely convergent; in particular,

$$\sum_{n=1}^{\infty} |Q(f, \varphi_n)|^2 \leq Q(f).$$

Proof. For $c_n = Q(f, \varphi_n)$ ($n = 1, 2, \dots$) we have, as is well known:^{T11}

$$\begin{aligned} Q\left(\sum_{n=1}^N c_n \varphi_n - f\right) &= Q(f) - \sum_{n=1}^N 2\Re Q(f, c_n \varphi_n) \\ &\quad + \sum_{m,n=1}^N Q(c_m \varphi_m, c_n \varphi_n) \\ &= Q(f) - \sum_{n=1}^N 2\Re [\overline{c_n} Q(f, \varphi_n)] \end{aligned}$$

^{T11}In the second line of this expansion the original prints $\varphi(f, \varphi_n)$ where the bilinear form $Q(f, \varphi_n)$ is required.

$$\begin{aligned}
& + \sum_{m,n=1}^N c_m \overline{c_n} Q(\varphi_m, \varphi_n) \\
& = Q(f) - 2 \sum_{n=1}^N |c_n|^2 + \sum_{n=1}^N |c_n|^2 \\
& = Q(f) - \sum_{n=1}^N |c_n|^2.
\end{aligned}$$

Since the left side is always ≥ 0 , it follows that

$$\sum_{n=1}^N |c_n|^2 \leq Q(f),$$

and thus the convergence of the series

$$\sum_{n=1}^{\infty} |c_n|^2 = \sum_{n=1}^{\infty} |Q(f, \varphi_n)|^2$$

is proved, as is the fact that its sum is $\leq Q(f)$ (the second assertion).

The first follows from the evident relations

$$\begin{aligned}
\Re[Q(f, \varphi_n) \overline{Q(g, \varphi_n)}] &= \left| Q\left(\frac{f+g}{2}, \varphi_n\right) \right|^2 - \left| Q\left(\frac{f-g}{2}, \varphi_n\right) \right|^2, \\
\Im[Q(f, \varphi_n) \overline{Q(g, \varphi_n)}] &= \left| Q\left(\frac{f+ig}{2}, \varphi_n\right) \right|^2 - \left| Q\left(\frac{f-ig}{2}, \varphi_n\right) \right|^2,
\end{aligned}$$

since the sums of the right sides converge absolutely.

Theorem 3. Let $\varphi_1, \varphi_2, \dots$ be a normalized orthogonal system. The series

$$\sum_{n=1}^{\infty} c_n \varphi_n$$

converges¹⁹ if and only if $\sum_{n=1}^{\infty} |c_n|^2$ is finite.

¹⁹Observe that this is convergence in $\overline{\mathfrak{H}}$! Thus, if we consider, for example, a continuous realization $\mathfrak{H}$ of $\overline{\mathfrak{H}}$ (the space of all functions f defined in Ω with finite $\int_{\Omega} |f|^2 dv$), this is not pointwise convergence there, but convergence in the mean!

Proof. By E, convergence of $\sum_{n=1}^{\infty} c_n \varphi_n$ means that for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that $N \leq m \leq n$ implies^{T12}

$$\sqrt{Q\left(\sum_{p=1}^n c_p \varphi_p - \sum_{p=1}^m c_p \varphi_p\right)} \leq \varepsilon.$$

But

$$\begin{aligned} Q\left(\sum_{p=1}^n c_p \varphi_p - \sum_{p=1}^m c_p \varphi_p\right) &= Q\left(\sum_{p=m+1}^n c_p \varphi_p\right) \\ &= \sum_{p,q=m+1}^n c_p \overline{c_q} Q(\varphi_p, \varphi_q) \\ &= \sum_{p=m+1}^n |c_p|^2 \\ &= \sum_{p=1}^n |c_p|^2 - \sum_{p=1}^m |c_p|^2, \end{aligned}$$

so we have precisely the convergence condition for $\sum_{p=1}^{\infty} |c_p|^2$.

Corollary. For this f we have $Q(f, \varphi_p) = c_p$.

Proof. In any case, for $N \geq p$,

$$Q\left(\sum_{n=1}^N c_n \varphi_n, \varphi_p\right) = \sum_{n=1}^N c_n Q(\varphi_n, \varphi_p) = c_p.$$

Now Q , by virtue of

$$|Q(f', g')| \leq \sqrt{Q(f')Q(g')}$$

and its bilinearity, is continuous in f', g' . We may therefore let $N \rightarrow \infty$, obtaining $Q(f, \varphi_p) = c_p$.

Theorem 4. Let $\varphi_1, \varphi_2, \dots$ be a normalized orthogonal system. For every f , the series

$$f' = \sum_{n=1}^{\infty} c_n \varphi_n, \quad c_n = Q(f, \varphi_n),$$

converges; and $f - f'$ is orthogonal to all $\varphi_1, \varphi_2, \dots$

^{T12}The lower index of the second partial sum in the original is printed as $n = 1$, although its summand uses p . It has been written $p = 1$ here.

Proof. This follows directly from [Theorems 2 and 3](#).

Theorem 5. Each of the following three conditions is both necessary and sufficient for a normalized orthogonal system $\varphi_1, \varphi_2, \dots$ to be complete:

α . The linear manifold spanned by $\varphi_1, \varphi_2, \dots$ is everywhere dense.

β . For all f ,

$$f = \sum_{n=1}^{\infty} c_n \varphi_n, \quad c_n = Q(f, \varphi_n).$$

γ . For all f, g ,

$$Q(f, g) = \sum_{n=1}^{\infty} Q(f, \varphi_n) \overline{Q(g, \varphi_n)}.$$

Proof. First, completeness implies β , since the $f - f'$ of [Theorem 4](#) must vanish. Second, β implies α , since f is an accumulation point of the $\sum_{p=1}^N c_p \varphi_p$, all of which belong to the linear manifold spanned by $\varphi_1, \varphi_2, \dots$. Third, β implies γ , since

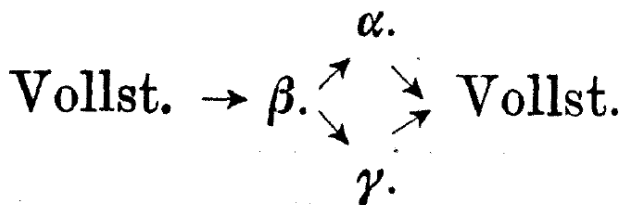
$$Q\left(\sum_{n=1}^N c_n \varphi_n - f\right) = Q(f) - \sum_{n=1}^N |c_n|^2,$$

and as $N \rightarrow \infty$ the left side tends to 0 (since, as remarked above, $Q(f)$ is continuous), so that

$$\sum_{n=1}^{\infty} |c_n|^2 = Q(f), \quad \sum_{n=1}^{\infty} |Q(f, \varphi_n)|^2 = Q(f).$$

If we substitute successively $(f + g)/2$, $(f - g)/2$ and $(f + ig)/2$, $(f - ig)/2$ for f , and subtract in each pair, the assertion follows as in [Theorem 2](#). Fourth, α implies completeness: for if f is orthogonal to all $\varphi_1, \varphi_2, \dots$, it is also orthogonal to the whole linear manifold spanned by them, and, since this is everywhere dense, to itself as well; that is, $Q(f) = 0$, $f = 0$. Fifth, γ implies completeness: if f is orthogonal to all $\varphi_1, \varphi_2, \dots$, then γ applied to f, f gives $Q(f) = 0$, $f = 0$.

We thus have the following logical scheme:^{T13}



^{T13}“Vollst.” in the reproduced original diagram abbreviates *Vollständigkeit*, “completeness.”

Hence these four statements are equivalent.

Theorem 6. For every sequence $f_1, f_2, \dots$ there is a normalized orthogonal system $\varphi_1, \varphi_2, \dots$ (both sequences may terminate after finitely many terms) which spans the same linear manifold.

Proof. Let g_1 be the first nonzero element of $f_1, f_2, \dots$, g_2 the first element of $f_1, f_2, \dots$ not of the form $a_1 g_1$, g_3 the first element of $f_1, f_2, \dots$ not of the form $a_1 g_1 + a_2 g_2$, and so on. The $g_1, g_2, \dots$ are plainly linearly independent and span the same linear manifold as $f_1, f_2, \dots$. By the familiar procedure of “orthogonalization,” one converts $g_1, g_2, \dots$ into a normalized orthogonal system.²⁰

Corollary. There exist complete normalized orthogonal systems.

Proof. Choose $f_1, f_2, \dots$ everywhere dense by D, and replace it by a normalized orthogonal system according to Theorem 6; this is complete by Theorem 5 α .

If we now take any complete normalized orthogonal system $\varphi_1, \varphi_2, \dots$, and assign to every f of $\bar{\mathfrak{H}}$ the sequence $c_1, c_2, \dots$ ($c_n = Q(f, \varphi_n)$, $n = 1, 2, \dots$), then $c_1, c_2, \dots$ belongs to $\mathfrak{H}_0$ by Theorem 2, and in turn determines f by Theorem 5 β . This one-to-one mapping of $\bar{\mathfrak{H}}$ covers the whole of $\mathfrak{H}_0$ by Theorem 3. The invariance under it of the operations $+$ and multiplication by a is trivial; the invariance of Q follows from Theorem 5 γ .

It is thus indeed proved that every space $\bar{\mathfrak{H}}$ with properties A–E must agree in all its properties with the ordinary Hilbert space $\mathfrak{H}_0$. (This mapping of $\bar{\mathfrak{H}}$ onto $\mathfrak{H}_0$ is, of course, analogous to that of $\mathfrak{H}$ onto $\mathfrak{H}_0$ sketched in § IV^{T14}.)

²⁰The procedure is as follows:

$$\begin{aligned}\gamma_1 &= g_1, \\ \varphi_1 &= \frac{1}{\sqrt{Q(\gamma_1)}} \gamma_1, \\ \gamma_2 &= g_2 - Q(g_2, \varphi_1) \varphi_1, \\ \varphi_2 &= \frac{1}{\sqrt{Q(\gamma_2)}} \gamma_2, \\ \gamma_3 &= g_3 - Q(g_3, \varphi_1) \varphi_1 - Q(g_3, \varphi_2) \varphi_2, \\ \varphi_3 &= \frac{1}{\sqrt{Q(\gamma_3)}} \gamma_3, \\ &\dots\end{aligned}$$

The $\varphi_1, \varphi_2, \dots$ plainly span the same linear manifold as the $g_1, g_2, \dots$, and are normalized orthogonal.

^{T14}The original refers to § III; the mapping by expansion coefficients is sketched in § IV.

Operator calculus

VII. The considerations of the preceding sections allow us to use the abstract (complex) Hilbert space as the basis of our further investigations. Recourse to its various realizations (the discontinuous one, as well as the various continuous ones; cf. the beginning of § V) will at first (in § VII–§ XI, § XIII) prove wholly unnecessary: all our general developments are independent of them. Only afterwards, in the physical applications, shall we have to concern ourselves with them.

The first thing we must develop in $\bar{\mathfrak{H}}$ is the formal operator calculus: we know (cf. the end of § II) its fundamental importance for quantum mechanics.

A function in $\bar{\mathfrak{H}}$ which is defined at certain points of $\bar{\mathfrak{H}}$ (possibly at all of them), and whose values are points of $\bar{\mathfrak{H}}$, is an *operator*. An operator T is *linear* if, first, it is defined on a linear manifold (which need by no means be closed), and if, secondly, we always have

$$T(a_1 f_1 + a_2 f_2 + \cdots + a_k f_k) = a_1 T f_1 + a_2 T f_2 + \cdots + a_k T f_k$$

($a_1, a_2, \dots, a_k$ complex constants).

As already noted in § V, B^{T15}, the expressions “continuous” and “bounded in the sphere of radius 1 about the origin” have a meaning for operators; for linear operators they evidently mean the same thing.²¹ Hilbert’s customary term for this is “bounded.”

If we consider the discontinuous realization $\mathfrak{H}_0$, then every bounded linear operator corresponds to an infinite matrix, and the associated bilinear form (in infinitely many variables) belongs to that class of bounded bilinear forms for which Hilbert completely solved the eigenvalue problem (cf. note 12). But the operators of interest in quantum mechanics do not belong to this type (cf. the same note).

^{T15}The original refers to § IV; the metric terminology recalled here is introduced under axiom B in § V.

²¹The second condition says: $Q(f) \leq 1$ implies $Q(Tf) \leq C$ (C a constant). Since $T(af) = aTf$, this entails

$$\begin{aligned} Q(Tf) &\leq CQ(f), \\ Q(Tf - Tg) &\leq CQ(f - g), \end{aligned}$$

that is, continuity. Conversely, for continuous T there is an $\varepsilon > 0$ such that $Q(f) \leq \varepsilon$ implies $Q(Tf) \leq 1$; from this it follows easily that, for $Q(f) \leq 1$, we have $Q(Tf) \leq 1/\varepsilon$.

In the continuous realizations $\mathfrak{H}$, certain bounded operators T do possess the integral-kernel representation analogous to matrices,

$$Tf(P) = \int_{\Omega} \Phi(P, Q)f(Q) dv_Q$$

(integration is with respect to Q); yet even the very simplest operators lack it, for example the “identity operator” (which takes f into f). We shall in fact conduct our considerations without making use of matrix or integral-kernel representations.

It is necessary to recall the definitions of a few simple operators and rules of operator calculation. We have

$$0f = 0, \quad 1f = f$$

(thus 0 occurs in three forms: as a number, as a point of the space $\overline{\mathfrak{H}}$, and as the zero operator; there should be no danger of confusion). Further,

$$\begin{aligned} (aT)f &= aTf & (a \text{ a complex constant}), \\ (R + T)f &= Rf + Tf, \\ (RT)f &= R(Tf). \end{aligned}$$

It is generally known that this operator addition is commutative and associative, that both multiplications are distributive and associative, but in general not commutative, and that 0, 1 play the roles of zero and unity.

Finally, we observe: if T is a linear operator, and for an f of $\overline{\mathfrak{H}}$ there is an f^* such that we always have (whenever Tg has a meaning)

$$Q(f^*, g) = Q(f, Tg),$$

then we denote this f^* by T^*f . It is uniquely determined if the g for which Tg has a meaning are everywhere dense: for if f_1^*, f_2^* had this property, we should have

$$Q(f_1^*, g) = Q(f_2^*, g), \quad Q(f_1^* - f_2^*, g) = 0$$

for an everywhere dense set of g , hence for all g , hence

$$Q(f_1^* - f_2^*) = 0, \quad f_1^* = f_2^*.$$

From now on we impose this requirement (of having a meaning for an everywhere dense set of g) on all operators to be considered; it will indeed be satisfied by all those of interest to us.²²

²²Its significance may be clarified by the following two examples.

As one sees at once, T^* is likewise a linear operator (we assume of it, too, that it has a meaning for an everywhere dense set of f). Its defining equation is

$$Q(f, Tg) = Q(T^*f, g).$$

On the basis of the properties of Q (§ V, B)^{T16}, one easily proves (with the necessary assumptions concerning meaningfulness):

$$\begin{aligned}(aT)^* &= \bar{a} T^*, \\ (R + T)^* &= R^* + T^*, \\ (RT)^* &= T^* R^*, \\ T^{**} &= T.\end{aligned}$$

A linear operator T is called *symmetric* (this is really complex Hermitian symmetry) if $T = T^*$. With the aid of the above equations one verifies immediately the following statements.

If T is symmetric, then aT is symmetric if and only if a is real. If R, T are symmetric, so is $R + T$; RT , on the other hand, is symmetric if and only if R and T commute ($RT = TR$).

$0, 1$ are both symmetric.

VIII. After these general considerations we turn to an important special class of operators: the *elementary operators*.²³ They are defined as follows.

Definition. A symmetric linear operator E (which has a meaning everywhere) is an *elementary operator* (abbreviated E. op.) if $E^2 = E$. One sees immediately that $0, 1$ are E. op. Further, together with E , $1 - E$ is always an E. op., since

$$(1 - E)^2 = 1 - 2E + E^2 = 1 - 2E + E = 1 - E.$$

Let Ω be the interval $0, 1$, and T differentiation. Tf does not always have a meaning (not all $f(x)$ are differentiable), but these f are everywhere dense: every function can already be approximated arbitrarily closely by polynomials (which are differentiable functions!), in the mean.

Let Ω be the interval $-\infty, +\infty$, and T multiplication by x . Tf does not always have a meaning ($\int_{-\infty}^{\infty} |f(x)|^2 dx$ may be finite without $\int_{-\infty}^{\infty} x^2 |f(x)|^2 dx$ being so), but these f are everywhere dense: they include all f which vanish identically outside an interval (arbitrarily large, but finite).

^{T16}The original prints “§ IV, B”; the intended reference is axiom B in § V.

²³The word is modeled on Hilbert’s designation for the analogous bilinear forms: “elementary form” [*Einzelform*].

We shall now derive a number of theorems about E. op. Even though we shall not need all of them below, they are suited to putting the nature of this fundamental concept in the proper light. We are chiefly concerned with the following two facts.

An E. op. takes certain f into f and others into 0; the former constitute its *interior*, the latter its *exterior*. Every f of $\overline{\mathfrak{H}}$ can be decomposed in exactly one way as $g + h$, with g in the interior and h in the exterior of E : Ef then arises from f by omitting the “exterior component” h .

There is a certain ordering by magnitude among E. op.: namely, according to the extent of their interiors (or, inversely, according to the extent of their exteriors).

With these results in mind, the discussions which now follow may also be skipped.

(The definition just given of the interior and exterior of E will be used in them.) Evidently,

$$\begin{aligned} Q(Ef, Eg) &= Q(f, E^*Eg) = Q(f, E^2g) = Q(f, Eg), \\ Q(Ef, Eg) &= Q(E^*Ef, g) = Q(E^2f, g) = Q(Ef, g), \end{aligned}$$

and in particular

$$Q(f, Ef) = Q(Ef).$$

It follows that:

Theorem 1. We always have $0 \leq Q(Ef) \leq Q(f)$.

Proof. The first inequality is trivial; the second holds because $1 - E$ is also an E. op.:

$$\begin{aligned} Q(f, Ef) &= Q(f) - Q(f, (1 - E)f), \\ Q(Ef) &= Q(f) - Q((1 - E)f) \leq Q(f). \end{aligned}$$

Thus every E. op. is continuous; indeed all E. op. are even uniformly continuous.

Theorem 2. If E, F are E. op., then EF is an E. op. if and only if they commute ($EF = FE$). $E + F$ is one if and only if $EF = 0$ (or, equivalently, $FE = 0$). $F - E$ is one if and only if $EF = E$ (or, equivalently, $FE = E$).

Proof. For EF , the condition $EF = FE$ is already necessary and sufficient for symmetry; but since it entails

$$(EF)^2 = EF EF = E E F F = E^2 F^2 = EF,$$

this case is settled.

For $E + F, F - E$, symmetry is always present; only the relation $T^2 = T$ remains to be examined.

Take $E + F$ first. If $E + F$ is an E. op., then

$$\begin{aligned} Q(f, Ef) + Q(f, Ff) &= Q(f, (E + F)f), \\ Q(Ef) + Q(Ff) &= Q((E + F)f) \leq Q(f); \end{aligned}$$

thus, for $Ef = f$,

$$Q(Ff) \leq 0, \quad Ff = 0.$$

Now $EEg = Eg$ always holds, hence $FEg = 0$; thus $FE = 0$. Conversely, if $FE = 0$, then also $EF = 0$ (because FE is an E. op., so that F, E commute), hence

$$\begin{aligned} (E + F)^2 &= E^2 + EF + FE + F^2 \\ &= E + 0 + 0 + F = E + F, \end{aligned}$$

and consequently $E + F$ is an E. op.

That is, $FE = 0$ is necessary and sufficient; and since the roles of E, F are entirely symmetric, the same holds for $EF = 0$.

Now take $F - E$. It is an E. op. if and only if $1 - (F - E) = E + (1 - F)$ is one; but for this (since $E, 1 - F$ are E. op.) the necessary and sufficient condition is

$$E(1 - F) = 0, \quad E = EF,$$

or, equivalently,

$$(1 - F)E = 0, \quad E = FE.$$

We introduce the following terminology: if $E + F$ is an E. op., E, F are *disjoint*; if $F - E$ is one, then $E \leq F$.^{T17} (Theorem 2 thus supplies simple necessary and sufficient conditions for both.) One sees at once that $\leq$ is an ordering relation for E. op.; that is, $E \leq E$; $E \leq F, F \leq E$ imply $E = F$; $E \leq F, F \leq G$ imply $E \leq G$. 0 precedes, and 1 follows, all E . $E \leq F$ is equivalent to $E, 1 - F$ being disjoint, or to $1 - F \leq 1 - E$.

We shall now examine the interior and exterior of E more closely. That both are closed linear manifolds follows from the linearity and continuity of E . Further, one sees at once that they simply exchange places on passing from E to $1 - E$.

^{T17}The original prints $E - F$ in this sentence. It must read $F - E$, as follows from Theorem 2 immediately above and from the subsequent characterization of $E \leq F$.

The interior of 1 and the exterior of 0 contain all f of $\overline{\mathfrak{H}}$; the exterior of 1 or the interior of 0, however, contains only 0.

Theorem 3. The f lying in the interior of E are precisely the Eg (g arbitrary); those lying in the exterior of E are precisely the $(1 - E)g$.

Proof. The second assertion follows from the first by replacing E by $1 - E$. We prove the first as follows. If f lies in the interior of E , then it is Ef , hence has the desired form; if $f = Eg$, then

$$Ef = EEg = Eg = f,$$

that is, f lies in the interior of E .

Theorem 4. f lies in the interior or exterior of E if and only if $Q(Ef) = Q(f)$ or $Q(Ef) = 0$, respectively.

Proof. The second assertion is trivial ($Q(Ef) = 0$ means $Ef = 0$); the first follows from it by interchanging E and $1 - E$.

Theorem 5. Every f can be decomposed in one and only one way into two summands $g + h$, with g, h in the interior and exterior of E , respectively; this decomposition is $Ef + (1 - E)f$.

Proof. That $Ef + (1 - E)f$ does what is required is clear; it is unique because the interior and exterior of E are linear manifolds with no common points other than 0.

Theorem 6. The interior or exterior of E consists of precisely those elements of $\overline{\mathfrak{H}}$ which are orthogonal to every element of the exterior or interior of E , respectively.

Proof. If f, g lie in the interior and exterior of E , respectively, then they are orthogonal:

$$Q(f, g) = Q(Ef, g) = Q(f, Eg) = Q(f, 0) = 0.$$

Our assertion follows evidently from this and [Theorem 5](#).

Theorem 7. E, F are disjoint if and only if every f in the interior of E lies in the exterior of F .

Proof. The condition is necessary and sufficient, for it says: for all f in the interior of E , $Ff = 0$; that is, always $FEg = 0$, that is, $FE = 0$, hence E and F are disjoint.

Theorem 8. $E \leq F$ is equivalent to the interior of E being part of the interior of F , or to the exterior of F being part of the exterior of E .

Proof. $E \leq F$ means that E and $1 - F$ are disjoint; apply [Theorem 7](#) to $E, 1 - F$, and to $1 - F, E$.

Theorem 9. $E \leq F$ is equivalent to $Q(Ef) \leq Q(Ff)$ for all f .

Proof. This is sufficient, for it implies: if f lies in the exterior of F , then

$$Q(Ef) \leq Q(Ff) = 0, \quad Ef = 0,$$

that is, f lies also in the exterior of E ; thus $E \leq F$. And it is necessary, since $E \leq F$ implies

$$Q(Ef) = Q(EFf) \leq Q(Ff).$$

We are now in a position to formulate the eigenvalue problem for general symmetric operators with the aid of E. op.; and, in contrast to § III, in such a way that the continuous spectrum also receives its due.

The eigenvalue problem

IX. As already mentioned, the most immediate formulation of the eigenvalue problem does not satisfy us. It reads as follows.

Let T be a symmetric linear operator; find all (real) numbers l and all $f \neq 0$ of $\bar{\mathfrak{H}}$ for which

$$Tf = lf.$$

(The l are the eigenvalues, the f the eigenfunctions.)

This formulation has two essential defects.

First, the eigenfunctions f are not uniquely determined objects, even when they are normalized by $Q(f) = 1$: a factor of absolute value 1 remains arbitrary; indeed, in the case of multiple eigenvalues (that is, an l with several linearly independent associated eigenfunctions), even an arbitrary orthogonal transformation of the eigenfunctions among themselves remains possible (§ I, ϑ).

Secondly, this approach fails for the continuous spectrum. If we return to the realizations $\mathfrak{H}_0, \mathfrak{H}$, we find: eigensequences $x_1, x_2, \dots$ or eigenfunctions $f(P)$ may indeed exist, but they do not belong to $\mathfrak{H}_0$ or $\mathfrak{H}$, respectively; that is,

$$\sum_{n=1}^{\infty} |x_n|^2 \quad \text{or} \quad \int_{\Omega} |f(P)|^2 dv$$

is infinite.²⁴ We must therefore seek another formulation.

²⁴This suggests the objection that, in delimiting $\mathfrak{H}_0$ or $\mathfrak{H}$ (thus $\bar{\mathfrak{H}}$), we have arbitrarily admitted too little: to stay, for example, with $\mathfrak{H}$, we ought also to admit functions with infinite

Let us try to approximate $\bar{\mathfrak{H}}$ by spaces of known constitution. For this purpose we take the discontinuous realization $\mathfrak{H}_0$: the space of all sequences $x_1, x_2, \dots$ with finite $\sum_{n=1}^{\infty} |x_n|^2$. We may regard it as the limiting case of the 1-, 2-, 3-, ...-dimensional (complex) Euclidean spaces, when the number of dimensions increases beyond all bounds.

Thus let $\mathfrak{R}^{(k)}$ be the k -dimensional (complex) Euclidean space: the space of all groups of numbers $x_1, x_2, \dots, x_k$. It is natural to define $Q(x, y)$ in $\mathfrak{R}^{(k)}$ as $\sum_{n=1}^k x_n \bar{y}_n$ (that is, as the inner product of the vectors x and y); then every linear operator A in $\mathfrak{R}^{(k)}$ is described by a matrix $\{a_{\mu\nu}\}$, with A taking the vector x into the vector y according to

$$y_\nu = \sum_{\mu=1}^k a_{\mu\nu} x_\mu \quad (\nu = 1, 2, \dots, k)$$

$$\int_{\Omega} |f(P)|^2 dv.$$

In our opinion, however, this objection is not valid: we shall show that in quantum mechanics it is not the eigenfunctions that matter at all, but quite other characteristics of the continuous spectrum, for whose description our $\bar{\mathfrak{H}}$ supplies precisely the most suitable framework. Moreover, admitting all $f(P)$ would still not be enough; two examples will now make this clear: in one of them the stated enlargement of the function space helps, in the other it does not.

First: let Ω be the numerical interval $-\infty, \infty$, and T the symmetric operator well known from quantum mechanics,

$$p = \frac{h}{2\pi i} \frac{d}{dx} \cdots$$

As one sees at once, the spectrum is the entire interval $-\infty, \infty$, and the eigenvalue l has the eigenfunction $e^{2\pi i(l/h)x}$. Since

$$\int_{-\infty}^{\infty} \left| e^{2\pi i(l/h)x} \right|^2 dx = \int_{-\infty}^{\infty} dx = \infty,$$

it does not belong to $\mathfrak{H}$.

Secondly: let Ω be as before, and T the likewise very important symmetric operator

$$q = x \cdots$$

It is evident (and will be rigorously justified in §X) that again every l from $-\infty$ to ∞ is an eigenvalue, only the eigenfunction would have to vanish for all $x \neq l$. Since such a function acts like 0 in every integral, we must regard it as 0; that is, there are no eigenfunctions. (One can, of course, regard Dirac's $\delta(x - l)$ —cf. §I, ζ —as an eigenfunction, but it is “improper.”)

Thus one sees that it is not straightforward to delimit a mathematically unobjectionable domain of functions in which the continuous spectrum could be treated directly.

and symmetry is equivalent to

$$a_{\mu\nu} = \overline{a_{\nu\mu}}.$$

As is well known, the Hermitian form associated with A ,

$$A(x | y) = Q(Ax | y) = Q(x | Ay) = \sum_{\mu, \nu=1}^k a_{\mu\nu} x_{\mu} \overline{y_{\nu}},$$

can always be brought into “principal-axis form”:

$$A(x | y) = \sum_{p=1}^k l_p (\alpha_{p1}x_1 + \cdots + \alpha_{pk}x_k) \overline{(\alpha_{p1}y_1 + \cdots + \alpha_{pk}y_k)},$$

where the matrix $\{\alpha_{\mu\nu}\}$ is orthogonal (unitary), that is, it takes the “unit form”

$$Q(x | y) = \sum_{n=1}^k x_n \overline{y_n}$$

into itself:

$$\sum_{n=1}^k x_n \overline{y_n} = \sum_{p=1}^k (\alpha_{p1}x_1 + \cdots + \alpha_{pk}x_k) \cdot \overline{(\alpha_{p1}y_1 + \cdots + \alpha_{pk}y_k)}.$$

The “eigenvalues” $l_1, l_2, \dots, l_k$ are, as is well known, uniquely determined apart from their order, but not the “eigenvectors” $\alpha_{1\nu}, \alpha_{2\nu}, \dots, \alpha_{k\nu}$ ($\nu = 1, 2, \dots, k$): in each of them at least a factor of absolute value 1 (the “phase”) remains arbitrary, and when several eigenvalues l_p coincide (“degeneracy”), even an orthogonal (unitary) transformation of the associated eigenvectors among themselves remains possible.

Nevertheless, it is easy to obtain a unique normal form for $A(x | y)$.

Let the distinct values among $l_1, l_2, \dots, l_k$ be denoted by $L_1, L_2, \dots, L_p$ ($p \leq k$), arranged in order of magnitude.^{T18} Further, put

$$E(l; x | y) = \sum_{l_r \leq l} (\alpha_{r1}x_1 + \cdots + \alpha_{rk}x_k)$$

^{T18}Several index slips in this finite-dimensional discussion have been regularized: the original interchanges p, q as the number of distinct eigenvalues, prints L_p instead of l_p in the condition on the sum defining $E(l; x | y)$, gives an infinite upper limit in the later formula for $E(l; x | y)$ when $l \geq L_p$, and repeats $\mu\nu$ on the conjugate side of the symmetry condition for $e_{\mu\nu}(l)$ instead of reversing the indices. The summation index in the preceding coordinate formula has also been made consistent.

$$\cdot (\overline{\alpha_{r1}y_1 + \cdots + \alpha_{rk}y_k}).$$

As one easily convinces oneself, the L_q and the $E(l; x | y)$ are uniquely determined despite the above indeterminacy of the l_p and $\alpha_{\mu\nu}$. We call the matrix associated with $E(l; x | y)$

$$E(l) = \{e_{\mu\nu}(l)\}.$$

Evidently,

$$A(x | y) = \sum_{q=1}^p L_q \{E(L_q; x | y) - E(L_{q-1}; x | y)\}$$

(L_0 denotes any number $< L_1$). Since $E(l; x | y)$, regarded as a function of l , is constant on the intervals

$$l < L_1, \quad L_{q-1} \leq l < L_q \quad (q = 2, \dots, p), \quad L_p \leq l,$$

we may also write this as

$$A(x | y) = \int_{-\infty}^{\infty} l dE(l; x | y),$$

$$Q(x | Ay) = \int_{-\infty}^{\infty} l dQ(x | E(l)y).$$

(The integral $\int_{-\infty}^{\infty}$ is a so-called Stieltjes integral; see [Appendix 3](#) for further details.)

Here the $E(l; x | y)$ are so-called elementary forms; that is, they are symmetric ($e_{\mu\nu}(l) = \overline{e_{\nu\mu}(l)}$), and the matrices $E(l)$ satisfy the following relation:²⁵

$$E(l)^2 = E(l).$$

We may therefore regard the associated operators

$$E(l)x = y,$$

$$y_\nu = \sum_{\mu=1}^k e_{\mu\nu}(l)x_\mu$$

²⁵That $E(l; x | y)$ is an elementary form is easily verified by direct calculation from its having the form

$$\sum L(x)\overline{L(y)},$$

where the $L(x)$ are mutually orthogonal normalized linear expressions in $x_1, x_2, \dots, x_k$.

as E. op. in the space $\mathfrak{K}^{(k)}$.

Further, one sees immediately that for $l < L_1$ or $l \geq L_p$, respectively,

$$E(l; x | y) = 0 \quad \text{or} \quad E(l; x | y) = \sum_{q=1}^k x_q \overline{y_q},$$

that is, $E(l)$ is zero or unity, respectively; and that $l \leq l'$ implies

$$E(l; x | x) \leq E(l'; x | x).$$

Thus the matrix $E(l)$ is constant everywhere except at the eigenvalues, where it makes discontinuous jumps; the same holds for $E(l; x | x)$, which is in addition monotonically nondecreasing.

But this formulation can easily be transferred directly to the space $\mathfrak{H}_0$, and thus to $\overline{\mathfrak{H}}$ itself. For a symmetric linear operator T we again seek a family of E. op. $E(l)$ such that we always have

$$Q(f, Tg) = \int_{-\infty}^{\infty} l dQ(f, E(l)g).$$

We express this fact briefly by

$$Tg = \int_{-\infty}^{\infty} l d(E(l)g), \quad T = \int_{-\infty}^{\infty} l dE(l).$$

Here, for $l \leq l'$, we always have $Q(E(l)f) \leq Q(E(l')f)$, that is, $E(l) \leq E(l')$ (corresponding to the monotonicity of $E(l; x | x)$). $E(l)$ had at most k jump points, but since k must become infinite for $\mathfrak{K}^{(k)}$ to approximate $\mathfrak{H}_0$, we can demand nothing analogous of $E(l)$. The fact that $E(l)$ starts as zero (small l) and ends as unity (large l), however, admits the natural interpretation: for $l \rightarrow +\infty$ or $l \rightarrow -\infty$, $E(l)f \rightarrow f$ or $E(l)f \rightarrow 0$, respectively.

Finally, the jumps of $E(l)$ always lie to the left of the jump point (that is, it is continuous from the right, since it is a sum $\sum_{l_p \leq l}$, not $\sum_{l_p < l}$); we impose the corresponding requirement on $E(l)$: for $l' \rightarrow l$ with $l' > l$, we must always have $E(l')f \rightarrow E(l)f$.

Let us collect our conditions.

If T is a symmetric linear operator in $\overline{\mathfrak{H}}$, we say that T is represented in *eigenvalue form* if we have found a family of E. op. $E(l)$ (for every real l) satisfying the following conditions:

1. $Q(f, Tg) = \int_{-\infty}^{\infty} l dQ(f, E(l)g)$, or, briefly,

$$Tg = \int_{-\infty}^{\infty} l d\{E(l)g\}, \quad T = \int_{-\infty}^{\infty} l dE(l).$$

2a. $l \leq l'$ implies $E(l) \leq E(l')$.

2b. For $l \rightarrow +\infty$ or $l \rightarrow -\infty$, $E(l)f \rightarrow f$ or $E(l)f \rightarrow 0$, respectively.

2c. For $l' \rightarrow l$, with $l' > l$, $E(l')f \rightarrow E(l)f$.

We also call $E(l)$ the *resolution of the identity* associated with T .²⁶

X. The definition just given of the eigenvalue representation of a symmetric linear operator naturally still requires certain critical considerations.

First, it is not immediately clear whether every symmetric linear operator admits an eigenvalue representation, and whether it determines this representation uniquely. For bounded operators, Hilbert's theorems always give one and only one such representation (cf. [note 12](#)); for unbounded operators, at least uniqueness is established.²⁷

Furthermore, it is desirable to interpret the elementary operators $E(l)$ directly.

For this purpose, consider the simple case in which T has a point spectrum: the eigenvalues $l_1, l_2, \dots$, with the corresponding eigenfunctions $\varphi_1, \varphi_2, \dots$

As is well known, $\varphi_1, \varphi_2, \dots$ form a complete normalized orthogonal system (for example, in a continuous realization, and therefore also in $\overline{\mathfrak{H}}$); hence

$$f = \sum_{n=1}^{\infty} Q(f, \varphi_n) \varphi_n,$$

²⁶It was essentially in this form that Hilbert solved the eigenvalue problem for bounded bilinear forms. We consistently disregard, however, the separation of the point spectrum from the continuous spectrum customary in the literature.

The operator

$$T = \int_{-\infty}^{\infty} l dE(l)$$

does not, of course, always have a meaning; one can show that Tf exists (in $\overline{\mathfrak{H}}$) if and only if the number

$$\int_{-\infty}^{\infty} l^2 dQ(E(l)f)$$

is finite.

²⁷For the case of real symmetric operators, the author has been able to show that one and only one solution always exists; for complex (Hermitian) symmetric operators, the same is to be conjectured, but certain difficulties stand in the way of a proof. See the paper, to appear in the *Mathematische Annalen*, mentioned in [note 12](#).

and, since they are eigenfunctions,

$$Tf = \sum_{n=1}^{\infty} Q(f, \varphi_n) l_n \varphi_n.$$

(Since we are concerned merely with a preliminary orientation, we do not examine the convergence questions more closely.) It follows further that

$$Tf = \int_{-\infty}^{\infty} l \, d \left[\sum_{l_n \leq l} Q(f, \varphi_n) \varphi_n \right].$$

But this is the desired eigenvalue representation, with

$$E(l)f = \sum_{l_n \leq l} Q(f, \varphi_n) \varphi_n;$$

$E(l)$ plainly has [properties 1–2 of § IX](#).

Thus the operator $E(l)$ does the following: expand f in eigenfunctions of T ($\varphi_1, \varphi_2, \dots$) and cut off all terms with eigenvalues $> l$. The interior of $E(l)$ therefore consists of all functions f whose expansion contains only eigenfunctions with eigenvalues $\leq l$; that is, the set of linear aggregates of all eigenfunctions with eigenvalues $\leq l$.

This is also the meaning of the $E(l)$ when there is a continuous spectrum: the interior of $E(l)$ (which, by [§ VIII^{T19}](#), determines $E(l)$) consists of all linear aggregates (which, in the case of a continuous spectrum, may also be integrals) of all eigenfunctions of T with eigenvalues $\leq l$.²⁸

This is of course not exact, but in many cases it gives a guide to finding the $E(l)$ (since one knows or conjectures the eigenfunctions even of the continuous spectrum); the [exact definition at the end of § IX](#) then allows one to verify whether the $E(l)$ so obtained are the correct ones.

^{T19}The original refers to § VI; the uniqueness assertion here follows from [Theorems 5 and 6 of § VIII](#).

²⁸How eigenfunctions not belonging to $\mathfrak{H}$ (that is, those with infinite $\int_{\Omega} |\varphi|^2 \, dv$) can give rise to linear aggregates belonging to $\mathfrak{H}$ is shown by the first example in [note 24](#). The eigenfunctions $e^{2\pi i(l/h)x}$ are not square-integrable, whereas, for example, the linear aggregates

$$\int_{l_1}^{l_2} e^{2\pi i(l/h)x} \, dl = \frac{h}{2\pi i} \left(e^{2\pi i(l_2/h)x} - e^{2\pi i(l_1/h)x} \right) \frac{1}{x}$$

are; here the last factor, $1/x$, brings about square-integrability.

We shall give two examples of this last phenomenon. We consider a continuous realization; let Ω be the interval $-\infty, \infty$. Let T be the operator

$$p = \frac{h}{2\pi i} \frac{d}{dx} \cdots, \quad \text{or, respectively,} \quad q = x \cdots,$$

as in [note 24](#).

In the first case we have the eigenfunctions $e^{2\pi i(l/h)x}$; the eigenfunction expansion of a function f is therefore the Fourier integral:^{T20}

$$f(x) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-2\pi i(l/h)z} f(z) dz \right] e^{2\pi i(l/h)x} dl.$$

$E(l)$ effects the “cutting off of the remainder”:

$$E(l)f(x) = \int_{-\infty}^l \left[\int_{-\infty}^{\infty} e^{-2\pi i(l'/h)z} f(z) dz \right] e^{2\pi i(l'/h)x} dl'.$$

Indeed,

$$\begin{aligned} & \int_{-\infty}^{\infty} l d\{E(l)f(x)\} \\ &= \int_{-\infty}^{\infty} l d \left\{ \int_{-\infty}^l \left[\int_{-\infty}^{\infty} e^{-2\pi i(l'/h)z} f(z) dz \right] \right. \\ & \quad \left. \cdot e^{2\pi i(l'/h)x} dl' \right\} \\ &= \int_{-\infty}^{\infty} l \left[\int_{-\infty}^{\infty} e^{-2\pi i(l/h)z} f(z) dz \right] e^{2\pi i(l/h)x} dl \\ &= \frac{h}{2\pi i} f'(x). \end{aligned}$$

In the second case, eigenfunctions would have to be functions that are nonzero only for $x = l$, which is impossible. Nevertheless, one may then expect the linear aggregates of all eigenfunctions belonging to eigenvalues $\leq l$ simply to be the functions that can be nonzero only for $x \leq l$. Thus $E(l)$ is to be defined by

$$E(l)f(x) = \begin{cases} f(x), & \text{if } x \leq l, \\ 0, & \text{if } x > l. \end{cases}$$

^{T20}With the exponential convention displayed here and ordinary measures dz and dl , Fourier inversion requires a factor $1/h$. The original omits that normalization factor here and in the following projection formula and calculation; those formulas are retained as printed.

Here one verifies quite easily that

$$\int_{-\infty}^{\infty} l d\{E(l)f(x)\} = xf(x).$$

Finally, we consider one more property of the eigenvalue representation. It permits a rather simple representation of the powers of T . Namely,

$$T^n = \int_{-\infty}^{\infty} l^n dE(l).$$

For $n = 0$ this is trivial ($T^0 = 1$), and for $n = 1$ it is the [definition](#); we prove it for $n = 2$. The proof for all other n proceeds in the same way (that is, one passes from n to $n + 1$ just as we pass from 1 to 2).

We have

$$\begin{aligned} Q(f, T^2 g) &= Q(Tf, Tg) = \int_{-\infty}^{\infty} l dQ(Tf, E(l)g) \\ &= \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^{\infty} l' dQ(E(l')f, E(l)g) \right] \\ &= \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^{\infty} l' dQ(f, E(l')E(l)g) \right] \\ &= \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^{\infty} l' dQ(f, E(\min(l, l'))g) \right]. \end{aligned}$$

For $l' > l$, $\min(l', l)$ is constant ($= l$), so it suffices to extend the inner integral over $-\infty, l$.^{[29](#),[T21](#)}

$$Q(f, T^2 g) = \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^l l' dQ(f, E(l')g) \right]$$

²⁹For Stieltjes integrals the relation

$$\int_A^B u(l) d \left[\int_A^l v(l') dw(l') \right] = \int_A^B u(l)v(l) dw(l)$$

holds; it corresponds to the relation

$$\int_A^B u(l) \frac{d}{dl} \left[\int_A^l v(l) dl \right] dl = \int_A^B u(l)v(l) dl$$

in ordinary integration.

^{T21}In note 29, the original has an additional dl after the outer Stieltjes differential; that redundant factor has been omitted.

$$= \int_{-\infty}^{\infty} l \cdot l \, dQ(f, E(l)g) = \int_{-\infty}^{\infty} l^2 \, dQ(f, E(l)g),$$

Or, in our abbreviated notation,

$$T^2 = \int_{-\infty}^{\infty} l^2 \, dE(l),$$

as was to be proved.

We further introduce the following notation: for $l \leq l'$ we have $E(l) \leq E(l')$, so $E(l') - E(l)$ is an elementary operator. We call it $E(l, l') = E(I)$, where I is the interval l, l' .

The Absolute Value of an Operator

XI. ^{T22} Before passing to the physical applications, we must develop one last category of concepts that we shall use.

Let A be a linear operator (which, like A^* , is meaningful on an everywhere dense subset of $\overline{\mathfrak{H}}$; cf. § VII). We take two complete normalized orthogonal systems $\varphi_1, \varphi_2, \dots$ and $\psi_1, \psi_2, \dots$, such that $A\psi_\nu$ is meaningful,³⁰ and put

$$[A; \varphi_\mu, \psi_\nu] = \sum_{\mu, \nu=1}^{\infty} |Q(\varphi_\mu, A\psi_\nu)|^2$$

(this sum is finite or infinite, but always meaningful, since only nonnegative terms occur).

Then

$$\sum_{\mu=1}^{\infty} |Q(\varphi_\mu, A\psi_\nu)|^2 = Q(A\psi_\nu),$$

$$[A; \varphi_\mu, \psi_\nu] = \sum_{\nu=1}^{\infty} Q(A\psi_\nu),$$

^{T22}The original prints “IX” here; the section is numbered XI in accordance with its position and the subsequent cross-references.

³⁰By choosing an everywhere dense sequence $f_1, f_2, \dots$ from the everywhere dense set on which Af is meaningful, and applying [Theorem 6 of § VI](#), one obtains the desired orthogonal system $\psi_1, \psi_2, \dots$

that is, the dependence of $[A; \varphi_\mu, \psi_\nu]$ on the φ_μ is only apparent. But because

$$\sum_{\mu, \nu=1}^{\infty} |Q(\varphi_\mu, A\psi_\nu)|^2 = \sum_{\mu, \nu=1}^{\infty} |Q(A^*\varphi_\mu, \psi_\nu)|^2 = \sum_{\mu, \nu=1}^{\infty} |Q(\psi_\nu, A^*\varphi_\mu)|^2,$$

$$[A; \varphi, \psi] = [A^*; \psi, \varphi]$$

(if all $A^*\varphi_\mu$ are meaningful; cf. [note 30](#)), the same holds for the ψ_ν , and consequently $[A; \varphi, \psi]$ depends on A alone:

$$[A; \varphi, \psi] = [A].$$

The last equation therefore has the consequence

$$[A] = [A^*].$$

We call the quantity $\sqrt{[A]}$ the *absolute value* of the operator A ; we now wish to examine it somewhat more closely.

First, we determine what $[A]$ is in the various realizations of $\bar{\mathfrak{H}}$.

In the discontinuous realization $\mathfrak{H}_0$, A can be represented by a matrix^{T23}

$$\{a_{\mu\nu}\}, \quad a_{\mu\nu} = \overline{a_{\nu\mu}} \quad (\mu, \nu = 1, 2, \dots).$$

The points

$$(1, 0, 0, \dots), \quad (0, 1, 0, \dots), \quad (0, 0, 1, \dots), \quad \dots$$

form, as is easily seen, a complete normalized orthogonal system. A transforms them, respectively, into

$$(a_{1,1}, a_{2,1}, a_{3,1}, \dots), \quad (a_{1,2}, a_{2,2}, a_{3,2}, \dots),$$

$$(a_{1,3}, a_{2,3}, a_{3,3}, \dots), \quad \dots$$

On taking the former points as $\psi_1, \psi_2, \dots$, we obtain^{T24}

$$[A] = \sum_{\nu=1}^{\infty} Q(A\psi_\nu) = \sum_{\nu=1}^{\infty} \left[\sum_{\mu=1}^{\infty} |a_{\mu\nu}|^2 \right] = \sum_{\mu, \nu=1}^{\infty} |a_{\mu\nu}|^2.$$

Thus $[A]$ is the sum of the squared absolute values of all elements of the matrix.

^{T23}The symmetry condition on the matrix is printed in the original, although A has been introduced as a general linear operator.

^{T24}In the middle member of this equality, the original omits the absolute-value signs around $a_{\mu\nu}$; they have been supplied to agree with the defining expression and the final member.

In the continuous realization $\mathfrak{H}$, we shall calculate $[A]$ only under the assumption that the operator A can be represented by an integral kernel Φ :

$$Af(P) = \int_{\Omega} \Phi(P, Q) f(Q) dv_Q.$$

Then

$$\begin{aligned} [A] &= \sum_{\nu=1}^{\infty} Q(A\psi_{\nu}) = \sum_{\nu=1}^{\infty} \int_{\Omega} |A\psi_{\nu}(P)|^2 dv_P \\ &= \sum_{\nu=1}^{\infty} \int_{\Omega} \left| \int_{\Omega} \Phi(P, Q) \psi_{\nu}(Q) dv_Q \right|^2 dv_P \\ &= \int_{\Omega} \left[\sum_{\nu=1}^{\infty} \left| \int_{\Omega} \Phi(P, Q) \psi_{\nu}(Q) dv_Q \right|^2 \right] dv_P \\ &= \int_{\Omega} \left[\int_{\Omega} |\Phi(P, Q)|^2 dv_Q \right] dv_P = \int_{\Omega} \int_{\Omega} |\Phi(P, Q)|^2 dv_P dv_Q. \end{aligned}$$

Thus $[A]$ is the integral of the squared absolute value of the integral kernel.

Second, we derive the most important properties of $[A]$.

Theorem 1. We always have $[A] \geq 0$, and it is zero only if $A = 0$. (Consequently, the same holds for $\sqrt{[A]}$.)

Proof. $[A] \geq 0$ is clear. From $[A] = 0$ it follows that

$$\sum_{\nu=1}^{\infty} Q(A\psi_{\nu}) = 0, \quad Q(A\psi_1) \leq 0, \quad A\psi_1 = 0.$$

If f is any element of $\overline{\mathfrak{H}}$ for which Af is meaningful, then either $f = 0$, $Af = 0$; or $f \neq 0$, and for $\varphi = \frac{1}{\sqrt{Q(f)}}f$ we have $Q(\varphi) = 1$, so $\varphi = \psi_1$ can be completed to a complete normalized orthogonal system $\psi_1, \psi_2, \dots$ (all $A\psi_{\nu}$ meaningful!). Thus $A\varphi = 0$, $Af = 0$.

Consequently, A must be zero.

Theorem 2. We always have

$$\begin{aligned} \sqrt{[aA]} &= |a| \sqrt{[A]}, \\ \sqrt{[A+B]} &\leq \sqrt{[A]} + \sqrt{[B]}, \\ \sqrt{[AB]} &\leq \sqrt{[A]} \sqrt{[B]}. \end{aligned}$$

Proof. The first formula is trivial. The second follows from

$$Q((A + B)\psi_\nu) - Q(A\psi_\nu) - Q(B\psi_\nu) = 2\Re Q(A\psi_\nu, B\psi_\nu),$$

whose absolute value is therefore

$$\leq 2\sqrt{Q(A\psi_\nu)Q(B\psi_\nu)}.$$

After summation over $\nu = 1, 2, \dots$, we obtain^{31, T25}

$$\begin{aligned} [A + B] - [A] - [B] &\leq 2 \sum_{\nu=1}^{\infty} \sqrt{Q(A\psi_\nu)Q(B\psi_\nu)} \\ &\leq 2 \sqrt{\sum_{\nu=1}^{\infty} Q(A\psi_\nu) \sum_{\nu=1}^{\infty} Q(B\psi_\nu)} \\ &= 2\sqrt{[A][B]}, \\ [A + B] &\leq [A] + [B] + 2\sqrt{[A][B]} = (\sqrt{[A]} + \sqrt{[B]})^2, \\ \sqrt{[A + B]} &\leq \sqrt{[A]} + \sqrt{[B]}. \end{aligned}$$

The third is proved as follows:

$$\begin{aligned} [AB] &= \sum_{\mu, \nu=1}^{\infty} |Q(\varphi_\mu, AB\psi_\nu)|^2 = \sum_{\mu, \nu=1}^{\infty} |Q(A^*\varphi_\mu, B\psi_\nu)|^2 \\ &\leq \sum_{\mu, \nu=1}^{\infty} Q(A^*\varphi_\mu)Q(B\psi_\nu) = \sum_{\mu=1}^{\infty} Q(A^*\varphi_\mu) \sum_{\nu=1}^{\infty} Q(B\psi_\nu) \\ &= [A^*][B] = [A][B], \\ \sqrt{[AB]} &\leq \sqrt{[A]}\sqrt{[B]}. \end{aligned}$$

Theorem 3. The equation

$$[A + B] = [A] + [B]$$

³¹The inequality

$$\sqrt{a_1 b_1} + \dots + \sqrt{a_n b_n} \leq \sqrt{(a_1 + \dots + a_n)(b_1 + \dots + b_n)}$$

holds, as is well known, for all nonnegative a_n, b_n .

^{T25}In the second line of the estimate for $[A + B] - [A] - [B]$, the original omits Q before $B\psi_\nu$ in the second sum under the square root; it has been supplied.

holds as soon as one of the following four equations is satisfied:

$$AB^* = 0, \quad A^*B = 0, \quad BA^* = 0, \quad B^*A = 0.$$

Proof. For $A^*B = 0$ we have

$$\begin{aligned} Q((A+B)\psi_\nu) - Q(A\psi_\nu) - Q(B\psi_\nu) &= 2\Re Q(A\psi_\nu, B\psi_\nu) \\ &= 2\Re Q(\psi_\nu, A^*B\psi_\nu) = 0, \end{aligned}$$

and hence, after summation over $\nu = 1, 2, \dots$,

$$[A+B] - [A] - [B] = 0, \quad [A+B] = [A] + [B].$$

We may replace A, B by A^*, B^* ; the sufficient condition $A^*B = 0$ then becomes $A^{**}B^* = AB^* = 0$. Furthermore, we may interchange A, B , which gives $B^*A = 0$ and $BA^* = 0$.

Clearly,

$$\begin{aligned} \sum_{\mu, \nu=1}^{\infty} |Q(A\varphi_\mu, B\psi_\nu)|^2 &= \sum_{\mu, \nu=1}^{\infty} |Q(\varphi_\mu, A^*B\psi_\nu)|^2 \\ &= [A^*B]. \end{aligned}$$

The expression on the left (apparently dependent on A, B , the φ and the ψ) therefore depends on A^*B alone. It will later play an important role, so we introduce for it the separate symbol $[A, B]$. Thus

$$[A, B] = [A^*B].$$

In particular, $(A^*B)^* = B^*A^{**} = B^*A$ gives

$$[A, B] = [B, A].$$

From [Theorem 2](#) it follows that

$$[aA, bB] = |ab|^2[A, B].$$

By [Theorem 3](#), we certainly have

$$[A, B+C] = [A, B] + [A, C]$$

if one of the following four conditions is satisfied:

$$A^*B(A^*C)^* = A^*BC^*A = 0,$$

$$\begin{aligned}
(A^*B)^*A^*C &= B^*AA^*C = 0, \\
A^*C(A^*B)^* &= A^*CB^*A = 0, \\
(A^*C)^*A^*B &= C^*AA^*B = 0.
\end{aligned}$$

Thus, in any event, $BC^* = 0$ or $CB^* = 0$ is sufficient.

We shall chiefly have to deal with expressions $[E, F]$, where E, F are elementary operators. If E, F commute, then EF is also an elementary operator, and $[E, F]$ has a simple geometrical meaning.

For

$$[E, F] = [E^*F] = [EF],$$

so we are concerned with $[E]$, where E is an elementary operator. It is easy to give a complete normalized orthogonal system consisting of ψ_ν all lying in the interior or exterior of E ;^{32, T26} then

$$\begin{aligned}
[E] &= \sum_{\nu=1}^{\infty} Q(E\psi_\nu) \\
&= \sum_{\psi_\nu \text{ in the interior of } E} Q(\psi_\nu) + \sum_{\psi_\nu \text{ in the exterior of } E} Q(0) \\
&= \sum_{\psi_\nu \text{ in the interior of } E} 1 \\
&= \text{the number of } \psi_\nu \text{ in the interior of } E.
\end{aligned}$$

But this is plainly the number of dimensions of the interior of E . Consequently, for example, $[1] = \infty$ (because the interior of 1 comprises all of $\tilde{H}$), and $[0] = 0$.

Theorems 1 and 2 show that $\sqrt{[A]}$ can indeed be regarded as an absolute value of operators. However, it is plainly usable only in the immediate neighborhood of 0: $\sqrt{[1]}$ is already infinite.

³²Let $f_1, f_2, \dots$ and $g_1, g_2, \dots$ be sequences everywhere dense in the interior and exterior of E , respectively. By applying **Theorem 6 of § VI**, let them give rise to the normalized orthogonal systems $\varrho_1, \varrho_2, \dots$ and $\sigma_1, \sigma_2, \dots$. By **Theorem 6 of § VIII**, $\varrho_1, \sigma_1, \varrho_2, \sigma_2, \dots$ is also a normalized orthogonal system, and by **Theorem 5 of § VIII** it spans an everywhere dense linear manifold, that is, it is complete.

^{T26}In note 32, the original prints “§ XIII” for the reference to Theorem 5; the intended section is VIII.

The Statistical Ansatz of Quantum Mechanics

XII. We are now in a position to undertake our actual program: the mathematically unobjectionable unification of statistical quantum mechanics. For this purpose we shall first consider a case as simple as possible, with results unambiguously established, and translate the (known) results into our mode of expression: this will give us a guide to the path to be followed in general.

For this purpose we consider a case of the continuous realization that is as simple as possible.

Let Ω be k -dimensional Euclidean space, and consider a quantum-mechanical system whose Schrödinger equation is (cf. § II)

$$H\psi - l\psi = 0.$$

As is well known,³³ the symmetric linear operator H arises as follows: take the classical-mechanical expression for the energy as a function of the coordinates $q_1, q_2, \dots, q_k$ and momenta $p_1, p_2, \dots, p_k$, and replace each q_μ by the operator $q_\mu \dots$, and each p_μ by the operator $\frac{\hbar}{2\pi i} \frac{\partial}{\partial q_\mu} \dots$. (There is a certain ambiguity here, since the q_μ, p_ν , as ordinary numbers, commute under multiplication, whereas the operators $q_\mu \dots$ and $\frac{\hbar}{2\pi i} \frac{\partial}{\partial q_\mu} \dots$ do not. There is, to be sure, a certain restriction on H , namely that it must be symmetric; but this alone does not suffice to determine it uniquely. Here lies one of the essential gaps in quantum mechanics.)

We now assume that there is only a nondegenerate point spectrum (that is, one consisting entirely of simple eigenvalues), and denote the eigenvalues by $l_1, l_2, \dots$, and the corresponding (normalized) eigenfunctions by $\varphi_1, \varphi_2, \dots$.

We have already [determined in § X](#) the resolution of the identity $E(l)$ belonging to this H . It is

$$E(l)f = \sum_{l_n \leq l} Q(f, \varphi_n) \varphi_n.$$

We further need the resolution of the identity $F_\mu(l)$ belonging to the operator $q_\mu \dots$. This is a slight generalization of the [last example, likewise considered in § X](#) (where we had only $k = 1$), namely,

$$F_\mu(l)f(q_1, \dots, q_k) = \begin{cases} f(q_1, \dots, q_k), & \text{for } q_\mu \leq l, \\ 0, & \text{for } q_\mu > l. \end{cases}$$

³³See, for example, [note 3](#).

(This follows from exactly the same qualitative consideration as there; and again the defining relation

$$q_\mu f(q_1, \dots, q_k) = \int_{-\infty}^{\infty} l d\{F_\mu(l) f(q_1, \dots, q_k)\}$$

is verified without difficulty.)

The probability ansatz of Pauli and Dirac valid in this case is as follows:³⁴ if the system is in the n th quantum state (l_n, φ_n) , then the probability that the coordinates $q = (q_1, \dots, q_k)$ lie in the k -dimensional rectangular box K (we write dq for $dq_1 \cdots dq_k$) is

$$\int_K |\varphi_n(q)|^2 dq.$$

We can generalize this ansatz somewhat further. If we know only that the energy lies in the interval I , let this probability be

$$\sum_{l_n \in I} \int_K |\varphi_n(q)|^2 dq$$

(apart from normalization factors). Indeed, if only one eigenvalue lies in I , this amounts to the earlier assertion; and if several lie there, it follows from that assertion if we regard the individual (nondegenerate) quantum states as a priori equally probable (as is generally customary).

But if K and I are defined, respectively, by the inequalities

$$q'_1 < q_1 \leq q''_1, \quad q'_2 < q_2 \leq q''_2, \quad \dots, \quad q'_k < q_k \leq q''_k, \\ l' < l \leq l'',$$

we can also write this expression as follows:

$$\begin{aligned} & \sum_{l_n \in I} \int_K |\varphi_n(q)|^2 dq \\ &= \sum_{l_n \in I} \int_{q'_1}^{q''_1} \cdots \int_{q'_k}^{q''_k} |\varphi_n(q_1, \dots, q_k)|^2 dq_1 \cdots dq_k \\ &= \sum_{l_n \in I} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \\ & \quad \cdot |F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) \varphi_n(q_1, \dots, q_k)|^2 dq_1 \cdots dq_k \end{aligned}$$

³⁴See, for example, the paper by Jordan cited in [note 7](#).

$$\begin{aligned}
&= \sum_{l_n \in I} \int_{\Omega} |F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) \varphi_n(q)|^2 dq \\
&= \sum_{n=1}^{\infty} \int_{\Omega} |F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) E(l', l'') \varphi_n(q)|^2 dq \\
&= \sum_{n=1}^{\infty} Q(F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) E(l', l'') \varphi_n) \\
&= [F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) E(l', l'')].
\end{aligned}$$

Now the $F_{\mu}(q'_{\mu}, q''_{\mu})$ all commute with one another, as is easily seen, and hence one concludes further that this is

$$\begin{aligned}
&= [(F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k))^*, E(l', l'')] \\
&= [F_k(q'_k, q''_k) \cdots F_1(q'_1, q''_1), E(l', l'')] \\
&= [F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k), E(l', l'')].
\end{aligned}$$

That is: if the energy lies in I , the (relative) probability that q_1 lies in $J_1, \dots, q_k$ in J_k is

$$[E(I), F_1(J_1) \cdots F_k(J_k)].$$

But according to an ansatz of Jordan, conversely, the quantity

$$\sum_{l_n \in I} \int_K |\varphi_n(q)|^2 dq$$

is also to be regarded as the (relative) probability that l_n , that is, the energy, lies in I if q lies in K (that is, q_1 in $J_1, \dots, q_k$ in J_k); cf. [note 7](#). More precisely, the Pauli–Jordan assertion refers only to the limiting case of infinitely small K (where, after division by the proportionality factor “volume of K ,” the probability

$$\sum_{l_n \in I} |\varphi_n(q)|^2$$

remains). Nevertheless, a correct result also emerges in the opposite limiting case, $K = \Omega$ (that is, the whole space):

$$\sum_{l_n \in I} \int_{\Omega} |\varphi_n(q)|^2 dq = \sum_{l_n \in I} 1 = \text{the number of } l_n \text{ in } I.$$

That is: all quantized (nondegenerate) states are a priori equally probable, and unquantized states are impossible (both are among the fundamental assumptions of quantum theory).

Thus we obtain here: if q_1 lies in $J_1, \dots, q_k$ in J_k , then the (relative) probability that l_n (the energy) lies in I is the same expression as before, which we now wish to write as follows:

$$[F_1(J_1) \cdots F_k(J_k), E(I)].$$

XIII. The results just obtained suggest the following assumption:

Let $R_1, R_2, \dots, R_i$ and $S_1, S_2, \dots, S_j$ be $i + j$ symmetric linear operators representing certain physically meaningful quantities. (We cannot discuss here in greater detail precisely what this latter concept, “representing a quantity by an operator,” which is entirely fundamental in present-day quantum mechanics, means. Compare the discussion at the beginning of the preceding section concerning the relation between the classical-mechanical Hamiltonian function and the “energy operator” H .) Let the resolutions of the identity belonging to R_μ ($\mu = 1, 2, \dots, i$) and S_ν ($\nu = 1, 2, \dots, j$) be $E_\mu(l)$ and $F_\nu(l)$, respectively.

We assume that all the $E_\mu(l)$ commute with one another, and likewise all the $F_\nu(l)$. We call this the complete commutativity of $R_1, R_2, \dots, R_i$ with one another, and of $S_1, S_2, \dots, S_j$, respectively. (For the complete commutativity of two operators T', T'' , their ordinary commutativity, $T'T'' = T''T'$, is certainly necessary, as is easily shown; it is also sufficient as soon as at least one of them is bounded. If both are unbounded, certain difficulties of a formal nature arise, into which we do not wish to enter here. For all operators occurring in quantum mechanics, the two notions are the same.)

We now make the following physical assumption:

If the quantities represented by $R_1, \dots, R_i$ have assumed values lying in the intervals $I_1, \dots, I_i$, respectively, then the (relative) probability that the quantities represented by $S_1, \dots, S_j$ assume values lying in the intervals $J_1, \dots, J_j$, respectively, is

$$[E_{I_1} \cdots E_{I_i}, F_{J_1} \cdots F_{J_j}] = [E_{I_1} \cdots E_{I_i} F_{J_1} \cdots F_{J_j}].$$

(The latter equality holds because the E_{I_μ} all commute with one another.)

We shall now draw some consequences from this assumption in order to demonstrate its usefulness.

α . Both the premises $(R_1, \dots, R_i)$ and the conclusions $(S_1, \dots, S_j)$ may be permuted among themselves in any way without changing the probability distribution. This follows from the commutativity of the $E_\mu(l)$ and the $F_\nu(l)$, respectively, and hence also of their differences $E_\mu(I_\mu)$ and $F_\nu(J_\nu)$, with one another.

β . Interchanging all the premises with all the conclusions changes nothing. (That is, the probability distribution is as though it arose from a priori probabilities.) This follows from the universally valid formula $[A, B] = [B, A]$.

γ . Vacuous premises or conclusions may be added or omitted at will (that is, those for which the interval is $-\infty, \infty$). For they merely cause a factor

$$E_\mu(-\infty, \infty) \quad \text{or} \quad F_\nu(-\infty, \infty) = 1 - 0 = 1$$

to appear or disappear.

δ . The multiplication law of probabilities does not hold in general (a weaker law does hold, corresponding to Jordan's "composition of probability amplitudes"; compare [note 7](#). We shall not enter into it further here). This is not particularly surprising, since the dependence relations of our probabilities can be arbitrarily complicated; moreover, we are dealing with relative probabilities.

ε . The addition law of probabilities holds. (This also holds in ordinary probability theory irrespective of the dependence relations.) We must show that $J'_j + J''_j = J_j$ implies

$$\begin{aligned} & [E_1(I_1) \cdots E_i(I_i) F_1(J_1) \cdots F_j(J'_j)] \\ & + [E_1(I_1) \cdots E_i(I_i) F_1(J_1) \cdots F_j(J''_j)] \\ & = [E_1(I_1) \cdots E_i(I_i) F_1(J_1) \cdots F_j(J_j)]. \end{aligned}$$

(By α and β , we may restrict ourselves to the case in which the last interval, J_j , is subdivided.) We may also write this as

$$[AF_j(J'_j)] + [AF_j(J''_j)] = [AF_j(J_j)].$$

By [Theorem 3, § XI](#), this is certainly the case if

$$\begin{aligned} & AF_j(J'_j) + AF_j(J''_j) = AF_j(J_j), \\ & AF_j(J'_j)(AF_j(J''_j))^* = AF_j(J'_j)F_j(J''_j)A^* = 0. \end{aligned}$$

The relation $F_j(J'_j) + F_j(J''_j) = F_j(J_j)$ entails the first, but also the second, because $F_j(J'_j)$ and $F_j(J''_j)$ are then disjoint (compare [Theorem 2, § VIII](#)). This is clear: if, for example, J'_j, J''_j, J_j are the intervals $l', l''; l'', l'''$; and l', l''' , then

$$\begin{aligned} F_j(J'_j) &= F_j(l'') - F_j(l'), \\ F_j(J''_j) &= F_j(l''') - F_j(l''), \end{aligned}$$

$$F_j(J_j) = F_j(l''') - F_j(l').$$

ϑ. Our expression for probability is invariant under canonical transformations. By a canonical transformation we understand the following (compare the references in [note 7](#)):

Let U be a linear operator with the property

$$UU^* = U^*U = 1;$$

we then call U orthogonal. The canonical transformation consists in replacing every linear operator R by URU^* .

The operations aR , $R + S$, RS , R^* are evidently invariant under this transformation; so too are the properties of symmetry and of being an elementary operator, and, between elementary operators, the relations $\leq$ and disjointness. Furthermore, Q is invariant when $\overline{\mathfrak{H}}$ is mapped onto itself by U :

$$Q(Uf, Ug) = Q(U^*Uf, g) = Q(f, g).$$

Thus, together with $\varphi_1, \varphi_2, \dots$, the system $U\varphi_1, U\varphi_2, \dots$ is also a complete normalized orthogonal system, from which the invariance of $[A]$ follows. The eigenvalue representation ([§ IX](#)) is evidently invariant as well.

Since none of the concepts we use changes, the same holds for the (relative) probability based on them.

Applications

XIV. We now consider some physical applications.

First, the case of the Schrödinger equations. This was already discussed for nondegenerate systems in [§ XII](#), and if we also admit degeneracies, that is, multiple eigenvalues, none of the results there changes. The a priori probabilities of the individual energy levels, for example, are (compare the notation of [§ XII](#))

$$\begin{aligned} & [F_1(-\infty, \infty) \cdots F_k(-\infty, \infty), E(I)] \\ & = [1, E(I)] = [E(I)], \end{aligned}$$

which is the dimension of the space of $f(q)$ lying in the interior of $E(I)$. But, as we know, the interior of $E(I)$ contains all linear aggregates

$$a_1\varphi_{\nu_1}(q) + a_2\varphi_{\nu_2}(q) + \cdots,$$

where $l_{\nu_1}, l_{\nu_2}, \dots$ are the eigenvalues of H lying in I (multiple eigenvalues counted according to their multiplicity), and $\varphi_{\nu_1}, \varphi_{\nu_2}, \dots$ the corresponding eigenfunctions. Thus the dimension in question is the number of eigenvalues in I .

This result therefore means: the a priori probability of a quantized state is the multiplicity of the eigenvalue lying in it, and nonquantized states are impossible.

Second, we consider the case of precise, causal dependence. We put $i = 1$, $j = 1$, and assume that a certain quantity is given, while another is sought that is a function of it, that is, is causally determined by it. Let the operators belonging to them be R and S , respectively, with $S = f(R)$, where $f(x)$ is a real function. We assume that $f(x)$ increases monotonically.

(The restrictions $i = 1$, $j = 1$, as well as the monotonicity of $f(x)$, are unnecessary; we impose them only to simplify the calculation, which is intended merely for orientation.)

Let the resolution of the identity belonging to R be $E(l)$:

$$R = \int_{-\infty}^{\infty} l dE(l).$$

As we [showed in § X](#), we then have

$$R^n = \int_{-\infty}^{\infty} l^n dE(l),$$

and hence

$$\begin{aligned} S = f(R) &= \int_{-\infty}^{\infty} f(l) dE(l) \\ &= \int_{-\infty}^{\infty} l' dE(g(l')), \end{aligned}$$

if g is the inverse of f . Thus $E(g(l))$ is the resolution of the identity belonging to S .

If J is the interval l', l'' , we understand by $g(J)$ the interval $g(l'), g(l'')$. Thus

$$F(J) = E(g(J)),$$

and consequently^{T27}

$$[E(I), F(J)] = [E(I)F(J)]$$

^{T27}In the original, the first product on the right in the following chain is printed as $E(J)F(J)$; the interval index has been corrected to I .

$$= [E(I)E(g(J))] = [E(I \cdot g(J))].$$

(By $I \cdot g(J)$ we understand the common part of the intervals I and $g(J)$; it is easy to verify the relation $E(I)E(J') = E(I \cdot J')$, which always holds.)

If the intervals $I, g(J)$ have no common part, this is zero; but $I \cdot g(J)$ is empty precisely when there is no x in I for which $f(x)$ belongs to J . Thus values contradicting the causal constraint do not occur.

If $I \cdot g(J)$ is not empty, it is the interval I' of all x in I for which $f(x)$ lies in J , that is, those admissible under the causal constraint. Thus the probability is $[E(I')]$, the dimension of the space of f in the interior of $E(I')$, or, as we know, the number of eigenvalues of R lying in I' .

That is: within the causal constraints, the customary quantum-theoretical result is obtained when R is quantized. (Incidentally, this probability is evidently ∞ if R has a continuous spectrum and that spectrum extends into I' .)

With these two examples we believe we have shown that statistical quantum mechanics, despite its probabilistic character, is quite capable of making precise and rigorous assertions whenever there is occasion to do so: for example, in the case of absolutely precise quantum prohibitions and causal constraints.

Finally, third, we consider Born's quantum-mechanical treatment of time-dependent systems.³⁵

The Hamiltonian function of the system depends explicitly on time, and so does the corresponding operator $H(t)$ (time is treated as a number, rather than as a "quantity," that is, as an operator). Let there be no degeneracies; let the eigenfunctions at time t_0 be

$$\varphi_1^{(0)}, \varphi_2^{(0)}, \dots,$$

and those at time t be

$$\varphi_1^{(t)}, \varphi_2^{(t)}, \dots$$

Then, according to Born,³⁶ the probability that the system is in the ν th state at time t , if it was in the μ th state at time t_0 , is $|c_{\mu,\nu}(t)|^2$, where $c_{\mu,\nu}(t)$ is the μ th

³⁵I owe this example to a remark by Mr. L. Nordheim.

³⁶*Zeitschr. für Physik*, vol. 38, p. 803, vol. 40, p. 167 (1926).

expansion coefficient of $\varphi_\nu^{(t)}$ with respect to $\varphi_1^{(0)}, \varphi_2^{(0)}, \dots$:^{T28}

$$\varphi_\nu^{(t)} = \sum_{\mu=1}^{\infty} c_{\mu,\nu}(t) \varphi_\mu^{(0)}.$$

Indeed, let $E_0(l)$ be the resolution of the identity belonging to $H(t_0)$, and $E_t(l)$ that belonging to $H(t)$. As we saw in § X, the interior of $E_0(l)$ and $E_t(l)$ consists of all linear aggregates of the $\varphi_\mu^{(0)}$ and $\varphi_\nu^{(t)}$, respectively, whose eigenvalues are $\leq l$; hence the interior of $E_0(I)$ and $E_t(I)$ consists of all linear aggregates of the $\varphi_\mu^{(0)}$ and $\varphi_\nu^{(t)}$, respectively, whose eigenvalues lie in I .

Now let I_μ contain a single eigenvalue of $H(0)$, the μ th, and let J_ν contain a single eigenvalue of $H(t)$, the ν th. Then the interiors of $E_0(I_\mu)$ and $E_t(J_\nu)$ consist of the multiples of $\varphi_\mu^{(0)}$ and $\varphi_\nu^{(t)}$, respectively, so that

$$\begin{aligned} E_0(I_\mu)f &= Q(f, \varphi_\mu^{(0)})\varphi_\mu^{(0)}, \\ E_t(J_\nu)f &= Q(f, \varphi_\nu^{(t)})\varphi_\nu^{(t)}. \end{aligned}$$

We now calculate $[E_0(I_\mu), E_t(J_\nu)]$, using $\varphi_1^{(0)}, \varphi_2^{(0)}, \dots$ as the complete normalized orthogonal system:

$$\begin{aligned} [E_0(I_\mu), E_t(J_\nu)] &= [E_t(J_\nu)E_0(I_\mu)] \\ &= \sum_{\varrho=1}^{\infty} Q(E_t(J_\nu)E_0(I_\mu)\varphi_\varrho^{(0)}) \\ &= Q(E_t(J_\nu)\varphi_\mu^{(0)}) \\ &= Q(Q(\varphi_\mu^{(0)}, \varphi_\nu^{(t)})\varphi_\nu^{(t)}) \\ &= |Q(\varphi_\mu^{(0)}, \varphi_\nu^{(t)})|^2, \end{aligned}$$

which is precisely Born's expression, since

$$c_{\mu,\nu}(t) = Q(\varphi_\nu^{(t)}, \varphi_\mu^{(0)}) = \overline{Q(\varphi_\mu^{(0)}, \varphi_\nu^{(t)})}.$$

^{T28}The indices in this displayed expansion have been made consistent with the preceding definition and the coefficient formula below. The original prints $\varphi_\mu^{(t)} = \sum_{\nu=1}^{\infty} c_{\mu,\nu}(t)\varphi_\nu^{(0)}$. In the formula for $E_t(J_\nu)f$ below, the original likewise prints $\varphi_\mu^{(t)}$ as the second argument of Q ; this has been corrected to $\varphi_\nu^{(t)}$.

Summary

XV. As can be seen, our assumption does justice to the dual nature of events (continuous–discontinuous) by associating with each quantity, or with the symmetric linear operator corresponding to it, a resolution of the identity 1 into elementary operators $E(l', l'') = E(l'') - E(l')$ ($l' \leq l''$).

The function $E(l)$ is monotonically nondecreasing (in the sense in which this is to be understood for elementary operators), and this monotonicity may display all the features known from atoms: its increase from 0 (for $l = -\infty$) to 1 (for $l = \infty$) may take place in individual jumps (quantized states) or continuously (nonquantized states), and intervals of constancy may occur between them (forbidden states). The assertion “the quantity to which R belongs has a value in the interval $l' < x \leq l''$ ” is represented in our calculus by the elementary operator $E(l', l'')$.

If several assertions are made: “the values of the quantities $R_1, R_2, \dots, R_i$ lie in the intervals $I_1, I_2, \dots, I_i$, and those of the quantities $S_1, S_2, \dots, S_j$ lie in the intervals $J_1, J_2, \dots, J_j$,” then the product

$$E_1(I_1) \cdots E_i(I_i) F_1(J_1) \cdots F_j(J_j)$$

must be formed (in this transferred sense, the multiplication law of probabilities holds!); its squared absolute value

$$\begin{aligned} & [E_1(I_1) \cdots E_i(I_i) F_1(J_1) \cdots F_j(J_j)] \\ &= [E_1(I_1) \cdots E_i(I_i), F_1(J_1) \cdots F_j(J_j)] \end{aligned}$$

is then the (relative) probability of their coexistence. Notice that a separation between cause and effect need not be present a priori in this scheme at all; it is subsequently brought about of itself by the commutativity relations: if all the $E_\mu(I_\mu)$ commute with one another on the one hand, and all the $F_\nu(J_\nu)$ on the other, then the product automatically falls into these two groups. Within each group the order of the factors is immaterial (α in § XIII), as is the order of the two groups as wholes (β in § XIII), that is, what is designated as cause and what as effect.

Of course, the commutativity relations need not determine the separation between cause and effect uniquely: if, for example, an E_ϱ commutes both with all the E_μ and with all the F_ν , it may be assigned to either group at will.

On the other hand, they exclude certain divisions from the outset: it is not possible to force, in all cases, the absolutely precise causal connection generally

present for commuting quantities (compare the second example of § XIV) by observing all coordinates and their conjugate momenta, according to the procedure of classical mechanics, and making all of them “causes,” that is, R_μ ’s. (Dirac was the first to point out the inadmissibility of this procedure.) For, as is well known, the operators of a coordinate and its conjugate momentum do not commute, namely $q_\mu \cdot \dots$ and $\frac{h}{2\pi i} \frac{\partial}{\partial q_\mu} \cdot \dots$; they will therefore always separate automatically: one must be cause, the other effect (or, respectively, what is observed and what is predicted). Even when everything has been observed, it is useless: quantum mechanics (and it comprises approximately everything exact that we know about atoms today) does not address this question at all!

In conclusion, it should be mentioned that the applicability of our method is probably not exhausted by the foregoing. We intend to return to this on another occasion, as well as to certain formal mathematical questions that have remained unsettled here.

Appendices

1)

Let w be an eigenvalue, and $x_1, x_2, \dots$ a sequence transformed by H into w times itself. Let application of S^{-1} transform it into $y_1, y_2, \dots$; then our condition is equivalent to saying that $S^{-1}HS$ transforms $y_1, y_2, \dots$ into w times itself.

Now $S^{-1}HS = W$ takes $y_1, y_2, \dots$ into $w_1 y_1, w_2 y_2, \dots$, so that $y_\mu = 0$ unless $w_\mu = w$. That is: if w differs from all of $w_1, w_2, \dots$, then all the y_μ , and consequently all the x_μ as well, are zero; that is, w is not an eigenvalue. If, on the other hand, w equals certain w_μ , say $w_{\mu'}, w_{\mu''}, \dots$, then $y_{\mu'}, y_{\mu''}, \dots$ may be arbitrary (all other y_μ are zero). The sequence $x_1, x_2, \dots$ is obtained from $y_1, y_2, \dots$ by transformation with S , so that^{T29}

$$x_\mu = \sum_{\nu=1}^{\infty} s_{\mu\nu} y_\nu = s_{\mu\mu'} y_{\mu'} + s_{\mu\mu''} y_{\mu''} + \dots$$

That is, $x_1, x_2, \dots$ can indeed be any linear combination of the columns

$$s_{1\mu'}, s_{2\mu'}, \dots; \quad s_{1\mu''}, s_{2\mu''}, \dots; \quad \dots,$$

and nothing else.

^{T29}The summation index in the original is printed as μ although the summand is indexed by ν ; the bound has been corrected to ν .

2)

We wish to show that the space $\mathfrak{H}$ of all functions $f(P)$ defined on Ω (Ω being any k -dimensional surface in l -dimensional space, P an arbitrary point on Ω , and dv the differential element on Ω) for which $\int_{\Omega} |f(P)|^2 dv$ is finite satisfies conditions A–E of § V^{T30}. For ordinary Hilbert space $\mathfrak{H}_0$ (all sequences $x_1, x_2, \dots$ with finite $\sum_{n=1}^{\infty} |x_n|^2$), the proof is unnecessary: for, by § VI, every space with properties A–E has the properties of $\mathfrak{H}_0$, and therefore $\mathfrak{H}_0$ has properties A–E.

The definition of the operations af and $f + g$ is evident; it is clear that af belongs to $\mathfrak{H}$ whenever f does,^{T31} and that $f + g$ belongs to $\mathfrak{H}$ whenever f, g do, by the inequality in note 16. As motivated in § III, we must define $Q(f, g)$ as $\int_{\Omega} f(P)\overline{g(P)} dv$. This is a finite number for all f, g in $\mathfrak{H}$, since

$$|f(P)\overline{g(P)}| \leq \frac{1}{2}|f(P)|^2 + \frac{1}{2}|g(P)|^2,$$

and $\int_{\Omega} |f(P)|^2 dv$ and $\int_{\Omega} |g(P)|^2 dv$ are finite.

Let us now go through conditions A–E!

A, B. These are evidently satisfied.

C. We choose k regions $\mathfrak{M}_1, \mathfrak{M}_2, \dots, \mathfrak{M}_k$ on Ω with no points in common; let $f_p(P)$ be 1 or 0 according as P belongs to $\mathfrak{M}_p$ or not ($p = 1, 2, \dots, k$). The f_p ($p = 1, 2, \dots, k$) are linearly independent: for if

$$a_1 f_1 + \dots + a_k f_k = 0,$$

then, in particular, throughout $\mathfrak{M}_p$,

$$a_1 f_1(P) + \dots + a_k f_k(P) = 0,$$

that is, $a_p = 0$ ($p = 1, 2, \dots, k$).

D. A general proof of this property for all surfaces Ω is hardly possible without a closer examination of the precise concept of a general surface and of the so-called Lebesgue measure.³⁷ We shall not undertake this here, but shall merely give sequences $f_1, f_2, \dots$ dense everywhere in $\mathfrak{H}$ in two characteristic cases.

^{T30}The original references in this paragraph are § IV for the axioms and § V for the characterization theorem; the intended sections are V and VI, respectively.

^{T31}Here the original prints $\mathfrak{H}_0$ for the function space $\mathfrak{H}$; the subscript has been removed.

³⁷Compare, for example, Carathéodory, *Vorlesungen über reelle Funktionen*. Berlin–Leipzig 1918, chapters V–IX.

First, let Ω be the “ n -dimensional rectangular box” of all points $x_1, x_2, \dots, x_n$ with

$$0 \leq x_\nu \leq 1 \quad (\nu = 1, 2, \dots, n).$$

Then $\mathfrak{H}$ is the space of all (complex) functions f defined on Ω with finite integral of the squared absolute value. We may therefore expand such functions $f(P) = f(x_1, x_2, \dots, x_n)$ in Fourier series:^{T32}

$$f(x_1, x_2, \dots, x_n) = \sum_{r_1, r_2, \dots, r_n = -\infty}^{\infty} c_{r_1, r_2, \dots, r_n} e^{2\pi i(r_1 x_1 + r_2 x_2 + \dots + r_n x_n)}.$$

The partial sum

$$\sum_{r_1, r_2, \dots, r_n = -N}^N c_{r_1, r_2, \dots, r_n} e^{2\pi i(r_1 x_1 + r_2 x_2 + \dots + r_n x_n)}$$

now converges in the mean to f as $N \rightarrow \infty$.³⁸

Thus, arbitrarily near every f in $\mathfrak{H}$ there is a function of the form

$$\sum_{r_1, r_2, \dots, r_n = -N}^N c_{r_1, r_2, \dots, r_n} e^{2\pi i(r_1 x_1 + r_2 x_2 + \dots + r_n x_n)},$$

and hence also one with rational $c_{r_1, r_2, \dots, r_n}$. But, as is well known, these can be written in the form of a sequence.³⁹

Second, let Ω be the “ n -dimensional (real) space” of all points $x_1, x_2, \dots, x_n$ without restriction, and let $\mathfrak{H}$ be defined correspondingly. Consider any (differentiable) function $\varphi(x)$ mapping the interval $0, 1$ onto the interval $-\infty, \infty$, and let $\psi(y)$ be its inverse (for example, $\ln \frac{x}{1-x}$ and $\frac{e^y}{e^y+1}$).

Then in general^{T33}

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

^{T32}The lower summation limit in the original is printed as $+\infty$; it has been corrected to $-\infty$.

³⁸I shall give a general and direct construction of a sequence dense everywhere in $\mathfrak{H}$ in the paper of mine mentioned in [note 12](#).

³⁹As is well known, all finite tuples of rational numbers (which are what we are concerned with here) can be written in the form of a sequence.

^{T33}The last argument of f in the first integral has been normalized to x_n , consistently with the n variables and the differentials that follow.

$$= \int_0^1 \cdots \int_0^1 f(\varphi(u_1), \dots, \varphi(u_n)) \varphi'(u_1) \cdots \varphi'(u_n) du_1 \cdots du_n;$$

consequently, we need only choose $f_1, f_2, \dots$ dense everywhere in the $\mathfrak{H}$ of the first example; the $g_1, g_2, \dots$ given by

$$g_\mu(x_1, \dots, x_n) = \frac{f_\mu(\psi(x_1), \dots, \psi(x_n))}{\varphi'(\psi(x_1)) \cdots \varphi'(\psi(x_n))}$$

are then evidently dense everywhere in the present example.

E. Let $f_1, f_2, \dots$ be a sequence of functions in $\mathfrak{H}$, and suppose that for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that $N \leq m \leq n$ implies

$$\int_{\Omega} |f_m(P) - f_n(P)|^2 dv \leq \varepsilon.$$

We choose a monotonically increasing sequence of numbers $N_1, N_2, \dots$ such that

$$N_\nu \geq N \left(\frac{1}{8^\nu} \right).$$

Then

$$\int_{\Omega} |f_{N_{\nu+1}}(P) - f_{N_\nu}(P)|^2 dv \leq \frac{1}{8^\nu}.$$

Thus the points P at which

$$|f_{N_{\nu+1}}(P) - f_{N_\nu}(P)| \geq \frac{1}{2^\nu}$$

form a surface portion of measure (area⁴⁰) $\leq \frac{1}{2^\nu}$.

The points for which the inequalities

$$\begin{aligned} |f_{N_{\nu+1}}(P) - f_{N_\nu}(P)| &\leq \frac{1}{2^\nu}, \\ |f_{N_{\nu+2}}(P) - f_{N_{\nu+1}}(P)| &\leq \frac{1}{2^{\nu+1}}, \\ |f_{N_{\nu+3}}(P) - f_{N_{\nu+2}}(P)| &\leq \frac{1}{2^{\nu+2}}, \\ &\dots \end{aligned}$$

do not hold therefore have measure

$$\leq \frac{1}{2^\nu} + \frac{1}{2^{\nu+1}} + \frac{1}{2^{\nu+2}} + \cdots = \frac{1}{2^{\nu-1}}.$$

⁴⁰By area we must, strictly speaking, understand Lebesgue measure (compare [note 37](#)).

But wherever the above inequalities hold, the sequence $f_{N_1}(P), f_{N_2}(P), \dots$ is evidently convergent; thus the points where it is not form a surface portion of measure $\leq \frac{1}{2^{\nu-1}}$. This holds for all ν , so it has measure zero.

The limit of $f_{N_\nu}(P)$ therefore exists everywhere (with the possible exception of a set of measure zero); denote it by $f(P)$.

If $m \geq N = N(\varepsilon)$, then for $N_\nu \geq m$, that is, for all N_ν with at most finitely many exceptions, we have

$$\int_{\Omega} |f_m(P) - f_{N_\nu}(P)|^2 dv \leq \varepsilon;$$

we may therefore let $\nu \rightarrow \infty$:

$$\int_{\Omega} |f_m(P) - f(P)|^2 dv \leq \varepsilon.$$

That is, f belongs to $\mathfrak{H}$, and the sequence $f_1, f_2, \dots$ converges to f .

3)

The Stieltjes integral

$$\int_a^b f(x) d\varphi(x)$$

is defined when a, b is a finite or infinite interval, $f(x)$ is a function continuous in this interval (including at a and b !), and $\varphi(x)$ is a monotonically nondecreasing function there (finite also at a and b), or the difference of two such functions (a function of bounded variation). Analogously to the Riemann integral, it is defined as the limit of the expressions

$$\sum_{n=1}^N f(x'_n)(\varphi(x_n) - \varphi(x_{n-1}))$$

with

$$a = x_0 \leq x'_1 \leq x_1 \leq x'_2 \leq x_2 \leq \dots \leq x_{N-1} \leq x'_N \leq x_N = b,$$

as the subdivision $x_0, x_1, \dots, x_N$ of the interval a, b becomes arbitrarily fine.^{41, T34}

⁴¹Stieltjes, *Recherches sur les fractions continues*, *Annales de la Faculté des sciences de Toulouse*, 1894/95, chapter VI. A brief account is given, for example, by Carleman, *Équations Intégrales à noyau réel singulier*, Upsala 1923, pp. 7–9.

^{T34}The subdivision indices have been made consistent with the upper summation limit N ; the last primed index, printed as $N - 1$ in the original, has been corrected to N .

Without entering further into this, we give the following geometric illustration (for monotone φ):

In the x, y -plane draw the curve

$$\begin{aligned} x &= \varphi(u) & (a \leq u \leq b), \\ y &= f(u) \end{aligned}$$

(if x falls into a gap because $\varphi(u)$ jumps discontinuously over it, join the neighboring $f(u)$ horizontally). Then

$$\int_a^b f(u) d\varphi(u)$$

is the area of the region between this curve, the x -axis, and the abscissae $x = \varphi(a)$ and $x = \varphi(b)$ (compare [figures a–b](#) and [c](#)).

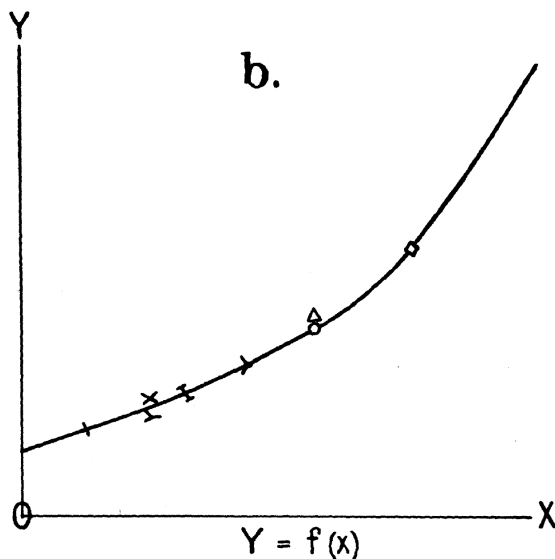
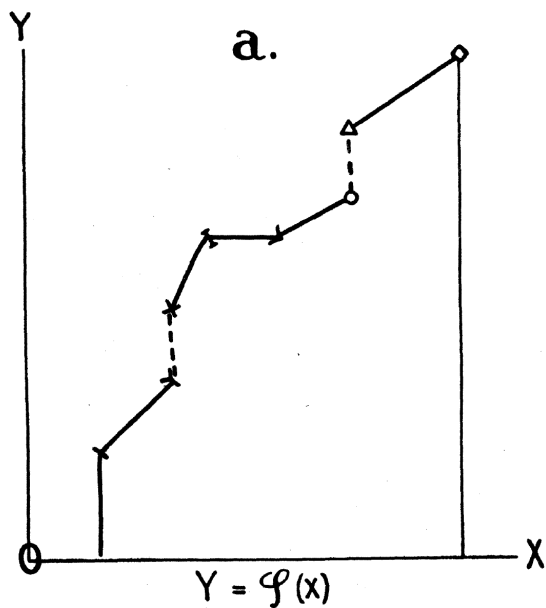


Figure 1: Figures a–b.

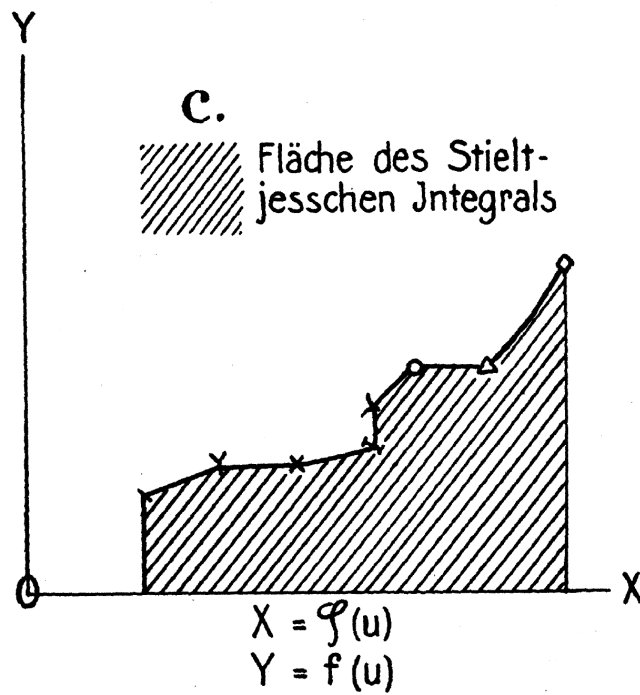


Figure 2: Figure c. The legend reads: "Area of the Stieltjes integral."