History of Physics (20)

What states are: an appraisal from recent history

Jan Lacki, Uni Genève
History of Physics (20)

What states are: an appraisal from recent history

Jan Lacki, Uni Genève

There has been recently a discussion among some SPS board members about the inclusion of philosophical matters in the activities of the "History of Physics" section. This seems sensible: did not an important 20th century historian and philosopher of science, Norwood Russell Hanson, once state: "Philosophy of Science without History of Science is empty, History of Science without Philosophy of Science is blind"? 1 I fully agree with this statement which characterizes perfectly my own approach to historical research. Now, the Board discussion is not yet over as many issues have to be clarified, to start with exactly what kind of "philosophy" does one want to consider (one devoted to strictly epistemological issues, or one aiming more at social matters), and how it should concretely affect the activities of the HoP section. The answers will very much depend on the future section leader who will take over my duties after the next General Assembly. In meanwhile let me take the opportunity of this interest in philosophy to deal in this article with a too often overlooked matter, one however that at second sight is highly non trivial and definitely philosophical. I want to deal with the notion of "physical state".

Now, when studying physics, one rarely meets an explicit definition of what a state is. In many ways, the concept of state of a physical system is very close to what in mathematics is called a "primitive notion". Such a notion is not defined with the help of an explicit definition but receives its meaning from the way it is put into work in postulates (because of this, the latter are often considered as being just implicit definitions) and propositions making a given theory. We take the notion of state as granted, intuitively clear, but cause of this, the latter are often considered as being just implicit definitions) and propositions making a given theory.

Let me then consider the history of quantum mechanics and the notion of state. Bohr's quantum theory of the Hydrogen atom. As is well known, Bohr proposed in 1913 to quantize the orbits available to the electron in terms of an integer quantum number \( n \) which enabled him to derive the wave lengths of the Balmer series of spectral lines. 2 To cope with the problem of electromagnetic energy radiated by the orbiting electron and leading to the instability and eventual destruction of the atom, Bohr flatly postulated that such a process was not taking place as long as the electron was on one of the allowed orbits. The notion of stationary state emerged this way. Now, at the time, the stationary states are conceptually not yet the quantum states we are used to. They are rather stationary processes with paradoxical mechanical and electromagnetic properties. This is also true of generalisations of Bohr's quantization to more complex orbits or complex atoms (the old quantum theory). One will have to wait for the rise of a genuine quantum mechanics to see progress on the road to quantum states.

And here comes a surprise. The advent of matrix mechanics in the second half of 1925 provided the first form of genuine quantum mechanics and a newcomer to the history of quantum theory could at this point expect that the notion of quantum state came at the same time. This is not the case, quite the contrary. Matrix mechanics according to its founding fathers, Heisenberg, Born and Jordan, is a formalism where each physical observable is associated to an array of coefficients. It was Born who recognized, shortly after Heisenberg's breakthrough, that these arrays were obeying the algebraic rules of matrices, starting a process of systematization of Heisenberg's initial results which ended up in a joint paper together with his assistant Jordan and Heisenberg entitled "On quantum mechanics II". 3 The matrix

1 Another 20th century major philosopher of science, Imre Lakatos, made the same declaration. See N. R. Hanson, "The Irrelevance of History of Science to Philosophy of Science", The Journal of Philosophy, vol. 59 (1962), 574–86; also Imre Lakatos, "History of Science and Its Rational Reconstructions", in Boston Studies in the Philosophy of Science, vol. viii, pp. 91–108. Engrossing discussions are to be found in Mauskopf, Seymour & Schmaltz, Tad (Eds.). Integrating History and Philosophy of Science: Problems and Prospects, Boston Studies in the Philosophy and History of Science, 2012, Springer.

2 Heisenberg was putting the observables (frequencies and intensities of spectral lines) at the heart of his new matrix mechanics. The exchange occurred during Heisenberg's lecture in Berlin in 1926.

3 See my articles in the SPG Mitteilungen Nr. 40 (July 2013) p. 52. and Nr. 41 (November 2013), p. 28, for details.

mechanics formalism explained in the paper and the rules to put it to work are deceptively similar to what any student of present day quantum mechanics learns. Indeed, any observable $A$ corresponds to a matrix $A$ the coefficients of which carry physical information (transition amplitudes). The dynamical equation is

$$\frac{dA}{dt} = i[H,A]$$

and it can be trivially solved provided the Hamiltonian matrix $H$ is in diagonal form. One finds then on the diagonal the energies of the stationary states and the time dependence of the matrix elements of $A$.

I said that the similarity of Heisenberg's matrix mechanics with present day quantum mechanics was deceptive. Indeed, there is a major difference of our views with the early understanding of the pioneers of matrix mechanics on what, mathematically speaking, the matrices really are. They recognized the arrays of numbers as matrices only insofar the latter obeyed the algebraic rules of the kind. But these arrays were, in the founding fathers minds, totally deprived of any status of operators acting in a linear space, mapping vectors on vectors. Accordingly, there are no states in matrix mechanics: the matrices are not "acting" on anything, there are no columns of coefficients which could be considered as the mathematical representatives of a quantum state. How is this so? One has to realize that the so-called "intrinsic view" on linear algebra objects such as matrices, column vectors, etc. was then quite exotic, even among professional mathematicians. The view that these are merely coordinate expressions of intrinsic objects in a linear space was rare and certainly almost unknown among physicists which had already a hard time just to cope with the algebraic manipulations of (infinite) matrices. We then understand where does the so-called "Heisenberg picture" come from. Only that in the early months of matrix mechanics, these "representation" did not result from a choice to make observables evolve in time instead of states, because there was no choice as there were simply no states in the theory.

Was Schrödinger's wave mechanics more successful at identifying/representing states? As is well known, shortly after the publishing of the joint paper by Born, Jordan and Heisenberg, Erwin Schrödinger released a tetralogy of papers explaining an alternative quantum formalism based on the use of partial differential equations. He introduced first the stationary equation and then the time dependent one, solving which lead him to the formulation of Hamiltonian mechanics. Without going into details, let me explain the problem in following terms.

The quantization procedure consisted in replacing the intrinsic meaning of its solutions could be put aside. And, last but not least, neither Schrödinger nor the orthodox people using wave equations were thinking of $\psi$ as something being acted upon (save for differential operators, see below). This remark is important, because with hindsight, we understand that $\psi$ is what Heisenberg matrices, viewed as operators, are acting upon. But this understanding was reached only at the end of 1926, after months of a deep conceptual effort prompted by the need to understand the equivalence of matrix and wave mechanics in spite of their radically different mathematical apparatuses. Before I elaborate on this, let me sum up the situation as of early 1926. There are no states in matrix mechanics, but one has matrices albeit not yet interpreted as operators. There are wave functions in wave mechanics but they are not acted upon save for differential operators associated to the Hamiltonian, and one does not think of them as candidates for (vector) variables in matrix operations.

I was slightly inexact when I stated that the eigenfunctions of the wave equation were of no direct use. Indeed, as early as spring 1926, Schrödinger explicitly constructed Heisenberg's matrices out of the solutions to his wave equation. He showed this by computing what we today call the matrix elements of the differential operator corresponding to the observable. The reader should be careful here: in spite of a procedure which is totally coherent with our contemporary view on what quantum mechanics is, Schrödinger's differential operators were not thought as a different "representation" of observables (i.e. different from their matrix form). Moreover, Schrödinger phrased his construction so as to suggest that his wave mechanics was more fundamental than Heisenberg's (the rivalry between wave and matrix formalisms was then rather fierce!). Starting with Schrödinger's proof of the problem of clarifying the equivalence of wave and matrix mechanics using so different formalisms became important. But this was not the only pressing theoretical problem. A very important and crucial one for the development of the theory was to understand the analogue of classical canonical transformations as this was the key to express the same quantum problem in different sets of coordinates. Without going into details, let me explain the problem in following terms. Consider a system whose dynamics are expressed classically with the help of a Hamiltonian $H(p,q)$, where the $q$ are the generalized coordinates and the momenta. The quantization procedure consisted in replacing

5 This analogy has a long history going back to the investigations of Hamilton on the similarities between Fermat and Least Action principles which lead him to the formulation of Hamiltonian mechanics.

6 First of all, the wave function is defined over the configuration space of the system and not over 3d-space (save for the one particle case).
q, and p, with matrices obeying canonical commutation relations \( q \rightarrow q \), \( p \rightarrow p \), \( \hat{p} \hat{q} - \hat{q} \hat{p} = -i\hbar \delta \). Then, as already stated, the dynamical problem was solved once one could diagonalize the resulting Hamiltonian matrix \( \hat{H}(\hat{p}, \hat{q}) \). Now, classically, one knows how to describe the same system with different choices of coordinates: one uses a canonical transformation which, for a change of coordinates

\[ q \rightarrow \mathcal{Q}(\rho, \kappa); \rho \rightarrow \mathcal{P}(\rho, \kappa), \]

enables to obtain a new Hamiltonian \( \mathcal{K}(\mathcal{P}, \mathcal{Q}) \). Now, a haunting question in the first months of quantum mechanics was the following: if one goes through the quantization procedure for both Hamiltonians \( \mathcal{H} \) and \( \mathcal{K} \), and then computes the spectra of the resulting operators, the eigenvalues better be the same (we are describing the same quantum system). But what secures mathematically this equality? Are there conditions on the classical canonically related Hamiltonians so that their quantized forms have same spectra of eigenvalues? A kind of quantum analogue of a classical canonical transformation was needed.

I have no place here to go into the details of the various, sometimes surprising answers this question rose 7, so let me just explain what this questioning ended in. In a beautiful paper "Winkelvariable und kanonische Transformationen in Undulationsmechanik" 8, Fritz London was the first to reach the remarkable conclusion that Schrödinger wave functions were the objects on which acted (loosely speaking) the matrices of Heisenberg. London's insight was crucial as it triggered the understanding that quantum mechanics involved matrices of Heisenberg. London's insight was crucial as it triggered the understanding that quantum mechanics involved some operator calculus in a linear space. This made London's insight was crucial as it triggered the understanding that quantum mechanics involved some operator calculus in a linear space. This made London the pioneer of what was soon called "transformation theory", some operator calculus in a linear space. This made London the pioneer of what was soon called "transformation theory", as developed by Jordan and especially by Dirac by the end of 1926 8. Basically, this is already quantum mechanics as we learn it today, where the wave function is understood roughly as providing the (continuous) state coordinates corresponding to the (Hilbert space) "basis" of position eigenfunctions which the following identity clearly shows

\[ \psi(x) = \int \psi(y) \delta(y-x) \, dy \]

Because of its use of delta-functions and their derivatives, this formulation is mathematically ill-defined 10, but it was good enough for physicists' needs and made the success of Dirac's bra-ket formalism. But the concept of quantum state was not yet sufficiently (I mean physically) clarified.

Just a couple of months after Dirac-Jordan transformation theory Johann von Neumann revisited the topic and explained the equivalence between matrix and wave mechanics with full mathematical rigor. To do so, he defined the concept of abstract Hilbert space and showed that the deep reason of the equivalence between matrix and wave mechanics was that the latter were just two isomorphic realizations of the same abstract structure. He explained all this in very clear terms in the introduction to his celebrated book, *Mathematische Grundlagen der Quantenmechanik* published in 1932.

The transition to von Neumann's contribution to quantum mechanics is purposeful. In his book, von Neumann not only developed a full-fledged theory of self-adjoint unbounded operators in a Hilbert space which is exactly what is needed to deal with the Hamiltonian operators and the associated wave equation (spectral theory), but he also ventured into some considerations on the consistency and completeness of quantum mechanics and on the measurement process. In chapter four, one finds in particular von Neumann's celebrated and hotly debated "proof" of the non-existence of hidden variables. The nature of the proof, its meaning and its failure are extensively discussed in the historical and philosophical literature and need not be considered here in full detail 11. But, in this context, von Neumann proposed an illuminating analysis of how one should understand states and we shall now focus on this.

Chapter four of von Neumans's book considers the problem of the probabilistic nature of the predictions obtainable from quantum mechanics. Von Neumann assumed first the formalism together with the usual rules for using operators and vectors in some Hilbert space. In this context he generalized Born's probability-of-presence interpretation of the wave function for operators \( R_1, ..., R_i \) associated to a given set of "quantities" (observables) 12 (p. 200-1):

**Proposition 1 (P) 13:**

The probability that in a state \( \phi \) the quantities with the operators \( R_1, ..., R_i \) take on values from the respective intervals \( I_1, ..., I_i \) is

\[ \| E_i(I_1) E_i(I_2) ... E_i(I_i) \phi \| \]

where \( E_i(I_1), ..., E_i(I_i) \) are the resolutions of the identity belonging to \( R_1, ..., R_i \) respectively.

The spectral "resolution of the identity" for \( R_i \) yields operators \( E_i(I) \) which are roughly projection operators on the subspace corresponding to eigenvalues of \( R_i \) in the interval \( I_i \). This result is valid only in the case where the quantities \( R_1, ..., R_i \) are commuting, so that the product \( E_i(I_1) E_i(I_2) ... E_i(I_i) \) is a projection and the probability becomes then:

\[ \| E_i(I_1) E_i(I_2) ... E_i(I_i) \phi \| \]

where the bracket \( \langle ..., ... \rangle \) is the inner product in the Hilbert space. Von Neumann showed next that the statement \( \mathbf{P} \) can be obtained from two assumptions of simpler form. The first is (p. 203):

**Proposition 2 (E2):**

For the expectation value of [a physical quantity] \( \mathcal{R} \), [corresponding to an operator \( \mathcal{R} \)] in the state \( \phi \); we have

\[ \text{Exp}(\mathcal{R}; \phi) = \langle \mathcal{R} \phi, \phi \rangle. \]

7 For details, see my paper "The puzzle of canonical transformations in early quantum mechanics", *Studies in the History and Philosophy of Modern Physics*, vol. 35 (2004), pp. 317-344. Today we understand that this analogue is a unitary transformation in the Hilbert state space.


9 For context and details, see my paper "The puzzle..." cited in 7.

10 Unless one develops a full-fledged theory of delta-functions and derivatives in terms of distributions.


12 I am quoting from the English translation of his book published in 1955 with the title *Mathematical Foundations of Quantum Mechanics*. I have changed sometimes von Neumann's notations. Also, I tend to use loosely "observable" for von Neumann's "quantity".

13 The letter in parantheses corresponds to von Neumann's original labeling. See sections II.5-II.9 of von Neumann's book.
and (p. 206)

**Proposition 3 (F)**

If the quantity $R$ has the operator $R$, then the quantity $f(R)$ must have the operator $f(R)$ [where $f$ is an arbitrary function].

According to von Neumann, these two assumptions can be considered "as the foundations on which the entire theory is built". They make manifest the departure with respect to classical determinism, because, as von Neumann points out, "$P$ is statistical, and not causal, i.e., it does not tell us what values $R_1, R_2, ..., R_N$ have in the state $\phi$, but only with what probability they take on all possible values; under the proviso that the observables are mutually compatible: "The problem of $P$ cannot be answered for arbitrary quantities $R_1, R_2, ..., R_N$ but only for those whose operators $A_1, A_2, ..., A_N$ commute with one another" (p. 206).

Comparing classical determinism against the probabilistic character of $P$, von Neumann introduced then the hidden variables hypothesis (p. 209):

If we want to explain the non-causal character of the connection between $\phi$ and the values of the physical quantities following the pattern of classical mechanics, then this interpretation is clearly the proper one: In reality, $\phi$ does not determine the state exactly. In order to know this state absolutely, additional numerical data are necessary. That is, the system has other characteristics or coordinates in addition to $\phi$. If we were to know all of these, then we could give the values of all physical quantities exactly and with certainty. On the other hand, with the use of $\phi$ alone, just as in classical mechanics when only some of the $q_1, q_2, ..., q_n$, $p_1, p_2, ..., p_n$ are known, only statistical pronouncements are possible. [...].

It is customary to call these hypothetical additional coordinates "hidden parameters". [...].

We shall show later [...] that an introduction of hidden parameters is certainly not possible without a basic change in the present theory.

Let us pause for a moment. As we see, the physical concept of state in the lines above is primitive and only its representation in terms of a vector in the Hilbert space is given. But in the sequel, von Neumann does provide an explicit physical characterization of what states are.

To do so, he considers, after R. von Mises [14], the so-called "collectives" or "ensembles" composed of many identical systems $S$ on which one will measure various quantities $R$. On such ensembles $S = \{S_1, S_2, ..., S_N\}$ (p. 298),

we do not measure the "value" of a quantity $R$ but its distribution of values: i.e., for each interval $a' < a < a''$ ($a', a''$ given, $a' < a''$) the number of those among the $S_1, ..., S_N$ for which the value of $R$ lies in the interval – dividing this number by $N$ we obtain the probability function $w(a', a'') = w(a'') - w(a')$.

Now (p. 301):

For such ensembles, it is not surprising that a physical quantity $R$ does not have a sharp value, i.e., that its distribution function does not consist of a single value $a_+$ but that several values or intervals of values are possible, and that a positive dispersion exists.

The lack of a sharp value (i.e. the presence of dispersion) can a priori be traced to the two following reasons. According to the first, the dispersion is the result of the inhomogeneity of the ensemble, i.e.

"The individual systems $S_1, ..., S_N$ of our ensemble [are] in different states [my emphasis] so that the ensemble $\{S_1, ..., S_N\}$ is defined by their relative frequencies. [...] we do not know in which state we are measuring, and therefore we cannot predict the results".

The other possibility is that

"All individual systems $S_1, ..., S_N$ are in the same state, but the laws of nature are not causal. Then the cause of the dispersions is [...] nature itself, which has disregard ed the 'principle of sufficient cause'" (p. 302).

We have reached the desired characterization of states. Under the assumption of determinism, systems reacting in an identical way to all possible measurements are by definition declared to be in the same state. If there is dispersion for a given observable, it has to be necessarily attributed the an inhomogeneity of the ensemble. When causality is on the other hand, not warranted, different responses to measurements may still be compatible with the identity of the states making the ensemble. Those rejecting this possibility on the ground that a different reaction to the same measurement denotes by definition a difference in measured states must then assume, in the case of quantum mechanics, that the wave function description is not complete and that "other variables must exist, 'the hidden parameters' [...]" (p. 303).

Von Neumann's characterization of states lets one conceive a practical procedure where more and more homogeneous sub-ensembles are obtained using measurements as sieves (filters). But in quantum situations there is a fundamental obstacle because one measurement can undo what the preceding has enabled to achieve: "A possibility [...] of resolving [the ensemble] into several differently constituted parts (according to the various states of its elements) [with respect] to the various values which $R$ has in the ensemble" (p. 303) cannot be realized because it necessitates that the partial stages of this resolving procedure have to be preserved under further manipulation and this is not warranted since the peculiarities of the quantum measurements are such that any measurement changes (disturbs) the measured system. Thus (p. 305),

we have no method which would make it always possible to resolve further the dispersing ensembles (without a change of their elements) or to penetrate to those homogeneous ensembles which no longer have dispersion.

This situation is for instance well illustrated in the case of the values of the spin projections on orthogonal axes using successive measurements with the help of Stern-Gerlach apparatuses.

**But one can at this point consider a theoretical version of sieves.** Indeed, argues von Neumann, there remains the (purely) theoretical possibility where one could assume, if

---

only in principle, that it is possible to represent each ensemble in which there is a quantity \( R \) with dispersion by the mixture of two (or more) ensembles different from one another and from it. In terms of the distribution function \( w(a) \) this reads (p. 306):

If [the ensemble] \( \{ S_1, \ldots, S_n \} \) were the mixture (sum) of \( \{ S_1', \ldots, S_m' \} \) and \( \{ S_1'', \ldots, S_p'' \} \), the probability function \( w_{a}(a) \) for each quantity \( R \) could be expressed with the aid of the probability functions \( w_{a}^{*}(a), w_{a}^{**}(a) \) of the two subensembles,

\[
\begin{align*}
    w_{a}(a) &= \alpha w_{a}^{*}(a) + \beta w_{a}^{**}(a), \\
    \alpha > 0, \beta > 0, \alpha + \beta &= 1.
\end{align*}
\]

where \( \alpha = P/N; \beta = Q/N; (N = P + Q) \) are independent of \( R \).

This gives rise to the following theoretical problem (p. 306):

If in an ensemble with the probability functions \( w_{a}^{*}(a) \) there exist quantities \( R \) with dispersion \( \{ \ldots \} \) are there two other ensembles with the probability functions \( w_{a}^{*}(a) \) and \( w_{a}^{**}(a) \) respectively, such that for all \( R \), [the equality] (M1) holds?

Even if we cannot practically resolve ensembles with dispersion into homogeneous subensembles to identify states as what the systems in such subensembles have in common, we can do so provided we have a theory. Indeed, we can study the above question for probability functions \( w_{a}^{*}(a) \). To do so, we have first to identify what such functions mathematically are, we have to define the class of functions which can be considered as \( w_{a}(a) \) functions for some ensemble: this is where a theory yielding information on measurement results on the systems considered is needed.

In the case when the theory provides us from the start with a definition of what the mathematical representatives of states are such a construction is superfluous, but think of a theory where only the information on measurements is provided, a theory which does not deal a priori with states (such is the case of matrix mechanics!). In such a situation we shall nonetheless be able to define what states are in terms of a correspondence with \( w_{a} \)-functions which cannot be written further as linear combinations (M1). Now, this is just what I alluded to in the beginning: the theory decides what states are!

Let us see how it works in von Neumann’s case. It turns out that it is mathematically more convenient to study, instead of probability functions \( w_{a}(a) \), the associated expectation value function \( \text{Exp} \) defined over quantities \( R \):

\[ \text{Exp}(R) = \sum (w_{a}(a) - w_{a}(a-\cdot)) a, \]

Reciprocally, it is easy to see that the knowledge of \( \text{Exp} \) for all observables enables to reconstruct all the probability functions \( w_{a}(a) \). Hence (p. 307):

To each ensemble there belongs one such \( \text{Exp} \) function which is defined for all physical quantities \( R \) [...] and which takes on real numbers as values, and which, conversely, completely characterizes the ensemble in all its statistical properties. [...] Of course we must still find out which properties an \( R \)-function must possess in order to be the \( \text{Exp}(R) \) of a suitable ensemble.

\[ \text{Exp} \] functions provide an equivalent means to study the question of the homogeneity of the ensembles (p. 306):

An ensemble is dispersion free if in it, for each \( R \),

\[ \text{Exp}((R - \text{Exp}(R))^2) = \text{Exp}(R^2) - (\text{Exp}(R))^2 \]

equal to zero, i.e.,

\[ \text{Exp}(R^2) = (\text{Exp}(R))^2 \]

If this is not the case, is it always possible to find other ensembles with \( \text{Exp}(R) \), \( \text{Exp}^*(R) \); \( \text{Exp}(R) \neq \text{Exp}^*(R) \), such that

\[
\begin{align*}
    \text{Exp}(R) &= \alpha \text{Exp}^*(R) + \beta \text{Exp}^*(R), \\
    \alpha > 0, \beta > 0, \alpha + \beta &= 1
\end{align*}
\]

always holds \((\alpha, \beta\)-independent of \( R \))?

In order to characterize further the \( \text{Exp} \) functions, von Neumann postulated after much considerations (see p. 308) the following requirements \( ^{16} \):

1. For a finite or infinite number of quantities \( a, b, c, \ldots \), and real numbers \( \alpha, \beta, \gamma, \ldots \) (the sum of which converges) one has

\[ \text{Exp}(\alpha a + \beta b + \gamma c + ...) = \alpha \text{Exp}(a) + \beta \text{Exp}(b) + \gamma \text{Exp}(c) + \ldots \]

2. For a quantity \( a \) which takes no negative values, one has

\[ \text{Exp}(a) \geq 0 \]

Von Neumann reexpressed as well the conditions of no dispersion and homogeneity (p. 312):

- An \( R \)-function which is an \( \text{Exp}(R) \) is said to be dispersion-free if \( \text{Exp}(1) \neq 0 \) and is finite, so that we can assume \( \text{Exp}(1) = 1 \); then \( \text{Exp}(R^2) = [\text{Exp}(R)]^2 \) is characteristic.
- An \( R \)-function which is an \( \text{Exp}(R) \) is said to be homogeneous or pure if, when \( \text{Exp}(R) = \text{Exp}^*(R) + \text{Exp}^*(R) \), one has

\[ \text{Exp}^*(R) = c' \text{Exp}(R), \text{Exp}^*(R) = c' \text{Exp}(R) \]

as a consequence (where \( c', c'' \) are constants, \( c' + c'' = 1 \) and \( c' > 0, c'' > 0 \)).

The definitions above are yet general. In order to consider the statistics of quantities within a specific theory, one must "know" the physical quantities [...] as well as the functional relations existing among them". In the quantum case one assumes that each physical quantity of a system defines a unique Hermitian operator. Von Neumann assumed furthermore that this correspondence is one-to-one, and that the following rules hold:

I: Let \( A, B, \ldots \) (a finite or infinite number of them) and \( \alpha A + \beta B + \ldots \) be [Hermitian] operators. If \( A, B, \ldots \) represent the quantities \( A, B, \ldots \), then \( \alpha A + \beta B + \ldots \) represents the quantity \( \alpha A + \beta B + \ldots \).

II: For a [Hermitian] operator \( A \) and a real-valued function \( f(x) \) defined for all real values \( x \), if \( A \) represents the quantity \( A \), then \( f(A) \) represents the quantity \( f(A) \).

15 That a physical quantity \( R \) has a sharp value \( a \) corresponds to the probability function \( w_{a}(a) \) with

\[ \begin{cases} w_{a}(a) = 1, \text{for } a \geq a, \\ w_{a}(a) = 0, \text{for } a < a, \end{cases} \]

footnote p. 301.

16 These postulates were obtained already in an earlier paper which contains much of the material used in chapter IV of the book.
Von Neumann considered that these two conditions are quite plausible and need no further discussion. He could then readily show that (for any quantity $R$) one has:  
$$\text{Exp}(R) = \text{Tr}(UR),$$
where the Hermitian operator $U$ does not depend on $R$ and is entirely determined by the function Exp, i.e. the ensemble one considers. The functions Exp associated to the various ensembles, and hence the ensembles themselves, can then be classified according to the corresponding $U$ operators. Von Neumann called those operators "statistical" $^{24}$. Now, the states correspond (one-to-one) to ensembles which are homogeneous, ensembles which cannot be further decomposed as mixtures of subensembles. Which statistical operators (which ensembles) correspond to such cases? It turns out that they correspond to statistical operators $U = P_r$, with $P_r$ a projection operator on the subspace generated by $\phi$, a (unit norm) vector in the Hilbert space. We have recovered and justified in this way the basic postulate of the quantum mechanics formalism: states are described by vectors of the Hilbert space.

What is odd about quantum mechanics is that even in the case of homogeneous ensembles (all the members of the ensemble are in the same "state"), there is dispersion at least for one observable. Von Neumann showed actually that there are no dispersion-free non-trivial ensembles: the condition of non-dispersion for any $R$,  
$$\text{Exp}(R) = \text{Tr}(UR) = \text{Exp}(R)^2 = \text{Tr}(UR)^2$$
has as sole solutions $U = 0$ and $U = 1$, but the first corresponds to a trivial ensemble and the second to a non-normalizable case and anyway leading to dispersion. This is why von Neumann concluded (p. 323):

Hence, within the limits of our conditions, the decision is made and it is against causality; because all ensembles have dispersions, even the homogeneous [...] It should be noted that we need not go any further into the mechanism of the "hidden parameters", since we now know that the established results of quantum mechanics can never be re-derived with their help. In fact, we have even ascertained that it is impossible that the same physical quantities exist with the same function connections, if other variables (i.e., "hidden parameters") should exist in addition to the wave functions. Nor would it help if there existed other, as yet undiscovered physical quantities, in addition to those represented by the operators in quantum mechanics, because the relations assumed by quantum mechanics (i.e., I, II) would have to fail already for the by now known quantities, those that we discussed above. It is therefore not, as is often assumed, a question of a re-interpretation of quantum mechanics, – the present system of quantum mechanics would have to be objectively false, in order that another description of the elementary processes than the statistical one be possible.

For completeness, let's point out that however "natural" von Neumann's choice of postulates for $Exp$ functions might appear at first sight, it proved retrospectively "Hilbert space theory laden" in the sense that states other than the usual quantum ones represented by Hilbert space rays were ruled out ab initio. Soon after the publication of von Neumann's book and since then various authors expressed recurrently their worry about a possible circularity in von Neumann's proof $^{18}$. The situation was exposed very clearly in a paper of John Bell $^{19}$ where the latter observed that condition  
$$\text{Exp}(R + S) = \text{Exp}(R) + \text{Exp}(S)$$
is ruling out dispersion-free states because, if one assumes that the sole values obtained during measurement can only be eigenvalues of operators, so that (over dispersion free states) $\text{Exp}(R + S)$, $\text{Exp}(R)$, $\text{Exp}(S)$ correspond to eigenvalues, then the condition above claims that the eigenvalues of a sum $R + S$ are the sum of eigenvalues of $R$ and $S$ which is mathematically false. However (remarkably!), on a standard quantum mechanical state (Hilbert space vector $\varphi$), the condition is true because of the linearity of the inner product:  
$$\text{Exp}(R + S) = \langle (R + S)\varphi, \varphi \rangle = \langle R\varphi, \varphi \rangle + \langle S\varphi, \varphi \rangle$$
\[= \text{Exp}(R) + \text{Exp}(S)\]
Von Neumann's proof triggered a considerable debate, especially between those trying to improve the original proof and those who considered it deeply irrelevant. The tenants of the orthodox interpretation of quantum mechanics defending that the theory yields a complete description of physical reality were happy enough with it or its avatars. Others, on the contrary, were reassured to learn that von Neumann result was not ruling out hidden variables in absolute terms $^{20}$. Their views went vindicated when David Bohm proposed (1952) his celebrated re-interpretation of quantum mechanics in terms of hidden variables. Be it as it may, it is plain that beyond this debate, von Neumann's proof was instrumental in triggering a new stage of the axiomatization of quantum mechanics. Indeed, the necessity to understand properly the proof's shortcomings and the possible ways to improve its scope is at the roots of the majority of further developments going well into our present time $^{21}$. All this goes much beyond the primary aim of my paper. I wanted to show how the concept of state needs much clarification before one can consider it sound and precise, and how this implies to put to work the physical theory pertaining to systems one wishes to describe. States do not come a priori, they are not directly measurable, only observables are, and one needs a theory to single them out.

18 For an account see Jammer 1974 quoted above, and the more recent one by James Cushing, Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony, Chicago, 1994


20 A close look at von Neumann's proof shows that what he actually proved was a kind of incompatibility of hidden variables with the Hilbert space formalism of quantum mechanics; the last quote above makes this clear.

21 It is worth quoting the work of Jauch and Piron who developed (1963) an algebraic, lattice-theoretical, approach to the axiomatization of quantum mechanics with the help of which they claimed to obtain a circularity-free no-hidden variables proof. This was the beginning of the so-called "Geneva School" in quantum mechanics.