(Never) Mind your p's and q's: Von Neumann versus Jordan on the Foundations of Quantum Theory.

Anthony Duncan [a], Michel Janssen [b,*]

[a] *Department of Physics and Astronomy, University of Pittsburgh*

[b] *Program in the History of Science, Technology, and Medicine, University of Minnesota*

Abstract

In early 1927, Pascual Jordan (1927b) published his version of what came to be known as the Dirac-Jordan statistical transformation theory. Later that year and partly in response to Jordan, John von Neumann (1927a) published the modern Hilbert space formalism of quantum mechanics. Central to both formalisms are expressions for conditional probabilities of finding some value for one quantity given the value of another. Beyond that Jordan and von Neumann had very different views about the appropriate formulation of problems in the new theory. For Jordan, unable to let go of the analogy to classical mechanics, the solution of such problems required the identification of sets of canonically conjugate variables, i.e., p's and q's. Jordan (1927e) ran into serious difficulties when he tried to extend his approach from quantities with fully continuous spectra to those with wholly or partly discrete spectra. For von Neumann, not constrained by the analogy to classical physics and aware of the daunting mathematical difficulties facing the approach of Jordan (and, for that matter, Dirac (1927)), the solution of a problem in the new quantum mechanics required only the identification of a maximal set of commuting operators with simultaneous eigenstates. He had no need for p's and q's. Related to their disagreement about the appropriate general formalism for the new theory, Jordan and von Neumann stated the characteristic new rules for probabilities in quantum mechanics somewhat differently. Jordan (1927b) was the first to state those rules in full generality, von Neumann (1927a) rephrased them and then sought to derive them from more basic considerations (von Neumann, 1927b). In this paper we reconstruct the central arguments of these 1927 papers by Jordan and von Neumann and of a paper on Jordan's approach by Hilbert, von Neumann, and Nordheim (1928). We highlight those elements in these papers that bring out the gradual loosening of the ties between the new quantum formalism and classical mechanics.

Key words: Pascual Jordan, John von Neumann, transformation theory, probability amplitudes, canonical transformations, Hilbert space, spectral theorem

1 Introduction

In 1954, Max Born was awarded part of the Nobel Prize in physics for "his fundamental research in quantum mechanics, especially for his statistical interpretation of the wave function." In analogy to Albert Einstein's proposal to interpret the electromagnetic field as a "ghost field" (*Gespensterfeld*) or "guiding field" (*Führungsfeld*) for light quanta, Born (1926b, p. 804), in a paper on the quantum mechanics of collision processes, proposed to interpret the de-Broglie-Schrödinger matter waves as a ghost or guiding field for electrons. Concretely, Born (1926b, p. 805) suggested that, given a large number of systems in a superposition $\psi(q) = \sum_n c_n \psi_n(q)$ of energy eigenstates $\psi_n(q)$, the fraction of systems in an eigenstate $\psi_n(q)$ is given by the absolute square $|c_n|^2$ of the complex expansion coefficients c_n.[1] In a preliminary announcement of the results of the paper, Born (1926a, p. 865) gave the example of the scattering of an electron by an atom. After the interaction of electron and atom, he noted, the system will be in a superposition of states with the electron flying off in different directions. He interpreted the absolute square of the expansion coefficients as the probability that the electron flies off in a particular direction.[2]

Referring to Born's notion of a ghost field, Wolfgang Pauli (1927a, p. 83), in a footnote in a paper on gas degeneracy and paramagnetism, proposed that, given a system of f degrees of freedom in an energy eigenstate $\psi(q)$ $(q \equiv (q_1, \ldots, q_f))$, the probability of finding that system at a position between q and $q + dq$ is given by $|\psi(q)|^2 dq$. By the time this paper appeared in print, Pascual Jordan (1927b), citing Born's two 1926 papers and Pauli's forthcoming 1927 paper, had already published a far-reaching generalization of Pauli's proposal.[3] Jordan did so in a paper entitled "On a new founda-

⋆ This paper was written as part of a joint project in the history of quantum physics of the *Max Planck Institut für Wissenschaftsgeschichte* and the *Fritz-Haber-Institut* in Berlin.

* Corresponding author. Address: Tate Laboratory of Physics, 116 Church St. SE, Minneapolis, MN 55455, USA, Email: janss011@umn.edu

[1] It is tempting to read 'fraction of systems in an eigenstate' as 'fraction of systems *found* in an eigenstate *upon measurement of the energy*', but Born did not explicitly say this. A careful distinction between pure and mixed states had yet to be made (von Neumann, 1927b, see Section 6).

[2] Born famously only added in a footnote that this probability is not given by these coefficients themselves but by their absolute square.

[3] In a semi-popular article in *Die Naturwissenschaften*, "The development of the new quantum mechanics," Jordan (1927f, Part 2, p. 647) emphasized that the statistical interpretations of Born and Pauli "related but nevertheless initially independent of one another." Although his own probabilistic interpretation owes more to Pauli than to Born, Jordan gave full credit to his Ph.D. advisor Born for the

tion (*neue Begründung*) of quantum mechanics" submitted to *Zeitschrift für Physik* in December 1926 (Jordan, 1927b). We will refer to this paper as *Neue Begründung* I to distinguish it from *Neue Begründung* II, submitted in June 1927 to the same journal, a sequel in which Jordan tried both to simplify and to generalize his theory (Jordan, 1927e).[4] The theory presented in these *Neue Begründung* papers is Jordan's version of what came to be known as transformation theory, or, in older literature, as the Dirac-Jordan (statistical) transformation theory. Independently of Jordan, Paul Dirac (1927) published very similar ideas, which also form the basis of his later book (Dirac, 1930). Dirac's paper is much clearer than Jordan's and uses a vastly superior notation, but Jordan actually brought out the statistical foundation of quantum theory much more explicitly than Dirac.

The central quantities in Jordan's formalism are what he called, following a suggestion by Pauli and in analogy with wave amplitudes, "probability amplitudes" (Jordan, 1927b, p. 811). In *Neue Begründung* I, Jordan restricted himself to quantities with completely continuous spectra, clearly laboring under the illusion that it would be relatively straightforward to generalize his formalism to cover quantities with wholly or partly discrete spectra as well. The probability amplitude $\varphi(a, b)$ between two quantum-mechanical quantities $\hat{a}$ and $\hat{b}$ with fully continuous spectra is a complex function, the absolute square of which, $|\varphi(a, b)|^2$, multiplied by da gives the conditional probability $\Pr(a|b)$ for finding a value between a and $a + da$ for $\hat{a}$ given that the system under consideration has been found to have the value b for the quantity $\hat{b}$.[5]

Jordan did not think of his quantum-mechanical quantities as operators acting in an abstract Hilbert space, but, like Schrödinger, he did associate them with differential operators acting in a function space. An eigenfunction $\psi_E(q)$ for a one-dimensional system with energy eigenvalue E of the time-independent Schrödinger equation is an example of a probability amplitude. The quantities

basic idea of a probabilistic interpretation of the Schrödinger wave function. In his *Habilitationsvortrag*, for instance, Jordan (1927d, 107) praised the "very clear and impressive way" in which Born had introduced this interpretation. In his article in *Die Naturwissenschaften*, Jordan (1927f, Part 2, p. 645) also mentioned the importance of two papers on quantum jumps, one by himself (Jordan, 1927a) and one by Heisenberg (1927a), for the development of the probabilistic interpretation of quantum mechanics.

[4] Jordan also published a preliminary version of *Neue Begründung* I (under the same title) in the Proceedings of the Göttingen Academy (Jordan, 1927c).

[5] This is the first of several instances where we will enhance Jordan's own notation. In *Neue Begründung* I, Jordan used different letters for quantities and their values. We will almost always use the same letter for a quantity and its values and use a hat to distinguish the former from the latter. The main exception will be the Hamiltonian $\hat{H}$ and the energy eigenvalues E. The notation $Pr(a|b)$ is strictly our own and is not used in the sources we discuss.

$\hat{a}$ and $\hat{b}$ in this case are the position $\hat{q}$ and the Hamiltonian $\hat{H}$, respectively. Hence $|\psi_E(q)|^2 dq = |\varphi(q, E)|^2 dq$ gives the conditional probability $\Pr(q|E)$ that $\hat{q}$ has a value between q and $q + dq$ given that $\hat{H}$ has the value E. This is the special case of Jordan's interpretation that Pauli had hit upon (in f dimensions). Jordan also recognized that in quantum mechanics the usual addition and multiplication rules of probability apply to the probability *amplitudes* rather than to the probabilities themselves. Again crediting Pauli—although Born (1926b, p. 804) already talked about the "interference of … "probability waves" "—Jordan (1927b, p. 812) used the phrase "interference of probabilities" for this phenomenon. Jordan was thus the first, at least in print and in full generality, to recognize the peculiar behavior of probabilities in the new quantum mechanics.

Given his proximity to David Hilbert in Göttingen (Jordan had been Richard Courant's assistant before becoming Born's), it should come as no surprise that Jordan took an axiomatic approach in his *Neue Begründung* papers. He first introduced his probability amplitudes and the rules they ought to obey in a series of postulates (the formulation and even the number of these postulates varies) and then developed a formalism realizing these postulates.

A clear description of the task at hand can be found in a paper by Hilbert, Lothar Nordheim, and the other main protagonist of our story, John von Neumann. This paper grew out of Hilbert's course on quantum mechanics in 1926/1927 for which Nordheim prepared most of the notes. [6] The course concluded with an exposition of Jordan's *Neue Begründung* (Sauer and Majer, 2009, pp. 698–706). The notes for this part of the course formed the basis for a paper, which was submitted in April 1927 but not published until 1928. In the introduction, the authors described the strategy for formulating the theory:

> One imposes certain physical requirements on these probabilities, which are suggested by earlier experience and developments, and the satisfaction of which calls for certain relations between the probabilities. Then, secondly, one searches for a simple analytical apparatus in which quantities occur that satisfy these relations exactly (Hilbert, von Neumann, and Nordheim, 1928, p. 2–3; cf. Lacki, 2000, p. 296).

From a modern point of view, the "simple analytical apparatus" in this case is supplied by the Hilbert space formalism. Jordan's probability amplitudes $\varphi(a, b)$ are the 'inner products' $\langle a|b\rangle$ of the 'eigenvectors' $|a\rangle$ and $|b\rangle$ of the operators $\hat{a}$ and $\hat{b}$. [7] The reason we used scare quotes in the preceding sentence

[6] See p. 13 of the transcript of the interview with Nordheim for the Archive of the History of Quantum Physics (AHQP) (Duncan and Janssen, 2009, p. 361).

[7] We will not introduce a special notation to distinguish between a physical quantity and the operator acting in Hilbert space representing that quantity. In most cases it will be clear from context whether $\hat{a}$ stands for a quantity or an operator.

is that for quantities with completely continuous spectra, to which Jordan restricted himself in *Neue Begründung* I, the 'eigenvectors' of the corresponding operators are *not* elements of Hilbert space. That in modern quantum mechanics they are nonetheless routinely treated *as if* they are vectors in Hilbert space with inner products such as $\langle a|b\rangle$ is justified by the spectral theorem for the relevant operators.

Of course, neither the spectral theorem nor the notions of an abstract Hilbert space and of operators acting in it were available when Jordan and Dirac published their respective versions of transformation theory in 1927. The Hilbert space formalism and the spectral theorem were only published later that year, by von Neumann (1927a). Even though Dirac (1927) introduced the notation (a/b) for what Jordan wrote as $\varphi(a, b)$, Dirac, like Jordan, did *not* at that time conceive of these quantities as 'inner products' of two more elementary quantities. We are not sure about Dirac but for Jordan probability amplitudes remained the fundamental quantities (Duncan and Janssen, 2009, p. 361).

Once the 'inner-product' structure of probability amplitudes is recognized and justified with the help of the spectral theorem, Jordan's basic axioms about the addition and multiplication of probability amplitudes are seen to reduce to statements about orthogonality and completeness familiar from elementary quantum mechanics. For example, for quantities $\hat{a}$, $\hat{b}$, and $\hat{c}$ with purely continuous spectra, Jordan's postulates demand that the probability amplitudes $\varphi(a, c)$, $\psi(a, b)$ and $\chi(b, c)$ satisfy the relation $\varphi(a, c) = \int db\, \psi(a, b)\, \chi(b, c)$. With the identification of probability amplitudes with 'inner products' of 'eigenvectors' (appropriately normalized, such that, e.g., $\langle a|a'\rangle = \delta(a - a')$, where $\delta(x)$ is the Dirac delta function), the familiar completeness relation, $\langle a|c\rangle = \int db\, \langle a|b\rangle\langle b|c\rangle$, which holds on account of the spectral decomposition $\hat{b} = \int db\, b\, |b\rangle\langle b|$, guarantees that $\varphi(a, c) = \int db\, \psi(a, b)\, \chi(b, c)$. In this sense, the Hilbert space formalism thus provides a realization of Jordan's postulates.

In the absence of the Hilbert space formalism and the spectral theorem, Jordan relied on the formalism of canonical transformations to develop the analytical apparatus realizing his axiomatic scheme. Canonical transformations had been central to the development of matrix mechanics (Born and Jordan, 1925; Born, Heisenberg, and Jordan, 1926). Prior to *Neue Begründung* I, Jordan (1926a,b) had actually published two important papers on the topic (Lacki, 2004; Duncan and Janssen, 2009). His starting point in *Neue Begründung* I was the assumption that the probability amplitude $\rho(p, q)$, where $\hat{q}$ is some generalized coordinate and $\hat{p}$ its conjugate momentum, is of the simple form $e^{-ipq/\hbar}$. As an aside we note that this is precisely the point at which Planck's constant enters into Jordan's formalism. What the assumption $\rho(p, q) = e^{-ipq/\hbar}$ tells us, as Jordan pointed out, is that "[f]or a given value of $\hat{q}$ all possible values of $\hat{p}$ are *equally probable*" (Jordan, 1927b, p. 814; emphasis in the original;

hats added).[8] In the parlance of modern quantum information theory, this is the statement that $\{|p\rangle\}$ and $\{|q\rangle\}$ are *mutually unbiased bases.* Jordan wrote down two differential equations trivially satisfied by this special probability amplitude. He then considered canonical transformations to other canonically conjugate variables $\hat{P}$ and $\hat{Q}$ and derived differential equations for arbitrary probability amplitudes starting from the equations for $\rho(p, q)$. In this way, he claimed, one could recover both the time-independent and the time-dependent Schrödinger equations as examples of such equations.

Both claims are problematic. The recovery of the time-dependent Schrödinger equation requires that we look upon the time t not as a parameter as we would nowadays but as an operator to be expressed in terms of the operators $\hat{p}$ and $\hat{q}$. More importantly, Jordan's construction only gets us to the time-independent Schrödinger equation for Hamiltonians with a fully continuous spectrum. In *Neue Begründung* I, Jordan deliberately restricted himself to quantities with completely continuous spectra but initially he was confident that his approach could easily be generalized to quantities with wholly or partly discrete spectra as well. He eventually had to accept that this generalization fails. The problem, as he himself recognized in *Neue Begründung* II, is that two quantities $\hat{\alpha}$ and $\hat{p}$ related to each other via a canonical transformation (implemented by the similarity transformation $\hat{\alpha} = T\hat{p}T^{-1}$) always have the same spectrum.[9] Hence, no canonical transformation that can be implemented in this way can take us from quantities such as $\hat{p}$ and $\hat{q}$ with a completely continuous spectrum to a Hamiltonian with a wholly or partly discrete spectrum.

The quantities $\varphi(a, b)$ in Jordan's formalism do double duty as probability amplitudes and as integral kernels of canonical transformations. The quantity $\psi(a, b)$ in the integral expression $\varphi(a, c) = \int db\, \psi(a, b)\, \chi(b, c)$ given above illustrates this dual role. The latter aspect was emphasized by Dirac and gave transformation theory its name. The 'mind your p's and q's' part of the title of our paper refers to the central role of canonical transformations and conjugate variables in Jordan's formalism. Even if we accept the restriction to quantities with fully continuous spectra for the moment, Jordan could not quite get his formalism to work, at least not at the level of generality he had hoped for. In hindsight, we can see that one of the major obstacles was that canonical transformations from one set of conjugate variables to another, although they do preserve the spectra, do *not* always preserve the Hermitian character of the operators associated with these variables in quantum mechanics (Duncan and

[8] As has been pointed out by several commentators, the Dirac-Jordan transformation theory played a key role in Heisenberg's (1927b) formulation of his uncertainty relations later that same year. See Duncan and Janssen (2009, p. 361, note 37) for references to and quotations from the relevant sections of Jammer (1966), Beller (1985, 1999), and Darrigol (1992).

[9] See Eq. (84) at the end of Section 2.3 for a simple proof of this claim in modern language.

Janssen, 2009, secs. 5–6). Jordan initially introduced what he called a "supplementary amplitude" (*Ergänzungsamplitude*) to clear this hurdle but, following the lead of Hilbert, von Neumann, and Nordheim (1928), who put great emphasis on the importance of Hermiticity for the probability interpretation of the formalism, he dropped this notion in *Neue Begründung* II (Jordan, 1927e, pp. 5–6). This meant, however, that he now somewhat artificially had to restrict the canonical transformations he used to ones associated with unitary operators.

In the modern Hilbert space formalism, the integral kernels of canonical transformations in Jordan's formalism are replaced by unitary operators. There is no need anymore for considering canonical transformations nor, for that matter, for sorting quantities into pairs of conjugate variables. Jordan's reliance on canonical transformations and conjugate variables became even more strained in *Neue Begründung* II, when he tried to extend his approach to quantities with partly or wholly discrete spectra. He had a particularly hard time dealing with the purely discrete spectrum of the recently introduced spin observable.

At the end of their exposition of Jordan's theory, written before Jordan wrote *Neue Begründung* II, Hilbert, von Neumann, and Nordheim (1928, p. 30) emphasized the mathematical difficulties with Jordan's approach (some of which they had caught, some of which they too had missed), announced that they might return to these on another occasion, and made a tantalizing reference to the first of three papers on quantum mechanics that von Neumann would publish in 1927 in the Proceedings of the Göttingen Academy (von Neumann, 1927a,b,c). This trilogy formed the basis for his famous book (von Neumann, 1932). The first of these papers, "Mathematical foundations (*Mathematische Begründung*) of quantum mechanics," is the one in which von Neumann introduced the Hilbert space formalism and the spectral theorem. One might therefore expect at this juncture that von Neumann would simply make the point that we made above, namely that the Hilbert space formalism provides the natural implementation of Jordan's axiomatic scheme and that the spectral theorem can be used to address the most glaring mathematical problems with this implementation. Von Neumann, however, did nothing of the sort.

Von Neumann was sharply critical of the Dirac-Jordan transformation theory. As he put it in the introduction of his 1932 book: "Dirac's method does not meet the demands of mathematical rigor in any way—not even when it is reduced in the natural and cheap way to the level that is common in theoretical physics" (von Neumann, 1932, p. 2; our emphasis). He went on to say that "the correct formulation is *not just a matter of making Dirac's method mathematically precise and explicit* but right from the start calls for a different approach related to Hilbert's spectral theory of operators" (ibid., our emphasis). Von Neumann only referred to Dirac in this passage, but as co-author of the paper with Hilbert and Nordheim mentioned above, he was thoroughly familiar with

Jordan's closely related work as well. He also clearly appreciated the difference in emphasis between Dirac and Jordan. Talking about the Schrödinger wave function in the introduction of the second paper of his 1927 trilogy, he wrote: "Dirac interprets it as a row of a certain transformation matrix, Jordan calls it a probability amplitude" (von Neumann, 1927b, p. 246).[10] In the opening paragraph of this article, von Neumann contrasted wave mechanics with "transformation theory" or "statistical theory," once again reflecting the difference in emphasis between Dirac and Jordan. Yet, despite his thorough understanding of it, von Neumann did not care for the Dirac-Jordan approach.

Von Neumann's best-known objection concerns the inevitable use of delta functions in the Dirac-Jordan approach. However, von Neumann also objected to the use of probability amplitudes. Jordan's basic amplitude, $\rho(p, q) = e^{-ipq/\hbar}$, is not in the space L^2 of square-integrable functions that forms one realization of abstract Hilbert space. Moreover, probability amplitudes are only determined up to a phase factor, which von Neumann thought particularly unsatisfactory. "It is true that the probabilities appearing as end results are invariant," he granted in the introduction of his paper, "but it is unsatisfactory and unclear why this detour through the unobservable and non-invariant is necessary" (von Neumann, 1927a, p. 3). So, rather than following the Jordan-Dirac approach and looking for ways to mend its mathematical shortcomings, a highly non-trivial task given that many of the 'vectors' whose 'inner products' give Jordan's probability amplitudes are not elements of Hilbert space, von Neumann, as indicated in the passage from his 1932 book quoted above, adopted an entirely new approach. He generalized Hilbert's spectral theory of operators[11] to provide a formalism for quantum mechanics that is very different from the one proposed by Jordan and Dirac.[12]

The only elements that von Neumann took from the Dirac-Jordan transformation theory were, first, Jordan's fundamental insight that quantum mechanics is ultimately a set of rules for conditional probabilities $\Pr(a|b)$, and second, the fundamental assumption that such probabilities are given by the absolute square of the corresponding probability amplitudes, which essentially boils down to the Born rule. Interestingly, von Neumann (1927a, pp. 43–44) mentioned Pauli, Dirac, and Jordan in this context, but not Born. Von Neumann derived a new expression for conditional probabilities in quantum mechanics that avoids probability amplitudes altogether and instead sets them equal to the trace of products of projection operators, as they are now called. Von

[10] Discussing Jordan's approach in his first paper, von Neumann (1927a) referred to "transformation operators (the integral kernels of which are the "probability amplitudes")" (p. 3).
[11] See Steen (1973) for a brief history of spectral theory.
[12] In his book, von Neumann (1932, p. 1) nonetheless used the term "transformation theory" to describe both his own theory and the theory of Dirac and Jordan.

Neumann used the term *Einzeloperator* or its abbreviation *E. Op.* instead. The probability $\Pr(a|b)$, e.g., is given by $\mathrm{Tr}(\hat{E}(a)\hat{F}(b))$, where $\hat{E}(a)$ and $\hat{F}(b)$ are projection operators onto, in Dirac notation, the 'eigenvectors' $|a\rangle$ and $|b\rangle$ of the operators $\hat{a}$ and $\hat{b}$, respectively. Unlike probability amplitudes, these projection operators do not have any phase ambiguity. This is easily seen in Dirac notation. The projection operator $\hat{E}(a) = |a\rangle\langle a|$ does not change if the ket $|a\rangle$ is replaced by $e^{i\vartheta}|a\rangle$ and the bra $\langle a|$ accordingly by $e^{-i\vartheta}\langle a|$. We should emphasize, however, that, just as Jordan and Dirac with their probability amplitudes/transformation functions $\langle a|b\rangle$, von Neumann did not think of his projection operators as constructed out of bras and kets, thus avoiding the problem that many of them are not in Hilbert space.

Toward the end of his paper, von Neumann (1927a, pp. 46–47) noted that his trace expression for conditional probabilities is invariant under "canonical transformations." What von Neumann called canonical transformations, however, are not Jordan's canonical transformations but simply, in modern terms, unitary transformations. Such transformations automatically preserve Hermiticity and the need for something like Jordan's *Ergänzungsamplitude* simply never arises. Von Neumann noted that his trace expression for conditional probabilities does not change if the projection operators $\hat{E}$ and $\hat{F}$ are replaced by $\hat{U}\hat{E}\hat{U}^\dagger$ and $\hat{U}\hat{F}\hat{U}^\dagger$, where $\hat{U}$ is an arbitrary unitary operator ($\hat{U}^\dagger = \hat{U}^{-1}$). In von Neumann's approach, as becomes particularly clear in his second paper of 1927 (see below), one also does not have to worry about sorting variables into sets of mutually conjugate ones. This then is what the 'never mind your p's and q's' part of the title of our paper refers to. By avoiding conjugate variables and canonical transformations, von Neumann completely steered clear of the problem that ultimately defeated Jordan's approach, namely that canonical transformations can never get us from $\hat{p}$'s and $\hat{q}$'s with fully continuous spectra to quantities with wholly or partly discrete spectra, such as the Hamiltonian.

In *Mathematische Begründung*, von Neumann not only provided an alternative to Jordan's analysis of probabilities in quantum mechanics, he also provided an alternative to the Dirac-Jordan transformation-theory approach to proving the equivalence of matrix mechanics and wave mechanics (von Neumann, 1927a, p. 14). This is where von Neumann put the abstract notion of Hilbert space that he introduced in his paper to good use. He showed that matrix mechanics and wave mechanics correspond to two instantiations of abstract Hilbert space, the space l^2 of square-summable sequences and the space L^2 of square-integrable functions, respectively (Dieudonné, 1981, p. 172). As von Neumann reminded his readers, well-known theorems due to Parseval and Riesz and Fisher had established that l^2 and L^2 are isomorphic.[13]

[13] In 1907–1908, Erhard Schmidt, a student of Hilbert who got his Ph.D. in 1905, fully worked out the theory of l^2 and called it 'Hilbert space' (Steen, 1973, p. 364). In a paper on canonical transformations, Fritz London (1926b, p. 197) used the

In his second 1927 paper, “Probability-theoretical construction (*Wahrscheinlichkeitstheoretischer Aufbau*) of quantum mechanics,” von Neumann (1927b) freed himself even further from relying on the Dirac-Jordan approach. In *Mathematische Begründung* he had accepted the Born rule and recast it in the form of his trace formula. In *Wahrscheinlichkeitstheoretischer Aufbau* he sought to derive this trace formula, and thereby the Born rule, from more fundamental assumptions about probability. Von Neumann started by introducing probabilities in terms of selecting members from large ensembles of systems. He then made two very general and *prima facie* perfectly plausible assumptions about expectation values of quantities defined on such ensembles (von Neumann, 1927b, pp. 246–250). From those assumptions, some assumptions about the repeatability of measurements (von Neumann, 1927b, pp. 271), and key features of his Hilbert space formalism (especially some assumptions about the association of observables with Hermitian operators), von Neumann did indeed manage to recover the Born rule. Admittedly, the assumptions needed for this result are not as innocuous as they look at first sight. They are essentially the same as those that go into von Neumann's infamous no-hidden-variable proof (Bell, 1966; Bacciagaluppi and Crull, 2009).

Along the way von Neumann (1927b, p. 253) introduced what we now call a density operator to characterize the ensemble of systems he considered. He found that the expectation value of an observable represented by some operator $\hat{a}$ in an ensemble characterized by $\hat{\rho}$ is given by $\mathrm{Tr}(\hat{\rho}\,\hat{a})$, where we used the modern notation $\hat{\rho}$ for the density operator (von Neumann used the letter U). This result holds both for what von Neumann (1927b) called a “pure” (*rein*) or “uniform” (*einheitlich*) ensemble (p. 255), one consisting of identical systems in identical states, and for what he called a “mixture” (*Gemisch*) (p. 265). So the result is more general than the Born rule, which obtains only in the former case. Von Neumann went on to show that the density operator for a uniform ensemble is just the projection operator onto the ray in Hilbert space corresponding to the state of all systems in this ensemble. However, he found it unsatisfactory to characterize the state of a physical system by specifying a ray in Hilbert space. “Our knowledge of a system,” von Neumann (1927b, p. 260) wrote, “is never described by the specification of a state … but, as a rule, by the results of experiments performed on the system.” In this spirit, he considered the simultaneous measurement of a maximal set of commuting operators and constructed the density operator for an ensemble where what is known is that the corresponding quantities have values in certain intervals. He showed that such measurements can fully determine the state and that the density operator in that case is once again the projection operator onto the corresponding ray in Hilbert space.

Von Neumann thus arrived at the typical quantum-mechanical way of conceiv-

term ‘Hilbert space’ for L^2 (Duncan and Janssen, 2009, p. 356, note 12).

ing of a physical problem nowadays, which is very different from the classical way to which Jordan was still wedded in *Neue Begründung*. In classical mechanics, as well as in the Dirac-Jordan transformation theory, at least in its original 1927 form, the full description of a physical system requires the specification of a complete set of p's and q's. In quantum mechanics, as first made clear in von Neumann's *Wahrscheinlichkeitstheoretischer Aufbau*, it requires the specification of the eigenvalues of all the operators in a maximal set of commuting operators for the system. In other words, the 'never mind your p's and q's' part of the title of our paper carried the day.

In the balance of this paper we cover the developments sketched above in greater detail.[14] We give self-contained reconstructions of the central arguments and derivations in the papers documenting these developments, which, somewhat incredibly, took place over the span of just one year, between late 1926 and late 1927. To make these arguments and derivations easier to follow for a modern reader, we will translate them all into Dirac notation.

We start (in Section 2) with Jordan's *Neue Begründung* I, which was submitted in late 1926 and appeared early in 1927 (Jordan, 1927b). We also take into account a 9-page synopsis of this paper presented to the Göttingen Academy in January 1927 (Jordan, 1927c) and his discussion of *Neue Begründung* in the final section of a semi-popular article on the development of the new quantum mechanics (Jordan, 1927f, Part 2, pp. 646–648; cf. note 2). In these papers Jordan only dealt with quantities with completely continuous spectra, while suggesting that the generalization to ones with partly or wholly discrete spectra would be straightforward (Jordan 1927b, p. 811, 816; 1927c, p. 161). We cover Jordan's postulates for his probability amplitudes and his use of canonical transformations between pairs of conjugate variables to derive the equations satisfied by these amplitudes.

In Section 3, we discuss the paper submitted in April 1927 by Hilbert, von Neumann, and Nordheim (1928). The authors of this paper had the advantage of having read the paper in which Dirac (1927) presented his version of transformation theory. Jordan only read Dirac's paper, submitted in late 1926, when he was correcting the page proofs of *Neue Begründung* I (Jordan, 1927b, p. 809; note added in proof). Although we occasionally refer to Dirac's work, both his original paper on transformation theory and the book based on it (Dirac, 1930), and avail ourselves freely of his bra and ket notation, our focus in this paper is on the group working in Göttingen, especially Jordan and von Neumann. As mentioned earlier, von Neumann must have become thoroughly familiar with the Dirac-Jordan transformation theory as one of Hilbert's junior co-authors of the *Grundlagen* paper.

[14] For other discussions of these developments, see, e.g., Jammer (1966, Ch. 6) and Mehra and Rechenberg (2000, Chs. I & III).

In Section 4, we consider Jordan's *Neue Begründung* II, submitted in June 1927 and written in part in response to criticism of *Neue Begründung* I in the *Grundlagen* paper by Hilbert, von Neumann, and Nordheim (1928) and in the *Mathematische Begründung* paper of von Neumann (1927a). Since von Neumann introduced an entirely new approach, we deviate slightly from the chronological order of these papers, and discuss *Mathematische Begründung* after *Neue Begründung* II. In the abstract of the latter, Jordan (1927e, p. 1) promised "a simplified and generalized presentation of the theory developed in [*Neue Begründung*] I." Drawing on Dirac (1927), Jordan simplified his notation somewhat, although he also added some new and redundant elements to it. Most importantly, however, the crucial generalization to quantities with partly or wholly discrete spectra turned out to be far more problematic than he had suggested in *Neue Begründung* I. Rather than covering *Neue Begründung* II in detail, we highlight the problems Jordan ran into, especially in his attempt to deal with spin in his new formalism.

In Sections 5 and 6, we turn to the first two papers of von Neumann's trilogy on quantum mechanics of 1927. In Section 5, on *Mathematische Begründung* (von Neumann, 1927a), we focus on von Neumann's criticism of the Dirac-Jordan transformation theory, his proof of the equivalence of wave mechanics and matrix mechanics based on the isomorphism between L^2 and l^2, and his derivation of the trace formula for probabilities in quantum mechanics. We do not cover the introduction of his Hilbert space formalism, which takes up a large portion of his paper. This material is covered in any number of modern books on functional analysis.[15] In Section 6, on *Wahrscheinlichkeits-theoretischer Aufbau* (von Neumann, 1927c), we likewise focus on the overall argument of the paper, covering the derivation of the trace formula from some basic assumptions about the expectation value of observables in an ensemble of identical systems, the introduction of density operators, and the specification of pure states through the values of a maximal set of commuting operators.

In Section 7, we summarize the transition from Jordan's quantum-mechanical formalism rooted in classical mechanics (mind your p's and q's) to von Neumann's quantum-mechanical formalism which no longer depends on classical mechanics for its formulation (never mind your p's and q's). As a coda to our story, we draw attention to the reemergence of the canonical formalism, its generalized coordinates and conjugate momenta, even for spin-$\frac{1}{2}$ particles, in quantum field theory.

[15] See, e.g., Prugovecki (1981), or, for a more elementary treatment, which will be sufficient to follow our paper, Dennery and Krzywicki (1996, Ch. 3).

2 Jordan's *Neue Begründung* I (December 1926)

Neue Begründung I was submitted to *Zeitschrift für Physik* on December 18, 1926 and published January 18, 1927 (Jordan, 1927b). It consists of two parts. In Part One (*I. Teil*), consisting of secs. 1–2 (pp. 809–816), Jordan laid down the postulates of his theory. In Part Two (*II. Teil*), consisting of secs. 3–7, he presented the formalism realizing these postulates. In the abstract of the paper, Jordan announced that his new theory would unify all earlier formulations of quantum theory:

> The four forms of quantum mechanics that have been developed so far—matrix theory, the theory of Born and Wiener, wave mechanics, and q-number theory—are contained in a more general formal theory. Following one of Pauli's ideas, one can base this new theory on a few simple fundamental postulates (*Grundpostulate*) of a statistical nature (Jordan, 1927b, p. 809).

As we already mentioned in the introduction, Jordan claimed that he could recover both the time-dependent and the time-independent Schrödinger equation as special cases of the differential equations he derived for the probability amplitudes central to his formalism. This is the basis for his claim that wave mechanics can be subsumed under his new formalism. Nowhere in the paper did he show explicitly how matrix mechanics is to be subsumed under the new formalism. Perhaps Jordan felt that this did not require a special argument as the new formalism naturally grew out of matrix mechanics and his own contributions to it (Jordan, 1926a,b). However, as emphasized repeatedly already, Jordan (1927b) restricted himself to quantities with purely continuous spectra in *Neue Begründung* I, so the formalism as it stands is not applicable to matrix mechanics. Dirac (1927) faced the same problem with his version of the Dirac-Jordan transformation theory, but other than that Jordan's new formalism can also be seen as a natural extension of Dirac's (1925) q-number theory. It is only toward the end of his paper (sec. 6) that Jordan turned to the operator theory of Born and Wiener (1926). In our discussion of *Neue Begründung* I, we omit this section along with some mathematically intricate parts of secs. 3 and 5 that are not necessary for understanding the paper's overall argument. We will also draw the veil of charity over the last part of Jordan's production, the ill-conceived account of quantum jumps that takes up sec. 7.

Although we will not cover Jordan's unification of the various forms of quantum theory in any detail, we will cover (in Section 5) von Neumann's criticism of the Dirac-Jordan way of proving the equivalence of matrix mechanics and wave mechanics as a prelude to his own proof based on the isomorphism of l^2 and L^2 (von Neumann, 1927a). In our discussion of *Neue Begründung* I in this section, we focus on the portion of Jordan's paper that corresponds to the last sentence of the abstract, which promises a statistical foundation of quantum mechanics. Laying this foundation actually takes up most of the paper (secs. 1–2, 4–5).

2.1 Jordan's postulates for probability amplitudes

The central quantities in *Neue Begründung* I are generalizations of Schrödinger energy eigenfunctions which Jordan called "probability amplitudes." He attributed both the generalization and the term to Pauli. Jordan referred to a footnote in a forthcoming paper by Pauli (1927a, p. 83, note) proposing, in Jordan's terms, the following interpretation of the energy eigenfunctions $\varphi_n(q)$ (where n labels the different energy eigenvalues) of a system (in one dimension): "If $\varphi_n(q)$ is normalized, then $|\varphi_n(q)|^2 dq$ gives the probability that, if the system is in the state n, the coordinate $[\hat{q}]$ has a value between q and $q + dq$" (Jordan, 1927b, p. 811). A probability amplitude such as this one for position and energy can be introduced for any two quantities.[16]

In *Neue Begründung* I, Jordan focused on quantities with completely continuous spectra. He only tried to extend his approach, with severely limited success, to partly or wholly discrete spectra in *Neue Begründung* II (see Section 4). For two quantities $\hat{x}$ and $\hat{y}$ that can take on a continuous range of values x and y, respectively,[17] there is a complex probability amplitude $\varphi(x, y)$ such that $|\varphi(x, y)|^2\, dx$ gives the probability that $\hat{x}$ has a value between x and $x + dx$ given that $\hat{y}$ has the value y.

In modern Dirac notation $\varphi(x, y)$ would be written as $\langle x|y\rangle$. Upon translation into this modern notation, many of Jordan's expressions turn into instantly recognizable expressions in modern quantum mechanics and we shall frequently provide such translations to make it easier to read Jordan's text. We must be careful, however, not to read too much into it. First of all, von Neumann had not yet introduced the abstract notion of Hilbert space when Jordan and Dirac published their theories in 1917, so neither one thought of probability amplitudes as 'inner products' of 'vectors' in Hilbert space at the time. More importantly, for quantities $\hat{x}$'s and $\hat{y}$'s with purely continuous spectra (e.g. position or momentum of a particle in an infinitely extended region), the 'vectors' , $|x\rangle$ and $|y\rangle$ are *not* elements of Hilbert space, although an inner

[16] It is unclear exactly how much of Jordan's statistical foundation of quantum mechanics is due to Pauli. See Duncan and Janssen (2009, p. 359) for discussion and further references.

[17] Recall that this is our notation (cf. note 5): Jordan used different letters for quantities and their numerical values. For instance, he used q (with value x) and β (with value y) for what we called $\hat{x}$ and $\hat{y}$, respectively (Jordan, 1927b, p. 813)

product $\langle x|y\rangle$ can be defined in a generalized sense (as a distribution) as an integral of products of continuum normalized wave functions, as is routinely done in elementary quantum mechanics. The fact that continuum eigenstates can be treated *as though* they are indeed states in a linear space satisfying completeness and orthogonality relations which are continuum analogs of the discrete ones which hold rigorously in a Hilbert space is, as we will see later, just the von Neumann spectral theorem for self-adjoint operators with a (partly or wholly) continuous spectrum.

In the introductory section of *Neue Begründung* I, Jordan (1927b, p. 811) listed two postulates, labeled I and II. Only two pages later, in sec. 2, entitled "Statistical foundation of quantum mechanics," these two postulates are superseded by a new set of four postulates, labeled A through D. [18] In *Neue Begründung* II, Jordan (1927e, p. 6) presented yet another set of postulates, three this time, labeled I through III (see Section 4). [19] The exposition of Jordan's theory by Hilbert, von Neumann, and Nordheim (1928), written in between *Neue Begründung* I and II, starts from six postulates, labeled I through VI (see Section 3). We will start from Jordan's four postulates of *Neue Begründung* I, which we paraphrase and comment on below, staying close to Jordan's own text but using the notation introduced above to distinguish between quantities and their numerical values.

Postulate A. For two mechanical quantities $\hat{q}$ and $\hat{\beta}$ that stand in a definite kinematical relation to one another there are two complex-valued functions, $\varphi(q,\beta)$ and $\psi(q,\beta)$, such that $\varphi(q,\beta)\psi^*(q,\beta)dq$ gives the probability of finding a value between q and $q+dq$ for $\hat{q}$ given that $\hat{\beta}$ has the value β. The function $\varphi(q,\beta)$ is called the probability amplitude, the function $\psi(q,\beta)$ is called the "supplementary amplitude" (*Ergänzungsamplitude*).

Comments: As becomes clear later on in the paper, "mechanical quantities that stand in a definite kinematical relation to one another" are quantities that can be written as functions of some set of generalized coordinates and their conjugate momenta. In his original postulate I, Jordan (1927b, p. 162) wrote that "$\varphi(q,\beta)$ is independent of the mechanical nature (the Hamiltonian) of the system and is determined only by the kinematical relation between q and β."

[18] In the short version of *Neue Begründung* I presented to the Göttingen Academy on January 14, 1927, Jordan (1927c, p. 162) only introduced postulates I and II. In this short version, Jordan (1927c, p. 163) referred to "a soon to appear extensive paper in *Zeitschrift für Physik*" (i.e., Jordan, 1927b).

[19] In his semi-popular account of the development of the new quantum mechanics, Jordan (1927f, p. 648; cf. note 2 above), after explaining the basic notion of a probability amplitude (cf. Postulate A below), listed only two postulates, or axioms as he now called them, namely "the assumption of probability interference" (cf. Postulate C below) and the requirement that there is a canonically conjugate quantity $\hat{p}$ for every quantum-mechanical quantity $\hat{q}$ (cf. Postulate D below).

Hilbert et al. made this into a separate postulate, their postulate V: "A further physical requirement is that the probabilities only depend on the functional nature of the quantities $F_1(pq)$ and $F_2(pq)$, i.e., on their kinematical connection [*Verknüpfung*], and not for instance on additional special properties of the mechanical system under consideration, such as for example its Hamiltonian" (Hilbert, von Neumann, and Nordheim, 1928, p. 5). With $\varphi(q,\beta) = \langle q|\beta\rangle$, the statement about the kinematical nature of probability amplitudes translates into the observation that they depend only on the inner-product structure of Hilbert space and not on the Hamiltonian governing the time evolution of the system under consideration. [20]

It turns out that for all quantities represented, in modern terms, by Hermitian operators, the amplitudes $\psi(q,\beta)$ and $\varphi(q,\beta)$ are equal to one another. At this point, however, Jordan wanted to leave room for quantities represented by non-Hermitian operators. This is directly related to the central role of canonical transformations in his formalism. As Jordan (1926a,b) had found in a pair of papers published in 1926, canonical transformations need not be unitary and therefore do not always preserve the Hermiticity of the conjugate variables one starts from (Duncan and Janssen, 2009). The *Ergänzungsamplitude* does not appear in the presentation of Jordan's formalism by Hilbert, von Neumann, and Nordheim (1928). [21] In *Neue Begründung* II, Jordan (1927e, p. 3) restricted himself to Hermitian quantities and silently dropped the *Ergänzungsamplitude*. We return to the *Ergänzungsamplitude* in Section 2.4 below, but until then we will simply set $\psi(q,\beta) = \varphi(q,\beta)$ everywhere.

Postulate B. The probability amplitude $\bar{\varphi}(\beta,q)$ is the complex conjugate of the probability amplitude $\varphi(q,\beta)$. In other words, $\bar{\varphi}(\beta,q) = \varphi^*(q,\beta)$. This implies a symmetry property of the probabilities themselves: the probability density $|\bar{\varphi}(\beta,q)|^2$ for finding the value β for $\hat{\beta}$ given the value q for $\hat{q}$ is equal to the probability density $|\varphi(q,\beta)|^2$ for finding the value q for $\hat{q}$ given the value β for $\hat{\beta}$.

Comment. This property is immediately obvious once we write $\phi(q,\beta)$ as $\langle q|\beta\rangle$ with the interpretation of $\langle q|\beta\rangle$ as an 'inner product' in Hilbert space (but recall that one has to be cautious when dealing with quantities with continuous spectra).

Postulate C. The probabilities combine through interference. In sec. 1, Jordan

[20] In the AHQP interview with Jordan, Kuhn emphasized the importance of this aspect of Jordan's formalism: "The terribly important step here is throwing the particular Hamiltonian function away and saying that the relationship is only in the kinematics" (session 3, p. 15).

[21] Both ψ and φ are introduced in the lectures by Hilbert on which this paper is based but they are set equal to one another almost immediately and without any further explanation (Sauer and Majer, 2009, p. 700).

(1927b, p. 812) already introduced the phrase "interference of probabilities" to capture the striking feature in his quantum formalism that the probability *amplitudes* rather than the probabilities themselves follow the usual composition rules for probabilities. Let F_1 and F_2 be two outcomes [*Tatsachen*] for which the amplitudes are φ_1 and φ_2. If F_1 and F_2 are mutually exclusive, $\varphi_1 + \varphi_2$ is the amplitude for the outcome 'F_1 or F_2'. If F_1 and F_2 are independent, $\varphi_1\varphi_2$ is the amplitude for the outcome 'F_1 and F_2'.

Consequence. Let $\varphi(q, \beta)$ be the probability amplitude for the outcome F_1 of finding the value q for $\hat{q}$ given the value β for $\hat{\beta}$. Let $\chi(Q, q)$ be the probability amplitude for the outcome F_2 of finding the value Q for $\hat{Q}$ given the value q for $\hat{q}$. Since F_1 and F_2 are independent, Jordan's multiplication rule tells us that the probability amplitude for 'F_1 and F_2' is given by the product $\chi(Q, q)\, \varphi(q, \beta)$. Now let $\Phi(Q, \beta)$ be the probability amplitude for the outcome F_3 of finding the value Q for $\hat{Q}$ given the value β for $\hat{\beta}$. According to Jordan's addition rule, this amplitude is equal to the 'sum' of the amplitudes for 'F_1 and F_2' for all different values of q. Since $\hat{q}$ has a continuous spectrum, this 'sum' is actually an integral. The probability amplitude for F_3 is thus given by [22]

$$\Phi(Q, \beta) = \int \chi(Q, q)\, \varphi(q, \beta)\, dq. \quad \text{[NB1, sec. 2, Eq. 14]} \tag{1}$$

Special case. If $\hat{Q} = \hat{\beta}$, the amplitude $\Phi(\beta', \beta'')$ becomes the Dirac delta function. Jordan (1927b, p. 814) introduced the notation $\delta_{\beta'\beta''}$ even though β' and β'' are continuous rather than discrete variables [NB1, sec. 2, Eq. 16]. In a footnote he conceded that this is mathematically dubious. In *Neue Begründung* II, Jordan (1927e, p. 5) used the delta function that Dirac (1927, pp. 625–627) had meanwhile introduced in his paper on transformation theory. Here and in what follows we will give Jordan the benefit of the doubt and assume the normal properties of the delta function.

Using that the amplitude $\chi(\beta', q)$ is just the complex conjugate of the amplitude $\varphi(q, \beta')$, we arrive at the following expression for $\Phi(\beta', \beta'')$:

$$\Phi(\beta', \beta'') = \int \varphi(q, \beta'')\, \varphi^*(q, \beta')\, dq = \delta_{\beta'\beta''}. \quad \text{[NB1, sec. 2, Eqs. 15, 16, 17]} \tag{2}$$

Comment. Translating Eqs. (1)–(2) above into Dirac notation, we recognize

[22] We include the numbers of the more important equations in the original papers in square brackets. 'NB1' refers to *Neue Begründung* I (Jordan, 1927b). Since the numbering of equations starts over a few times in this paper (see note 36), we will often include the section number as well.

them as familiar *completeness* and *orthogonality* relations:[23]

$$\langle Q|\beta\rangle = \int \langle Q|q\rangle\langle q|\beta\rangle\, dq, \quad \langle\beta'|\beta''\rangle = \int \langle\beta'|q\rangle\langle q|\beta''\rangle\, dq = \delta(\beta' - \beta''). \qquad (3)$$

Since the eigenvectors $|q\rangle$ of the operator $\hat{q}$ are not in Hilbert space, the spectral theorem, first proven by von Neumann (1927a), is required for the use of the resolution of the unit operator $\hat{1} = \int dq|q\rangle\langle q|$.

Postulate D. For every $\hat{q}$ there is a conjugate momentum $\hat{p}$. Before stating this postulate, Jordan offered a new definition of what it means for $\hat{p}$ to be the conjugate momentum of $\hat{q}$. If the amplitude $\rho(p, q)$ of finding the value p for $\hat{p}$ given the value q for $\hat{q}$ is given by

$$\rho(p, q) = e^{-ipq/\hbar}, \quad \text{[NB1, sec. 2, Eq. 18]} \qquad (4)$$

then $\hat{p}$ is the conjugate momentum of $\hat{q}$.

Anticipating a special case of Heisenberg's (1927b) uncertainty principle, Jordan (1927b, p. 814) noted that Eq. (4) implies that "[f]or a given value of $\hat{q}$ all possible values of $\hat{p}$ are *equally probable*."

For $\hat{p}$'s and $\hat{q}$'s with completely continuous spectra, Jordan's definition of when $\hat{p}$ is conjugate to $\hat{q}$ is equivalent to the standard one that the operators $\hat{p}$ and $\hat{q}$ satisfy the commutation relation $[\hat{p}, \hat{q}] \equiv \hat{p}\,\hat{q} - \hat{q}\,\hat{p} = \hbar/i$. This equivalence, however, presupposes the usual association of the differential operators $(\hbar/i)\,\partial/\partial q$ and 'multiplication by q' with the quantities $\hat{p}$ and $\hat{q}$, respectively. As we emphasized in the introduction, Jordan did not think of these quantities as operators acting in an abstract Hilbert space, but he did associate them (as well as any other quantity obtained through adding and multiplying $\hat{p}$'s and $\hat{q}$'s) with the differential operators $(\hbar/i)\,\partial/\partial q$ and q (and combinations of them). The manipulations in Eqs. (19ab)–(24) of *Neue Begründung* I, presented under the subheading "Consequences" (*Folgerungen*) immediately following postulate D, are meant to show that this association follows from his postulates (Jordan, 1927b, pp. 814–815). Using modern notation, we reconstruct Jordan's rather convoluted argument. As we will see, the argument as it stands does not work, but a slightly amended version of it does.

The probability amplitude $\langle p|q\rangle = e^{-ipq/\hbar}$, Jordan's $\rho(p, q)$, trivially satisfies the following pair of equations:

$$\left(p + \frac{\hbar}{i}\frac{\partial}{\partial q}\right)\langle p|q\rangle = 0, \qquad \text{[NB1, sec. 2, Eq. 19a]} \qquad (5)$$

[23] The notation $\langle Q|\beta\rangle$ for $\Phi(Q, \beta)$ etc. obviates the need for different letters for different probability amplitudes that plagues Jordan's notation.

$$\left(\frac{\hbar}{i}\frac{\partial}{\partial p} + q\right)\langle p|q\rangle = 0. \qquad \text{[NB1, sec. 2, Eq. 19b]} \tag{6}$$

Unless we explicitly say otherwise, expressions such as $\langle a|b\rangle$ are to be interpreted as our notation for Jordan's probability amplitudes $\varphi(a, b)$ and *not* as inner products of vectors $|a\rangle$ and $|b\rangle$ in Hilbert space.

Following Jordan (NB1, sec. 2, Eqs. 20–22), we define the map T, which takes functions f of p and turns them into functions Tf of Q (the value of a new quantity $\hat{Q}$ with a fully continuous spectrum):

$$T: \quad f(p) \quad \rightarrow \quad (Tf)(Q) \equiv \int \langle Q|p\rangle f(p)dp. \tag{7}$$

In other words, $T\ldots = \int dp\, \langle Q|p\rangle \ldots$ [NB1, Eq. 21]. For the special case that $f(p) = \langle p|q\rangle$, we get:

$$\left(T\langle p|q\rangle\right)(Q) = \int \langle Q|p\rangle\langle p|q\rangle dp = \langle Q|q\rangle, \tag{8}$$

where we used completeness, one of the consequences of Jordan's postulate C (cf. Eqs. (1)–(3)). In other words, T maps $\langle p|q\rangle$ onto $\langle Q|q\rangle$:[24]

$$\langle Q|q\rangle = T\,\langle p|q\rangle. \quad \text{[NB1, sec. 2, Eq. 22]} \tag{9}$$

Likewise, we define the inverse map T^{-1}, which takes functions F of Q and turns them into functions $T^{-1}F$ of p:[25]

$$T^{-1}: \quad F(Q) \quad \rightarrow \quad (T^{-1}F)(p) \equiv \int \langle p|Q\rangle F(Q)dQ. \tag{10}$$

In other words, $T^{-1}\ldots = \int dQ\, \langle p|Q\rangle \ldots$[26] For the special case that $F(Q) =$

[24] At this point, Jordan's notation, $\varphi(x, y) = T.\rho(x, y)$ [NB1, sec. 2, Eq. 22], gets particularly confusing as the x on the left-hand side and the x on the right-hand side refer to values of different quantities. The same is true for the equations that follow [NB1, Eqs. 23ab, 24ab].

[25] To verify that T^{-1} is indeed the inverse of T, we take $F(Q)$ in Eq. (10) to be $(Tf)(Q)$ in Eq. (7). In that case we get:

$$\begin{aligned}(T^{-1}Tf)(p) &= \int \langle p|Q\rangle \left(\int \langle Q|p'\rangle f(p')dp'\right) dQ \\ &= \iint \langle p|Q\rangle\langle Q|p'\rangle f(p')\, dQ\, dp' \\ &= \int \langle p|p'\rangle f(p')\, dp' = f(p).\end{aligned}$$

[26] Jordan (1927b, p. 815, note) used the *Ergänzungsamplitude* to represent T^{-1} in this form.

$\langle Q|q\rangle$ we get (again, by completeness):

$$\left(T^{-1}\langle Q|q\rangle\right)(p) = \int \langle p|Q\rangle\langle Q|q\rangle dQ = \langle p|q\rangle, \tag{11}$$

or, more succinctly,

$$\langle p|q\rangle = T^{-1}\langle Q|q\rangle. \tag{12}$$

Applying T to the left-hand side of Eq. (5) [NB1, Eq. 19a], we find:

$$T\left(\left(p + \frac{\hbar}{i}\frac{\partial}{\partial q}\right)\langle p|q\rangle\right) = Tp\langle p|q\rangle + \frac{\hbar}{i}\frac{\partial}{\partial q}T\langle p|q\rangle = 0, \tag{13}$$

where we used that differentiation with respect to q commutes with applying T (which only affects the functional dependence on p). Using that $\langle p|q\rangle = T^{-1}\langle Q|q\rangle$ (Eq. (12)) and $T\langle p|q\rangle = \langle Q|q\rangle$ (Eq. (9)), we can rewrite Eq. (13) as:[27]

$$\left(TpT^{-1} + \frac{\hbar}{i}\frac{\partial}{\partial q}\right)\langle Q|q\rangle = 0. \quad \text{[NB1, sec. 2, Eq. 23a]} \tag{14}$$

Similarly, applying T to the left-hand side of Eq. (6) [NB1, Eq. 19b], we find:

$$T\left(\left(\frac{\hbar}{i}\frac{\partial}{\partial p} + q\right)\langle p|q\rangle\right) = T\frac{\hbar}{i}\frac{\partial}{\partial p}\langle p|q\rangle + qT\langle p|q\rangle = 0, \tag{15}$$

where we used that multiplication by q commutes with applying T. Once again using that $\langle p|q\rangle = T^{-1}\langle Q|q\rangle$ and $T\langle p|q\rangle = \langle Q|q\rangle$, we can rewrite this as:[28]

$$\left(T\frac{\hbar}{i}\frac{\partial}{\partial p}T^{-1} + q\right)\langle Q|q\rangle = 0. \quad \text{[NB1, sec. 2, Eq. 23b]} \tag{16}$$

Eqs. (14) and (16) [NB1, Eqs. 23ab] gave Jordan a representation of the quantities $\hat{p}$ and $\hat{q}$ in the Q-basis. The identification of $\hat{p}$ in the Q-basis is straightforward. The quantity p in Eq. (5) [NB1, Eq. 19a] turns into the quantity TpT^{-1} in Eq. (14), [NB1, Eq. 23a]. This is just what Jordan had come to expect on the basis of his earlier use of canonical transformations (see Section 2.2 below). The identification of $\hat{q}$ in the Q-basis is a little trickier. Eq. (6) [NB1, Eq. 19b] told Jordan that the position operator in the original p-basis is $-(\hbar/i)\,\partial/\partial p$ (note the minus sign). This quantity turns into $-T\,(\hbar/i)\,\partial/\partial p\,T^{-1}$ in Eq. (16) [NB1, Eq. 23b]. This then should be the representation of $\hat{q}$ in the new Q-basis, as Jordan stated right below this last equation: "With respect to (*in Bezug*

[27] There is a sign error in NB1, sec. 2, Eq. 23a: $-TxT^{-1}$ should be TxT^{-1}.
[28] There is a sign error in NB1, sec. 2, Eq. 23b: $-y$ should be y.

auf) the fixed chosen quantity $[\hat{Q}]$ every other quantity $[\hat{q}]$ corresponds to an operator $[-T\,(\hbar/i)\,\partial/\partial p\,T^{-1}]$" (Jordan, 1927b, p. 815).[29]

With these representations of his quantum-mechanical quantities $\hat{p}$ and $\hat{q}$, Jordan could now define their addition and multiplication through the corresponding addition and multiplication of the differential operators representing these quantities.

Jordan next step was to work out what the differential operators TpT^{-1} and $-T(\hbar/i)\,\partial/\partial pT^{-1}$, representing $\hat{p}$ and $\hat{q}$ in the Q-basis, are in the special case that $\hat{Q} = \hat{q}$. In that case, Eqs. (14) and (16) [NB1, Eqs. 23ab] turn into:

$$\left(TpT^{-1} + \frac{\hbar}{i}\frac{\partial}{\partial q}\right)\langle q'|q\rangle = 0, \tag{17}$$

$$\left(T\frac{\hbar}{i}\frac{\partial}{\partial p}T^{-1} + q\right)\langle q'|q\rangle = 0. \tag{18}$$

On the other hand, $\langle q'|q\rangle = \delta(q' - q)$. So $\langle q'|q\rangle$ trivially satisfies:

$$\left(\frac{\hbar}{i}\frac{\partial}{\partial q'} + \frac{\hbar}{i}\frac{\partial}{\partial q}\right)\langle q'|q\rangle = 0, \qquad \text{[NB1, sec. 2, Eq. 24a]} \tag{19}$$

$$(-q' + q)\,\langle q'|q\rangle = 0. \qquad \text{[NB1, sec. 2, Eq. 24b]} \tag{20}$$

Comparing Eqs. (19)–(20) with Eqs. (17)–(18), we arrive at

$$TpT^{-1}\langle q'|q\rangle = \frac{\hbar}{i}\frac{\partial}{\partial q'}\langle q'|q\rangle, \tag{21}$$

$$-T\frac{\hbar}{i}\frac{\partial}{\partial p}T^{-1}\langle q'|q\rangle = q'\langle q'|q\rangle. \tag{22}$$

Eq. (21) suggests that TpT^{-1}, the momentum $\hat{p}$ in the q-basis acting on the q' variable, is just $(\hbar/i)\,\partial/\partial q'$. Likewise, Eq. (22) suggests that $-T\,(\hbar/i)\,\partial/\partial p\,T^{-1}$, the position $\hat{q}$ in the q-basis acting on the q' variable, is just multiplication by q'. As Jordan put it in a passage that is hard to follow because of his confusing notation:

> Therefore, as a consequence of (24) [our Eqs. (19)–(20)], the operator x [multiplying by q' in our notation] is assigned (*zugeordnet*) to the quantity [*Grösse*] Q itself [$\hat{q}$ in our notation]. One sees furthermore that the operator

[29] Because of the sign error in Eq. (16) [NB1, Eq. 23b], Jordan set $\hat{q}$ in the Q-basis equal to $T((\hbar/i)\,\partial/\partial p)T^{-1}$.

$\varepsilon\,\partial/\partial x$ [$(\hbar/i)\,\partial/\partial q'$ in our notation] corresponds to the momentum P [$\hat{p}$] belonging to Q [$\hat{q}$] (Jordan, 1927b, p. 815).[30]

It is by this circuitous route that Jordan arrived at the usual functional interpretation of coordinate and momentum operators in the Schrödinger formalism. Jordan (1927b, pp. 815–816) emphasized that the association of $(\hbar/i)\,\partial/\partial q$ and q with $\hat{p}$ and $\hat{q}$ can easily be generalized. Any quantity (*Grösse*) obtained through multiplication and addition of $\hat{q}$ and $\hat{p}$ is associated with the corresponding combination of differential operators q and $(\hbar/i)\,\partial/\partial q$.

Jordan's argument as it stands fails. We cannot conclude that two operations are identical from noting that they give the same result when applied to one special case, here the delta function $\langle q'|q\rangle = \delta(q'-q)$ (cf. Eqs. (21)–(22)). We need to show that they give identical results when applied to an *arbitrary* function. We can easily remedy this flaw in Jordan's argument, using only the kind of manipulations he himself used at this point (though we will do so in modern notation). We contrast this proof in the spirit of Jordan with a modern proof showing that Eqs. (14) and (16) imply that $\hat{p}$ and $\hat{q}$, now understood in the spirit of von Neumann as operators acting in an abstract Hilbert space, are represented by $(\hbar/i)\,\partial/\partial q$ and q, respectively, in the q-basis. The input for the proof in the spirit *à la* Jordan are his postulates and the identification of the differential operators representing momentum and position in the Q-basis as TpT^{-1} and $-T\,(\hbar/i)\,\partial/\partial p\,T^{-1}$, respectively (cf. our comments following Eq. (16)). The input for the proof *à la* van Neumann are the inner-product structure of Hilbert space and the spectral decomposition of the operator $\hat{p}$. Of course, von Neumann (1927a) only introduced these elements *after* Jordan's *Neue Begründung* I.

Closely following Jordan's approach, we can show that Eqs. (14) and (16) [NB1, Eqs. 23ab] imply that, for arbitrary functions $F(Q)$, if Q is set equal to q,

$$(TpT^{-1}F)(q) = \frac{\hbar}{i}\frac{\partial}{\partial q}F(q), \tag{23}$$

$$\left(-T\frac{\hbar}{i}\frac{\partial}{\partial p}T^{-1}F\right)(q) = qF(q). \tag{24}$$

Since F is an arbitrary function, the problem we noted with Eqs. (21)–(22) is solved. Jordan's identification of the differential operators representing momentum and position in the q-basis does follow from Eqs. (23)–(24).

To derive Eq. (23), we apply T, defined in Eq. (7), to $p\,(T^{-1}F)(p)$. We then use the definition of T^{-1} in Eq. (10) to write $(TpT^{-1}F)(Q)$ as:

[30] We remind the reader that Jordan used the term 'operator' [*Operator*] *not* for an operator acting in an abstract Hilbert space but for the differential operators $(\hbar/i)\,\partial/\partial x$ and (multiplying by) x and for combinations of them.

$$
\begin{aligned}
(TpT^{-1}F)(Q) &= \int \langle Q|p\rangle \; p \; (T^{-1}F)(p) \; dp \\
&= \int \langle Q|p\rangle \; p \left[\int \langle p|Q'\rangle F(Q') dQ' \right] dp \\
&= \iint \langle Q|p\rangle \; p \; \langle p|Q'\rangle F(Q') \; dp \, dQ'.
\end{aligned} \tag{25}
$$

We now set $\hat{Q} = \hat{q}$, use Eq. (5) to substitute $-(\hbar/i)\,\partial/\partial q'\,\langle p|q'\rangle$ for $p\langle p|q'\rangle$, and perform a partial integration:

$$
\begin{aligned}
(TpT^{-1}F)(q) &= \iint \langle q|p\rangle \; p \; \langle p|q'\rangle F(q') \; dp \, dq' \\
&= \iint \langle q|p\rangle \left(-\frac{\hbar}{i} \frac{\partial}{\partial q'} \langle p|q'\rangle \right) F(q') \; dp \, dq' \\
&= \iint \langle q|p\rangle \langle p|q'\rangle \frac{\hbar}{i} \frac{dF(q')}{dq'} \; dp \, dq'.
\end{aligned} \tag{26}
$$

On account of completeness and orthogonality (see Eq. (3) [NB1, Eqs. 14–17]), the right-hand side reduces to $(\hbar/i)\,F'(q)$. This concludes the proof of Eq. (23).

To derive Eq. (24), we similarly apply T to $-(\hbar/i)\,\partial/\partial p\,(T^{-1}F)(p)$:

$$
\begin{aligned}
\left(-T \frac{\hbar}{i} \frac{\partial}{\partial p} T^{-1} F \right)(Q) &= - \int \langle Q|p\rangle \; \frac{\hbar}{i} \frac{\partial}{\partial p} \, (T^{-1}F)(p) \, dp \\
&= - \int \langle Q|p\rangle \; \frac{\hbar}{i} \frac{\partial}{\partial p} \left[\int \langle p|Q'\rangle F(Q') dQ' \right] dp \\
&= - \iint \langle Q|p\rangle \; \frac{\hbar}{i} \frac{\partial}{\partial p} \; \langle p|Q'\rangle F(Q') \; dp \, dQ'.
\end{aligned} \tag{27}
$$

We now set $\hat{Q} = \hat{q}$ and use Eq. (6) to substitute $q'\langle p|q'\rangle$ for $-(\hbar/i)\,\partial/\partial p\,\langle p|q'\rangle$:

$$
\left(-T \frac{\hbar}{i} \frac{\partial}{\partial p} T^{-1} F \right)(q) = \iint \langle q|p\rangle \, q' \, \langle p|q'\rangle \; F(q') \, dp \, dq' = q \, F(q), \tag{28}
$$

where in the last step we once again used completeness and orthogonality. This concludes the proof of Eq. (24).

We now turn to the modern proofs. It is trivial to show that the representation of the position operator $\hat{q}$ in the q-basis is simply multiplication by the eigenvalues q. Consider an arbitrary eigenstate $|q\rangle$ of position with eigenvalue q, i.e., $\hat{q}\,|q\rangle = q\,|q\rangle$. It follows that $\langle Q|\,\hat{q}\,|q\rangle = q\langle Q|q\rangle$, where $|Q\rangle$ is an arbitrary eigenvector of an arbitrary Hermitian operator $\hat{Q} = \hat{Q}^\dagger$ with eigenvalue Q. The complex conjugate of this last relation,

$$
\langle q|\,\hat{q}\,|Q\rangle = q\,\langle q|Q\rangle, \tag{29}
$$

is just the result we wanted prove.

It takes a little more work to show that Eq. (14) [NB1, Eq. 23a] implies that the representation of the momentum operator $\hat{p}$ in the q-basis is $(\hbar/i)\,\partial/\partial q$. Consider Eq. (25) for the special case $F(Q) = \langle Q|q\rangle$:

$$TpT^{-1}\langle Q|q\rangle = \iint \langle Q|p\rangle\ p\ \langle p|Q'\rangle\langle Q'|q\rangle\ dp\,dQ'. \tag{30}$$

Recognizing the spectral decomposition $\int dp\,p\,|p\rangle\langle p|$ of $\hat{p}$ in this equation, we can rewrite it as:

$$TpT^{-1}\langle Q|q\rangle = \int \langle Q|\,\hat{p}\,|Q'\rangle\langle Q'|q\rangle\,dQ' = \langle Q|\,\hat{p}\,|q\rangle, \tag{31}$$

where in the last step we used the decomposition $\int |Q'\rangle\langle Q'|\,dQ'$ of the unit operator. Eq. (14) tells us that

$$TpT^{-1}\langle Q|q\rangle = -\frac{\hbar}{i}\frac{\partial}{\partial q}\langle Q|q\rangle. \tag{32}$$

Setting the complex conjugates of the right-hand sides of these last two equations equal to one another, we arrive at:

$$\langle q|\,\hat{p}\,|Q\rangle = \frac{\hbar}{i}\frac{\partial}{\partial q}\langle q|Q\rangle, \tag{33}$$

which is the result we wanted to prove. Once again, the operator $\hat{Q}$ with eigenvectors $|Q\rangle$ is arbitrary. If $\hat{Q}$ is the Hamiltonian and $\hat{q}$ is a Cartesian coordinate, $\langle q|Q\rangle$ is just a Schrödinger energy eigenfunction.

With these identifications of $\hat{p}$ and $\hat{q}$ in the q-basis we can finally show that Jordan's new definition of conjugate variables in Eq. (4) [NB1, Eq. 18] reduces to the standard definition, $[\hat{p}, \hat{q}] = \hbar/i$, at least for quantities with completely continuous spectra. Letting $[(\hbar/i)\,\partial/\partial q, q]$ act on an arbitrary function $f(q)$, one readily verifies that the result is $(\hbar/i)\,f(q)$. Given the association of $(\hbar/i)\,\partial/\partial q$ and q with the quantities $\hat{p}$ and $\hat{q}$ that has meanwhile been established, it follows that these quantities indeed satisfy the usual commutation relation

$$[\hat{p}, \hat{q}] = \hat{p}\,\hat{q} - \hat{q}\,\hat{p} = \frac{\hbar}{i}. \quad \text{[NB1, Eq. 25]} \tag{34}$$

This concludes Part I (consisting of secs. 1–2) of *Neue Begründung* I. Jordan wrote:

> This is the content of the new theory. The rest of the paper will be devoted, through a mathematical discussion of these differential equations [NB1, Eqs. 23ab, our Eqs. (14)–(16), and similar equations for other quantities], on the one hand, to proving that our postulates are mathematically consistent

[*widerspruchsfrei*] and, on the other hand, to showing that the earlier forms [*Darstellungen*] of quantum mechanics are contained in our theory (Jordan, 1927b, p. 816).

In this paper we focus on the first of these tasks, which amounts to providing a realization of the postulates discussed in this secton.

2.2 Canonical transformations in classical mechanics, the old quantum theory and matrix mechanics

Given the central role of canonical transformations in *Neue Begründung*, we insert a brief subsection to review the use of canonical transformations in the developments leading up to it. [31] Canonical transformations in classical physics are transformations of the position and conjugate momentum variables (q, p) that preserve the form of Hamilton's equations,

$$\dot{q} = \frac{\partial H(p,q)}{\partial p}, \quad \dot{p} = -\frac{\partial H(p,q)}{\partial q}. \tag{35}$$

Following Jordan (1927b, p. 810) in *Neue Begründung* I, we assume that the system is one-dimensional. For convenience, we assume that the Hamiltonian $H(p, q)$ does not explicitly depend on time. The canonical transformation to new coordinates and momenta (Q, P) is given through a *generating function*, which is a function of one of the old and one of the new variables. For a generating function of the form $F(q, P)$, for instance, [32] we find the equations for the canonical transformation $(q, p) \rightarrow (Q, P)$ by solving the equations

$$p = \frac{\partial F(q,P)}{\partial q}, \quad Q = \frac{\partial F(q,P)}{\partial P} \tag{36}$$

for $Q(q, p)$ and $P(q, p)$. This transformation preserves the form of Hamilton's equations: [33]

$$\dot{Q} = \frac{\partial \hat{H}(P,Q)}{\partial P}, \quad \dot{P} = -\frac{\partial \hat{H}(P,Q)}{\partial Q}, \tag{37}$$

where the Hamiltonians $H(p, q)$ and $\overline{H}(P, Q)$ are numerically equal to one another but given by different functions of their respective arguments. One way to solve the equations of motion is to find a canonical transformation such that, in terms of the new variables, the Hamiltonian depends only on

[31] This subsection is based on Duncan and Janssen (2009, sec. 2).
[32] In the classification of Goldstein *et al.* (2002, p. 373, table 9.1), this corresponds to a generating function of type 2, $F_2(q, P)$. The other types depend on (q, Q), (p, Q), or (p, P). Which type one chooses is purely a matter of convenience and does not affect the physical content.
[33] For elementary discussion, see, e.g., Duncan and Janssen (2007, Pt. 2, sec. 5.1).

momentum, $\overline{H}(P,Q) = \overline{H}(P)$. Such variables are called *action-angle variables* and the standard notation for them is (J, w). The basic quantization condition of the old quantum theory of Bohr and Sommerfeld restricts the value of a set of action variables for the system under consideration to integral multiples of Planck's constant, $J = nh$ $(n = 0, 1, 2, \ldots)$. Canonical transformations to action-angle variables thus played a central role in the old quantum theory. With the help of them, the energy spectrum of the system under consideration could be found.

In classical mechanics, canonical transformations preserve the so-called Poisson bracket, $\{p, q\} = 1$. For any two phase-space functions $G(p,q), H(p,q)$ of the pair of canonical variables (p,q), the Poisson bracket is defined as

$$\{G(p,q), H(p,q)\} \equiv \frac{\partial G(p,q)}{\partial p}\frac{\partial H(p,q)}{\partial q} - \frac{\partial H(p,q)}{\partial p}\frac{\partial G(p,q)}{\partial q}. \tag{38}$$

For $G(p,q) = p$ and $H(p,q) = q$, this reduces to $\{p, q\} = 1$. We now compute the Poisson bracket $\{P, Q\}$ of a new pair of canonical variables related to (p,q) by the generating function $F(q, P)$ as in Eq. (36):

$$\{P(p,q), Q(p,q)\} = \frac{\partial P(p,q)}{\partial p}\frac{\partial Q(p,q)}{\partial q} - \frac{\partial Q(p,q)}{\partial p}\frac{\partial P(p,q)}{\partial q}. \tag{39}$$

By the usual chain rules of partial differentiation, we have

$$\left.\frac{\partial Q}{\partial p}\right|_q = \left.\frac{\partial^2 F}{\partial P^2}\right|_q \left.\frac{\partial P}{\partial p}\right|_q, \tag{40}$$

$$\left.\frac{\partial Q}{\partial q}\right|_p = \frac{\partial^2 F}{\partial q \partial P} + \left.\frac{\partial^2 F}{\partial P^2}\right|_q \left.\frac{\partial P}{\partial q}\right|_p. \tag{41}$$

Substituting these two expressions into Eq. (39), we find

$$\begin{aligned}\{P(p,q), Q(p,q)\} = &\left.\frac{\partial P}{\partial p}\right|_q \left(\frac{\partial^2 F}{\partial q \partial P} + \left.\frac{\partial^2 F}{\partial P^2}\right|_q \left.\frac{\partial P}{\partial q}\right|_p\right) \\ &- \left(\left.\frac{\partial^2 F}{\partial P^2}\right|_q \left.\frac{\partial P}{\partial p}\right|_q\right) \left.\frac{\partial P}{\partial q}\right|_p \\ = &\frac{\partial^2 F}{\partial q \partial P} \left.\frac{\partial P}{\partial p}\right|_q. \end{aligned} \tag{42}$$

The final line is identically equal to 1, as

$$\frac{\partial^2 F}{\partial q \partial P} = \left.\frac{\partial p}{\partial P}\right|_q = \left(\left.\frac{\partial P}{\partial p}\right|_q\right)^{-1}. \tag{43}$$

This shows that the Poisson bracket $\{p, q\} = 1$ is indeed invariant under canonical transformations.

In matrix mechanics a canonical transformation is a transformation of the matrices (q, p) to new matrices (Q, P) preserving the canonical commutation relations

$$[p, q] \equiv pq - qp = \frac{\hbar}{i} \tag{44}$$

that replace the Poisson bracket $\{p, q\} = 1$ in quantum mechanics. Such transformations are of the form

$$P = TpT^{-1}, \quad Q = TqT^{-1}, \quad \overline{H} = THT^{-1}, \tag{45}$$

where $\overline{H}$ is obtained by substituting TpT^{-1} for p and TqT^{-1} for q in the operator H given as a function p and q. One easily recognizes that this transformation does indeed preserve the form of the commutation relations (44): $[P, Q] = \hbar/i$. Solving the equations of motion in matrix mechanics boils down to finding a transformation matrix T such that the new Hamiltonian $\overline{H}$ is diagonal. The diagonal elements, $\overline{H}_{mm}$, then give the (discrete) energy spectrum.

In two papers before *Neue Begründung*, Jordan (1926a,b) investigated the relation between the matrices T implementing canonical transformations in matrix mechanics and generating functions in classical mechanics. He showed that the matrix T corresponding to a generating function of the form [34]

$$F(p, Q) = \sum_n f_n(p) g_n(Q), \tag{46}$$

is given by

$$T(q, p) = \exp \frac{i}{\hbar} \left\{ (p, q) - \sum_n (f_n(p),\, g_n(q)) \right\}, \tag{47}$$

where the notation $(.\,,.)$ in the exponential signals an ordering such that, when the exponential is expanded, all p's are put to the left of all q's in every term of the expansion. [35]

When he wrote *Neue Begründung* I in late 1926, Jordan was thus steeped in the use of canonical transformations, both in classical and in quantum physics. When Kuhn asked Jordan about his two papers on the topic (Jordan, 1926a,b) during an interview in 1963 for the AHQP, Jordan told him:

> Canonical transformations in the sense of Hamilton-Jacobi were ... our daily bread in the preceding years, so to tie in the new results with those as

[34] In the classification of Goldstein *et al.* (cf. note 32), this corresponds to a generating function of type 3, $F_3(p, Q)$.
[35] See Duncan and Janssen (2009, pp. 355–356) for a reconstruction of Jordan's proof of this result.

closely as possible—that was something very natural for us to try (AHQP interview with Jordan, session 4, p. 11).

2.3 The realization of Jordan's postulates: probability amplitudes and canonical transformations

At the beginning of sec. 4 of *Neue Begründung* I, "General comments on the differential equations for the amplitudes," Jordan announced:

> To prove that our postulates are mathematically consistent, we want to give a new foundation of the theory—independently from the considerations in sec. 2—based on the differential equations which appeared as end results there (Jordan, 1927b, p. 821).

He began by introducing the canonically conjugate variables $\hat{\alpha}$ and $\hat{\beta}$, satisfying, by definition, the commutation relation $[\hat{\alpha}, \hat{\beta}] = \hbar/i$. They are related to the basic variables $\hat{p}$ and $\hat{q}$, for which the probability amplitude, according to Jordan's postulates, is $\langle p|q\rangle = e^{-ipq/\hbar}$ (see Eq. (4)), via

$$\hat{\alpha} = f(\hat{p}, \hat{q}) = T\hat{p}T^{-1}, \tag{48}$$

$$\hat{\beta} = g(\hat{p}, \hat{q}) = T\hat{q}T^{-1}. \tag{49}$$

with $T = T(\hat{p}, \hat{q})$ [NB1, sec. 4, Eq. 1].[36] Note that the operator $T(\hat{p}, \hat{q})$ defined here is different from the operator $T \ldots = \int dp\, \langle Q|p\rangle \ldots$ defined in sec. 2 (see Eq. (7), Jordan's Eq. (21)). The $T(\hat{p}, \hat{q})$ operator defined in sec. 4 is a similarity transformation operator implementing the canonical transformation from the pair $(\hat{p}, \hat{q})$ to the pair $(\hat{\alpha}, \hat{\beta})$. We will see later that there is an important relation between the T operators defined in secs. 2 and 4.

Jordan now posited the fundamental differential equations for the probability amplitude $\langle q|\beta\rangle$ in his theory:[37]

$$\left\{ f\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) + \frac{\hbar}{i}\frac{\partial}{\partial \beta} \right\} \langle q|\beta\rangle = 0, \qquad \text{[NB1, sec. 4, Eq. 2a]} \tag{50}$$

$$\left\{ g\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) - \beta \right\} \langle q|\beta\rangle = 0. \qquad \text{[NB1, sec. 4, Eq. 2b]} \tag{51}$$

[36] The numbering of equations in *Neue Begründung* I starts over in sec. 3, the first section of Part Two (*II. Teil*), and then again in sec. 4. Secs. 6 and 7, finally, have their own set of equation numbers.

[37] He introduced separate equations for the *Ergänzungsamplitude* [NB1, sec. 4, Eqs. 3ab] (see Eqs. (90)–(91) below). We ignore these additional equations for the moment but will examine them for some special cases in sec. 2.4.

These equations have the exact same form as Eqs. (14)–(16) [NB1, sec. 2, Eqs. 23ab], with the understanding that the operator T is defined differently. As Jordan put it in the passage quoted above, he took the equations that were the end result in sec. 2 as his starting point in sec. 4.

Before turning to Jordan's discussion of these equations, we show that they are easily recovered in the modern Hilbert space formalism. The result of the momentum operator $\hat{\alpha}$ in Eq. (48) acting on eigenvectors $|\beta\rangle$ of its conjugate operator $\hat{\beta}$ in Eq. (49) is, as we saw in Section 2.1:[38]

$$\hat{\alpha}|\beta\rangle = -\frac{\hbar}{i}\frac{\partial}{\partial\beta}|\beta\rangle. \tag{52}$$

Taking the inner product of these expressions with $|q\rangle$ and using that $\hat{\alpha} = f(\hat{p}, \hat{q})$, we find that

$$-\frac{\hbar}{i}\frac{\partial}{\partial\beta}\langle q|\beta\rangle = \langle q|\hat{\alpha}|\beta\rangle = \langle q|f(\hat{p}, \hat{q})|\beta\rangle. \tag{53}$$

Since $\hat{p}$ and $\hat{q}$ are represented by the differential operators $(\hbar/i)\,\partial/\partial q$ and q, respectively, in the q-basis, we can rewrite this as

$$\langle q|f(\hat{p}, \hat{q})|\beta\rangle = f\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right)\langle q|\beta\rangle. \tag{54}$$

Combining these last two equations, we arrive at Eq. (50). Likewise, using that $\hat{\beta}|\beta\rangle = \beta|\beta\rangle$ and that $\hat{\beta} = g(\hat{p}, \hat{q})$, we can write the inner product $\langle q|\hat{\beta}|\beta\rangle$ as

$$\langle q|\hat{\beta}|\beta\rangle = \beta\langle q|\beta\rangle = g\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right)\langle q|\beta\rangle, \tag{55}$$

where in the last step we used the representation of $\hat{p}$ and $\hat{q}$ in the q-basis. From this equation we can read off Eq. (51).

We turn to Jordan's discussion of Eqs. (50)–(51) [NB1, sec. 4, Eqs. 2ab]. As he pointed out:

> As is well-known, of course, one cannot in general simultaneously impose two partial differential equations on one function of two variables. We will prove, however, in sec. 5: the presupposition—which we already made—that $\hat{\alpha}$ and $\hat{\beta}$ are connected to $\hat{p}$ and $\hat{q}$ via a canonical transformation (1) [our

[38] The complex conjugate of Eq. (33) can be written as

$$\langle Q|\,\hat{p}\,|q\rangle = -\frac{\hbar}{i}\frac{\partial}{\partial q}\langle Q|q\rangle = \langle Q|\left(-\frac{\hbar}{i}\frac{\partial}{\partial q}\right)|q\rangle.$$

Since this holds for arbitrary $|Q\rangle$, it follows that $\hat{p}\,|q\rangle = -(\hbar/i)(\partial/\partial q)|q\rangle$. This will be true for any pair of conjugate variables.

Eqs. (48)–(49)] is the necessary and sufficient condition for (2) [our Eqs. (50)–(51)] to be solvable (Jordan, 1927b, p. 822; hats added).

In sec. 5, "Mathematical theory of the amplitude equations," Jordan (1927b, pp. 824–828) made good on this promise. To prove that the "presupposition" is sufficient, he used canonical transformations to explicitly construct a simultaneous solution of the pair of differential equations (50)–(51) for probability amplitudes (ibid., pp. 824–825, Eqs. 9–17). He did this in two steps.

(1) He showed that the sufficient condition for $\langle Q|\beta\rangle$ to be a solution of the amplitude equations in the Q-basis, given that $\langle q|\beta\rangle$ is a solution of these equations in the q-basis, is that $(\hat{p}, \hat{q})$ and $(\hat{P}, \hat{Q})$ are related by a canonical transformation.
(2) He established a starting point for generating such solutions by showing that a very simple canonical transformation (basically switching $\hat{p}$ and $\hat{q}$) turns the amplitude equations (50)–(51) into a pair of equations immediately seen to be satisfied by the amplitude $\langle q|\beta\rangle = e^{iq\beta/\hbar}$.

With these two steps Jordan had shown that the assumption that $\hat{P}$ and $\hat{Q}$ are related to $\hat{p}$ and $\hat{q}$ through a canonical transformation is indeed a sufficient condition for the amplitude equations (50)–(51) [NB1, sec. 4, Eqs. 2ab] to be simultaneously solvable. We cover this part of Jordan's argument in detail.

The proof that this assumption is necessary as well as sufficient is much more complicated (Jordan, 1927b, pp. 825–828, Eqs. 18–34). The mathematical preliminaries presented in sec. 3 of *Neue Begründung* I (ibid., pp. 816–821) are needed only for this part of the proof in sec. 5. We will cover neither this part of sec. 5 nor sec. 3.

However, we do need to explain an important result that Jordan derived in sec. 5 as a consequence of this part of his proof (ibid., p. 828, Eqs. 35–40): Canonical transformations $T(\hat{p}, \hat{q})$ as defined above (see Eqs. (48)–(49) [NB1, sec. 4, Eq. 1]), which are differential operators once $\hat{p}$ and $\hat{q}$ have been replaced by their representations $(\hbar/i)\,\partial/\partial q$ and q in the q-basis, can be written in the form of the integral operators T defined in sec. 2 (see Eq. (7) [NB1, sec. 2, Eq. 21]).

This result is central to the basic structure of Jordan's theory and to the logic of his *Neue Begründung* papers. It shows that Jordan's probability amplitudes do double duty as integral kernels of the operators implementing canonical transformations. As such, Jordan showed, they satisfy the completeness and orthogonality relations required by postulate C (see Eqs. (1)–(3) [NB1, sec. 2, Eqs. 14–17]). To paraphrase the characterization of Jordan's project by Hilbert et al. that we already quoted in the introduction, Jordan postulated certain relations between his probability amplitudes in Part One of his paper and then, in Part Two, presented "a simple analytical apparatus in which quantities occur that satisfy these relations exactly" (Hilbert, von Neumann, and Nordheim, 1928, p. 2). These quantities, it turns out, are the integral kernels of canonical transformations. Rather than following Jordan's own proof of this key result, which turns on properties of canonical transformations, we present a modern proof, which turns on properties of Hilbert space and the spectral theorem.

But first we show, closely following Jordan's own argument in sec. 5 of *Neue Begründung* I, how to construct a simultaneous solution of the differential equations (50)–(51) [NB1, sec. 4, Eqs. 2ab] for the amplitudes. Suppose we can exhibit just one case of a canonical transformation $(\hat{p},\hat{q}) \to (\hat{\alpha},\hat{\beta})$ (Eqs. (48)–(49) [NB1, sec. 4, Eq. 1]) where the amplitude equations manifestly have a unique simultaneous solution. Any other canonical pair can be arrived at from the pair $(\hat{p},\hat{q})$ via a new transformation function $S(\hat{P},\hat{Q})$, in the usual way

$$\hat{p} = S\hat{P}S^{-1}, \quad \hat{q} = S\hat{Q}S^{-1}. \tag{56}$$

with $S = S(\hat{P},\hat{Q})$ [NB1, sec. 5, Eq. 10]. The connection between the original pair $(\hat{\alpha},\hat{\beta})$ and the new pair $(\hat{P},\hat{Q})$ involves the composite of two canonical transformations [NB1, sec. 5, Eq. 11]:

$$\hat{\alpha} = f(\hat{p},\hat{q}) = f(S\hat{P}S^{-1}, S\hat{Q}S^{-1}) \equiv F(\hat{P},\hat{Q}), \tag{57}$$

$$\hat{\beta} = g(\hat{q},\hat{q}) = g(S\hat{P}S^{-1}, S\hat{Q}S^{-1}) \equiv G(\hat{P},\hat{Q}). \tag{58}$$

In the new Q-basis, the differential equations (50)–(51) [NB1, sec. 4, Eqs. 2ab] for probability amplitudes take the form

$$\left\{ F\left(\frac{\hbar}{i}\frac{\partial}{\partial Q}, Q\right) + \frac{\hbar}{i}\frac{\partial}{\partial \beta} \right\} \langle Q|\beta\rangle = 0, \quad \text{[NB1, sec. 5, Eq. 12a]} \tag{59}$$

$$\left\{ G\left(\frac{\hbar}{i}\frac{\partial}{\partial Q}, Q\right) - \beta \right\} \langle Q|\beta\rangle = 0. \quad \text{[NB1, sec. 5, Eq. 12b]} \tag{60}$$

Jordan now showed that

$$\langle Q|\beta\rangle = \left\{ S\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) \langle q|\beta\rangle \right\}_{q=Q} \qquad \text{[NB1, sec. 5, Eq. 13]} \tag{61}$$

is a simultaneous solution of the amplitude equations (59)–(60) in the Q-basis if $\langle q|\beta\rangle$ is a simultaneous solution of the amplitude equations (50)–(51) in the q-basis. Using the operator S and its inverse S^{-1}, we can rewrite the latter as [39]

$$S\left\{ f\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) + \frac{\hbar}{i}\frac{\partial}{\partial \beta} \right\} S^{-1} S\,\langle q|\beta\rangle = 0, \qquad \text{[NB1, sec. 5, Eq. 14a]} \tag{62}$$

[39] This step is formally the same as the one that got us from Eqs. (5)–(6) [NB1, sec. 2, Eqs. 19ab] to Eqs. (14)–(16) [NB1, sec. 2, Eqs. 23ab].

$$S\left\{g\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)-\beta\right\}S^{-1}S\,\langle q|\beta\rangle=0, \qquad \text{[NB1, sec. 5, Eq. 14b]} \tag{63}$$

both taken, as in Eq. (61), at $q = Q$. Written more carefully, the first term in curly brackets in Eq. (62), sandwiched between S and S^{-1}, is

$$S\,f\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)\,S^{-1}=\left\{S\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)\,f\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)\,S^{-1}\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)\right\}_{q=Q}. \tag{64}$$

With the help of Eq. (57), this can further be rewritten as

$$S(P,Q)\,f(P,Q)\,S(P,Q)^{-1}|_{P=\frac{\hbar}{i}\frac{\partial}{\partial Q}}=F\left(\frac{\hbar}{i}\frac{\partial}{\partial Q},Q\right). \tag{65}$$

The second term in curly brackets in Eq. (62), sandwiched between S and S^{-1}, is simply equal to

$$S\,\frac{\hbar}{i}\frac{\partial}{\partial\beta}\,S^{-1}=\frac{\hbar}{i}\frac{\partial}{\partial\beta}, \tag{66}$$

as S does not involve β. Using Eqs. (61) and (64)–(66), we can rewrite Eq. (62) [NB1, sec. 5, Eq. 14a] as

$$\left\{F\left(\frac{\hbar}{i}\frac{\partial}{\partial Q},Q\right)+\frac{\hbar}{i}\frac{\partial}{\partial\beta}\right\}\langle Q|\beta\rangle=0, \tag{67}$$

which is just Eq. (59) [NB1, sec. 5, Eq. 12a]. A completely analogous argument establishes that Eq. (63) [NB1, sec. 5, Eq. 14b] reduces to Eq. (60) [NB1, sec. 5, Eq. 12b]. This concludes the proof that $\langle Q|\beta\rangle$ is a solution of the amplitude equations in the new Q-basis, if $\langle q|\beta\rangle$, out of which $\langle Q|\beta\rangle$ was constructed with the help of the operator S implementing a canonical transformation, is a solution of the amplitude equations in the old q-basis.

As S is completely general, we need only exhibit a single valid starting point, i.e., a pair (f, g) and an amplitude $\langle q|\beta\rangle$ satisfying the amplitude equations in the q-basis (Eqs. (50)–(51) [NB1, sec. 4, Eqs. 2ab]), to construct general solutions of the amplitude equations in some new Q-basis (Eqs. (59)–(60) [NB1, sec. 5, Eqs. 12ab]). The trivial example of a canonical transformation switching the roles of coordinate and momentum does the trick (cf. Eqs. (57)–(58) [NB1, sec. 5, Eq. 11]):

$$\hat{\alpha}=f(\hat{p},\hat{q})=-\hat{q}, \quad \beta=g(\hat{p},\hat{q})=\hat{p}. \quad \text{[NB1, sec. 5, Eq. 15]} \tag{68}$$

In that case, Eqs. (50)–(51) become [NB1, sec. 5, Eq. 16]

$$\left\{q-\frac{\hbar}{i}\frac{\partial}{\partial\beta}\right\}\langle q|\beta\rangle=0, \tag{69}$$

$$\left\{\frac{\hbar}{i}\frac{\partial}{\partial q}-\beta\right\}\langle q|\beta\rangle=0. \tag{70}$$

Except for the minus signs, these equations are of the same form as the trivial equations (5)–(6) [NB1, sec. 2, Eqs. 19ab] for $\langle p|q\rangle$, satisfied by the basic amplitude $\langle p|q\rangle = e^{-ipq/\hbar}$. In the case of Eqs. (69)–(70), the solution is:

$$\langle q|\beta\rangle = e^{i\beta q/\hbar}. \quad \text{[NB1, sec. 5, Eq. 17]} \tag{71}$$

This establishes that the canonical nature of the transformation to the new variables is a sufficient condition for the consistency (i.e. simultaneous solvability) of the pair of differential equations (59)–(60) [NB1, sec. 5, Eq. 12ab] for the probability amplitudes.

Jordan (1927b, pp. 825–828) went on to prove the converse, i.e., that the canonical connection is also a necessary condition for the consistency of Eqs. (59)–(60). This is done, as Jordan explained at the top of p. 827 of his paper, by explicit construction of the operator S (in Eq. (61)), given the validity of Eqs. (59)–(60). We skip this part of the proof.

Jordan (1927b, p. 828) then used some of the same techniques to prove a key result in his theory. As mentioned above, we will appeal to the modern Hilbert space formalism and the spectral theorem to obtain this result. Once again consider Eq. (61):

$$\langle Q|\beta\rangle = \left\{ S\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) \langle q|\beta\rangle \right\}_{q=Q}. \quad \text{[NB1, sec. 5, Eq. 13]} \tag{72}$$

This equation tells us that the differential operator $S((\hbar/i)\,\partial/\partial q, q)$ maps arbitrary states $\langle q|\beta\rangle$ in the q-basis (recall that $\hat{\beta}$ can be any operator) onto the corresponding states $\langle Q|\beta\rangle$ in the Q-basis. The spectral theorem, which gives us the resolution $\int dq|q\rangle\langle q|$ of the unit operator, tells us that this mapping can also be written as

$$\langle Q|\beta\rangle = \int dq\, \langle Q|q\rangle\langle q|\beta\rangle. \tag{73}$$

Schematically, we can write

$$S\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) \ldots = \int dq\, \langle Q|q\rangle \ldots \tag{74}$$

In other words, the probability amplitude $\langle Q|q\rangle$ is the integral kernel for the integral representation of the canonical transformation operator $S((\hbar/i)\,\partial/\partial q, q)$. Using nothing but the properties of canonical transformations and his differential equations for probability amplitudes (Eqs. (50)–(51) [NB1, sec. 4, Eqs. 2ab]), Jordan (1927b, p. 828) derived an equation of exactly the same form as Eq. (74), which we give here in its original notation:

$$T\left(\varepsilon\frac{\partial}{\partial x}, x\right) = \int dx . \varphi(y,x) \ldots . \quad \text{[NB1, sec. 5, Eq. 40]} \tag{75}$$

Jordan claimed that Eqs. (50)–(51) [NB1, sec. 4, Eqs. 2ab] contain both the time-independent and the time-dependent Schrödinger equations as special cases. The time-independent Schrödinger equation is a special case of Eq. (51):

> If in (2b) we take β to be the energy W, and g to be the Hamiltonian function $H(p, q)$ of the system, we obtain the Schrödinger wave equation, which corresponds to the classical Hamilton-Jacobi equation. With (2b) comes (2a) as a second equation. In this equation we need to consider f to be the time t (as a function of p and q) (Jordan, 1927b, p. 822).

Actually, the variable conjugate to $\hat{H}$ would have to be *minus* $\hat{t}$. For $\hat{\alpha} = f(\hat{p}, \hat{q}) = -\hat{t}$ and $\hat{\beta} = g(\hat{p}, \hat{q}) = \hat{H}$ (with eigenvalues E), Eqs. (50)–(51) become:

$$\left\{ \hat{t} - \frac{\hbar}{i} \frac{\partial}{\partial E} \right\} \langle q|E \rangle = 0, \tag{76}$$

$$\left\{ \hat{H} - E \right\} \langle q|E \rangle = 0. \tag{77}$$

If $\langle q|E \rangle$ is set equal to $\psi_E(q)$, Eq. (77) is indeed just the time-independent Schrödinger equation.

Jordan likewise claimed that the time-dependent Schrödinger equation is a special case of Eq. (50)

> if for β we choose the time t [this, once again, should be $-t$], for g [minus] the time $t(p, q)$ as function of p, q, and, correspondingly, for f the Hamiltonian function $H(p, q)$ (Jordan, 1927b, p. 823).

This claim is more problematic. For $\hat{\alpha} = f(\hat{p}, \hat{q}) = \hat{H}$ (eigenvalues E) and $\hat{\beta} = g(\hat{p}, \hat{q}) = -\hat{t}$, Eqs. (50)–(51) [NB1, sec. 4, Eqs. 2ab] become:

$$\left\{ \hat{H} - \frac{\hbar}{i} \frac{\partial}{\partial t} \right\} \langle q|t \rangle = 0, \tag{78}$$

$$\left\{ \hat{t} - t \right\} \langle q|t \rangle = 0. \tag{79}$$

If $\langle q|t \rangle$ is set equal to $\psi(q, t)$, Eq. (78) turns into the time-dependent Schrödinger equation. However, time is a parameter in quantum mechanics and *not* an operator $\hat{t}$ with eigenvalues t and eigenstates $|t\rangle$.[40]

This also makes Eqs. (76) and (79) problematic. Consider the former. For a free particle, the Hamiltonian is $\hat{H} = \hat{p}^2/2m$, represented by $((\hbar/i)\partial/\partial q)^2/2m$ in the q-basis. The solution of Eq. (77),

$$\langle q|E \rangle = e^{i\sqrt{2mE}q/\hbar}, \tag{80}$$

[40] There is an extensive literature on this subject. For an introduction to this issue, see, e.g., Hilgevoord (2002) and the references therein.

is also a solution of Eq. (76) as long as we define $\hat{t} \equiv m\hat{q}\hat{p}^{-1}$, as suggested by the relation $q = (p/m)t$, Note, however, that we rather arbitrarily decided on this particular ordering of the non-commuting operators $\hat{p}$ and $\hat{q}$. Using that

$$\frac{\hbar}{i}\frac{\partial}{\partial q}\, e^{i\sqrt{2mE}q/\hbar} = \sqrt{2mE}\, e^{i\sqrt{2mE}q/\hbar}, \tag{81}$$

we find that $\hat{t}\,\langle q|E\rangle = m\hat{q}\hat{p}^{-1}\langle q|E\rangle$ is given by:

$$m\hat{q}\left(\frac{\hbar}{i}\frac{\partial}{\partial q}\right)^{-1} e^{i\sqrt{2mE}q/\hbar} = \frac{mq}{\sqrt{2mE}}\, e^{i\sqrt{2mE}q/\hbar}. \tag{82}$$

This is indeed equal to $(\hbar/i)\partial/\partial E\,\langle q|E\rangle$ as required by Eq. (76):

$$\frac{\hbar}{i}\frac{\partial}{\partial E}\, e^{i\sqrt{2mE}q/\hbar} = \frac{mq}{\sqrt{2mE}}\, e^{i\sqrt{2mE}q/\hbar}. \tag{83}$$

So with $\hat{t} \equiv m\hat{q}\hat{p}^{-1}$, both Eq. (50) and Eq. (51) [NB1, sec. 4, Eqs. 2ab] hold in the special case of a free particle. It is not at all clear, however, whether this will be true in general.

It is probably no coincidence that we can get Jordan's formalism to work, albeit with difficulty, for a free particle where the energy spectrum is fully continuous. Recall that, in *Neue Begründung* I, Jordan restricted himself to quantities with completely continuous spectra. As he discovered when he tried to generalize his formalism to quantities with partly or wholly discrete spectra in *Neue Begründung* II, this restriction is not nearly as innocuous as he made it sound in *Neue Begründung* I.

Consider the canonical transformation $\hat{\alpha} = T\hat{p}T^{-1}$ (Eq. (48) [NB1, sec. 4, Eq. 1]) that plays a key role in Jordan's construction of the model realizing his axioms. Consider (in modern terms) an arbitrary eigenstate $|p\rangle$ of the operator $\hat{p}$ with eigenvalue p, i.e., $\hat{p}|p\rangle = p|p\rangle$. It only takes one line to show that then $T|p\rangle$ is an eigenstate of $\hat{\alpha}$ with the same eigenvalue p:

$$\hat{\alpha}\, T|p\rangle = T\hat{p}T^{-1}T|p\rangle = T\hat{p}|p\rangle = p\, T|p\rangle. \tag{84}$$

In other words, the operators $\hat{\alpha}$ and $\hat{p}$ connected by the canonical transformation $\hat{\alpha} = T\hat{p}T^{-1}$ have the same spectrum. This simple observation, more than anything else, reveals the limitations of Jordan's formalism. It is true, as Eq. (77) demonstrates, that his differential equations Eqs. (50)–(51) [NB1, sec. 4, Eqs. 2ab] for probability amplitudes contain the time-independent Schrödinger equation as a special case. However, since the energy spectrum is bounded from below and, in many interesting cases, at least partially discrete, it is impossible to arrive at the time-independent Schrödinger equation starting from the trivial equations (69)–(70) [NB1, sec. 5, Eq. 16] for the probability amplitude $e^{iq\beta/\hbar}$ between $\hat{q}$ and $\hat{\beta}$—recall that $\hat{\beta} = \hat{p}$ in this case (see Eq.

(68))—and performing some canonical transformation. As Eq. (84) shows, a canonical transformation cannot get us from $\hat{p}$'s and $\hat{q}$'s with completely continuous spectra to $\hat{\alpha}$'s and $\hat{\beta}$'s with partly discrete spectra. This, in turn, means that, in many interesting cases (i.e., for Hamiltonians with at least partly discrete spectra), the time-independent Schrödinger equation does *not* follow from Jordan's postulates. In Jordan's defense one could note at this point that this criticism is unfair as he explicitly restricted himself to quantities with fully continuous spectra in *Neue Begründung* I. However, as we will see when we turn to *Neue Begründung* II in Section 4, Jordan had to accept in this second paper that the extension of his general formalism to quantities with wholly or partly discrete spectra only served to drive home the problem and did nothing to alleviate it.

2.4 The confusing matter of the Ergänzungsamplitude

In this subsection, we examine the "supplementary amplitude" (*Ergänzungsamplitude*) $\psi(x, y)$ that Jordan introduced in *Neue Begründung* I in addition to the probability amplitude $\varphi(x, y)$.[41] Jordan's (1927b, p. 813) postulate I sets the conditional probability $\Pr(x|y)$ that $\hat{x}$ has a value between x and $x + dx$ given that $\hat{y}$ has the value y equal to:

$$\varphi(x, y)\psi^*(x, y)dx. \qquad \text{[NB1, sec. 2, Eq. 10]} \tag{85}$$

Jordan allowed the eigenvalues x and y to be complex. He stipulated that the "star" in $\psi^*(x, y)$ is to be interpreted in such a way that, when taking the complex conjugate of $\psi(x, y)$, one should retain x and y and *not* replace them, as the "star" would naturally suggest, by their complex conjugates. The rationale for this peculiar rule will become clear below.

For general complex amplitudes, Eq. (85) only makes sense as a positive real probability if the phases of $\varphi(x, y)$ and $\psi^*(x, y)$ exactly compensate, leaving only their positive absolute magnitudes (times the interval dx, as we are dealing with continuous quantities). Jordan certainly realized that in cases where the mechanical quantities considered were represented by self-adjoint operators, this duplication was unnecessary.[42] He seems to have felt the need, however, to advance a more general formalism, capable of dealing with the not uncommon circumstance that a canonical transformation of perfectly real (read "self-adjoint" in the quantum-mechanical case) mechanical quantities

[41] This subsection falls somewhat outside the main line of argument of our paper and can be skipped by the reader without loss of continuity.
[42] For instance, if $\hat{x}$ is a Cartesian coordinate and $\hat{y}$ is the Hamiltonian, the amplitudes $\varphi(x, y) = \psi(x, y)$ are just the Schrödinger energy eigenfunctions of the system in coordinate space.

actually leads to a new canonically conjugate, but *complex* (read "non-self-adjoint") pair of quantities. An early example of this can be found in London's (1926b) solution of the quantum harmonic oscillator by canonical transformation from the initial $(\hat{q}, \hat{p})$ coordinate-momentum pair to raising and lowering operators, which are obviously not self-adjoint (Duncan and Janssen, 2009, sec. 6.2, pp. 357–358).

Jordan could hardly have been aware at this stage of the complete absence of "nice" spectral properties in the general case of a non-self-adjoint operator, with the exception of a very special subclass to be discussed shortly. In contrast to the self-adjoint case, such operators may lack a complete set of eigenfunctions spanning the Hilbert space, or there may be an overabundance of eigenfunctions which form an "over-complete" set, in the sense that proper subsets of eigenfunctions may suffice to construct an arbitrary state. To the extent that eigenfunctions exist, the associated eigenvalues are in general complex, occupying some domain—of possibly very complicated structure—in the complex plane. In the case of the lowering operator in the simple harmonic oscillator, the spectrum occupies the entire complex plane! Instead, Jordan (1927b, p. 812) seems to have thought of the eigenvalue spectrum as lying on a curve even in the general case of arbitrary non-self-adjoint quantities.

There is one subclass of non-self-adjoint operators for which Jordan's attempt to deal with complex mechanical quantities can be given at least a limited validity. The spectral theorem usually associated with self-adjoint and unitary operators (existence and completeness of eigenfunctions) actually extends with full force to the larger class of normal operators $\hat{N}$, defined as satisfying the commutation relation $[\hat{N}, \hat{N}^\dagger] = 0$, which obviously holds for self-adjoint $(\hat{N} = \hat{N}^\dagger)$ and unitary $(\hat{N}^\dagger = \hat{N}^{-1})$ operators.[43] The reason that the spectral theorem holds for such operators is very simple: given a normal operator $\hat{N}$, we may easily construct a pair of *commuting* self-adjoint operators:

$$\hat{A} \equiv \frac{1}{2}(\hat{N} + \hat{N}^\dagger), \quad \hat{B} \equiv \frac{1}{2i}(\hat{N} - \hat{N}^\dagger). \tag{86}$$

It follows that $\hat{A} = \hat{A}^\dagger$, $\hat{B} = \hat{B}^\dagger$, and that $[\hat{A}, \hat{B}] = (i/2)\,[N, N^\dagger] = 0$. A well-known theorem assures us that a complete set of simultaneous eigenstates $|\lambda\rangle$ of $\hat{A}$ and $\hat{B}$ exist, where the parameter λ is chosen to label uniquely the state (we ignore the possibility of degeneracies here), with

$$\hat{A}\,|\lambda\rangle = \alpha(\lambda)\,|\lambda\rangle, \quad \hat{B}\,|\lambda\rangle = \beta(\lambda)\,|\lambda\rangle, \quad \hat{N}\,|\lambda\rangle = \zeta(\lambda)\,|\lambda\rangle, \tag{87}$$

where $\zeta(\lambda) \equiv \alpha(\lambda) + i\beta(\lambda)$ are the eigenvalues of $\hat{N} = \hat{A} + i\hat{B}$. Of course, there is no guarantee that $\alpha(\lambda)$ and $\beta(\lambda)$ are continuously connected (once we

[43] For discussion of the special case of finite Hermitian matrices, see Dennery and Krzywicki (1996, sec. 24.3, pp. 177–178). For a more general and more rigorous discussion, see von Neumann (1932, Ch. II, sec. 10).

eliminate the parameter λ), so the spectrum of $\hat{N}$ (the set of points $\zeta(\lambda)$ in the complex plane) may have a very complicated structure. For a normal operator, there at least exists the possibility though that the spectrum indeed lies on a simple curve, as assumed by Jordan. In fact, it is quite easy to construct an example along these lines, and to show that Jordan's two amplitudes, $\varphi(x,y)$ and $\psi(x,y)$, do exactly the right job in producing the correct probability density in the (self-adjoint) $\hat{x}$ variable for a given *complex* value of the quantity $\hat{y}$, in this case associated with a normal operator with a complex spectrum.

Our example is a simple generalization of one that Jordan (1927b, sec. 5, pp. 830–831) himself gave (in the self-adjoint case). For linear canonical transformations, the differential equations specifying the amplitudes $\varphi(x,y)$ and $\psi(x,y)$ [NB1, sec. 4, Eqs. 2ab and 3ab] are readily solved analytically. Thus, suppose that the canonical transformation from a self-adjoint conjugate pair $(\hat{p},\hat{q})$ to a new conjugate pair $(\hat{\alpha},\hat{\beta})$ is given by

$$\hat{\alpha} = a\,\hat{p} + b\,\hat{q}, \quad \hat{\beta} = c\,\hat{p} + d\,\hat{q}, \tag{88}$$

where the coefficients a,b,c,d must satisfy $ad - bc = 1$ for the transformation to be canonical, but may otherwise be complex numbers [cf. NB1, sec. 5, Eqs. 56–57]. The requirement that $\hat{\alpha}$ be a normal operator (i.e., $[\hat{\alpha},\hat{\alpha}^\dagger] = 0$) is easily seen to imply $a/a^* = b/b^*$. Thus, a and b have the same complex phase (which we may call $e^{i\vartheta}$). Likewise, normality of $\hat{\beta}$ implies that c and d have equal phase (say, $e^{i\chi}$). Moreover, the canonical condition $ad - bc = 1$ implies that the phases $e^{i\varphi}$ and $e^{i\chi}$ must cancel, so we henceforth set $\chi = -\vartheta$, and rewrite the basic canonical transformation as

$$\hat{\alpha} = \zeta(a\,\hat{p} + b\,\hat{q}), \quad \hat{\beta} = \zeta^*(c\,\hat{p} + d\,\hat{q}), \tag{89}$$

where $\zeta \equiv e^{i\vartheta}$, and a,b,c,d are now real and satisfy $ad - bc = 1$. We see that the spectrum of $\hat{\beta}$ lies along the straight line in the complex plane with phase $-\vartheta$ (as the operator $c\,\hat{p} + d\,\hat{q}$ is self-adjoint and therefore has purely real eigenvalues): the allowed values for $\hat{\beta}$ are such that $\zeta\beta$ is real.

It is now a simple matter to solve the differential equations for the amplitude $\varphi(q,\beta)$ and the supplementary amplitude $\psi(q,\beta)$ in this case. The general equations are (Jordan, 1927b, sec. 4, p. 821): [44]

[44] Jordan (1927b, p. 817) introduced the notation $F^\dagger$ for the adjoint of F in sec. 3 of his paper.

$$\left\{f\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)+\frac{\hbar}{i}\frac{\partial}{\partial\beta}\right\}\varphi(q,\beta)=0,\qquad \text{[NB1, sec. 4, Eq. 2a]} \tag{90}$$

$$\left\{g\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)-\beta\right\}\varphi(q,\beta)=0,\qquad \text{[NB1, sec. 4, Eq. 2b]} \tag{91}$$

$$\left\{f^{\dagger}\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)+\frac{\hbar}{i}\frac{\partial}{\partial\beta}\right\}\psi(q,\beta)=0,\qquad \text{[NB1, sec. 4, Eq. 3a]} \tag{92}$$

$$\left\{g^{\dagger}\left(\frac{\hbar}{i}\frac{\partial}{\partial q},q\right)-\beta\right\}\psi(q,\beta)=0.\qquad \text{[NB1, sec. 4, Eq. 3b]} \tag{93}$$

The differential operators in this case are (recall that $\hat{\alpha} = f(\hat{p},\hat{q})$ and $\hat{\beta} = g(\hat{p},\hat{q})$ (see Eqs. (48)–(49) [NB1, sec. 4, Eq. 1]):

$$\begin{aligned} f &= \zeta\left(a\,\frac{\hbar}{i}\frac{\partial}{\partial q}+b\,q\right), \qquad & f^{\dagger} &= \zeta^{*}\left(a\,\frac{\hbar}{i}\frac{\partial}{\partial q}+b\,q\right), \\ g &= \zeta^{*}\left(c\,\frac{\hbar}{i}\frac{\partial}{\partial q}+d\,q\right), \qquad & g^{\dagger} &= \zeta\left(c\,\frac{\hbar}{i}\frac{\partial}{\partial q}+d\,q\right), \end{aligned} \tag{94}$$

and for $\varphi(q,\beta)$ and $\psi(q,\beta)$ we find (up to an overall constant factor):

$$\begin{aligned} \varphi(q,\beta) &= \exp\left\{-\frac{i}{\hbar}\left(\frac{d}{2c}q^2-\frac{1}{c}\zeta\beta q+\frac{a}{2c}(\zeta\beta)^2\right)\right\}, \\ \psi(q,\beta) &= \exp\left\{-\frac{i}{\hbar}\left(\frac{d}{2c}q^2-\frac{1}{c}\zeta^{*}\beta q+\frac{a}{2c}(\zeta^{*}\beta)^2\right)\right\}. \end{aligned} \tag{95}$$

We note that the basic amplitude $\varphi(q,\beta)$ is a pure oscillatory exponential, as the combinations $\zeta\beta$ and the constants a, c, and d appearing in the exponent are all real, so the exponent is overall purely imaginary, and the amplitude has unit absolute magnitude. This is not the case for $\psi(q,\beta)$, due to the appearance of ζ^{*}, but at this point we recall that, according to Jordan's postulate A, the correct probability density is obtained by multiplying $\varphi(x,y)$ by $\psi^{*}(x,y)$, where the star symbol includes the instruction that the eigenvalue y of $\hat{y}$ is *not to be conjugated* (cf. our comment following Eq. (85)). This rather strange prescription is essential if we are to maintain consistency of the orthogonality property

$$\int \varphi(x,y'')\psi^{*}(x,y')dx = \delta_{y'y''} \tag{96}$$

with the differential equations (91) and (93) for the amplitudes [NB1, sec. 4, Eqs. 2b and 3b]. With this proviso, we find (recalling again that $\zeta\beta$ is real):

$$\psi^{*}(q,\beta) = \exp\left\{\frac{i}{\hbar}\left(\frac{d}{2c}q^2-\frac{1}{c}\zeta\beta q+\frac{a}{2c}(\zeta\beta)^2\right)\right\} = \bar{\varphi}(q,\beta), \tag{97}$$

where the bar now denotes conventional complex conjugation, and we see that the product $\varphi(q,\beta)\psi^*(q,\beta) = \varphi(q,\beta)\bar{\varphi}(q,\beta)$ is indeed real, and in fact, equal to unity, as we might expect in the case of a purely oscillatory wave function.

That Jordan's prescription for the construction of conditional probabilities cannot generally be valid in the presence of classical complex (or quantum-mechanically non-self-adjoint) quantities is easily verified by relaxing the condition of normal operators in the preceding example. In particular, we consider the example of the raising and lowering operators for the simple harmonic oscillator, obtained again by a complex linear canonical transformation of the $(\hat{q},\hat{p})$ canonical pair.[45] Now, as a special case of Eq. (88), we take

$$\hat{\alpha} = \frac{1}{\sqrt{2}}(\hat{p} + i\hat{q}) = f(\hat{p},\hat{q}), \quad \hat{\beta} = \frac{1}{\sqrt{2}}(i\hat{p} + \hat{q}) = g(\hat{p},\hat{q}), \tag{98}$$

which, though canonical, clearly does not correspond to normal operators, as the coefficients a, b (and likewise c, d) are now 90 degrees out of phase. Solving the differential equations for the amplitudes, Eqs. (90)–(93) [NB1, sec. 4, Eqs. 2ab, 3ab], we now find (up to an overall constant factor [cf. Eq. (95)]):

$$\begin{aligned} \varphi(q,\beta) &= \exp\left\{-\frac{1}{2\hbar}\left(q^2 - 2\sqrt{2}\beta q + \beta^2\right)\right\}, \\ \psi(q,\beta) &= \exp\left\{\frac{1}{2\hbar}\left(q^2 - 2\sqrt{2}\beta q + \beta^2\right)\right\} = \psi^*(q,\beta) = \frac{1}{\varphi(q,\beta)}, \end{aligned} \tag{99}$$

where β is an *arbitrary* complex number. In fact, the wave function $\varphi(q,\beta)$ is a *square-integrable* eigenfunction of $\hat{\beta}$ for an arbitrary complex value of β: it corresponds to the well-known "coherent eigenstates" of the harmonic oscillator, with the envelope (absolute magnitude) of the wave function executing simple harmonic motion about the center of the potential well with frequency ω (given the Hamiltonian $\hat{H} = \frac{1}{2}(\hat{p}^2 + \omega^2\hat{q}^2)$). The probability density in q of such a state for fixed β is surely given by the conventional prescription $|\varphi(q,\beta)|^2$. On the other hand, for complex β, the Jordan prescription requires us to form the combination (with the peculiar interpretation of the "star" in $\psi^*(q,\beta)$, in which β is *not* conjugated):

$$\varphi(q,\beta)\psi^*(q,\beta) = 1, \tag{100}$$

which clearly makes no physical sense in this case, as the state in question is a localized, square-integrable one. If Jordan's notion of an *Ergänzungsamplitude* is to have any nontrivial content, it would seem to require, at the very least,

[45] As mentioned above, London (1926b) had looked at this example of a canonical transformation (Duncan and Janssen, 2009, sec. 6.2, p. 358).

that the complex quantities considered fall into the very special category of normal operators after quantization.

In fact, as we will see shortly, in the paper by Hilbert, von Neumann, and Nordheim (1928) on Jordan's version of statistical transformation theory, the requirement of self-adjointness already acquires the status of a *sine qua non* for physical observables in quantum theory, and the concept of an *Ergänzungsamplitude* disappears even from Jordan's own treatment of his theory after *Neue Begründung* I.

3 Hilbert, von Neumann, and Nordheim's *Grundlagen* (April 1927)

In the winter semester of 1926/27, Hilbert gave a course entitled "Mathematical methods of quantum theory." The course consisted of two parts. The first part, "The older quantum theory," was essentially a repeat of the course that Hilbert had given under the same title in 1922/23. The second part, "The new quantum theory," covered the developments since 1925. As he had in 1922/23, Nordheim prepared the notes for this course, which were recently published (Sauer and Majer, 2009, pp. 504–707; the second part takes up pp. 609–707). At the very end (ibid., pp. 700–706), we find a concise exposition of the main line of reasoning of Jordan's *Neue Begründung* I.

This presentation served as the basis for a paper by Hilbert, von Neumann, and Nordheim (1928). As the authors explained in the introduction (ibid., pp. 1–2), "important parts of the mathematical elaboration" were due to von Neumann, while Nordheim was responsible for the final text (Duncan and Janssen, 2009, p. 361). The paper was submitted to the *Mathematische Annalen* April 6, 1927, but, for reasons not clear to us, was published only in 1928, i.e., well after the appearance of the trilogy by von Neumann (1927a,b,c) that rendered much of it obsolete. In this section we go over the main points of this three-man paper. [46]

In the lecture notes for Hilbert's course, Dirac is not mentioned at all, and even though in the paper it is acknowledged that Dirac (1927) had independently arrived at and published similar results, the focus continues to be on Jordan. There are only a handful of references to Dirac, most importantly in connection with the delta function and in the discussion of the Schrödinger equation for a Hamiltonian with a partly discrete spectrum (Hilbert, von Neumann, and Nordheim, 1928, p. 8 and p. 30, respectively). Both the lecture notes and the

[46] For other discussions of Hilbert, von Neumann, and Nordheim (1928), see Jammer (1966, 309–312) and Lacki (2000, pp. 295–300), who mainly focuses on its axiomatic structure.

paper stay close to the relevant sections of *Neue Begründung* I, but Hilbert and his collaborators did change Jordan's notation. Their notation is undoubtedly an improvement over his—not a high bar to clear—but the modern reader trying to follow the argument in these texts may still want to translate it into the kind of modern notation we introduced in Section 2. We will adopt the notation of Hilbert and his co-authors in this section, except that we will continue to use hats to distinguish (operators representing) mechanical quantities from their numerical values.

As we mentioned in Section 2.1, when we discussed postulates A through D of *Neue Begründung* I, Hilbert, von Neumann, and Nordheim (1928, pp. 4–5) based their exposition of Jordan's theory on six "physical axioms."[47] Axiom I introduces the basic idea of a probability amplitude. The amplitude for the probability that a mechanical quantity $\hat{F}_1(\hat{p}\,\hat{q})$ (some function of momentum $\hat{p}$ and coordinate $\hat{q}$) has the value x given that another such quantity $\hat{F}_2(\hat{p}\,\hat{q})$ has the value y is written as $\varphi(x\,y; \hat{F}_1\,\hat{F}_2)$.

Jordan's *Ergänzungsamplitude* still made a brief appearance in the notes for Hilbert's course (Sauer and Majer, 2009, p. 700) but is silently dropped in the paper. As we saw in Section 2, amplitude and supplementary amplitude are identical as long as we only consider quantities represented, in modern terms, by Hermitian operators. In that case, the probability $w(x\,y; \hat{F}_1\,\hat{F}_2)$ of finding the value x for $\hat{F}_1$ given the value y for $\hat{F}_2$ is given by the product of $\varphi(x\,y; \hat{F}_1\,\hat{F}_2)$ and its complex conjugate, which, of course, will always be a real number. Although they did not explicitly point out that this eliminates the need for the *Ergänzungsamplitude*, Hilbert, von Neumann, and Nordheim (1928, p. 17–25) put great emphasis on the restriction to Hermitian operators. Secs. 6–8 of their paper ("The reality conditions," "Properties of Hermitian operators," and "The physical meaning of the reality conditions") are devoted to this issue.

Axiom II corresponds to Jordan's postulate B and says that the amplitude for finding a value for $\hat{F}_2$ given the value of $\hat{F}_1$ is the complex conjugate of the amplitude of finding a value for $\hat{F}_1$ given the value of $\hat{F}_2$. This symmetry property entails that these two outcomes have the same probability. Axiom III is not among Jordan's postulates. It basically states the obvious demand that when $\hat{F}_1 = \hat{F}_2$, the probability $w(x\,y; \hat{F}_1\,\hat{F}_2)$ be either 0 (if $x \neq y$) or 1 (if $x = y$). Axiom IV corresponds to Jordan's postulate C and states that the amplitudes rather than the probabilities themselves follow the usual composition rules for probabilities (cf. Eqs. (1) and (3) in Section 2.1):

$$\varphi(x\,z; \hat{F}_1\,\hat{F}_3) = \int \varphi(x\,y; \hat{F}_1\,\hat{F}_2)\,\varphi(y\,z; \hat{F}_2\,\hat{F}_3)\,dy. \tag{101}$$

[47] In the lecture notes we find four axioms that are essentially the same as Jordan's four postulates (Sauer and Majer, 2009, pp. 700–701).

Though they did not use Jordan's phrase "interference of probabilities," the authors emphasized the central importance of this particular axiom:

> This requirement [Eq. (101)] is obviously analogous to the addition and multiplication theorems of ordinary probability calculus, except that in this case they hold for the amplitudes rather than for the probabilities themselves.
>
> The characteristic difference to ordinary probability calculus lies herein that initially, instead of the probabilities themselves, amplitudes occur, which in general will be complex quantities and only give ordinary probabilities if their absolute value is taken and then squared (Hilbert, von Neumann, and Nordheim, 1928, p. 5)

Axiom V, as we already mentioned in Section 2.1, makes part of Jordan's postulate A into a separate axiom. It demands that probability amplitudes for quantities $\hat{F}_1$ and $\hat{F}_2$ depend only on the functional dependence of these quantities on $\hat{q}$ and $\hat{p}$ and not on "special properties of the system under consideration, such as, for example, its Hamiltonian" (ibid., p. 5). Axiom VI, finally, adds another obvious requirement to the ones recognized by Jordan: that probabilities be independent of the choice of coordinate systems.

Before they introduced the axioms, in a passage that we quoted in the introduction, Hilbert, von Neumann, and Nordheim (1928, p. 2) had already explained that the task at hand was to find "a simple analytical apparatus in which quantities occur that satisfy" axioms I–VI. As we know from *Neue Begründung* I, the quantities that fit the bill are the integral kernels of certain canonical transformations, implemented as $T\hat{p}T^{-1}$ and $T\hat{q}T^{-1}$ (cf. Eqs. (48)–(49)). After introducing this "simple analytical apparatus" in secs. 3–4 ("Basic formulae of the operator calculus," "Canonical operators and transformations"), the authors concluded in sec. 5 ("The physical interpretation of the operator calculus"):

> *The probability amplitude $\varphi(x\,y; \hat{q}\,\hat{F})$ between the coordinate $\hat{q}$ and an arbitrary mechanical quantity $\hat{F}(\hat{q}\,\hat{p})$—i.e., for the situation that for a given value y of $\hat{F}$, the coordinate lies between x and $x + dx$—is given by the kernel of the integral operator that canonically transforms the operator $\hat{q}$ into the operator corresponding to the mechanical quantity $\hat{F}(\hat{q}\,\hat{p})$* (Hilbert, von Neumann, and Nordheim, 1928, p. 14; italics in the original, hats added).

They immediately generalized this definition to cover the probability amplitude between two arbitrary quantities $\hat{F}_1$ and $\hat{F}_2$. In sec. 3, the authors already derived differential equations for integral kernels $\varphi(x\,y)$ (ibid., pp. 10–11, Eqs. (19ab) and (21ab)). Given the identification of these integral kernels with probability amplitudes in sec. 5, these equation are just Jordan's fundamental differential equations for the latter (NB1, sec. 4, Eqs. (2ab); our Eqs. (50)–(51) in Section 2.3).

In sec. 4, they also stated the key assumption that any quantity of interest can be obtained through a canonical transformation starting from some canonically conjugate pair of quantities $\hat{p}$ and $\hat{q}$:

> We will assume that every operator $\hat{F}$ can be generated out of the basic operator $\hat{q}$ by a canonical transformation. This statement can also be expressed in the following way, namely that, given $\hat{F}$, the operator equation $T\hat{q}T^{-1}$ has to be solvable.
>
> The conditions that $\hat{F}$ has to satisfy for this to be possible will not be investigated here (Hilbert, von Neumann, and Nordheim, 1928, p. 12; hats added).

What this passage suggests is that the authors, although they recognized the importance of this assumption, did not quite appreciate that, as we showed at the end of Section 2.3, it puts severe limits on the applicability of Jordan's formalism. In the simple examples of canonical transformations ($\hat{F} = f(\hat{q})$ and $\hat{F} = \hat{p}$) that they considered in sec. 9 ("Application of the theory to special cases"), the assumption is obviously satisfied and the formalism works just fine (ibid., pp. 25–26). In sec. 10 ("The Schrödinger differential equations"), however, they set $\hat{F}$ equal to the Hamiltonian $\hat{H}$ and claimed that one of the differential equations for the probability amplitude $\varphi(x\,W; \hat{q}\,\hat{H})$ (where W is an energy eigenvalue) is the time-independent Schrödinger equation. As soon as the Hamiltonian has a wholly or partly discrete spectrum, however, there simply is no operator T such that $\hat{H} = T\hat{q}T^{-1}$.

In secs. 6–8, which we already briefly mentioned above, Hilbert, von Neumann, and Nordheim (1928, pp. 17–25) showed that the necessary and sufficient condition for the probability $w(x\,y; \hat{F}_1\,\hat{F}_2)$ to be real is that $\hat{F}_1$ and $\hat{F}_2$ are both represented by Hermitian operators. As we pointed out earlier, they implicitly rejected Jordan's attempt to accommodate $\hat{F}$'s represented by non-Hermitian operators through the introduction of the *Ergänzungsamplitude*. They also showed that the operator representing the canonical conjugate $\hat{G}$ of a quantity $\hat{F}$ represented by a Hermitian operator is itself Hermitian.

The authors ended their paper on a cautionary note emphasizing its lack of mathematical rigor. They referred to von Neumann's (1927a) forthcoming paper, *Mathematische Begründung*, for a more satisfactory treatment of the Schrödinger equation for Hamiltonians with a partly discrete spectrum. In the concluding paragraph, they warned the reader more generally:

> In our presentation the general theory receives such a perspicuous and formally simple form that we have carried it through in a mathematically still imperfect form, especially since a fully rigorous presentation might well be considerably more tedious and circuitous (Hilbert, von Neumann, and Nordheim, 1928, p. 30).

4 Jordan's *Neue Begründung* II (June 1927)

On June 3, 1927, while at Bohr's institute in Copenhagen on an International Education Board fellowship, Jordan (1927e) submitted *Neue Begründung* II to *Zeitschrift für Physik*. In the abstract he announced a "simplified and generalized" version of the theory presented in *Neue Begründung* I.[48]

One simplification was that Jordan, like Hilbert, von Neumann, and Nordheim (1928), dropped the *Ergänzungsamplitude* and restricted himself accordingly to physical quantities represented by Hermitian operators and to canonical transformations preserving Hermiticity. Another simplification was that he adopted Dirac's (1927) convention of consistently using the same letter for a mechanical quantity and its possible values, using primes to distinguish the latter from the former. When, for instance, the letter β is used for some quantity, its values are denoted as β', β'', etc. We will continue to use the notation $\hat{\beta}$ for the quantity (and the operator representing that quantity) and the notation $\beta, \beta', \ldots$ for its values. While this new notation for quantities and their values was undoubtedly an improvement, the new notation for probability amplitudes and for transformation operators with those amplitudes as their integral kernels is actually more cumbersome than in *Neue Begründung* I.

In the end, however, these new notational complications only affect the cosmetics of the paper. What is far more troublesome is that the generalization of the formalism promised in the abstract to handle cases with wholly or partly discrete spectra is much more problematic than Jordan suggested and, we argue, ultimately untenable. By the end of *Neue Begründung* II, Jordan is counting quantities nobody would think of as canonically conjugate (e.g., different components of spin) as pairs of conjugate variables and has abandoned the notion, central to the formalism of *Neue Begründung* I, that any quantity of interest (e.g, the Hamiltonian) is a member of a pair of conjugate variables connected to some initial pair of $\hat{p}$'s and $\hat{q}$'s by a canonical transformation. We can fairly characterize the state of affairs by saying that, although Jordan is still clinging to his p's and q's, they have effectively ceased to play any significant role in his formalism.

As we showed at the end of Section 2.3, the canonical transformation

$$\hat{\alpha} = T\hat{p}T^{-1}, \quad \hat{\beta} = T\hat{q}T^{-1} \tag{102}$$

(cf. Eqs. (48)–(49)) can never get us from a quantity with a completely continuous spectrum (such as position or momentum) to a quantity with a discrete

[48] Interviewed in 1963 by Kuhn, who complained about the "dreadful notation" of *Neue Begründung* I, Jordan said he just wanted to give a "prettier and clearer" exposition of the same material (Duncan and Janssen, 2009, p. 360).

spectrum (such as the Hamiltonian). In *Neue Begründung* II, Jordan (1927e, pp. 16–17) evidently recognized this problem even though it is not clear that he realized the extent to which this undercuts his entire approach.

The central problem is brought out somewhat indirectly in the paper. As Jordan (1927e, pp. 1–2) already mentioned in the abstract and then demonstrated in the introduction, the commutation relation, $[\hat{p}, \hat{q}] = \hbar/i$, for two canonically conjugate quantities $\hat{p}$ and $\hat{q}$ cannot hold as soon as the spectrum of one of them is partly discrete. Specifically, this means that action-angle variables $\hat{J}$ and $\hat{w}$, where the eigenvalues of the action variable $\hat{J}$ are restricted to integral multiples of Planck's constant, cannot satisfy the canonical commutation relation.

The proof of this claim is very simple (ibid., p. 2). Jordan considered a pair of conjugate quantities $\hat{\alpha}$ and $\hat{\beta}$ where $\hat{\beta}$ is assumed to have a purely discrete spectrum. We will see that one runs into the same problem as soon as one $\hat{\alpha}$ or $\hat{\beta}$ has a single discrete eigenvalue. Suppose $\hat{\alpha}$ and $\hat{\beta}$ satisfy the standard commutation relation:

$$[\hat{\alpha}, \hat{\beta}] = \frac{\hbar}{i}. \tag{103}$$

As Jordan pointed out, it then follows that an operator that is some function F of $\hat{\beta}$ satisfies [49]

$$[\hat{\alpha}, F(\hat{\beta})] = \frac{\hbar}{i} F'(\hat{\beta}). \tag{104}$$

Jordan now chose a function such that $F(\beta) = 0$ for all eigenvalues $\beta_1, \beta_2, \ldots$ of $\hat{\beta}$, while $F'(\beta) \neq 0$ at those same points. In that case, the left-hand side of Eq. (104) vanishes at all these points, whereas the right-hand side does not. Hence, Eq. (104) and, by *modus tollens*, Eq. (103) cannot hold.

Since Eq. (104) is an operator equation, we should, strictly speaking, compare the results of the left-hand side and the right-hand side acting on some state. To show that Eq. (104)—and thereby Eq. (103)—cannot hold, it suffices to show that it does not hold for one specific function F and one specific state $|\psi\rangle$. Consider the simple function $F_1(\beta) = \beta - \beta_1$, for which $F_1(\beta_1) = 0$ and $F_1'(\beta_1) = 1$, and the discrete (and thus normalizable) eigenstate $|\beta_1\rangle$ of the

[49] If the function $F(\beta)$ is assumed to be a polynomial, $\sum_n c_n \beta^n$, which is all we need for what we want to prove although Jordan (1927e, p. 2) considered a "fully transcendent function," Jordan's claim is a standard result in elementary quantum mechanics:

$$[\hat{\alpha}, F(\hat{\beta})] = [\hat{\alpha}, \sum_n c_n \hat{\beta}^n] = \sum_n c_n \, n \, \frac{\hbar}{i} \, \hat{\beta}^{n-1} = \frac{\hbar}{i} \frac{d}{d\hat{\beta}} \left(\sum_n c_n \hat{\beta}^n \right) = \frac{\hbar}{i} \, F'(\hat{\beta}),$$

where in the second step we repeatedly used that $[\hat{\alpha}, \hat{\beta}] = \hbar/i$ and that $[\hat{A}, \hat{B}\hat{C}] = [\hat{A}, \hat{B}]\hat{C} + \hat{B}[\hat{A}, \hat{C}]$ for any three operators $\hat{A}$, $\hat{B}$, and $\hat{C}$.

operator corresponding to the quantity $\hat{\beta}$. Clearly,

$$\langle\beta_1|[\hat{\alpha}, F_1(\hat{\beta})]|\beta_1\rangle = \langle\beta_1|[\hat{\alpha}, \hat{\beta} - \beta_1]|\beta_1\rangle = 0, \tag{105}$$

as $\hat{\beta}|\beta_1\rangle = \beta_1|\beta_1\rangle$, while

$$\langle\beta_1|(\hbar/i)F_1'(\hat{\beta})]|\beta_1\rangle = \frac{\hbar}{i}\langle\beta_1|\beta_1\rangle = \frac{\hbar}{i}, \tag{106}$$

as $F_1'(\hat{\beta})|\beta_1\rangle = F_1'(\beta_1)|\beta_1\rangle = |\beta_1\rangle$. This shows that the relation,

$$\langle\psi|[\hat{\alpha}, F(\hat{\beta})]|\psi\rangle = \langle\psi|\frac{\hbar}{i}F'(\hat{\beta})|\psi\rangle, \tag{107}$$

and hence Eqs. (103)–(104), cannot hold. The specific example $F_1(\beta) = \beta - \beta_1$ that we used above immediately makes it clear that the commutation relation $[\hat{\alpha}, \hat{\beta}] = \hbar/i$ cannot hold as soon as either one of the two operators has a single discrete eigenvalue.

Much later in the paper, in sec. 4 ("Canonical transformations"), Jordan (1927e, p. 16) acknowledged that it follows directly from this result that no canonical transformation can ever get us from a pair of conjugate variables $\hat{p}$'s and $\hat{q}$'s with completely continuous spectra to $\hat{\alpha}$'s and $\hat{\beta}$'s with partly discrete spectra. It is, after all, an essential property of canonical transformations that they preserve canonical commutation relations. From Eq. (102) and $[\hat{p}, \hat{q}] = \hbar/i$ it follows that

$$[\hat{\alpha}, \hat{\beta}] = [T\hat{p}T^{-1}, T\hat{q}T^{-1}] = T[\hat{p}, \hat{q}]T^{-1} = \hbar/i. \tag{108}$$

Since, as we just saw, only quantities with purely continuous spectra can satisfy this commutation relation, Eq. (108) cannot hold for $\hat{\alpha}$'s and $\hat{\beta}$'s with partly discrete spectra and such $\hat{\alpha}$'s and $\hat{\beta}$'s cannot possibly be obtained from $\hat{p}$ and $\hat{q}$ through a canonical transformation of the form (102).

We will discuss below how this obstruction affects Jordan's general formalism. When Jordan, in the introduction of *Neue Begründung* II, showed that no quantity with a partly discrete spectrum can satisfy a canonical commutation relation, he presented it not as a serious problem for his formalism but as an argument for the superiority of his alternative definition of conjugate variables in *Neue Begründung* I (Jordan, 1927b, p. 814, cf. Eq. (4)). In that definition $\hat{p}$ and $\hat{q}$ are considered canonically conjugate if the probability amplitude $\varphi(p, q)$ has the simple form $e^{-ipq/\hbar}$, which means that as soon as the value of one of the quantities $\hat{p}$ and $\hat{q}$ is known, all possible values of the other quantity are equiprobable. As we saw in Section 2.1, Jordan showed that for $\hat{p}$'s and $\hat{q}$'s with purely continuous spectra this implies that they satisfy $[\hat{p}, \hat{q}] = \hbar/i$ (cf. Eq. (34)), which is the standard definition of what it means for $\hat{p}$ and $\hat{q}$ to be conjugate variables. In *Neue Begründung* II, Jordan (1927e, p. 6, Eq. (C)) extended his alternative definition to quantities with wholly or partly discrete spectra, in which case the new definition, of course, no longer reduces to the standard one.

As Jordan (1927e) wrote in the opening paragraph, his new paper only assumes a rough familiarity with the considerations of *Neue Begründung* I. He thus had to redevelop much of the formalism of his earlier paper, while trying to both simplify and generalize it at the same time. In sec. 2 ("Basic properties of quantities and probability amplitudes"), Jordan began by restating the postulates to be satisfied by his probability amplitudes.

He introduced a new notation for these amplitudes. Instead of $\varphi(\beta, q)$ (cf. note 17) he now wrote $\Phi_{\alpha p}(\beta', q')$. The primes, as we explained above, distinguish values of quantities from those quantities themselves. The subscripts α and p denote which quantities are canonically conjugate to the quantities $\hat{\beta}$ and $\hat{q}$ for which the probability amplitude is being evaluated. As we will see below, one has a certain freedom in picking the $\hat{\alpha}$ and $\hat{p}$ conjugate to $\hat{\beta}$ and $\hat{q}$, respectively, and settling on a specific pair of $\hat{\alpha}$ and $\hat{p}$ is equivalent to fixing the phase ambiguity of the amplitude $\varphi(\beta, q)$ up to some constant factor. So for a given choice of $\hat{\alpha}$ and $\hat{p}$, the amplitude $\Phi_{\alpha p}(\beta', q')$ is essentially unique. In this way, Jordan (1927e, p. 20) could answer, at least formally, von Neumann's (1927a, p. 3) objection that probability amplitudes are not uniquely determined even though the resulting probabilities are. It is only made clear toward the end of the paper that this is the rationale behind these additional subscripts. Their only other role is to remind the reader that $\Phi_{\alpha p}(\beta', q')$ is determined not by one Schrödinger-type equation in Jordan's formalism but by a pair of such equations involving both canonically conjugate pairs of variables, $(\hat{p}, \hat{q})$ and $(\hat{\alpha}, \hat{\beta})$ (Jordan, 1927e, p. 20). As none of this is essential to the formalism, we will simply continue to use the notation $\langle \beta | q \rangle$ for the probability amplitude between the quantities $\hat{\beta}$ and $\hat{q}$.

Jordan also removed the restriction to systems of one degree of freedom that he had adopted for convenience in *Neue Begründung* I (Jordan, 1927b, p. 810). So $\hat{q}$, in general, now stands for $(\hat{q}_1, \ldots \hat{q}_f)$, where f is the number of degrees of freedom of the system under consideration. The same is true for other quantities. Jordan (1927e, pp. 4–5) spent a few paragraphs examining the different possible structures of the space of eigenvalues for such f-dimensional quantities depending on the nature of the spectrum of its various components—fully continuous, fully discrete, or combinations of both. He also introduced the notation $\delta(\beta' - \beta'')$ for a combination of the Dirac delta function and the Kronecker delta (or, as Jordan called the latter, the "Weierstrassian symbol").

In *Neue Begründung* II (Jordan, 1927e, p. 6), the four postulates of *Neue Begründung* I (see our discussion in Section 2.1) are replaced by three postulates or "axioms," as Jordan now also called them (perhaps), numbered with

Roman numerals. This may have been in deference to Hilbert, von Neumann, and Nordheim (1928), although they listed six such axioms (as we saw in Section 3). Jordan's new postulates or axioms do not include the key portion of postulate A of *Neue Begründung* I stating the probability interpretation of the amplitudes. That is relegated to sec. 5, "The physical meaning of the amplitudes" (Jordan, 1927e, p. 19). Right before listing the postulates, however, Jordan (1927e, p. 5) did mention that he will only consider "real (Hermitian) quantities," thereby obviating the need for the *Ergänzungsamplitude* and simplifying the relation between amplitudes and probabilities. There is no discussion of the *Ergänzungsamplitude* amplitude in the paper. Instead, following the lead of Hilbert, von Neumann, and Nordheim (1928), Jordan silently dropped it. It is possible that this was not even a matter of principle for Jordan but only one of convenience. Right after listing the postulates, he wrote that the restriction to real quantities is made only "on account of simplicity" (*der Einfachkeit halber*, ibid., p. 6).

Other than the probability-interpretation part of postulate A, all four postulates of *Neue Begründung* I return, generalized from one to f degrees of freedom and from quantities with completely continuous spectra to quantities with wholly or partly discrete spectra. Axiom I corresponds to the old postulate D. It says that for every generalized coordinate there is a conjugate momentum. Axiom II consists of three parts, labeled (A), (B), and (C). Part (A) corresponds to the old postulate B, asserting the symmetry property, which, in the new notation, becomes:

$$\Phi_{\alpha p}(\beta', q') = \Phi^*_{p\alpha}(q', \beta'). \tag{109}$$

Part (B) corresponds to the old postulate C, which gives the basic rule for the composition of probability amplitudes,

$$\overline{\sum_{q'}}\, \Phi_{\alpha p}(\beta', q')\Phi_{pP}(q', Q') = \Phi_{\alpha P}(\beta', Q'), \tag{110}$$

where the notation $\overline{\sum}_q$ indicates that, in general, we need a combination of integrals over the continuous parts of the spectrum of a quantity and sums over its discrete parts. In Eq. (110), $\overline{\sum}_q$ refers to an ordinary integral as the coordinate $\hat{q}$ has a purely continuous spectrum. Adopting this $\overline{\sum}$ notation, we can rewrite the composition rule (110) in the modern language introduced in Section 2 and immediately recognize it as a completeness relation (cf. Eqs. (1) and (3)):

$$\overline{\sum_{q}} \langle\, \beta|q\rangle\langle q|Q\rangle = \langle\beta|Q\rangle. \tag{111}$$

We can likewise formulate orthogonality relations, as Jordan (1927e, p. 7, Eq. (5)) did at the beginning of sec. 2 ("Consequences"):

$$\overline{\sum_{q}} \langle\, \beta|q\rangle\langle q|\beta'\rangle = \delta(\beta - \beta'), \quad \overline{\sum_{\beta}} \langle\, q|\beta\rangle\langle\beta|q'\rangle = \delta(q - q'). \tag{112}$$

Recall that $\delta(\beta-\beta')$ can be either the Dirac delta function or the Kronecker delta, as $\hat{\beta}$ can have either a fully continuous or a partly or wholly discrete spectrum. The relations in Eq. (112) can, of course, also be read as completeness relations, i.e., as giving two different resolutions,

$$\overline{\sum_q}|q\rangle\langle q|, \quad \overline{\sum_\beta}|\beta\rangle\langle\beta|, \tag{113}$$

of the unit operator. Part (C) of axiom II is the generalization of the definition of conjugate variables familiar from *Neue Begründung* I to f degrees of freedom and to quantities with wholly or partly discrete spectra. Two quantities $\hat{\alpha}=(\hat{\alpha}_1,\ldots,\hat{\alpha}_f)$ and $\hat{\beta}=(\hat{\beta}_1,\ldots,\hat{\beta}_f)$ are canonically conjugate to one another if

$$\Phi_{\alpha,-\beta}(\beta,\alpha)=C\,e^{i\left(\sum_{k=1}^{f}\beta_k\alpha_k\right)/\hbar}, \tag{114}$$

where C is a normalization constant. [50] Axiom III, finally, is essentially axiom III of Hilbert, von Neumann, and Nordheim (1928, p. 4), which was not part of *Neue Begründung* I and which says, in our notation, that $\langle\beta|\beta'\rangle=\delta(\beta-\beta')$, where, once again, $\delta(\beta-\beta')$ can be either the Dirac delta function or the Kronecker delta.

We need to explain one more aspect of Jordan's notation in *Neue Begründung* II. As we have seen in Sections 2 and 3, probability amplitudes do double duty as integral kernels of canonical transformations. Jordan (1927e, p. 6, Eqs. (A′)-(B′)) introduced the special notation $\Phi^{\beta q}_{\alpha p}$ to indicate that the amplitude $\Phi_{\alpha p}(\beta',q')$ serves as such an integral kernel, thinking of $\Phi^{\beta q}_{\alpha p}$ as 'matrices' with β and q as 'indices' that, in general, will take on both discrete and continuous values. We will continue to use the modern notation $\langle\beta|q\rangle$ both when we want to think of this quantity as a probability amplitude and when we want to think of it as a transformation 'matrix'. The notation of *Neue Begründung* II, like the modern notation, clearly brings out the double role of this quantity. In *Neue Begründung* I, we frequently encountered canonical transformations such as $\hat{\alpha}=T\hat{p}T^{-1}$, $\hat{\beta}=T\hat{p}T^{-1}$ (cf. Eqs. (48)–(49)). In *Neue Begründung* II, such transformations are written with $\Phi^{\beta q}_{\alpha p}$'s instead of T's. As we will explain in detail below, this conceals an important shift in Jordan's usage of such equations. This shift is only made explicit in sec. 4, which, as its title announces, deals specifically with "Canonical transformations." Up to that point, and especially in sec. 3, "The functional equations of the amplitudes," Jordan appears to be vacillating between two different interpretations of these canonical transformation equations, the one of *Neue Begründung* I in which $\hat{\alpha}$ and $\hat{\beta}$ are *new* conjugate variables different from the $\hat{p}$ and $\hat{q}$ we started from, and one, inspired by Dirac (1927), as Jordan (1927e, pp. 16–17) acknowledged

[50] Contrary to what Jordan (1927e, p. 7) suggested, the sign of the exponent in Eq. (114) agrees with the sign of the exponent in the corresponding formula in *Neue Begründung* I (Jordan, 1927b, p. 814, Eq. (18)).

in sec. 4, in which $\hat{\alpha}$ and $\hat{\beta}$ are still the *same* $\hat{p}$ and $\hat{q}$ but expressed with respect to a *new basis*.

Before he got into any of this, Jordan (1927e, sec. 2, pp. 8–10) examined five examples, labeled (a) through (e), of what he considered to be pairs of conjugate quantities and convinced himself that they indeed qualify as such under his new definition (114). More specifically, he checked in these five cases that these purportedly conjugate pairs of quantities satisfy the completeness or orthogonality relations (112). The examples include familiar pairs of canonically conjugate variables, such as action-angle variables (Jordan's example (c)), but also quantities that we normally would not think of as conjugate variables, such as different components of spin (a special case of example (e)). We take a closer look at these two specific examples.

In example (c), the allegedly conjugate variables are the angle variable $\hat{w}$ with a purely continuous spectrum and eigenvalues $w \in [0,1]$ (which means that the eigenvalues of a true angle variable $\hat{\vartheta} \equiv 2\pi w$ are $\vartheta \in [0, 2\pi]$) and the action variable $\hat{J}$ with a purely discrete spectrum and eigenvalues $J = C + nh$, where C is an arbitrary (real) constant and n is a positive integer. For convenience we set $C = -1$, so that $J = mh$ with $m = 0, 1, 2, \ldots$ The probability amplitude $\langle w|J\rangle$ has the form required by Jordan's definition (114) of conjugate variables, with $\hat{\alpha} = \hat{J}$, $\hat{\beta} = \hat{w}$, and $f = 1$:

$$\langle w|J\rangle = e^{iwJ/\hbar}. \tag{115}$$

We now need to check whether $\langle w|J\rangle$ satisfies the two relations in Eq. (112):

$$\int_0^1 dw\, \langle J_{n_1}|w\rangle\langle w|J_{n_2}\rangle = \delta_{n_1 n_2}, \quad \sum_{n=0}^{\infty} \langle w|J_n\rangle\langle J_n|w'\rangle = \delta(w - w'). \tag{116}$$

Using Eq. (115), we can write the integral in the first of these equations as:

$$\int_0^1 dw\, \langle w|J_{n_2}\rangle\langle w|J_{n_1}\rangle^* = \int_0^1 dw\, e^{2\pi i(n_2 - n_1)w} = \delta_{n_1 n_2}. \tag{117}$$

Hence the first relation indeed holds. We can similarly write the sum in the second relation as

$$\sum_{n=0}^{\infty} \langle w|J_n\rangle\langle w'|J_n\rangle^* = \sum_{n=0}^{\infty} e^{2\pi i n(w - w')}. \tag{118}$$

Jordan set this equal to $\delta(w - w')$. However, for this to be true the sum over n should have been from minus to plus infinity. [51] If the action-angle variables are $(\hat{L}_z, \hat{\varphi})$, the z-component of angular momentum and the azimuthal angle, the eigenvalues of $\hat{J}$ are, in fact, $\pm m\hbar$ with $m = 0, 1, 2, \ldots$, but if the action

[51] A quick way to see this is to consider the Fourier expansion of some periodic

variable is proportional to the energy, as it is in many applications in the old quantum theory (Duncan and Janssen, 2007, Pt. 2, pp. 628–629), the spectrum is bounded below. So even under Jordan's alternative definition of canonically conjugate quantities, action-angle variables do not always qualify. However, since action-angle variables do not play a central role in the *Neue Begründung* papers, this is a relatively minor problem, especially compared to the much more serious problems that Jordan ran into in his attempt to make his formalism applicable to quantities with wholly or partly discrete spectra.

In view of this attempt, the other example of supposedly conjugate quantities that we want to examine, a special case of Jordan's example (e), is of particular interest. Consider a quantity $\hat{\beta}$ with a purely discrete spectrum with N eigenvalues $0, 1, 2, \ldots, N-1$. Jordan showed that, if the completeness or orthogonality relations (112) are to be satisfied, the quantity $\hat{\alpha}$ conjugate to $\hat{\beta}$ must also have a discrete spectrum with N eigenvalues hk/N where $k = 0, 1, 2, \ldots, N-1$. We will check this for the special case that $N = 2$. As Jordan (1927e, pp. 9–10) noted, this corresponds to the case of electron spin. In sec. 6, "On the theory of the magnetic electron," he returned to the topic of spin, acknowledging a paper by Pauli (1927b) on the "magnetic electron," which he had read in manuscript (Jordan, 1927e, p. 21, note 2). Pauli's discussion of spin had, in fact, been an important factor prompting Jordan to write *Neue Begründung* II:

> But the magnetic electron truly provides a case where the older canonical commutation relations completely fail; the desire to fully understand the relations encountered in this case was an important reason for carrying out this investigation (Jordan, 1927e, p. 22).

The two conditions on the amplitudes $\langle\beta|\alpha\rangle = Ce^{i\beta\alpha/\hbar}$ (Eq. (114) for $f = 1$)

---

function $f(w)$:

$$f(w) = \sum_{n=-\infty}^{\infty} c_n \, e^{2\pi i n w}, \text{ with Fourier coefficients } c_n \equiv \int_0^1 dw' \, e^{-2\pi i n w'} \, f(w').$$

Substituting the expression for c_n back into the Fourier expansion of $f(w)$, we find

$$f(w) = \int_0^1 dw' \left(\sum_{n=-\infty}^{\infty} e^{2\pi i n (w-w')} \right) f(w'),$$

which means that the expression in parentheses must be equal to $\delta(w - w')$. Note that the summation index does not run from 0 to ∞, as in Eq. (118), but from $-\infty$ to $+\infty$.

to be verified in this case are:

$$\sum_{k=1}^{2} \langle \alpha_m | \beta_k \rangle \langle \beta_k | \alpha_n \rangle = \delta_{nm}, \quad \sum_{k=1}^{2} \langle \beta_m | \alpha_k \rangle \langle \alpha_k | \beta_n \rangle = \delta_{nm}. \tag{119}$$

Inserting the expression for the amplitudes, we can write the first of these relations as

$$\sum_{k=1}^{2} \langle \beta_k | \alpha_n \rangle \langle \beta_k | \alpha_m \rangle^* = \sum_{k=1}^{2} C^2 \, e^{i\beta_k(\alpha_n - \alpha_m)/\hbar} = C^2 \left(1 + e^{i(\alpha_n - \alpha_m)/\hbar} \right), \tag{120}$$

where in the last step we used that $\beta_1 = 0$ and $\beta_2 = 1$. The eigenvalues of $\hat{\alpha}$ in this case are $\alpha_1 = 0$ and $\alpha_2 = h/2$. For $m = n$, the right-hand side of Eq. (120) is equal to $2C^2$. For $m \neq n$, $e^{i(\alpha_n - \alpha_m)/\hbar} = e^{\pm i\pi} = -1$ and the right-hand side of Eq. (120) vanishes. Setting $C = 1/\sqrt{2}$, we thus establish the first relation in Eq. (119). A completely analogous argument establishes the second.

This example is directly applicable to the treatment of two arbitrary components of the electron spin, $\hat{\sigma} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$, say the x- and the z-components, even though the eigenvalues of both $\hat{\sigma}_x$ and $\hat{\sigma}_z$ are $(\frac{1}{2}, -\frac{1}{2})$ rather than $(0, h/2)$ for $\hat{\alpha}$ and $(0, 1)$ for $\hat{\beta}$ as in Jordan's example (e) for $N = 2$. We can easily replace the pair of spin components $(\hat{\sigma}_x, \hat{\sigma}_z)$ by a pair of quantities $(\hat{\alpha}, \hat{\beta})$ that do have the exact same eigenvalues as in Jordan's example:

$$\hat{\alpha} \equiv \frac{h}{2} \left(\hat{\sigma}_x + \frac{1}{2} \right), \quad \hat{\beta} \equiv \hat{\sigma}_z + \frac{1}{2}. \tag{121}$$

The amplitudes $\langle \beta | \alpha \rangle = (1/\sqrt{2}) \, e^{i\beta\alpha/\hbar}$ now express that if the spin in one direction is known, the two possible values of the spin in the directions orthogonal to that direction are equiprobable. Moreover, as we just saw, amplitudes $\langle \beta | \alpha \rangle$ satisfy the completeness or orthogonality relations (112). It follows that *any two orthogonal components of spin are canonically conjugate to one another on Jordan's new definition*! One can thus legitimately wonder whether this definition is not getting much too permissive. The main problem, however, with Jordan's formalism is not that it is asking too little of its conjugate variables, but rather that it is asking too much of its canonical transformations!

Canonical transformations enter into the formalism in sec. 3, where Jordan (1927e, pp. 13–16) introduced a simplified yet at the same time generalized version of equations (2ab) of *Neue Begründung* I for probability amplitudes (Jordan, 1927b, p. 821). They are simplified in that there are no longer additional equations for the *Ergänzungsamplitude* (ibid., Eqs. (3ab)). They are generalized in that they are no longer restricted to systems with only one degree of freedom and, much more importantly, in that they are no longer restricted to cases where all quantities involved have purely continuous spectra. Quantities with partly or wholly discrete spectra are now also allowed.

Recall how Jordan built up his theory in *Neue Begründung* I (cf. our discussion in Section 2.3). He posited a number of axioms to be satisfied by his probability amplitudes. He then constructed a model for these postulates. To this end he identified probability amplitudes with the integral kernels for certain canonical transformations. Starting with differential equations trivially satisfied by the amplitude $\langle p|q\rangle = e^{-ipq/\hbar}$ for some initial pair of conjugate variables $\hat{p}$ and $\hat{q}$, Jordan derived differential equations for amplitudes involving other quantities related to the initial ones through canonical transformations. As we already saw above, this approach breaks down as soon as we ask about the probability amplitudes for quantities with partly discrete spectra, such as, typically, the Hamiltonian.

Although Jordan (1927e, p. 14) emphasized that one has to choose initial $\hat{p}$'s and $\hat{q}$'s with "fitting spectra" (*passende Spektren*) and that the equations for the amplitudes are solvable only "if it is possible to find" such spectra, he did not state explicitly in sec. 3 that the construction of *Neue Begründung* I fails for quantities with discrete spectra.[52] That admission is postponed until the discussion of canonical transformations in sec. 4. At the beginning of sec. 3. the general equations for probability amplitudes are given in the form (Jordan, 1927e, p. 14, Eqs. (2ab)):

$$\Phi^{\beta q}_{\alpha p}\hat{B}_k - \hat{\beta}_k\Phi^{\beta q}_{\alpha p} = 0, \tag{122}$$

$$\Phi^{\beta q}_{\alpha p}\hat{A}_k - \hat{\alpha}_k\Phi^{\beta q}_{\alpha p} = 0, \tag{123}$$

where $\hat{A}_k$ and $\hat{B}_k$ are defined as [NB2, sec. 3, Eq. (1)]:

$$\hat{B}_k = \left(\Phi^{\beta q}_{\alpha p}\right)^{-1}\hat{\beta}\,\Phi^{\beta q}_{\alpha p}, \quad \hat{A}_k = \left(\Phi^{\beta q}_{\alpha p}\right)^{-1}\hat{\alpha}\,\Phi^{\beta q}_{\alpha p} \tag{124}$$

Jordan (1927e, pp. 14–15) then showed that the differential equations of *Neue Begründung* I are included in these new equations as a special case. Since there is only one degree of freedom in that case, we do not need the index k. We can also suppress all indices of $\Phi^{\beta q}_{\alpha p}$ as this is the only amplitude/transformation-matrix involved in the argument. So we have $\hat{A} = \Phi^{-1}\hat{\alpha}\,\Phi$ and $\hat{B} = \Phi^{-1}\hat{\beta}\,\Phi$. These transformations, however, are used very differently in the two installments of *Neue Begründung*. Although Jordan only discussed this change in sec. 4, he already alerted the reader to it in sec. 3, noting that "$\hat{B}$, $\hat{A}$ are the operators for $\hat{\beta}$, $\hat{\alpha}$ with respect to $\hat{q}$, $\hat{p}$" (Jordan, 1927e, p. 15)

Suppressing all subscripts and superscripts, we can rewrite Eqs. (122)–(123)

[52] In his *Mathematische Begründung*, von Neumann (1927a) had already put his finger on this problem: "A special difficulty with [the approach of] Jordan is that one has to calculate not just the transforming operators (the integral kernels of which are the "probability amplitudes"), but also the value-range onto which one is transforming (i.e., the spectrum of eigenvalues)" (p. 3).

as:

$$(\Phi\,\hat{B}\,\Phi^{-1} - \hat{\beta})\,\Phi = 0, \tag{125}$$
$$(\Phi\,\hat{A}\,\Phi^{-1} - \hat{\alpha})\,\Phi = 0. \tag{126}$$

Using that

$$\Phi\,\hat{A}\,\Phi^{-1} = \hat{\alpha} = f(\hat{p}, \hat{q}), \quad \Phi\,\hat{B}\,\Phi^{-1} = \hat{\beta} = g(\hat{p}, \hat{q}) \tag{127}$$

(Jordan, 1927e, p. 15, Eq. 8); that $\hat{p}$ and $\hat{q}$ in the q-basis are represented by $(\hbar/i)\partial/\partial q$ and multiplication by q, respectively; and that $\hat{\alpha}$ and $\hat{\beta}$ in Eqs. (125)–(126) are represented by $-(\hbar/i)\partial/\partial\beta$ and multiplication by β, respectively, we see that in this special case Eqs. (122)–(123) (or, equivalently, Eqs. (125)–(126)) reduce to [NB2, p. 15, Eqs. (9ab)]

$$\left(g\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) - \beta\right)\Phi = 0, \tag{128}$$

$$\left(f\left(\frac{\hbar}{i}\frac{\partial}{\partial q}, q\right) + \frac{\hbar}{i}\frac{\partial}{\partial\beta}\right)\Phi = 0, \tag{129}$$

which are just Eqs. (2a) of *Neue Begründung* I (Jordan, 1927b, p. 821; cf. Eqs. (50)–(51) with $\langle q|\beta\rangle$ written as Φ). This is the basis for Jordan's renewed claim that his general equations for probability amplitudes contain both the time-dependent and the time-independent Schrödinger equations as a special case (cf. our discussion at the end of Section 2.3). It is certainly true that, if the quantity $\hat{B}$ in Eq. (124) is chosen to be the Hamiltonian, Eq. (128) turns into the time-independent Schrödinger equation. However, there is no canonical transformation that connects this equation for $\psi_n(q) = \langle q|E\rangle$ to the equations trivially satisfied by $\langle p|q\rangle$ that formed the starting point for Jordan's construction of his formalism in *Neue Begründung* I.

Jordan (1927e, pp. 16–17) finally conceded this point in sec. 4 of *Neue Begründung* II. Following Dirac (1927), Jordan switched to a new conception of canonical transformations. Whereas before, he saw canonical transformations such as $\hat{\alpha} = T\hat{p}T^{-1}, \hat{\beta} = T\hat{q}T^{-1}$, as taking us from one pair of conjugate variables $(\hat{p}, \hat{q})$ to a *different* pair $(\hat{\alpha}, \hat{\beta})$, he now saw them as taking us from one particular *representation* of a pair of conjugate variables to a *different representation* of those *same* variables. The canonical transformation used in sec. 3, $\hat{A} = \Phi^{-1}\hat{\alpha}\,\Phi, \hat{B} = \Phi^{-1}\hat{\beta}\,\Phi$, is already an example of a canonical transformation in the new Dirac sense. By giving up on canonical transformations in the older sense, Jordan effectively abandoned the basic architecture of the formalism of *Neue Begründung* I.

This is how Jordan explained the problem at the beginning of sec. 4 of *Neue Begründung* II:

> Canonical transformations, the theory of which, as in classical mechanics, gives the natural generalization and the fundamental solution of the problem

> of the integration of the equations of motion, were originally [footnote citing Born, Heisenberg, and Jordan (1926)] conceived of as follows: the canonical quantities $\hat{q}$, $\hat{p}$ should be represented as functions of certain other canonical quantities $\hat{\beta}$, $\hat{\alpha}$:
>
> $$\hat{q}_k = G_k(\hat{\beta}, \hat{\alpha}), \quad \hat{p}_k = F_k(\hat{\beta}, \hat{\alpha}). \tag{1}$$
>
> On the assumption that canonical systems can be defined through the usual canonical commutation relations, a formal proof could be given [footnote referring to Jordan (1926a)] that for canonical $\hat{q}$, $\hat{p}$ and $\hat{\beta}$, $\hat{\alpha}$ equations (1), as was already suspected originally, can always be cast in the form
>
> $$\hat{q}_k = T\,\hat{\beta}_k\,T^{-1}, \quad \hat{p}_k = T\,\hat{\alpha}_k\,T^{-1}. \tag{2}$$
>
> However, since, as we saw, the old canonical commutation relations are not valid [cf. Eqs. (103)–(107)], this proof too loses its meaning; in general, one can *not* bring equations (1) in the form (2).
>
> Now a modified conception of canonical transformation was developed by Dirac [footnote citing Dirac (1927) and Lanczos (1926)]. According to Dirac, [canonical transformations] are not about representing certain canonical quantities as functions of other canonical quantities, but rather about switching, *without a transformation of the quantities themselves*, to a different *matrix representation* (Jordan, 1927e, pp. 16–17; emphasis in the original, hats added).

In modern terms, canonical transformations in the new Dirac sense transform the matrix elements of operators in one basis to matrix elements of operators in another basis. This works whether or not the operator under consideration is part of a pair of operators corresponding to canonically conjugate quantities. The notation $\Phi^{\beta q}_{\alpha p} = \langle \beta | q \rangle$ introduced in *Neue Begründung* II that replaces the notation T in *Neue Begründung* I for the operators implementing a canonical transformation nicely prepared us for this new way of interpreting such transformations. Consider the matrix elements of the position operator $\hat{q}$ (with a purely continuous spectrum) in the q-basis:

$$\langle q' | \hat{q} | q'' \rangle = q' \delta(q' - q''). \tag{130}$$

Now let $\hat{\beta}$ be an arbitrary self-adjoint operator. In general, $\hat{\beta}$ will have a spectrum with both continuous and discrete parts. Von Neumann's spectral theorem tells us that

$$\hat{\beta} = \sum_n \beta_n \, |\beta_n\rangle\langle\beta_n| + \int \beta \, |\beta\rangle\langle\beta| \, d\beta, \tag{131}$$

where sums and integrals extend over the discrete and continuous parts of the spectrum of $\hat{\beta}$, respectively. In Jordan's notation, the spectral decomposition of $\hat{\beta}$ can be written more compactly as: $\hat{\beta} = \overline{\sum} \, \beta \, |\beta\rangle\langle\beta|$. We now want to find the relation between the matrix elements of $\hat{q}$ in the β-basis and its matrix elements in the q-basis. Using the spectral decomposition $\int q\,|q\rangle\langle q|\,dq$ of $\hat{q}$, we can write:

$$\begin{aligned}\langle\beta'|\hat{q}|\beta''\rangle &= \int dq'\,\langle\beta'|q'\rangle q'\langle q'|\beta''\rangle \\ &= \int dq'dq''\,\langle\beta'|q''\rangle q''\delta(q'-q'')\langle q''|\beta''\rangle \\ &= \int dq'dq''\,\langle\beta'|q''\rangle\langle q''|\hat{q}|q'\rangle\langle q''|\beta''\rangle\end{aligned} \tag{132}$$

In the notation of *Neue Begründung* II, the last line would be the 'matrix multiplication', $\left(\Phi_{q\alpha}^{q\beta}\right)^{-1}\hat{q}\,\Phi_{q\alpha}^{q\beta}$. This translation into modern notation shows that Jordan's formalism, even with a greatly reduced role for canonical transformations, implicitly relies on the spectral theorem, which von Neumann (1927a) published in *Mathematische Begründung*, submitted just one month before *Neue Begründung* II. The most important observation in this context, however, is that an explicit choice of quantities $\hat{p}$ and $\hat{\alpha}$ conjugate to $\hat{q}$ and $\hat{\beta}$, respectively, is completely irrelevant for the application of the spectral theorem.

This gets us to the last aspect of *Neue Begründung* II that we want to discuss in this section, namely Jordan's response to von Neumann's criticism of the Dirac-Jordan transformation theory. A point of criticism we already mentioned is that the probability amplitude $\varphi(\beta,q)=\langle\beta|q\rangle$ is determined only up to a phase factor.

As we mentioned in the introduction, the projection operator $|a\rangle\langle a|$ does not change if the ket $|a\rangle$ is replaced by $e^{i\vartheta}|a\rangle$ and the bra $\langle a|$ accordingly by $e^{-i\vartheta}\langle a|$, where ϑ can be an arbitrary real function of a. Hence, the spectral decomposition of $\hat{\beta}$ in Eq. (131) does not change if $|\beta\rangle$ is replaced by $e^{-i\rho(\beta)/\hbar}|\beta\rangle$, where we have written the phase factor, in which ρ is an arbitrary real function of β, in a way that corresponds to Jordan's notation for the resulting phase ambiguity in the amplitude $\varphi(\beta,q)=\langle\beta|q\rangle$. Similarly, we note that the spectral decomposition of $\hat{q}$ does not change if we replace $|q\rangle$ by $e^{i\sigma(q)/\hbar}|q\rangle$, where σ is an arbitrary real function of q. However, with the changes $|\beta\rangle\to e^{-i\rho(\beta)/\hbar}|\beta\rangle$ and $|q\rangle\to e^{i\sigma(q)/\hbar}|q\rangle$, the amplitude $\varphi(\beta,q)=\langle\beta|q\rangle$ changes (Jordan, 1927e, p. 20):

$$\langle\beta|q\rangle \quad\longrightarrow\quad e^{i(\rho(\beta)+\sigma(q))/\hbar}\langle\beta|q\rangle. \tag{133}$$

Unlike projection operators, as von Neumann pointed out, probability amplitudes, are determined only up to such phase factors.

As we noted above, Jordan responded to this criticism by adding a dependence on quantities $\hat{\alpha}$ and $\hat{p}$ conjugate to $\hat{\beta}$ and $\hat{q}$, respectively, to the probability amplitude $\langle\beta|q\rangle$, thus arriving at the amplitudes $\Phi_{\alpha p}(\beta',q')$ of *Neue Begründung* II. It turns out that the phase ambiguity of $\langle\beta|q\rangle$ is equivalent to a certain freedom we have in the definition of the quantities $\hat{\alpha}$ and $\hat{p}$ conjugate to $\hat{\beta}$ and $\hat{q}$, respectively. The phase ambiguity, as we saw, can be characterized by the arbitrary functions $\rho(\beta)$ and $\sigma(q)$. Following Jordan, we will show that our freedom in the definition of $\hat{\alpha}$ and $\hat{p}$ is determined by the derivatives $\rho'(\beta)$ and $\sigma'(q)$ of those same functions. By considering amplitudes $\Phi_{\alpha p}(\beta', q')$ with uniquely determined $\hat{\alpha}$ and $\hat{p}$, Jordan could thus eliminate the phase ambiguity that von Neumann found so objectionable.

Following Jordan (1927e, p. 20), we establish the relation between these two elements of arbitrariness for the special case that all quantities involved have fully continuous spectra. In sec. 6 on spin, Jordan (1927e, pp. 21–25) tried to extend his argument to some special cases of discrete spectra. We will not discuss those efforts.

Consider two complete sets of eigenstates of $\hat{q}$, $\{|q\rangle_1\}$ and $\{|q\rangle_2\}$, related to one another via

$$|q\rangle_1 = e^{i\sigma(q)/\hbar}\,|q\rangle_2. \tag{134}$$

This translates into two different amplitudes that differ by that same phase factor: $\varphi_1(\beta, q) = e^{i\sigma(q)/\hbar}\,\varphi_2(\beta, q)$.[53] Suppose $\hat{p}_1$ is conjugate to $\hat{q}$ if we use the $|q\rangle_1$ set of eigenstates of $\hat{q}$. Since we restrict ourselves to quantities with fully continuous spectra, this means that $[\hat{p}_1, \hat{q}] = \hbar/i$. It also means, as we saw in Section 2.1, that

$$\hat{p}_1|q\rangle_1 = -\frac{\hbar}{i}\frac{\partial}{\partial q}|q\rangle_1. \tag{135}$$

Eigenstates $|p\rangle_1$ of $\hat{p}_1$ can be written as a Fourier series in terms of the $|q\rangle_1$ states:[54]

$$|p\rangle_1 = \int dq\, e^{ipq/\hbar}\,|q\rangle_1. \tag{136}$$

We can likewise construct a $\hat{p}_2$ conjugate to $\hat{q}$ if we use the $|q\rangle_2$ set of eigenstates

[53] An argument completely analogous to one we give for the relation between the phase factor $e^{i\sigma(q)/\hbar}$ and the definition of $\hat{p}$ can be be given for the relation between the phase factor $e^{i\rho(\beta)/\hbar}$ and the definition of $\hat{\alpha}$.

[54] One easily verifies that $|p\rangle_1$ is indeed an eigenstate of $\hat{p}_1$. The action of $\hat{p}_1$ on $|p\rangle_1$ can be written as:

$$\hat{p}_1|p\rangle_1 = \int dq\, e^{ipq/\hbar}\hat{p}_1|q\rangle_1 = -\int dq\, e^{ipq/\hbar}\frac{\hbar}{i}\frac{\partial}{\partial q}|q\rangle_1.$$

Partial integration gives:

$$\hat{p}_1|p\rangle_1 = \int dq\,\frac{\hbar}{i}\frac{\partial}{\partial q}\left(e^{ipq/\hbar}\right)|q\rangle_1 = p\int dq\, e^{ipq/\hbar}|q\rangle_1 = p|p\rangle_1,$$

which is what we wanted to prove.

of $\hat{q}$. Instead of Eqs. (135)–(136), we then have

$$\hat{p}_2|q\rangle_2 = -\frac{\hbar}{i}\frac{\partial}{\partial q}|q\rangle_2, \qquad |p\rangle_2 = \int dq\, e^{ipq/\hbar}\,|q\rangle_2. \tag{137}$$

The relation between these two different conjugate momenta, it turns out, is

$$\hat{p}_2 = \hat{p}_1 + \sigma'(\hat{q}). \tag{138}$$

Note that the commutator $[\hat{p}_1, \hat{q}]$ does not change if we add an arbitrary function of $\hat{q}$ to $\hat{p}_1$. To prove that Eq. (138) indeed gives the relation between $\hat{p}_1$ and $\hat{p}_2$, we show that $|p\rangle_2$ in Eq. (137) is indeed an eigenstate of $\hat{p}_2$ as defined in Eq. (138), using relation (134) between $|q\rangle_1$ and $|q\rangle_2$:

$$\begin{aligned}
\hat{p}_2|p\rangle_2 &= \int dq\, e^{ipq/\hbar}\,\hat{p}_2\,|q\rangle_2 \\
&= \int dq\, e^{ipq/\hbar}\,(\hat{p}_1 + \sigma'(\hat{q}))\,e^{-i\sigma(q)/\hbar}\,|q\rangle_1 \\
&= \int dq\, e^{i(pq-\sigma(q))/\hbar}\left(-\frac{\hbar}{i}\frac{\partial}{\partial q} + \sigma'(q)\right)|q\rangle_1 \\
&= \int dq\,(p - \sigma'(q) + \sigma'(q))\,e^{i(pq-\sigma(q))/\hbar}|q\rangle_1 \\
&= p\int dq\, e^{ipq/\hbar}\,|q\rangle_2 = p\,|p\rangle_2,
\end{aligned} \tag{139}$$

where in the fourth step we performed a partial integration. This proves that the ambiguity (138) in the $\hat{p}$ conjugate to $\hat{q}$ corresponds directly to the phase ambiguity (134) in the amplitude $\langle\beta|q\rangle$. Similarly, the ambiguity in the $\hat{\alpha}$ conjugate to $\hat{\beta}$, which is determined only up to a term $\rho'(\hat{\beta})$, corresponds directly to the phase ambiguity $e^{i\rho(\beta)/\hbar}$ in the amplitude $\langle\beta|q\rangle$. Hence, for specific conjugate variables $\hat{p}$ and $\hat{\alpha}$ conjugate to $\hat{q}$ and $\hat{\beta}$, the amplitude $\Phi_{\alpha p}(\beta', q')$ of *Neue Begründung* II is unique up to a constant phase factor (i.e., one that is not a function of q or β).

In addition to responding to von Neumann's criticism of his approach, Jordan (1927e, p. 20) also offered some criticism of von Neumann's approach. In particular, he complained that von Neumann showed no interest in either canonical transformations or conjugate variables. As we will see when we cover von Neumann's *Mathematische Begründung* in the next section, this is simply because von Neuman did not need either for his formulation of quantum mechanics. That formulation clearly did not convince Jordan. In fact, von Neumann's paper only seems to have increased Jordan's confidence in his own approach. After his brief discussion of *Mathematische Begründung*, he concluded: "It thus appears that the amplitudes themselves are to be considered the fundamental concept of quantum mechanics" (Jordan, 1927e, pp. 20–21).

5 Von Neumann's *Mathematische Begründung* (May 1927)

In the next two sections we turn our attention to the first two papers of the trilogy that von Neumann (1927a,b,c) published the same year as and partly in response to the papers by Dirac (1927) and Jordan (1927b) on transformation theory. This trilogy provided the backbone of his famous book published five years later (von Neumann, 1932). The first paper in the trilogy, *Mathematische Begründung*, was presented in the meeting of the Göttingen Academy of May 20, 1927. In this paper, von Neumann first introduced the Hilbert space formalism and the spectral theorem, at least for bounded operators, two contributions that have since become staples of graduate texts in quantum physics and functional analysis.[55] In part because of this greater familiarity but also because of its intrinsic clarity, von Neumann's *Mathematische Begründung* is much easier to follow for modern readers than Jordan's *Neue Begründung*. There is no need for us to cover it in as much detail as we did with Jordan's papers in Sections 2 and 4.

Mathematische Begründung is divided into nine parts, comprising 15 sections and two appendices:

(1) "Introduction," sec. I, pp. 1–4;
(2) "The Hilbert space," secs. II–VI, pp. 4–22;
(3) "Operator calculus," secs. VII–VIII, pp. 22–29;
(4) "The eigenvalue problem," sec. IX–X, pp. 29–37;
(5) "The absolute value of an operator," sec. "IX" (a typo: this should be XI), pp. 37–41;
(6) "The statistical assumption [*Ansatz*] of quantum mechanics," secs. XII–XIII, pp. 42–47;
(7) "Applications," sec. XIV, pp. 47–50;
(8) "Summary," sec. XV, pp. 50–51;
(9) "Appendices," pp. 51–57.

Abstract Hilbert space is introduced in secs. V–VI, the spectral theorem in secs. IX–X. After going over the introduction of the paper, we focus on parts of sec. IV and secs. IX–XIII.

In Sec. IV, von Neumann criticized the way in which wave mechanics and matrix mechanics are unified in the approach of Dirac and Jordan and presented his superior alternative approach to this unification, based on the isomorphism of two concrete instantiations of abstract Hilbert space $\overline{\mathfrak{H}}$, the space of square-summable sequences $\mathfrak{H}_0$ and the space of square-integrable functions $\mathfrak{H}$ (von Neumann, 1927a, the designations $\mathfrak{H}$, $\mathfrak{H}_0$, and $\overline{\mathfrak{H}}$ are introduced on pp. 14–15).

[55] In the latter category we already mentioned Prugovecki (1981) and Dennery and Krzywicki (1996, Ch. 3) (see note 15).

In modern notation, this is the isomorphism between l^2 and L^2.

Secs. IX–XIII contain von Neumann's criticism of Jordan's use of probability amplitudes and his derivation of an alternative formula for conditional probabilities in quantum mechanics in terms of projection operators. Unlike von Neumann, we present this derivation in Dirac notation.

In the introduction of *Mathematische Begründung*, von Neumann (1927a, pp. 1–3) gave a list of seven points, labeled α through ϑ (there is no point η), in which he took stock of the current state of affairs in the new quantum theory and identified areas where it ran into mathematical difficulties. We paraphrase these points. (α) Quantum theory describes the behavior of atomic systems in terms of certain eigenvalue problems. (β) This allows for a unified treatment of continuous and discontinuous elements in the atomic world. (γ) The theory suggests that the laws of nature are stochastic.[56] (δ) Returning to the formulation of the theory in terms of eigenvalue problems, von Neumann briefly characterized the different but equivalent ways in which such problems are posed in matrix mechanics and in wave mechanics. Both approaches have their difficulties. (ε) The application of matrix mechanics appears to be restricted to situations with purely discrete spectra. To deal with wholly or partly continuous spectra, one ends up using, side by side, matrices with indices taking on discrete values and "continuous matrices," i.e., the integral kernels of the Dirac-Jordan transformation theory, with 'indices' taking on continuous values. It is "very hard," von Neumann (1927a, p. 2) warned, to do this in a mathematically rigorous way. (ζ) These same problems start to plague the differential-operator approach of wave mechanics as soon as wave functions are interpreted as probability amplitudes. Von Neumann credited Born, Pauli, and Jordan with transferring the probability concepts of matrix mechanics to wave mechanics and Jordan with developing these ideas into a "closed system" (ibid.).[57] This system, however, faces serious mathematical objections because of the unavoidable use of improper eigenfunctions, such as the Dirac delta function, the properties of which von Neumann thought were simply "absurd" (ibid., p. 3). His final objection seems mild by comparison but weighed heavily for von Neumann: (ϑ) eigenfunctions in wave mechanics and probability amplitudes in transformation theory are determined only up to an arbitrary phase factor. The probabilities one ultimately is after in quantum theory do not depend on these phase factors and von Neumann therefore

[56] Parenthetically, von Neumann (1927a, p. 1) added an important qualification: "(at least the quantum laws known to us)." So, at this point, he left open the possibility that, at a deeper level, the laws would be deterministic again.

[57] In this context von Neumann (1927a, p. 2) referred to his forthcoming paper with Hilbert and Nordheim (1928). Oddly, von Neumann did not mention Dirac at this point, although Dirac is mentioned alongside Jordan in sec. XII (von Neumann, 1927a, p. 43) as well as in the second paper of the trilogy (von Neumann, 1927b, p. 245; see Section 6).

wanted to avoid them altogether.[58]

In sec. II, von Neumann set the different guises in which the eigenvalue problems appear in matrix and in wave mechanics side by side. In matrix mechanics, the problem is to find square-summable infinite sequences of complex numbers $\mathbf{v} = (v_1, v_2, \ldots)$ such that

$$\mathbf{Hv} = E\mathbf{v}, \tag{140}$$

where $\mathbf{H}$ is the matrix representing the Hamiltonian of the system in matrix mechanics, and where E is an energy eigenvalue. In wave mechanics, the problem is to find square-integrable complex-valued functions $f(x)$ such that

$$\hat{H} f(x) = E f(x), \tag{141}$$

where $\hat{H}$ is the differential operator, involving multiplication by x and differentiation with respect to x, that represents the Hamiltonian of the system in wave mechanics.

One way to unify these two approaches, von Neumann (1927a, pp. 10–11) pointed out at the beginning of sec. IV, is to look upon the discrete set of values $1, 2, 3 \ldots$ of the index i of the sequences $\{x_i\}_{i=1}^{\infty}$ in matrix mechanics and the continuous (generally multi-dimensional) domain Ω of the functions $f(x)$ in wave mechanics as two particular realizations of some more general space, which von Neumann called R. Following the notation of his book (von Neumann, 1932, sec. 4, pp. 15–16), we call the 'space' of index values Z. Eq. (140) can then be written as:[59]

$$\sum_{j \in Z} H_{ij} v_j = E v_i. \tag{142}$$

'Summation over Z' can be seen as one instantiation of 'integration over R;' 'integration over Ω' as another. In this way Eq. (141) can, at least formally, be subsumed under matrix mechanics. One could represent the operator $\hat{H}$ in Eq. (141) by the integral kernel $H(x, y)$ and write

$$\int_{\Omega} dy\, H(x, y) f(y) = E f(x). \tag{143}$$

Both the matrix H_{ij} and the integral kernel $H(x, y)$ can be seen as 'matrices' H_{xy} with indices $x, y \in R$. For H_{ij}, $R = Z$; for $H(x, y)$, $R = \Omega$. Von Neumann

[58] In *Neue Begründung* II, as we saw in Section 5, Jordan (1927e, p. 8) responded to this criticism by adding subscripts to the probability amplitudes for two quantities $\hat{\beta}$ and $\hat{q}$ indicating a specific choice of the canonically-conjugate quantities $\hat{\alpha}$ and $\hat{p}$ (see Eqs. (133)–(139)).

[59] We replaced von Neumann's (1927a, p. 10) x_i's by v_i's to avoid confusion with the argument(s) of the functions $f(x)$.

identified this way of trying to unify matrix and wave mechanics as Dirac's way (and, one may add, although he is not mentioned by name at this point: Jordan's way). Von Neumann rejected this approach. He dismissed the analogy between Z and Ω sketched above as "really superficial, as long as one sticks to the usual measure of mathematical rigor" (von Neumann, 1927a, p. 11).[60] He pointed out that even the simplest linear operator, the identity operator, does not have a proper integral-kernel representation. Its integral kernel is the improper Dirac delta function: $\int dy\, \delta(x-y) f(y) = f(x)$.

The appropiate analogy, von Neumann (1927a, pp. 11–14) argued, is not between Z and Ω, but between the space of square-summable sequences *over* Z and the space of square-integrable functions *over* Ω. In his book, von Neumann (1932, p. 16) used the notation F_Z and F_Ω for these two spaces.[61] In 1927, as mentioned above, he used $\mathfrak{H}_0$ and $\mathfrak{H}$, instead. Today they are called l^2 and L^2, respectively.[62] Von Neumann (1927a, pp. 12–13) reminded his readers of the "Parseval formula," which maps sequences in l^2 onto functions in L^2, and a "theorem of Fischer and F. Riesz," which maps functions in L^2 onto sequences in l^2.[63] The combination of these two results establishes that l^2 and L^2 are isomorphic. As von Neumann (1927a, p. 12) emphasized, these "mathematical facts that had long been known" could be used to unify matrix mechanics and wave mechanics in a mathematically impeccable manner. With a stroke of the pen, von Neumann thus definitively settled the issue of the equivalence of wave mechanics and matrix mechanics. *Anything that can be done in wave mechanics, i.e., in* L^2*, has a precise equivalent in matrix mechanics, i.e.,* l^2. This is true regardless of whether we are dealing with discrete spectra, continuous spectra, or a combination of the two.

In sec. V, von Neumann (1927a, pp. 14–18) introduced abstract Hilbert space, for which he used the notation $\overline{\mathfrak{H}}$, carefully defining it in terms of five axioms

[60] In the introduction, we already quoted some passages from the introduction of von Neumann's 1932 book in which he complained about the lack of mathematical rigor in Dirac's approach. After characterizing the approach in terms of the analogy between Z and Ω, he wrote: "It is no wonder that this cannot succeed without some violence to formalism and mathematics: the spaces Z and Ω are really very different, and every attempt, to establish a relation between them, must run into great difficulties" (von Neumann, 1932, p. 15).

[61] Jammer (1966, pp. 314–315) also used this 1932 notation in his discussion of von Neumann (1927a).

[62] Earlier in his paper, von Neumann (1927a, p. 7) remarked that what we now call l^2 was usually called "(complex) Hilbert space." Recall, however, that London (1926b, p. 197) used the term "Hilbert space" for L^2 (note 13).

[63] The paper cited by von Neumann (1927a, p. 13, note 15) is Riess (1907a). In his discussion of von Neumann's paper, Jammer (1966, pp. 314–315) cited Riess (1907a,b) and Fischer (1907).

labeled A through E.[64] In sec. VI, he added a few more definitions and then stated and proved six theorems about Hilbert space, labeled 1 through 6 (ibid., pp. 18–22). In sec. VII, he turned to the discussion of operators acting in Hilbert space (ibid., pp. 25). This will be familiar terrain for the modern reader and need not be surveyed in any more detail.

The same goes for sec. VIII, in which von Neumann introduced a special class of Hermitian operators. Their defining property is that they are idempotent: $\hat{E}^2 = \hat{E}$. Von Neumann called an operator like this an *Einzeloperator* or *E. Op.* for short (von Neumann, 1927a, p. 25).[65] They are now known, of course, as projection operators. In a series of theorems, numbered 1 through 9, von Neumann (1927a, pp. 25–29) proved some properties of such operators. For our purposes, it suffices to know that they are Hermitian and idempotent.

We do need to take a somewhat closer look at sec. IX. In this section, von Neumann (1927a, pp. 29–33) used projection operators to formulate the spectral theorem. Following von Neumann (1927a, p. 31), we start by considering a finite Hermitian operator $\hat{A}$ with a non-degenerate discrete spectrum (von Neumann, 1927a, sec. VIII, p. 31). Order its real eigenvalues a_i: $a_1 < a_2 < a_3 \ldots$ Let $|a_i\rangle$ be the associated normalized eigenvectors ($\langle a_i|a_j\rangle = \delta_{ij}$). Now introduce the operator $\hat{E}(l)$:[66]

$$\hat{E}(l) \equiv \sum_{(i|a_i \leq l)} |a_i\rangle\langle a_i|, \tag{144}$$

where, unlike von Neumann, we used modern Dirac notation. As we already noted in the introduction, there is no phase ambiguity in $\hat{E}(l)$. The operator stays the same if we replace $|a_i\rangle$ by $|a_i\rangle' = e^{i\varphi_i}|a_i\rangle$:

$$|a_i\rangle'\langle a_i|' = e^{i\varphi_i}|a_i\rangle\langle a_i|e^{-i\varphi_i} = |a_i\rangle\langle a_i|. \tag{145}$$

Of course, von Neumann did not think of an *E. Op.* as constructed out of bras and kets, just as Jordan did not think of a probability amplitude $\langle a|b\rangle$ as an inner product of $|a\rangle$ and $|b\rangle$.

The operator $\hat{E}(l)$ has the property:

$$\hat{E}(a_i) - \hat{E}(a_{i-1}) = |a_i\rangle\langle a_i|. \tag{146}$$

---

[64] In his book, von Neumann (1932) adopted the notation 'H. R.' (shorthand for *Hilbertscher Raum*) for $\overline{\mathfrak{H}}$.

[65] As he explains in a footnote, the term *Einzeloperator* is based on Hilbert's term *Einzelform* (von Neumann, 1927a, p. 25, note 23).

[66] Von Neumann initially the defined this operator in terms of its matrix elements $\langle v|\hat{E}(l)|w\rangle$ for two arbitrary sequences $\{v_i\}_{i=1}^{\kappa}$ and $\{w_i\}_{i=1}^{\kappa}$ (where we replaced von Neumann's x and y by v and w; cf. note 59). He defined (in our notation): $E(l; x|y) = \sum_{(i|a_i \leq l)} \langle v|a_i\rangle\langle a_i|w\rangle$ (von Neumann, 1927a, p. 31).

It follows that:

$$\hat{A} = \sum_i a_i(\hat{E}(a_i) - \hat{E}(a_{i-1})) = \sum_i a_i|a_i\rangle\langle a_i| \tag{147}$$

$\hat{E}(l)$ is piece-wise constant with jumps where l equals an eigenvalue. Hence we can write $\hat{A}$ as a so-called Stieltjes integral, which von Neumann discussed and illustrated with some figures in appendix 3 of his paper (von Neumann, 1927a, pp. 55–57):

$$\hat{A} = \int l\, d\hat{E}(l). \tag{148}$$

As von Neumann (1927a, p. 32) noted, these results (Eqs. (144)–(148)) can easily be generalized from finite Hermitian matrices and finite sequences to bounded Hermitian operators and the space $\mathfrak{H}_0$ or l^2 of infinite square-summable sequences. Since $\mathfrak{H}_0$ is just a particular instantiation of the abstract Hilbert space $\overline{\mathfrak{H}}$, it is clear that the same results hold for bounded Hermitian operators $\hat{T}$ in $\overline{\mathfrak{H}}$. After listing the key properties of $\hat{E}(l)$ for $\hat{T}$,[67] he concluded sec. IX writing: "We call $\hat{E}(l)$ the resolution of unity belonging to $\hat{T}$" (von Neumann, 1927a, p. 33).

In sec. X, von Neumann (1927a, pp. 33–37) further discussed the spectral theorem. Most importantly, he conceded that he had not yet been able to prove that it also holds for *unbounded* operators.[68] He only published the proof of this generalization in the paper in *Mathematische Annalen* mentioned

[67] As before (see note 66), he first defined the matrix elements $\langle f|\hat{E}(l)|g\rangle$ for two arbitrary elements f and g of Hilbert space. So he started from the relation

$$\langle f|\hat{T}|g\rangle = \int_{-\infty}^{\infty} l\, d\langle f|\hat{E}(l)|g\rangle,$$

and inferred from that, first, that $\hat{T}|g\rangle = \int_{-\infty}^{\infty} l\, d\{\hat{E}(l)|g\rangle\}$, and, finally, that $\hat{T} = \int_{-\infty}^{\infty} l\, d\hat{E}(l)$ (cf. Eq. (148)). Instead of the notation $\langle f|g\rangle$, von Neumann (1927a, p. 12) used the notation $Q(f, g)$ for the inner product of f and g (on p. 32, he also used $Q(f|g)$). So, in von Neumann's own notation, the relation he started from is written as $Q(f, Tg) = \int_{-\infty}^{\infty} l\, dQ(f, E(l)g)$.

[68] We remind the reader that a linear operator $\hat{A}$ in Hilbert space is bounded if there exists a positive real constant C such that $|\hat{A}f| < C|f|$ for arbitrary vectors f in the space (where $|...|$ indicates the norm of a vector, as induced from the defining inner-product in the space). If this is *not* the case, then there exist vectors in the Hilbert space on which the operator $\hat{A}$ is not well-defined, basically because the resultant vector has infinite norm. Instead, such unbounded operators are only defined (i.e., yield finite-norm vectors) on a proper subset of the Hilbert space, called the *domain* $\mathcal{D}(\hat{A})$ of the operator $\hat{A}$. The set of vectors obtained by applying $\hat{A}$ to all elements of its domain is called the *range* $\mathcal{R}(\hat{A})$ of $\hat{A}$. Multiplication of two unbounded operators evidently becomes a delicate matter insofar as the domain and ranges of the respective operators may not coincide.

earlier, which was submitted on December 15, 1928 (von Neumann, 1929). The key to the extension of the spectral theorem from bounded to unbounded operators is a so-called Cayley transformation (von Neumann, 1929, p. 80). Given unbounded Hermitian operator $\hat{R}$, introduce the operator $\hat{U}$ and its adjoint

$$\hat{U} = \frac{\hat{R} + i\hat{1}}{\hat{R} - i\hat{1}}, \qquad \hat{U}^\dagger = \frac{\hat{R} - i\hat{1}}{\hat{R} + i\hat{1}}, \tag{149}$$

where $\hat{1}$ is the unit operator. Since $\hat{R}$ is Hermitian, it only has real eigenvalues, so $(\hat{R} - i\hat{1})|\varphi\rangle \neq 0$ for any $|\varphi\rangle \in \overline{\mathfrak{H}}$. Since $\hat{U}$ is unitary ($\hat{U}\hat{U}^\dagger = 1$), the absolute value of all its eigenvalues equals 1. $\hat{U}$ is thus a bounded operator for which the spectral theorem holds. If it holds for $\hat{U}$, however, it must also hold for the original unbounded operator $\hat{R}$. The spectral decomposition of $\hat{R}$ is essentially the same as that of $\hat{U}$. In his book, von Neumann (1932, p. 80) gave Eq. (149), but he referred to his 1929 paper for a mathematically rigorous treatment of the spectral theorem for unbounded operators (von Neumann, 1932, p. 75, p. 246, note 95, and p. 244, note 78)

Sec. XI concludes the purely mathematical part of the paper. In this section, von Neumann (1927a, pp. 37-41) introduced the "absolute value" of an operator, an important ingredient, as we will see, in his derivation of his formula for conditional probabilities in quantum mechanics (see Eqs. (157)–(161) below).

In sec. XII, von Neumann (1927a, pp. 42–45) finally turned to the statistical interpretation of quantum mechanics. At the end of sec. I, he had already warned the reader that secs. II–XI would have a "preparatory character" and that he would only get to the real subject matter of the paper in secs. XII–XIV. At the beginning of sec. XII, the first section of the sixth part of the paper (see our table of contents above), on the statistical interpretation of quantum mechanics, he wrote: "We are now in a position to take up our real task, the mathematically unobjectionable unification of statistical quantum mechanics" (von Neumann, 1927a, p. 42). He then proceeded to use the spectral theorem and the projection operators $\hat{E}(l)$ of sec. IX to construct an alternative to Jordan's formula for conditional probabilities in quantum mechanics, which does not involve probability amplitudes. Recall von Neumann's objections to probability amplitudes (see Sections 1 and 4). First, Jordan's basic amplitudes, $\rho(p, q) = e^{-ipq/\hbar}$ (see Eq. (4)), which from the perspective of Schrödinger wave mechanics are eigenfunctions of momentum, are not square-integrable and hence not in Hilbert space (von Neumann, 1927a, p. 35). Second, they are only determined up to a phase factor (von Neumann, 1927a, p. 3, point ϑ). Von Neumann avoided these two problems by deriving an alternative formula which expresses the conditional probability $\Pr(a|b)$ in terms of projection operators associated with the spectral decomposition of the operators for the observables $\hat{a}$ and $\hat{b}$.

Von Neumann took over Jordan's basic statistical *Ansatz*. Consider a one-particle system in one dimension with coordinate q. Von Neumann (1927a, p. 43) considered the more general case with coordinates $q \equiv (q_1, \ldots, q_k)$. The probability of finding a particle in some region K if we know that its energy is E_n, i.e., if we know the particle is in the pure state $\psi_n(x)$ belonging to that eigenvalue, is given by (ibid.):[69]

$$\Pr(q \text{ in } K|E_n) = \int_K |\psi_n(q)|^2 dq. \tag{150}$$

Next, he considered the probability of finding the particle in some region K if we know that its energy is in some interval I that includes various eigenvalues of its energy, i.e., if the particle is in some mixed state where we only know that, with equal probability, its state is one of the pure states $\psi_n(x)$ associated with the eigenvalues within the interval I:[70]

$$\Pr(q \text{ in } K|E_n \text{ in } I) = \sum_{(n|E_n \,\text{in}\, I)} \int_K |\psi_n(q)|^2 dq. \tag{151}$$

The distinction between pure states (in Eq. (150)) and mixed states (in Eq. (151)) slipped in here was only made explicit in the second paper in the trilogy (von Neumann, 1927b). These conditional probabilities can be written in terms of the projection operators,

$$\hat{E}(I) \equiv \sum_{(n|E_n \,\text{in}\, I)} |\psi_n\rangle\langle\psi_n|, \quad \hat{F}(K) \equiv \int_K |q\rangle\langle q|\, dq, \tag{152}$$

that project arbitrary state vectors onto the subspaces of $\overline{\mathfrak{H}}$ spanned by 'eigenvectors' of the Hamiltonian $\hat{H}$ and of the position operator $\hat{q}$ with eigenvalues in the ranges I and K, respectively. The right-hand side of Eq. (151) can be rewritten as:

$$\sum_{(n|E_n \,\text{in}\, I)} \int_K \langle\psi_n|q\rangle\langle q|\psi_n\rangle\, dq. \tag{153}$$

We now choose an arbitrary orthonormal discrete basis $\{|\alpha\rangle\}_{\alpha=1}^{\infty}$ of the Hilbert space $\overline{\mathfrak{H}}$. Inserting the corresponding resolution of unity, $\hat{1} = \sum_\alpha |\alpha\rangle\langle\alpha|$, into Eq. (153), we find

$$\sum_\alpha \sum_{(n|E_n \,\text{in}\, I)} \int_K \langle\psi_n|\alpha\rangle\langle\alpha|q\rangle\langle q|\psi_n\rangle\, dq. \tag{154}$$

[69] The left-hand side is short-hand for: $\Pr(\hat{q}$ has value q in $K|\hat{H}$ has value $E_n)$. We remind the reader that the notation $\Pr(.,.)$ is ours and is not used in any of our sources.

[70] The left-hand side is short-hand for: $\Pr(\hat{q}$ has value q in $K|\hat{H}$ has value E_n in $I)$.

This can be rewritten as:

$$\sum_\alpha \langle\alpha| \left(\int_K |q\rangle\langle q|\, dq \;\cdot \sum_{(n|E_n \,\mathrm{in}\, I)} |\psi_n\rangle\langle\psi_n| \right) |\alpha\rangle. \tag{155}$$

This is nothing but the trace of the product of the projection operators $\hat{F}(K)$ and $\hat{E}(I)$ defined in Eq. (152). The conditional probability in Eq. (151) can thus be written as:

$$\Pr(x \text{ in } K|E_n \text{ in } I) = \sum_\alpha \langle\alpha|\hat{F}(K)\hat{E}(I)|\alpha\rangle = \mathrm{Tr}(\hat{F}(K)\hat{E}(I)). \tag{156}$$

This is our notation for what von Neumann (1927a, p. 45) wrote as [71]

$$[\hat{F}(K), \hat{E}(I)]. \tag{157}$$

He defined the quantity $[\hat{A}, \hat{B}]$—*not to be confused with a commutator*—as (ibid., p. 40):

$$[\hat{A}, \hat{B}] \equiv [\hat{A}^\dagger \hat{B}]. \tag{158}$$

For any operator $\hat{O}$, he defined the quantity $[\hat{O}]$, which he called the "absolute value" of $\hat{O}$, as (ibid., pp. 37–38): [72]

$$[\hat{O}] \equiv \sum_{\mu,\nu} |\langle\varphi_\mu|\hat{O}|\psi_\nu\rangle|^2, \tag{159}$$

where $\{|\varphi_\mu\rangle\}_{\mu=1}^\infty$ and $\{|\psi_\nu\rangle\}_{\nu=1}^\infty$ are two arbitrary orthonormal bases of $\overline{\mathfrak{H}}$. Eq. (159) can also be written as:

$$[\hat{O}] \equiv \sum_{\mu,\nu} \langle\varphi_\mu|\hat{O}|\psi_\nu\rangle\langle\psi_\nu|\hat{O}^\dagger|\varphi_\mu\rangle = \sum_\mu \langle\varphi_\mu|\hat{O}\hat{O}^\dagger|\varphi_\mu\rangle = \mathrm{Tr}(\hat{O}\hat{O}^\dagger), \tag{160}$$

where we used the resolution of unity, $\hat{1} = \sum_\nu |\psi_\nu\rangle\langle\psi_\nu|$, and the fact that $\mathrm{Tr}(\hat{O}) = \sum_\alpha \langle\alpha|\hat{O}|\alpha\rangle$ for any orthonormal basis $\{|\alpha\rangle\}_{\alpha=1}^\infty$ of $\overline{\mathfrak{H}}$. [73] Using the

[71] Since von Neumann (ibid., p. 43) chose K to be k-dimensional, he actually wrote: $[\hat{F}_1(J_1) \cdot \ldots \cdot \hat{F}_k(J_k), \hat{E}(I)]$ (ibid., p. 45; hats added). For the one-dimensional case we are considering, von Neumann's expression reduces to Eq. (157).

[72] Using the notation $Q(.,.)$ for the inner product (see note 67) and using A instead of O, von Neumann (1927a, p. 37) wrote the right-hand side of Eq. (159) as $\sum_{\mu,\nu=1}^\infty |Q(\phi_\mu, A\psi_\nu)|^2$.

[73] Eq. (160) shows that $[\hat{O}]$ is independent of the choice of the bases $\{|\varphi_\mu\rangle\}_{\mu=1}^\infty$ and $\{|\psi_\nu\rangle\}_{\nu=1}^\infty$. Von Neumann initially introduced the quantity $[\hat{O}; \varphi_\mu; \psi_\nu] \equiv \sum_{\mu,\nu} |\langle\varphi_\mu|\hat{O}|\psi_\nu\rangle|^2$ (ibid., p. 37). He then showed that this quantity does not actually depend on φ_μ and ψ_ν, renamed it $[\hat{O}]$ (see Eq. (159) and note 72), and called it the "absolute value of the operator" $\hat{O}$ (ibid., p. 38).

definitions of $[\hat{A}, \hat{B}]$ and $[\hat{O}]$ in Eqs. (158) and (160), with $\hat{A} = \hat{F}(K)$, $\hat{B} = \hat{E}(I)$, and $\hat{O} = \hat{F}(K)^\dagger \hat{E}(I)$, we can rewrite Eq. (157) as

$$[\hat{F}, \hat{E}] = [\hat{F}^\dagger \hat{E}] = \mathrm{Tr}((\hat{F}^\dagger \hat{E})(\hat{F}^\dagger \hat{E})^\dagger) = \mathrm{Tr}(\hat{F}^\dagger \hat{E} \hat{E}^\dagger \hat{F}), \qquad (161)$$

where, to make the equation easier to read, we temporarily suppressed the value ranges K and I of $\hat{F}(K)$ and $\hat{E}(I)$. Using the cyclic property of the trace, we can rewrite the final expression in Eq. (161) as $\mathrm{Tr}(\hat{F}\hat{F}^\dagger \hat{E}\hat{E}^\dagger)$. Since projection operators $\hat{P}$ are both Hermitian and idempotent, we have $\hat{F}\hat{F}^\dagger = \hat{F}^2 = \hat{F}$ and $\hat{E}\hat{E}^\dagger = \hat{E}^2 = \hat{E}$. Combining these observations and restoring the arguments of $\hat{F}$ and $\hat{E}$, we can rewrite Eq. (161) as:

$$[\hat{F}(K), \hat{E}(I)] = \mathrm{Tr}(\hat{F}(K)\hat{E}(I)), \qquad (162)$$

which is just the expression for $\Pr(x \text{ in } K | E_n \text{ in } I)$ that we found above (see Eq. (156)).

From $\mathrm{Tr}(\hat{F}\hat{E}) = \mathrm{Tr}(\hat{E}\hat{F})$, it follows that

$$\Pr(x \text{ in } K | E_n \text{ in } I) = \Pr(E_n \text{ in } I | x \text{ in } K), \qquad (163)$$

which is just the symmetry property imposed on Jordan's probability amplitudes in postulate B of *Neue Begründung* I (Jordan, 1927b, p. 813; see Section 2.1) and postulate II in *Neue Begründung* II (Jordan, 1927e, p. 6; see Section 4).

Von Neumann generalized Eq. (156) for a pair of quantities to one for a pair *of sets* of quantities such that the operators for all quantities in each set commute with those for all other quantities in that same set but not necessarily with those for quantities in the other set (von Neumann, 1927a, p. 45).[74] Let $\{\hat{R}_i\}_{i=1}^n$ and $\{\hat{S}_j\}_{j=1}^m$ be two such sets of commuting operators: $[\hat{R}_{i_1}, \hat{R}_{i_2}] = 0$ for all $1 \leq i_1, i_2 \leq n$; $[\hat{S}_{j_1}, \hat{S}_{j_2}] = 0$ for all $1 \leq j_1, j_2 \leq m$. Let $\hat{E}_i(I_i)$ $(i = 1, \ldots, n)$ and $\hat{F}_j(J_j)$ $(j = 1, \ldots, m)$ be the corresponding projection operators (cf. Eq. (152)). A straightforward generalization of von Neumann's trace formula (156) tells us that the probability that the $\hat{S}_j$'s have values in the intervals J_j provided that the $\hat{R}_i$'s have values in the intervals I_i is given by:

$$\Pr(\hat{S}_j\text{'s in } J_j\text{'s} | \hat{R}_i\text{'s in } I_i\text{'s}) = \mathrm{Tr}(\hat{E}_1(I_1) \ldots \hat{E}_n(I_n) \hat{F}_1(J_1) \ldots \hat{F}_m(J_m)). \qquad (164)$$

[74] Von Neumann distinguished between the commuting of $\hat{R}_i$ and $\hat{R}_j$ and the commuting of the corresponding projection operators $\hat{E}_i(I_i)$ and $\hat{E}_j(I_j)$. For *bounded* operators, these these two properties are equivalent. If both $\hat{R}_i$ and $\hat{R}_j$ are *unbounded*, von Neumann (1927a, p. 45) cautioned, "certain difficulties of a formal nature occur, which we do not want to go into here" (cf. note 68).

The outcomes '$\hat{R}_i$ in I_i' are called the "assertions" (*Behauptungen*) and the outcomes '$\hat{S}_j$ in J_j' are called the "conditions" (*Voraussetzungen*) (von Neumann, 1927a, p. 45). Because of the cyclic property of the trace, which we already invoked in Eq. (163), Eq. (164) is invariant under switching all assertions with all conditions. Since all $\hat{E}_i(I_i)$'s commute with each other and all $\hat{F}_j(J_j)$'s commute with each other, Eq. (164) is also invariant under changing the order of the assertions and changing the order of the conditions. These two properties are given in the first two entries of a list of five properties, labeled α through θ (there are no points ζ and η), of the basic rule (164) for probabilities in quantum mechanics (von Neumann, 1927a, pp. 45–47).

Under point δ, von Neumann pointed out that the standard multiplication law of probabilities does not hold in quantum mechanics. Parenthetically, he added, referring to Jordan (1927b) and Hilbert, von Neumann, and Nordheim (1928): "what does hold is a weaker law corresponding to the "superposition [*Zusammensetzung*] of probability amplitudes" in [the formalism of] Jordan, which we will not go into here" (von Neumann, 1927a, p. 46). Note that von Neumann did not use Jordan's phrase "interference of probabilities."

Under point ε, von Neumann (1927a) wrote: "The addition rule of probabilities is valid" (p. 46). In general, as Jordan (1927b, p. 18) made clear in *Neue Begründung* I (see Eq. (1) in Section 1), the addition rule does *not* hold in quantum mechanics. In general, in other words, $\Pr(A \text{ or } B) \neq \Pr(A) + \Pr(B)$, even if the outcomes A and B are mutually exclusive. Instead, Jordan pointed out, the addition rule, like the multiplication rule, holds for the corresponding probability *amplitudes*. Von Neumann, however, considered only a rather special case, for which the addition rule *does* hold for the probability themselves. Consider Eq. (156) for the conditional probability that we find a particle in some region K given that its energy E has a value in some interval I. Let the region K consist of two disjoint subregions, K' and K'', such that $K = K' \cup K''$ and $K' \cap K'' = \emptyset$. Given that the energy E lies in the interval I, the probability that the particle is *either* in K' *or* in K'', is obviously equal to the probability that it is in K. Von Neumann now noted that

$$\Pr(x \text{ in } K | E \text{ in } I) = \Pr(x \text{ in } K' | E \text{ in } I) + \Pr(x \text{ in } K'' | E \text{ in } I). \tag{165}$$

In terms of the trace formula (156), Eq. (165) becomes:

$$\mathrm{Tr}(\hat{F}(K)\hat{E}(I)) = \mathrm{Tr}(\hat{F}(K')\hat{E}(I)) + \mathrm{Tr}(\hat{F}(K'')\hat{E}(I)). \tag{166}$$

Similar instances of the addition rule obtain for the more general version of the trace formula in Eq. (164).

Under point ϑ, finally, we find the one and only reference to "canonical transformations" in *Mathematische Begründung*. Von Neumann (1927a, pp. 46–47) defined a canonical transformation as the process of subjecting *all* operators $\hat{A}$ to the transformation $\hat{U}\hat{A}\hat{U}^\dagger$, where $\hat{U}$ is some unitary operator. The absolute value squared $[\hat{A}]$ is invariant under such transformations. Recall $[\hat{A}] = \mathrm{Tr}(\hat{A}\hat{A}^\dagger)$. Now consider $[\hat{U}\hat{A}\hat{U}^\dagger]$:

$$[\hat{U}\hat{A}\hat{U}^\dagger] \equiv [\hat{U}\hat{A}\hat{U}^\dagger(\hat{U}\hat{A}\hat{U}^\dagger)^\dagger] = \mathrm{Tr}(\hat{U}\hat{A}\hat{U}^\dagger\hat{U}\hat{A}^\dagger\hat{U}^\dagger) = \mathrm{Tr}(\hat{A}\hat{A}^\dagger) = [\hat{A}]. \quad (167)$$

Traces of products of operators are similarly invariant. This definition of canonical transformations makes no reference in any way to sorting quantities into sets of conjugate variables.

6 Von Neumann's *Wahrscheinlichkeitstheoretischer Aufbau* (November 1927)

On November 11, 1927, about half a year after the first installment, *Mathematische Begründung* (von Neumann, 1927a), the second and the third installment of von Neumann's 1927 trilogy were presented to the Göttingen Academy (von Neumann, 1927b,c).[75] The second, *Wahrscheinlichkeitstheoretischer Aufbau*, is important for our purposes; the third, *Thermodynamik quantenmechanischer Gesamtheiten*, is not. In *Mathematische Begründung*, as we saw in Section 5, von Neumann had simply taken over the basic rule for probabilities in quantum mechanics as stated by Jordan, namely that probabilities are given by the absolute square of the corresponding probability amplitudes, the prescription now known as the Born rule. In *Wahrscheinlichkeitstheoretischer Aufbau*, he sought to derive this rule from more basic considerations.

In the introduction of the paper, von Neumann (1927b, p. 245) replaced the old opposition between "wave mechanics" and "matrix mechanics" by a new distinction between "wave mechanics" on the one hand and what he called "transformation theory" or "statistical theory," on the other. By this time, matrix mechanics and Dirac's q-number theory had morphed into the Dirac-Jordan statistical transformation theory. The two names von Neumann used for this theory reflect the difference in emphasis between Dirac (transformation theory) and Jordan (statistical theory).[76] Von Neumann mentioned Born,

[75] These three papers take up 57, 28, and 19 pages. The first installment is thus longer than the other two combined. Note that in between the first and the two later installments, *Neue Begründung* II appeared (see Section 4), in which Jordan (1927e) responded to von Neumann's criticism in *Mathematische Begründung*. Von Neumann made no comment on this response in these two later papers.

[76] In the introduction, we already quoted his observation about the Schrödinger wave function, $\langle q|E\rangle$ in our notation: "Dirac interprets it as a row of a certain transformation matrix, Jordan calls it a probability amplitude" (von Neumann, 1927b, p. 246, note 3).

Pauli, and London as the ones who had paved the way for the statistical theory and Dirac and Jordan as the ones responsible for bringing this development to a conclusion (ibid., p. 245; cf. note 57).[77]

Von Neumann was dissatisfied with the way in which probabilities were introduced in the Dirac-Jordan theory. He listed two objections. First, he felt that the relation between quantum probability concepts and ordinary probability theory needed to be clarified. Second, he felt that the Born rule was not well-motivated:

> The method hitherto used in statistical quantum mechanics was essentially *deductive*: the square of the norm of certain expansion coefficients of the wave function or of the wave function itself was fairly *dogmatically* set equal to a probability, and agreement with experience was verified afterwards. A systematic derivation of quantum mechanics from empirical facts or fundamental probability-theoretic assumptions, i.e., an *inductive* justification, was not given (von Neumann, 1927b, p. 246; our emphasis).

To address these concerns, von Neumann started by introducing probabilities in terms of selecting members from large ensembles of systems. He then presented his "inductive" derivation of his trace formula for probabilities (see Eq. (156)), which contains the Born rule as a special case, from two very general, deceptively innocuous, but certainly non-trivial assumptions about expectation values of properties of the systems in such ensembles. From those two assumptions, some key elements of the Hilbert space formalism introduced in *Mathematische Begründung*, and two assumptions about the repeatability of measurements not identified until the summary at the end of the paper (ibid., p. 271), Von Neumann indeed managed to recover Eq. (156) for probabilities. He downplayed his reliance on the formalism of *Mathematische Begründung* by characterizing the assumptions taken from it as "not very far going formal and material assumptions" (ibid., p. 246). He referred to sec. IX, the summary of the paper, for these assumptions at this point, but most of them are already stated, more explicitly in fact, in sec. II, "basic assumptions" (ibid.,

[77] He cited the relevant work by Dirac (1927) and Jordan (1927b,e). He did not give references for the other three authors but presumably was thinking of Born (1926a,b), Pauli (1927a), and London (1926b). The reference to London is somewhat puzzling. While it is true that London anticipated important aspects of the Dirac-Jordan transformation theory (see Lacki, 2004; Duncan and Janssen, 2009), the statistical interpretation of the formalism is not among those. Our best guess is that von Neumann took note of Jordan's repeated acknowledgment of London's paper (most prominently perhaps in footnote 1 of *Neue Begründung* I). In his book, von Neumann (1932, p. 2, note 2; the note itself is on p. 238) cited papers by Dirac (1927), Jordan (1927b), and London (1926b) in addition to the book by Dirac (1930) for the development of transformation theory. In that context, the reference to London is entirely appropriate.

pp. 249–252).

Consider an ensemble $\{\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3, \ldots\}$ of copies of a system $\mathfrak{S}$. Von Neumann wanted to find an expression for the expectation value $\mathcal{E}(\mathfrak{a})$ in that ensemble of some property $\mathfrak{a}$ of the system (we use $\mathcal{E}$ to distinguish the expectation value from the projection operator E). He made the following basic assumptions about $\mathcal{E}$ (von Neumann, 1927b, pp. 249–250):

A. *Linearity*: $\mathcal{E}(\alpha\,\mathfrak{a} + \beta\,\mathfrak{b} + \gamma\,\mathfrak{c} + \ldots) = \alpha\,\mathcal{E}(\mathfrak{a}) + \beta\,\mathcal{E}(\mathfrak{b}) + \gamma\,\mathcal{E}(\mathfrak{c}) + \ldots$ (where α, β, and γ are real numbers).[78]
B. *Positive-definiteness.* If the quantity $\mathfrak{a}$ never takes on negative values, then $\mathcal{E}(\mathfrak{a}) \geq 0$.

To this he added two formal assumptions (ibid., p. 252):

C. *Linearity of the assignment of operators to quantities.* If the operators $\hat{S}$, $\hat{T}$, ... represent the quantities $\mathfrak{a}$, $\mathfrak{b}$, ..., then $\alpha\hat{S} + \beta\hat{T} + \ldots$ represents the quantity $\alpha\,\mathfrak{a} + \beta\,\mathfrak{b} + \ldots$[79]
D. If the operator $\hat{S}$ represents the quantity $\mathfrak{a}$, then $f(\hat{S})$ represents the quantity $f(\mathfrak{a})$.

In sec. IX, the summary of his paper, von Neumann once again listed the assumptions that go into his derivation of the expression for $\mathcal{E}(\mathfrak{a})$. He wrote:

> The goal of the present paper was to show that quantum mechanics is not only compatible with the usual probability calculus, but that, if it [i.e., ordinary probability theory]—along with a few plausible factual [*sachlich*] assumptions—is taken as given, it [i.e., quantum mechanics] is actually the only possible solution. The assumptions made were the following:
> 1. Every measurement changes the measured object, and two measurements therefore always disturb each other—except when they can be replaced by a single measurement.

[78] Here von Neumann appended a footnote in which he looked at the example of a harmonic oscillator in three dimensions. The same point can be made with a one-dimensional harmonic oscillator with position and momentum operators $\hat{q}$ and $\hat{p}$, Hamiltonian $\hat{H}$, mass m, and characteristic angular frequency ω: "The three quantities $[\hat{p}/2m, m\omega^2\hat{q}^2/2, \hat{H} = \hat{p}/2m + m\omega^2\hat{q}^2/2]$ have very different spectra: the first two both have a continuous spectrum, the third has a discrete spectrum. Moreover, no two of them can be measured simultaneously. Nevertheless, the sum of the expectation values of the first two equals the expectation value of the third" (ibid., p. 249). While it may be reasonable to impose condition (A) on directly measurable quantities, it is questionable whether this is also reasonable for hidden variables (see note 83).
[79] In von Neumann's own notation, the operator $\hat{S}$ and the matrix S representing that operator are both written simply as S.

2. However, the change caused by a measurement is such that the measurement itself retains its validity, i.e., if one repeats it immediately afterwards, one finds the same result.

In addition, a formal assumption:

3. Physical quantities are to be described by functional operators in a manner subject to a few simple formal rules.

These principles already inevitably entail quantum mechanics and its statistics (von Neumann, 1927b, p. 271).

Assumptions A and B of sec. II are not on this new list in sec. IX. Presumably, this is because they are part of ordinary probability theory. Conversely, assumptions 1 and 2 of sec. IX are not among the assumptions A–D of sec. II. These two properties of measurements, as we will see below, are guaranteed in von Neumann's formalism by the idempotency of the projection operators associated with those measurements.[80] Finally, the "simple formal rules" referred to in assumption 3 are spelled out in assumptions C–D.

We go over the main steps of von Neumann's proof, which is laid out clearly and in detail in his paper. Instead of the general Hilbert space $\bar{\mathfrak{H}}$, von Neumann considered $\mathfrak{H}_0$, i.e., l^2 (von Neumann, 1927b, p. 253; cf. von Neumann, 1927a, pp. 14–15 (see Section 5)). Consider the components $s_{\mu\nu}$ of some infinite-dimensional Hermitian matrix S (with matrix elements $s_{\mu\nu} = s^*_{\nu\mu}$) representing an Hermitian operator $\hat{S}$. This operator, in turn, represents some measurable quantity $\mathfrak{a}$. The matrix S can be written as a linear combination of three types of infinite-dimensional matrices labeled A, B, and C. To show what these matrices look like, we write down their finite-dimensional counterparts:

$$A_\mu \equiv \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 0 & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}, B_{\mu\nu} \equiv \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & 0 & & 1 & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & 1 & & 0 & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}, C_{\mu\nu} \equiv \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & 0 & & i & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & -i & & 0 & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}.$$

The A_μ's have 1 in the μ^{th} row and the μ^{th} column and 0's everywhere else. The $B_{\mu\nu}$'s ($\mu < \nu$) have 1 in the μ^{th} row and the ν^{th} column and in the ν^{th} row and the μ^{th} column and 0's everywhere else. The $C_{\mu\nu}$'s ($\mu < \nu$) have i in

[80] In a footnote in the introduction of *Thermodynamik quantenmechanischer Gesamtheiten*, von Neumann (1927c) reiterated these two assumptions and commented: "1. corresponds to the explanation given by Heisenberg for the a-causal behavior of quantum physics; 2. expresses that the theory nonetheless gives the appearance of a kind of causality" (p. 273, note 2). Von Neumann cited Heisenberg (1927b), the paper on the uncertainty principle (submitted March 23, 1927).

the μ^{th} row and the ν^{th} column and $-i$ in the ν^{th} row and the μ^{th} column and 0's everywhere else.

The matrix S can be written as a linear combination of A, B, and C:

$$S = \sum_{\mu} s_{\mu\mu} \cdot A_{\mu} + \sum_{\mu<\nu} \operatorname{Re} s_{\mu\nu} \cdot B_{\mu\nu} + \sum_{\mu<\nu} \operatorname{Im} s_{\mu\nu} \cdot C_{\mu\nu}, \tag{168}$$

where $\operatorname{Re} s_{\mu\nu}$ and $\operatorname{Im} s_{\mu\nu}$ and the real and imaginary parts of $s_{\mu\nu}$, respectively.

Using von Neumann's linearity assumption (A), we can write the expectation value of S in the ensemble $\{\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3, \ldots\}$ as:

$$\mathcal{E}(S) = \sum_{\mu} s_{\mu\mu} \cdot \mathcal{E}(A_{\mu}) + \sum_{\mu<\nu} \operatorname{Re} s_{\mu\nu} \cdot \mathcal{E}(B_{\mu\nu}) + \sum_{\mu<\nu} \operatorname{Im} s_{\mu\nu} \cdot \mathcal{E}(C_{\mu\nu}). \tag{169}$$

Now define the matrix U (associated with some operator $\hat{U}$) with diagonal components $u_{\mu\mu} \equiv \mathcal{E}(A_{\mu})$ and off-diagonal components ($\mu < \nu$):

$$u_{\mu\nu} \equiv \frac{1}{2} \left(\mathcal{E}(B_{\mu\nu}) + i\, \mathcal{E}(C_{\mu\nu}) \right), \quad u_{\nu\mu} \equiv \frac{1}{2} \left(\mathcal{E}(B_{\mu\nu}) - i\, \mathcal{E}(C_{\mu\nu}) \right). \tag{170}$$

Note that this matrix is Hermitian: $u^*_{\mu\nu} = u_{\nu\mu}$. With the help of this matrix U, the expectation value of S can be written as (von Neumann, 1927b, p. 253):

$$\mathcal{E}(S) = \sum_{\mu\nu} s_{\mu\nu}\, u_{\nu\mu}. \tag{171}$$

To verify this, we consider the sums over $\mu = \nu$ and $\mu \neq \nu$ separately. For the former we find

$$\sum_{\mu} s_{\mu\mu}\, u_{\mu\mu} = \sum_{\mu} s_{\mu\mu} \cdot \mathcal{E}(A_{\mu}). \tag{172}$$

For the latter, we have

$$\sum_{\mu\neq\nu} s_{\mu\nu}\, u_{\nu\mu} = \sum_{\mu<\nu} s_{\mu\nu}\, u_{\nu\mu} + \sum_{\mu>\nu} s_{\mu\nu}\, u_{\nu\mu}. \tag{173}$$

The second term can be written as $\sum_{\nu>\mu} s_{\nu\mu}\, u_{\mu\nu} = \sum_{\mu<\nu} s^*_{\mu\nu}\, u^*_{\nu\mu}$, which means that

$$\sum_{\mu\neq\nu} s_{\mu\nu}\, u_{\nu\mu} = \sum_{\mu<\nu} 2 \operatorname{Re} (s_{\mu\nu}\, u_{\nu\mu}). \tag{174}$$

Now write $s_{\mu\nu}$ as the sum of its real and imaginary parts and use Eq. (170) for $u_{\nu\mu}$:

$$\begin{aligned} \sum_{\mu\neq\nu} s_{\mu\nu}\, u_{\nu\mu} &= \sum_{\mu<\nu} \operatorname{Re} \{ (\operatorname{Re} s_{\mu\nu} + i \operatorname{Im} s_{\mu\nu}) \cdot (\mathcal{E}(B_{\mu\nu}) - i\mathcal{E}(C_{\mu\nu})) \} \\ &= \sum_{\mu<\nu} \operatorname{Re} s_{\mu\nu} \cdot \mathcal{E}(B_{\mu\nu}) + \sum_{\mu<\nu} \operatorname{Im} s_{\mu\nu} \cdot \mathcal{E}(C_{\mu\nu}). \end{aligned} \tag{175}$$

Adding Eq. (172) and Eq. (175), we arrive at

$$\sum_{\mu\nu} s_{\mu\nu}\, u_{\nu\mu} = \sum_{\mu} s_{\mu\mu} \cdot \mathcal{E}(A_\mu) + \sum_{\mu<\nu} \mathrm{Re}\, s_{\mu\nu} \cdot \mathcal{E}(B_{\mu\nu}) + \sum_{\mu<\nu} \mathrm{Im}\, s_{\mu\nu} \cdot \mathcal{E}(C_{\mu\nu}). \quad (176)$$

Eq. (169) tells us that the right-hand side of this equation is just $\mathcal{E}(S)$. This concludes the proof of Eq. (171), in which one readily recognizes the trace of the product of S and U:[81]

$$\mathcal{E}(S) = \sum_{\mu\nu} s_{\mu\nu}\, u_{\nu\mu} = \sum_{\mu} (SU)_{\mu\mu} = \mathrm{Tr}(US). \quad (177)$$

In other words, U is what is now called a *density matrix*, usually denoted by the Greek letter ρ. It corresponds to a density operator $\hat{U}$ or $\hat{\rho}$.

The matrix U characterizes the ensemble $\{\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3, \ldots\}$. Von Neumann (1927b, sec. IV, p. 255) now focused on "pure" (*rein*) or "uniform" (*einheitlich*) ensembles, in which every copy $\mathfrak{S}_i$ of the system is in the exact same state. Von Neumann characterized such ensembles as follows: one cannot obtain a uniform ensemble "by mixing (*vermischen*) two ensembles unless it is the case that both of these correspond to that same ensemble" (ibid., p. 256). He then proved an important theorem. Let the density operators $\hat{U}$, $\hat{U}^*$, and $\hat{U}^{**}$ correspond to the ensembles $\{\mathfrak{S}_i\}$, $\{\mathfrak{S}_j^*\}$, $\{\mathfrak{S}_k^{**}\}$, respectively. Suppose $\{\mathfrak{S}_i\}$ consists of $\eta \times 100\%$ $\{\mathfrak{S}_j^*\}$ and $\vartheta \times 100\%$ $\{\mathfrak{S}_k^{**}\}$. The expectation value of an arbitrary property represented by the operator $\hat{S}$ in $\{\mathfrak{S}_i\}$ is then given by

$$\mathcal{E}(\hat{S}) = \eta\, \mathcal{E}^*(\hat{S}) + \vartheta\, \mathcal{E}^{**}(\hat{S}), \quad (178)$$

where $\mathcal{E}^*$ and $\mathcal{E}^{**}$ refer to ensemble averages over $\{\mathfrak{S}_j^*\}$ and $\{\mathfrak{S}_k^{**}\}$, respectively. Using Eq. (177), we can write this as:

$$\mathrm{Tr}(\hat{U}\hat{S}) = \eta\, \mathrm{Tr}(\hat{U}^*\hat{S}) + \vartheta\, \mathrm{Tr}(\hat{U}^{**}\hat{S}). \quad (179)$$

Since $\hat{S}$ is arbitrary, it follows that $\hat{U}$, $\hat{U}^*$, and $\hat{U}^{**}$ satisfy

$$\hat{U} = \eta\, \hat{U}^* + \vartheta\, \hat{U}^{**}. \quad (180)$$

Von Neumann now proved a theorem pertaining to uniform ensembles (ibid., pp. 257–258). That $\hat{U}$ is the density operator for a uniform ensemble can be expressed by the following conditional statement: If $(\hat{U} = \hat{U}^* + \hat{U}^{**})$ then $(\hat{U}^* \propto \hat{U}^{**} \propto \hat{U})$. Von Neumann showed that this is equivalent to the statement that there is a unit vector $|\varphi\rangle$ such that $\hat{U}$ is the projection operator onto that

[81] Von Neumann (1927b, p. 255) only wrote down the first step of Eq. (177). It was only in the third installment of his trilogy, that von Neumann (1927c, p. 274) finally introduced the notation trace (*Spur*), which we used here and in Eq. (156) in Section 5.

vector, i.e., $\hat{U} = \hat{P}_\varphi = |\varphi\rangle\langle\varphi|$.[82] Written more compactly, the theorem says:

$$\{ (\hat{U} = \hat{U}^* + \hat{U}^{**}) \Rightarrow (\hat{U}^* \propto \hat{U}^{**} \propto \hat{U}) \} \Leftrightarrow \{ \exists\, |\varphi\rangle,\, \hat{U} = \hat{P}_\varphi = |\varphi\rangle\langle\varphi| \}. \quad (181)$$

The crucial input for the proof of the theorem is the inner-product structure of Hilbert space. The theorem implies two important results, which, given the generality of the assumptions going into the proof of the theorem, have the unmistakable flavor of a free lunch. First, pure dispersion-free states (or ensembles) correspond to unit vectors in Hilbert space.[83] Second, the expectation value of a quantity $\mathfrak{a}$ represented by the operator $\hat{S}$ in a uniform ensemble $\{\mathfrak{S}_i\}$ characterized by the density operator $\hat{U} = |\varphi\rangle\langle\varphi|$ is given by the trace of the product of the corresponding matrices:

$$\mathcal{E}(\hat{S}) = \mathrm{Tr}(\hat{U}\hat{S}) = \mathrm{Tr}(|\varphi\rangle\langle\varphi|\hat{S}) = \langle\varphi|\hat{S}|\varphi\rangle, \quad (182)$$

which is equivalent to the Born rule.

Von Neumann was still not satisfied. In sec. V, "Measurements and states," he noted that

> our knowledge of a system $\mathfrak{S}'$, the structure of a statistical ensemble $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \ldots\}$, is never described by the specification of a state—or even by the corresponding φ [i.e., the vector $|\varphi\rangle$]; but usually by the result of measurements performed on the system (von Neumann, 1927b, p. 260).

He considered the simultaneous measurement of a complete set of commuting operators and constructed a density operator for (the ensemble representing) the system on the basis of the outcomes of these measurements showing the measured quantities to have values in certain intervals. He showed that these measurements can fully determine the state and that the density operator in that case is the projection operator onto that state.

Let $\{\hat{S}_\mu\}$ ($\mu = 1, \ldots, m$) be a complete set of commuting operators with

[82] The notation $\hat{P}_\varphi$ (except for the hat) is von Neumann's own (ibid., p. 257).
[83] This is the essence of von Neumann's later no-hidden variables proof (von Neumann, 1932, Ch. 4, p. 171), which was criticized by John Bell (1966, pp. 1–5), who questioned the linearity assumption (A), $\mathcal{E}(\alpha\,\mathfrak{a} + \beta\,\mathfrak{b}) = \alpha\,\mathcal{E}(\mathfrak{a}) + \beta\,\mathcal{E}(\mathfrak{b})$ (see note 78). Bell argued, with the aid of explicit examples, that the linearity of expectation values was too strong a requirement to impose on hypothetical dispersion-free states (dispersion-free via specification of additional "hidden" variables). In particular, the dependence of spin expectation values on the (single) hidden variable in the explicit example provided by Bell is manifestly *nonlinear*, although the model reproduces exactly the standard quantum-mechanical results when one averages (uniformly) over the hidden variable. For recent discussion, see Bacciagaluppi and Crull (2009).

common eigenvectors, $\{|\sigma_n\rangle\}$ with eigenvalues $\lambda_\mu(n)$:

$$\hat{S}_\mu\,|\sigma_n\rangle = \lambda_\mu(n)\,|\sigma_n\rangle. \tag{183}$$

Now construct an operator $\hat{S}$ with those same eigenvectors and completely non-degenerate eigenvalues λ_n:

$$\hat{S}|\sigma_n\rangle = \lambda_n|\sigma_n\rangle, \tag{184}$$

with $\lambda_n \neq \lambda_{n'}$ if $n \neq n'$. Define the functions $f_\mu(\lambda_n) = \lambda_\mu(n)$. Consider the action of $f_\mu(\hat{S})$ on $|\sigma_n\rangle$:

$$f_\mu(\hat{S})\,|\sigma_n\rangle = f_\mu(\lambda_n)\,|\sigma_n\rangle = \lambda_\mu(n)\,|\sigma_n\rangle = \hat{S}_\mu\,|\sigma_n\rangle. \tag{185}$$

Hence, $\hat{S}_\mu = f_\mu(\hat{S})$. It follows from Eq. (185) that a measurement of $\hat{S}$ uniquely determines the state of the system. As von Neumann (1927b) put it: "In this way measurements have been identified that uniquely determine the state of [the system represented by the ensemble] $\mathfrak{S}'$" (p. 264).

As a concrete example, consider the bound states of a hydrogen atom. These states are uniquely determined by the values of four quantum numbers: the principal quantum number n, the orbital quantum number l, the magnetic quantum number m_l, and the spin quantum number m_s. These four quantum numbers specify the eigenvalues of four operators, which we may make dimensionless by suitable choices of units: the Hamiltonian in Rydberg units ($\hat{H}/\mathrm{Ry}$), the angular momentum squared ($\hat{L}^2/\hbar^2$), the z-component of the angular momentum ($\hat{L}_z/\hbar$), and the z-component of the spin ($\hat{\sigma}_z/\hbar$). In this case, in other words,

$$\{\hat{S}_\mu\}_{\mu=1}^4 = (\hat{H}/\mathrm{Ry}, \hat{L}^2/\hbar^2, \hat{L}_z/\hbar, \hat{\sigma}_z/\hbar). \tag{186}$$

The task now is to construct an operator $\hat{S}$ that is a function of the $\hat{S}_\mu$'s (which have rational numbers as eigenvalues) and that has a completely non-degenerate spectrum. One measurement of $\hat{S}$ then uniquely determines the (bound) state of the hydrogen atom. For example, choose α, β, γ, and δ to be four real numbers, incommensurable over the rationals (i.e., no linear combination of $\alpha, \beta, \gamma, \delta$ with rational coefficients vanishes), and define

$$\hat{S} = \alpha\hat{S}_1 + \beta\hat{S}_2 + \gamma\hat{S}_3 + \delta\hat{S}_4 \tag{187}$$

One sees immediately that the specification of the eigenvalue of $\hat{S}$ suffices to uniquely identify the eigenvalues of $\hat{H}, \hat{L}^2, \hat{L}_z$ and $\hat{\sigma}_z$.

Von Neumann thus arrived at the typical statement of a problem in modern quantum mechanics. There is no need anymore for $\hat{q}$'s and $\hat{p}$'s, where the $\hat{p}$'s do not commute with the $\hat{q}$'s. Instead one identifies a complete set of commuting operators. Since all members of the set commute with one another, they can all be viewed as $\hat{q}$'s. The canonically conjugate $\hat{p}$'s do not make an appearance.

To conclude this section, we want to draw attention to one more passage in *Wahrscheinlichkeitstheoretischer Aufbau*. Both Jordan and von Neumann considered conditional probabilities of the form

$$\Pr(\hat{a} \text{ has the value } a \,|\, \hat{b} \text{ has the value } b),$$

or, more generally,

$$\Pr(\hat{a} \text{ has a value in interval } I \,|\, \hat{b} \text{ has a value in interval } J).$$

To test the quantum-mechanical predictions for these probabilities one needs to prepare the system under consideration in a pure state in which $\hat{b}$ has the value b or in a mixed state in which $\hat{b}$ has a value in the interval I, and then measure $\hat{a}$. A question that has not been addressed so far is what happens *after* that measurement. Von Neumann did address just this question in the concluding section of *Wahrscheinlichkeitstheoretischer Aufbau*:

> A system left to itself (not disturbed by any measurements) has a completely causal time evolution [governed by the Schrödinger equation]. In the confrontation with experiments, however, the statistical character is unavoidable: for every experiment there is a state adapted [*angepaßt*] to it in which the result is uniquely determined (the experiment in fact produces such states if they were not there before); however, for every state there are "non-adapted" measurements, the execution of which demolishes [*zertrümmert*] that state and produces adapted states according to stochastic laws (von Neumann, 1927b, pp. 271–272)

As far as we know, this is the first time the infamous collapse of the state vector in quantum mechanics was mentioned in print.

7 Conclusion: Never mind your p's and q's

The postulates of Jordan's *Neue Begründung* papers amount to a clear and concise formulation of the fundamental tenets of the probabilistic interpretation of quantum mechanics. Jordan (1927b) was the first to state in full generality that probabilities in quantum mechanics are given by the absolute square of what he called probability amplitudes. He was also the first to recognize that, in quantum mechanics, the addition and multiplication rules of ordinary probability theory apply to these probability amplitudes and, at least in general, not to the probabilities themselves. He did not succeed, however, in constructing a satisfactory formalism capturing the quantum-mechanical laws governing these probabilities (such as the Schrödinger equation, time-dependent and time-independent) and the various relations between different probability amplitudes as given by the quantum theory that he was trying to axiomatize with his postulates.

As we argued in this paper, Jordan was lacking the requisite mathematical tools to do so, namely abstract Hilbert space and the spectral theorem for operators acting in Hilbert space. Instead, Jordan used the canonical formalism of classical mechanics in his attempt to construct a realization of his postulates. Jordan was steeped in this formalism, which had played a central role in the transition from the old quantum theory to matrix mechanics (Duncan and Janssen, 2007). This is also true for the further development of the new theory, to which Jordan had made a number of significant contributions (Duncan and Janssen, 2008, 2009). Most importantly in view of the project Jordan pursued in *Neue Begründung*, he had published two papers the year before (Jordan, 1926a,b), in which he investigated the implementation of canonical transformations in matrix mechanics (Lacki, 2004; Duncan and Janssen, 2009). As he put it in his AHQP interview (see Section 2.2 for the exact quotation), canonically conjugate variables and canonical transformations had thus been his "daily bread" in the years leading up to *Neue Begründung*.

Unfortunately, as we saw in Sections 2–4, this formalism—the $\hat{p}$'s and $\hat{q}$'s—proved ill-suited to the task at hand. As a result, Jordan ran into a number of serious problems. First, it turns out to be crucially important for the probability interpretation of the formalism that only Hermitian operators be allowed. Unfortunately, canonical transformations can turn $\hat{p}$'s and $\hat{q}$'s corresponding to Hermitian operators into new $\hat{P}$'s and $\hat{Q}$'s that do not correspond to Hermitian operators. Initially, Jordan hoped to make room for such non-Hermitian quantities in his formalism by introducing the so-called *Ergänzungsamplitude* (see Section 2.4). Eventually, following the lead of Hilbert, von Neumann, and Nordheim (1928), he dropped the *Ergänzungsamplitude*, which forced him to restrict the class of allowed canonical transformations rather arbitrarily to those preserving Hermiticity. The difficulties facing Jordan's approach became particularly severe when, in *Neue Begründung* II, Jordan (1927e) tried to extend his formalism, originally formulated only for quantities with purely continuous spectra, to quantities with wholly or partly discrete spectra. One problem with this extension was that, whereas canonical transformations do not necessarily preserve Hermiticity, they *do* preserve the spectra of the $\hat{p}$'s and $\hat{q}$'s to which they are applied. Hence, there is no canonical transformation, for instance, that connects the generalized coordinate $\hat{q}$, which has a continuous spectrum, to the Hamiltonian $\hat{H}$, which, in general, will have at least a partly discrete spectrum. As Jordan's construction of a realization of his postulates hinged on the existence of a canonical transformation connecting $\hat{q}$ and $\hat{H}$, this presented an insurmountable obstacle. The newly introduced spin variables further exposed the limitations of Jordan's canonical formalism. To subsume these variables under his general approach, Jordan had to weaken the definition of when two quantities are to be considered canonically conjugate to such an extent that the concept lost much of its meaning. Under the definition Jordan adopted in *Neue Begründung* II, any two of the three components $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$ of spin angular momentum are canonically conjugate to each other.

All these problems can be avoided if the canonical formalism of classical mechanics is replaced by the Hilbert space formalism, even though other mathematical challenges remain. When Jordan's probability amplitudes $\varphi(a, b)$ for the quantities $\hat{a}$ and $\hat{b}$ are equated with 'inner products' $\langle a|b\rangle$ of normalized 'eigenvectors' of the corresponding operators $\hat{a}$ and $\hat{b}$, the rules for such amplitudes as laid down in the postulates of Jordan's *Neue Begründung* are automatically satisfied. Probabilities are given by the absolute square of these inner products, and Jordan's addition and multiplication rules for probability amplitudes reduce to the familiar completeness and orthogonality relations in Hilbert space (see Section 2.1, Eq. (3)). Once the Hilbert space formalism is adopted, the need to sort quantities into $\hat{p}$'s and $\hat{q}$'s disappears. Canonical transformations, at least in the classical sense as understood by Jordan, similarly cease to be important. Instead of canonical transformations $(\hat{p}, \hat{q}) \rightarrow (T\hat{p}T^{-1}, T\hat{q}T^{-1})$ of pairs of canonically conjugate quantities, one now considers unitary transformations $\hat{A} \rightarrow U\hat{A}U^{-1}$ of individual Hermitian operators. Such transformations get us from one orthonormal basis of Hilbert space to another, preserving inner products as required by the probability interpretation of quantum theory.

The Hilbert space formalism was introduced by von Neumann (1927a) in *Mathematische Begründung*. However, von Neumann did not use this formalism to provide a realization of Jordan's postulates along the lines sketched in the preceding paragraph. As we saw in Section 5, Von Neumann had some fundamental objections to the approach of Jordan (and Dirac). The basic probability amplitude for $\hat{p}$ and $\hat{q}$ in Jordan's formalism, $\langle p|q\rangle = e^{-ipq/\hbar}$ (see Eq. (4)), is not a square-integrable function and is thus not an element of the space L^2 instantiating abstract Hilbert space. The delta function, which is unavoidable in the Dirac-Jordan formalism, is simply not a function at all. Moreover, von Neumann objected to the phase-ambiguity of the probability amplitudes.

Jordan's response to this last objection illustrates the extent to which he was still trapped in thinking solely in terms of $\hat{p}$'s and $\hat{q}$'s. In *Neue Begründung* II, he eliminated the phase-ambiguity of the probability amplitude for any two quantities by adding two indices indicating a specific choice of the quantities canonically conjugate to those two quantities (see Section 4, Eqs. (133)–(139)). Von Neumann's response to this same problem was very different and underscores that he was not wedded at all to the canonical formalism of classical mechanics. Von Neumann decided to avoid probability amplitudes altogether.

Instead he turned to projection operators in Hilbert space, which he used both to formulate the spectral theorem and to construct a new formula for conditional probabilities in quantum mechanics (see Eq. (156) and Eq. (164)).

Although von Neumann took Jordan's formula for conditional probabilities as his starting point and rewrote it in terms of projection operators, his final formula is more general than Jordan's in that it pertains both to pure and to mixed states. However, it was not until the next installment of his 1927 trilogy, *Wahrscheinlichkeitstheoretischer Aufbau*, that von Neumann (1927b) carefully defined the difference between pure and mixed states. In this paper, von Neumann freed his approach from reliance on Jordan's even further (see Section 6). He now derived his formula for conditional probabilities in terms of the trace of products of projection operators from the Hilbert space formalism, using a few seemingly innocuous assumptions about expectation values of observables of systems in an ensemble of copies of those systems characterized by a density operator. He then showed that the density operator for a uniform ensemble is just the projection operator onto the corresponding pure dispersion-free state. Such pure states can be characterized completely by the eigenvalues of a complete set of commuting operators. This led von Neumann to a new way of formulating a typical problem in quantum mechanics. Rather than identifying $\hat{p}$'s and $\hat{q}$'s for the system under consideration, he realized that it suffices to specify the values of a maximal set of commuting operators for the system. All operators in such sets can be thought of as $\hat{q}$'s. There is no need to find the $\hat{p}$'s canonically conjugate to these $\hat{q}$'s.

Coda: Return of the p's and q's in Quantum Field Theory

In this paper, we emphasized the difficulties engendered by Jordan's insistence on the primacy of canonical $(\hat{p}, \hat{q})$ variables in expressing the dynamics of general quantum-mechanical systems (see especially Section 4 on *Neue Begründung* II). These difficulties became particularly acute in the case of systems with observables with purely or partially discrete spectra, of which the most extreme case is perhaps the treatment of electron spin (see Eqs. (119)–(121)). Here, the arbitrary choice of two out of the three spin components to serve as a non-commuting canonically conjugate pair clearly reveals the artificiality of this program. In a sense, however, Jordan was perfectly right to insist on the importance of a canonical approach, even for particles with spin: his error was simply to attempt to impose this structure at the level of non-relativistic quantum theory, where electron spin appears as an essentially mysterious internal attribute which must be grafted on to the nonrelativistic kinematics (which does have a perfectly sensible canonical interpretation). Once electron spin was shown to emerge naturally at the relativistic level, and all aspects of the electron's dynamics subsumed in the behavior of a local relativistic field, canonical 'p & q' thinking could be reinstated in a perfectly natural way. This was first done explicitly by Heisenberg and Pauli (1929) in their seminal paper on Lagrangian field theory. In modern notation, they introduce a relativistically invariant action for a spin-$\frac{1}{2}$ field, as a spacetime integral of a Lagrangian:

$$S = \int \mathcal{L}\, d^4x = \int \overline{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) d^4x \tag{188}$$

Here the field $\psi(x)$ is a four-component field, with the $\gamma^\mu, \mu = 0, 1, 2, 3$ the 4x4 Dirac matrices. A conjugate momentum *field* $\pi^\psi(x) \equiv \partial\mathcal{L}/\partial\dot{\psi}$ is then defined in the standard fashion, with canonical equal-time *anticommutation* relations imposed (as indicated by earlier work of Jordan and Wigner) in order to insert the desired fermionic statistics of the particles described by the field,

$$\{\pi^\psi_m(\vec{x}, t), \psi_n(\vec{y}, t)\} = i\hbar\delta_{mn}\delta^3(\vec{x} - \vec{y}) \tag{189}$$

The transition from a Lagrangian to a Hamiltonian (density) is then carried out in the usual way

$$\mathcal{H} = \pi^\psi\dot{\psi} - \mathcal{L} = \bar{\psi}(i\vec{\gamma} \cdot \vec{\nabla} + m)\psi(x) \tag{190}$$

The spatial integral of this Hamiltonian energy density would within a few years be shown to describe exactly the free Hamiltonian for arbitrary multi-particle states of non-interacting electrons and positrons, degenerate in mass and each displaying the usual panoply of spin-$\frac{1}{2}$ behavior which had finally been deciphered, in the non-relativistic context, by the atomic spectroscopy of the mid to late 1920s. The essential point is that the relevant 'p'-s and 'q'-s appearing in the theory are not associated with the first-quantized wave-functions or state vectors appropriate for a non-relativistic treatment, but rather with the *fields* that must replace them once a fully relativistic theory takes center stage.

Acknowledgments

This paper originated in talks we gave at New Directions in the Foundations of Physics, Washington, D.C., April 30–May 2, 1910, and at the Third International Conference on the History of Quantum Physics (HQ3) at the *Max-Planck-Institut für Wissenschaftsgeschichte* (MPIWG), Berlin, June 28–July 2, 2010. We are grateful to Gonzalo Gimeno, Christoph Lehner, Charles Midwinter, Serge Rudaz, Rob "Ryno" Rynasiewicz, Jos Uffink, and Mercedes Xipell, for helpful discussion and references. Michel Janssen gratefully acknowledges support of the MPIWG. The research of Anthony Duncan is supported in part by the National Science Foundation under grant PHY-0854782.

References

Bacciagaluppi, Guido, and Crull, Elise (2009). Heisenberg (and Schrödinger, and Pauli) on hidden variables. *Studies in History and Philosophy of Modern Physics* 40: 374–382.

Bell, John. S. (1966). On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics* 38: 447–452. Reprinted in (Bell, 1987, pp. 1–13). Page references to this reprint.

Bell, John. S. (1987). *Speakable and unspeakable in quantum mechanics.* Cambridge: Cambridge University Press.

Beller, Mara (1985). Pascual Jordan's influence on the discovery of Heisenberg's indeterminacy principle. *Archive for History of Exact Sciences* 33: 337–349.

Beller, Mara (1999). *Quantum dialogue. The making of a revolution.* Chicago: University of Chicago Press.

Born, Max (1926a). Zur Quantenmechanik der Stoßvorgänge. Vorläufige Mitteilung. *Zeitschrift für Physik* 37: 863–867.

Born, Max (1926b). Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik* 38: 803–827.

Born, Max, Heisenberg, Werner, and Jordan, Pascual (1926). Zur Quantenmechanik II. *Zeitschrift für Physik* 35: 557–615. English translation in Van der Waerden (1968, pp. 321–385).

Born, Max, and Jordan, Pascual (1925). Zur Quantenmechanik. *Zeitschrift für Physik* 34: 858–888. English translation of chs. 1–3 in Van der Waerden (1968, pp. 277–306).

Born, Max, and Wiener, Norbert (1926), Eine neue Formulierung der Quantengesetze für periodische und nicht periodische Vorgänge. *Zeitschrift für Physik* 36: 174–187.

Darrigol, Olivier (1992). *From c-numbers to q-numbers: the classical analogy in the history of quantum theory.* Berkeley: University of California Press.

Dennery, Philippe, and Krzywicki, André (1996). *Mathematics for physicists.* New York: Dover.

Dieudonné, Jean Alexandre (1981). *History of functional analysis. North-holland mathematics studies.* Vol. 49. Amsterdam: North-Holland.

Dirac, Paul Adrien Maurice (1925). The fundamental equations of quantum mechanics. *Proceedings of Royal Society of London. Series A* 109, 642–653. Reprinted in Van der Waerden (1968, pp. 307–320)

Dirac, Paul Adrien Maurice (1927). The physical interpretation of the quantum dynamics. *Proceedings of the Royal Society of London. Series A* 113: 621–641.

Dirac, Paul Adrien Maurice (1930). *Principles of Quantum Mechanics.* Oxford: Clarendon.

Duncan, Anthony, and Janssen, Michel (2007). On the verge of *Umdeutung* in Minnesota: Van Vleck and the correspondence principle. 2 Pts. *Archive for History of Exact Sciences* 61: 553–624, 625–671.

Duncan, Anthony, and Janssen, Michel (2008). Pascual Jordan's resolution of the conundrum of the wave-particle duality of light. *Studies in History and Philosophy of Modern Physics* 39: 634–666.

Duncan, Anthony, and Janssen, Michel (2009). From canonical transformations to transformation theory, 1926–1927: The road to Jordan's *Neue Begründung. Studies in History and Philosophy of Modern Physics* 40: 352–362.

Fischer, Ernst (1907). Sur la convergence en moyenne. *Comptes Rendus* 144: 1022–1024.

Goldstein, Herbert, Poole, Charles P., and Safko, John L. (2002). *Classical mechanics.* 3rd ed. San Francisco: Addison Wesley.

Heisenberg, Werner (1925). Über die quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik* 33: 879–893. English translation in Van der Waerden (1968, pp. 261–276).

Heisenberg, Werner (1927a). Schwankungserscheinungen in der Quantenmechanik. *Zeitschrift für Physik* 40: 501–506.

Heisenberg, Werner (1927b). Über den anschaulichen Inhalt der quantentheoretischen Kinematic und Mechanik. *Zeitschrift für Physik* 43: 172–198.

Heisenberg, Werner, and Pauli, Wolfgang (1929). Zur Quantendynamik der Wellenfelder. *Zeitschrift für Physik* 56: 1–61.

Hilbert, David, von Neumann, John, and Nordheim, Lothar (1928). Über die Grundlagen der Quantenmechanik. *Mathematische Annalen* 98: 1–30.

Hilgevoord, Jan (2002). Time in quantum mechanics. *American Journal of Physics* 70: 301–306.

Jammer, Max (1966). *The conceptual development of quantum mechanics.* New York: McGraw-Hill.

Jordan, Pascual (1926a). Über kanonische Transformationen in der Quantenmechanik. *Zeitschrift für Physik* 37: 383–386.

Jordan, Pascual (1926b). Über kanonische Transformationen in der Quantenmechanik. II. *Zeitschrift für Physik* 38: 513–517.

Jordan, Pascual (1927a). Über quantenmechanische Darstellung von Quantensprüngen. *Zeitschrift für Physik* 40: 661–666.

Jordan, Pascual (1927b). Über eine neue Begründung der Quantenmechanik. *Zeitschrift für Physik* 40: 809–838.

Jordan, Pascual (1927c). Über eine neue Begründung der Quantenmechanik. *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* 161–169

Jordan, Pascual (1927d). Kausalität und Statistik in der modernen Physik. *Die Naturwissenschaften* 15: 105–110.

Jordan, Pascual (1927e). Über eine neue Begründung der Quantenmechanik II. *Zeitschrift für Physik* 44: 1–25.

Jordan, Pascual (1927f). Die Entwicklung der neuen Quantenmechanik. 2 Pts. *Die Naturwissenschaften* 15: 614–623, 636–649.

Lacki, Jan (2000). The early axiomatizations of quantum mechanics: Jordan, von Neumann and the continuation of Hilbert's program. *Archive for His-*

tory of Exact Sciences 54: 279–318.

Lacki, Jan (2004). The puzzle of canonical transformations in early quantum mechanics. *Studies in History and Philosophy of Modern Physics* 35: 317–344.

Lanczos, Kornel (1926). Über eine Feldmäßige Darstellung der neueren Quantenmechanik. *Zeitschrift für Physik* 35: 812–830.

London, Fritz (1926a). Über die Jacobischen Transformationen der Quantenmechanik. *Zeitschrift für Physik* 37: 915–925.

London, Fritz (1926b). Winkelvariable und kanonische Transformationen in der Undulationsmechanik. *Zeitschrift für Physik* 40: 193–210.

Mehra, Jagdish, and Rechenberg, Helmut (2000). *The historical development of quantum theory.* Vol. 6. *The completion of quantum mechanics 1926–1941.* Part I. *The probability interpretation and the statistical transformation theory, the physical interpretation, and the empirical and mathematical foundations of quantum mechanics 1926–1932.* New York, Berlin: Springer.

Pauli, Wolfgang (1927a). Über Gasentartung und Paramagnetismus. *Zeitschrift für Physik* 41: 81–102.

Pauli, Wolfgang (1927b). Zur Quantenmechanik des Magnetischen Elektrons. *Zeitschrift für Physik* 43: 601–623

Prugovecki, Eduard (1981). *Quantum mechanics in Hilbert space.* 2nd ed. New York: Academic Press.

Riess, Frédréric (1907a). Über orthogonale Funktionensysteme. *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* 116–122.

Riess, Frédréric (1907b). Sur les systèmes orthogonaux de fonctions. *Comptes Rendus* 144: 615–619.

Sauer, Tilman, and Majer, Ulrich, eds. (2009). *David Hilbert's Lectures on the Foundations of Physics, 1915–1927.* Berlin: Springer.

Steen, Lynn Arthur (1973). Highlights in the history of spectral theory. *The American Mathematical Monthly* 80: 359–381.

Van der Waerden, Bartel Leendert, ed. (1968). *Sources of quantum mechanics.* New York: Dover.

Von Neumann, John (1927a). Mathematische Begründung der Quantenmechanik. *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* 1–57.

Von Neumann, John (1927b). Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik. *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* 245–272.

Von Neumann, John (1927c). Thermodynamik quantenmechanischer Gesamtheiten. *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* 273–291.

Von Neumann, John (1929). Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren. *Mathematische Annalen* 102: 49–131.

Von Neumann, John (1932). *Mathematische Grundlagen der Quantenmechanik.* Berlin: Springer, 1932.