5

Square root of minus one, complex phases and Erwin Schrödinger

CHEN NING YANG

State University of New York

5.1. Introduction

In a lecture in April 1970 Dirac talked about the early days of quantum mechanics (Dirac, 1972). Among other topics he discussed noncommutative algebra, and added

> The question arises whether the noncommutation is really the main new idea of quantum mechanics. Previously I always thought it was but recently I have begun to doubt it and to think that maybe from the physical point of view, the noncommutation is not the only important idea and there is perhaps some deeper idea, some deeper change in our ordinary concepts which is brought about by quantum mechanics.

He then expanded on this subject and concluded

> So if one asks what is the main feature of quantum mechanics, I feel inclined now to say that it is not noncommutative algebra. It is the existence of probability amplitudes which underlie all atomic processes. Now a probability amplitude is related to experiment but only partially. The square of its modulus is something that we can observe. That is the probability which the experimental people get. But besides that there is a phase, a number of modulus unity which can modify without affecting the square of the modulus. And this phase is all important because it is the source of all interference phenomena but its physical significance is obscure. So the real genius of Heisenberg and Schrödinger, you might say, was to discover the existence of probability amplitudes containing this phase quantity which is very well hidden in nature and it is because it was so well hidden that people hadn't thought of quantum mechanics much earlier.

One may or may not agree with Dirac on the question of which was more important: the introduction of an amplitude with a phase or that of noncommutative algebra, but there is no doubt that both are revolutionary developments of profound significance in the physicists' description of nature.

Classical physics, that is the physics before 1925, used exclusively real quantities. This was true for mechanics, thermodynamics, electrodynamics – the whole of classical physics. To be sure, complex numbers were used in many places. For example, in solving a linear alternating current problem complex numbers were used. But after a solution had been found, one always took the real or imaginary part of the solution in order to obtain the *true* physical answer. So the use of complex numbers was as a computational aid, i.e. the physics was conceptually in terms of real numbers.

With matrix mechanics and wave mechanics, however, the situation dramatically changed. Complex numbers became a conceptual element of the very foundation of physics: the fundamental equations of matrix mechanics and of wave mechanics:

$$pq - qp = -i\hbar \tag{1.1}$$

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi \tag{1.2}$$

both explicitly contain the imaginary unit $i = \sqrt{-1}$. It is to be emphasized that the very meaning of these equations would be totally destroyed if one tries to get rid of i by writing (1.1) and (1.2) in terms of real and imaginary parts.

5.2. Complex numbers in matrix and wave mechanics

The following is a brief history of the entry of complex numbers in matrix mechanics and in wave mechanics.

Take matrix mechanics first. In the pioneering paper of Heisenberg (1925) a comparison was made between the Fourier transform of a dynamical quantity (which depends on one state and one Fourier multiplicity) and its 'quantum theoretical' correspondence (which depends upon two states). In this process Heisenberg very naturally was conceptually discussing *complex* Fourier amplitudes. In the subsequent two-man paper (Born and Jordan, 1925), (1.1) explicitly appeared for the first time in history. That was also the first time that the imaginary i entered physics in a fundamental way. A little later, in Dirac's first paper on quantum mechanics (Dirac, 1925), (1.1) again appeared, together with

$$\dot{q} = [q, H] = (qH - Hq)/(i\hbar) \tag{2.1}$$

which also explicitly contains i. These developments implied that complex numbers play an essential role in matrix mechanics. But there seemed to be little appreciation at the time that this was a major new development in physics – perhaps because matrix mechanics was so new and Fourier analysis so natural that the full implications of the entry of complex numbers was obscured by the great revolution that was taking place.

Now we turn to wave mechanics which was created* in a historical series of six papers (Schrödinger, 1926*a–f*) all written within the first six months of 1926 by Erwin Schrödinger. In the first five of these Schrödinger had in mind the factorization of his wave function into a real stationary function of x and a sinusoidal function of time (Schrödinger, 1926*c*)†.

That Schrödinger did this was not surprising, since he was thinking of a standing wave description of the electron, very much in analogy with a standing electromagnetic wave or a water wave. Such waves do have phases, but nevertheless they are described by real functions of space-time. In Schrödinger (1926*e*) for example, there appeared a footnote to the equation

$$\psi_n = \mathrm{e}^{-x^2/2} H_n(x)\, \mathrm{e}^{2\pi \mathrm{i} \nu_n t} \tag{2.2}$$

which reads 'i means $\sqrt{-1}$. On the right-hand side the real part is to be taken, *as usual*.' [my italics], revealing his general attitude on this matter, which was the same as in the usual *linear* circuit theory: ψ may be complex, but one always takes the real part in the end.

Of course, in his search for the relationship between matrix mechanics and wave mechanics Schrödinger unavoidably encountered $\mathrm{i} = \sqrt{-1}$, as for example in equation (20) Schrödinger (1926*c*). Whether this fact perturbed him we shall probably never know. But he must have been disturbed when he involved himself with a discussion of quadratic forms such as $\psi(\partial\bar{\psi}/\partial t)$ (which he did briefly in Schrödinger (1926*c*)) or $\psi\bar{\psi}$ (which he did sometime before June 6, 1926, see below).

On May 27, 1926, H. A. Lorentz, then 73 years old, wrote a long letter to Schrödinger thanking the latter for having sent him the proof sheets of three articles and raising a lot of questions, some quite general, others very specific, about wave mechanics. Two of these questions are relevant to our present discussion: (*a*) how to interpret the ψ function for two or more particles; (*b*) Lorentz's opinion that 'the true "equations of motion" ... [should not] contain E at all, but contain time derivatives instead.' Schrödinger answered on June 6 in an equally long letter of eight points. The first two points addressed the two questions of Lorentz's mentioned above.

About (*a*), Schrödinger said that he had abandoned the expression $\psi(\partial\bar{\psi}/\partial t)$ of his earlier manuscript (Schrödinger, 1926*f*), and was now

* Pais (1986) quotes Weyl as saying 'Schrödinger did his great work during a late erotic outburst in his life'.

† See remark in parentheses following equation (35).

focussing on $\psi\bar{\psi}$ for the electric charge density in real space. He then continued: 'What is unpleasant here, and indeed directly to be objected to, is the use of complex numbers. ψ is surely fundamentally a real function.' There followed an involved suggestion of how to generate a complex ψ from its real part ψ_r, a suggestion clearly not quite satisfactory to Schrödinger himself.

About (*b*), Schrödinger wrote down

$$-\hbar^2\ddot{\psi}=E^2\psi \tag{2.3}$$

and then eliminated E by using $H\psi=E\psi$ to obtain

$$-\hbar^2\ddot{\psi}=H^2\psi. \tag{2.4}$$

He added 'This might well be the *general wave equation* which *no longer* contains the integration constant E, but contains time derivatives instead.' He continued to think about this matter, and five days later, on June 11, in writing to Planck he said 'By the way, during the last few days another heavy stone has been rolled away from my heart... I had considerable anxiety over it... But it all resolved itself with unheard of simplicity and unheard of beauty.' What was this resolution? It was (2.4) above.

Why did Schrödinger not write down simply the correct time dependent equation (1.2) rather than the more complicated (2.4)? He certainly knew the simpler equation but chose to go to the more complicated second order equation*. Why? I suggest the answer is as follows.

Schrödinger did not want his wave equation to contain i, so in a way he eliminated it by going to the fourth order equation (2.4), utilizing $i^2=-1$. That he tried to avoid i was quite natural since he had started on wave mechanics in Schrödinger (1926*a*) by writing down the real Hamilton–Jacobi equation

$$H\left(q,\frac{\partial S}{\partial q}\right)=E,$$

together with

$$S=K\log\psi.$$

His ψ up to this point was real and time independent. Later in § 3 of Schrödinger (1926*a*) he wrote 'It is, of course, strongly suggested that we should try to connect the function ψ with some *vibration process* in the atom...'. Alas this was not a simple process, because Schrödinger had to

* In all five of Schrödinger's papers written before June 11, 1926 (Schrödinger, 1926*a*–*e*), (1.2) never appeared. Yet one finds such equations as (2.2), which implies that he knew $i\hbar\dot{\psi}_n=H\psi_n$. The confusing discussion in his June 6 letter to Lorentz about the real part ψ_r of ψ is very revealing in that it shows how Schrödinger was struggling to eliminate/define the imaginary part.

struggle with the question of what frequency to use for this vibration. The subsequent evolution of his thinking on this question is a very interesting topic, but is not the subject we are considering here. What is relevant to us now is the fact that Schrödinger had started his conceptualization of wave mechanics by envisaging a description of a vibration in terms of a real function of space-time. Later when he superposed ψ's he again meant to add real ψ's, each of which depends on time sinusoidally.

To return to Schrödinger's letter of June 11 to Planck, he further emphasized that in (2.4) one may 'let the potential energy be an explicit function of the time.' This turned out to be incorrect and Schrödinger realized this in the next ten days, during which he wrote Schrödinger (1926*f*) which was received by the publisher on June 23. It was in this paper that the concept was first stated that ψ is a complex function of space-time and satisfies the complex time evolution equation (1.2) which Schrödinger called the *true** wave equation, in contrast to $H\psi = E\psi$ which he called the vibration or amplitude equation.

I should emphasize that I do not infer from the chronology outlined above that Schrödinger's discovery in Schrödinger (1926*f*), that ψ should be complex, was started by Lorentz's letter of May 27, 1926, to him. That may be the case, but it also could be that after writing Schrödinger (1926*d*), which was on time independent perturbation theory, Schrödinger began to work on a perturbation theory where the perturbation is time dependent. He would then have to study the time evolution of the wave function ψ, and Lorentz's letter may have arrived in the midst of such a study. What one can be certain is that between June 11 and June 23 Schrödinger was finally convinced that ψ is complex.

A few days after Schrödinger (1926*f*) was submitted Born submitted the first of his two historic papers on the statistical interpretation of the wave function (Born, 1926*a*, *b*). It is interesting to notice that in the first of these two Born used a real wave function,

$$\sin \frac{2\pi}{\lambda} z$$

for the incoming wave and another real wave function

$$\sin k_{nm\tau}(\alpha x + \beta y + \gamma z + \delta)$$

for the scattered wave. Because everything was real, Born did not use 'absolute square' but only 'square' in the famous footnote (added to the first paper in proof) which Pais referred to as follows (Pais, 1986): 'that

* Schrödinger used *eigentlich*, which I translate as *true*. Elsewhere it has been translated as *real*, which is very confusing in the present context.

great novelty, the correct transition probability concept, entered physics by way of a footnote'. It was only in the second paper that Born used complex numbers for the incoming and outgoing waves.

5.3. Complex numbers in Weyl's gauge theory

We have traced above the entry of complex numbers into fundamental physics during the period 1925–6. In fact, several years before that, Schrödinger (1922) had published a most interesting paper entitled 'On a remarkable property of the quantum orbit of one electron', in which he had already mentioned the possibility of introducing an imaginary factor

$$\gamma = -i\hbar \tag{3.1}$$

into Weyl's 1918 gauge theory. He started with Weyl's 'world geometry', i.e. Weyl's 1918 gauge theory of electromagnetism, summarizing Weyl's idea in a *Streckenfaktor*:

$$\exp\left[-\frac{e}{\gamma}\int(V\,dt - \mathbf{A}\cdot d\mathbf{x})\right], \tag{3.2}$$

and went on to remark that for a hydrogen atom where $A = 0$, the expression in the exponential is equal to

$$-\gamma^{-1}e\bar{V}\tau$$

where τ is the period. For a Bohr orbit with quantum number n this is equal to $-\gamma^{-1}nh$, an integral multiple of $\gamma^{-1}h$. Schrödinger called this result remarkable and said he could not believe that it was without deep physical significance.

At the end of the paper Schrödinger mentioned two possible values for γ, $\gamma = e^2/c$, a real number, or $\gamma = -i\hbar$, i.e. (3.1) above. For the latter case, he remarked, the factor (3.2) becomes unity.

In his great papers of 1926 which created wave mechanics Schrödinger did not refer to this 1922 paper. But Raman and Forman (1969) in their historical research argued that this 1922 paper had in fact played an important role in 'Why was it Schrödinger who developed de Broglie's ideas?' Their thesis was later confirmed by Hanle (1977, 1979; see also Wessels, 1977), who found the following passage in a letter dated November 3, 1925 from Schrödinger to Einstein:

> The de Broglie interpretation of the quantum rules seems to me to be related in some ways to my note in the *Zs. f. Phys.* 12, 13, 1922, where a remarkable property of the Weyl 'gauge factor' $\exp[-\int \phi\, dx]$ along each quasi-period is shown. The mathematical situation is, as far as I can see, the same, only from me much more formal, less elegant and not really shown generally. Naturally de Broglie's consideration in the framework of his large theory is altogether of

far greater value than my single statement, which I did not know what to make of at first.

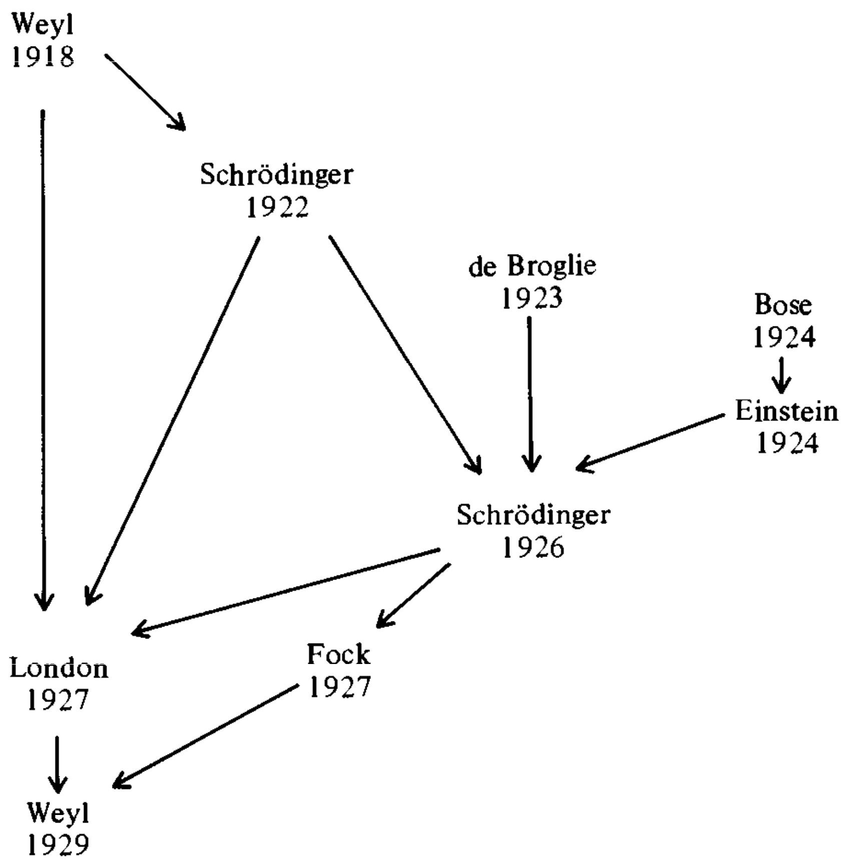
Thirteen days later, on November 16, 1925, Schrödinger wrote to Lande (Raman and Forman, 1969):

> Recently I have been deeply involved with Louis de Broglie's ingenious thesis. It's extraordinarily stimulating but nonetheless some of it is very hard to swallow. I have vainly attempted to make myself a picture of the phase wave of an electron in an elliptical orbit. The 'rays' are almost certainly neighboring Kepler ellipses of equal energy. That, however, gives horrible 'caustics' or the like as the wave front. At the same time, the length of the wave ought to be equal to [that of the orbit traced out by the electron in] one Zeeman or Stark cycle!

Schrödinger was by then evidently well on his way to the first great paper on wave mechanics which he submitted on January 27, 1926 (Schrödinger, 1926*a*)!

The Raman–Forman–Hanle thesis that Schrödinger's 1922 paper played an essential role in the creation of wave mechanics is clearly correct. We illustrate this fact by an arrow in Fig. 5.1.

Fig. 5.1. Flow of ideas relating to complex phases and gauge fields. The importance of the 1922 paper of Schrödinger was discovered by Raman and Forman (1969) and Hanle (1977, 1979).



Why then did Schrödinger not refer to his own 1922 paper in 1926? The reason is probably as follows: the 1922 paper discussed a factor (3.2) above, with which (3.1) becomes

$$\exp\left[ei\int A_\mu\,\mathrm{d}x^\mu/\hbar\right] \tag{3.3}$$

while the 1926 papers were related to de Broglie's idea which, for comparison, could be put in the form of a factor

$$\exp\left[\mathrm{i}\int p\cdot\mathrm{d}x/\hbar\right]. \tag{3.4}$$

The two are similar but not the same, and Schrödinger recognized that the relevant one to start wave mechanics from was (3.4) and not (3.3).

While Schrödinger was busy with developing wave mechanics, his 1922 paper caught the attentions of F. London who wrote to Schrödinger a very interesting letter reprinted in Raman and Forman (1969). We attach as an appendix a translation of this letter, which according to Raman and Forman, was written around December 10, 1926.

London developed further this thinking in a paper entitled 'Quantum mechanical meaning of the theory of Weyl' (London, 1927*a*; see also London, 1927*b*). A little earlier, Fock had published a paper which discussed invariance of wave equations (Fock, 1927). Both are somewhat confusing* as is natural in those early days of wave mechanics, but both contain the right idea that, in today's notation, electromagnetism enters in wave mechanics with an operator

$$(\partial_\mu - \mathrm{i}eA_\mu)$$

on ψ, which is the heart of the gauge principle (see Fig. 1). The definitive discussion of electromagnetism as a gauge theory came later, in an important paper (Weyl, 1929; see also Yang, 1986).

5.4. Modern consequences

The importance of the introduction of complex amplitudes with phases into physicists' description of nature was not fully appreciated until the 1970s when two developments took place: (1) all interactions were found to be some form of gauge field; and (2) gauge fields were found to be related to the mathematical concept of fibre bundles (Wu and Yang, 1975), each fibre being a complex phase or a more general phase. With these developments there

* In my article in *Ann. N.Y. Sci.* 294, 86 (1977) I had said that London's paper pointed out the similarity between Fock's work and the 1918 Weyl paper. This is wrong. I had misread the meaning of the footnote on p. 111 of W. Pauli, *Handbuch der Physik* Vol. 24, Part 1 (1933).

arose a basic tenet, of today's physics: *all fundamental forces are phase fields* (Yang, 1983). Thus the almost casual introduction in 1922 by Schrödinger of the imaginary unit i into (3.1) above has flowered into deep concepts that lie at the very foundation of our understanding of the physical world.

In 1975 Wu and I drew up a 'dictionary', reproduced here as Table 1, identifying physicists' terminology for gauge fields with mathematicians' terminology for fibre bundles (Wu and Yang, 1975).

There is in this 'dictionary' a blank space with a question mark because at that time the mathematicians had not studied the concept that corresponds to physicists' 'sources', i.e. density–current four-vector, a natural and fundamental concept in Maxwell's theory of electromagnetism. In the language of the mathematicians, this concept would have been written as

$$*\partial * f = J. \tag{4.1}$$

A sourceless case would satisfy

$$*\partial * f = 0. \tag{4.2}$$

The mathematicians have now studied (4.2) and the results have helped to resolve some deep and long standing problems in topology and differential geometry, providing a modern example, so abundant in past centuries but

Table 5.1. *Reproduction of 'dictionary' comparing terminologies in gauge field theory and fibre bundle theory*

Gauge field terminology	Bundle terminology
gauge (or global gauge)	principal coordinate bundle
gauge type	principal fibre bundle
gauge potential b_μ^k	connection on a principal fibre bundle
S_{ba}	transition function
phase factor Φ_{QP}	parallel displacement
field strength $f_{\mu\nu}^k$	curvature
source[a] J_μ^K	?
electromagnetism	connection on a $U_1(1)$ bundle
isotopic spin gauge field	connection on a SU_2 bundle
Dirac's monopole quantization	classification of $U_1(1)$ bundle according to first Chern class
electromagnetism without monopole	connection on a trivial $U_1(1)$ bundle
electromagnetism with monopole	connection on a nontrivial $U_1(1)$ bundle

[a] i.e. electric source; this is the generalization of the concept of electric charges and currents.

rare now, of how physics could supply powerful insights for the advance of mathematics (Freed and Uhlenbeck, 1984; Lawson, 1985).

5.5. Appendix

A letter from F. London to E. Schrödinger*
Translated into English by Prof. T. C. Meng.

Dear Professor,

I must have a serious word with you today. Are you acquainted with a certain Mr. Schrödinger, who in the year 1922 (*Zeits. fur Phys.*, 12) described a 'bemerkenswerte Eigenschaft der Quantenbahnen'? Are you acquainted with this man? What! You affirm that you know him very well, that you were even present when he did this work and that you were his accomplice in it? That is absolutely unheard of. So you already know for four years that in the continuous space-time in which atomic processes have to be studied, no rulers and clocks can be used to define an Einstien–Riemann metrical relationship (*Masszausammenhang*), once one has to see whether the general metrical principles which have been expressed by Weyl's theory of distance transference (*Streckenübertragung*) is perhaps helpful. And you have for four years very well noticed that they are even extremely helpful. Namely while usually nonsense emerges by using Weyl's distance transference [Einstein's objection (Yang, 1986), Weyl's very poor excuse (Weyl, 1968) with 'adjustment' (*Einstellung*)] you have shown that on the discrete physical orbits the scale unit (*Eicheinheit*) (with $\gamma = 2\pi i/h$) can be reproduced for spatially closed paths; and in fact you observed at that time for the nth orbit the scale unit swells and shrinks (*anschwillt und zusammenschrumpft*) exactly n times, just like the standing wave which describes the location of the charge. So you have shown that the theory of Weyl is only reasonable – i.e., it leads to a unique measure-determination – when one combines this theory with the quantum theory. Actually there is nothing else one can do, if the entire atomic world is a continuous space-time without any fixed point for identification. You knew this and said nothing and made no statement about it. This kind of thing has never happened before. You wrote very modestly in your paper (p. 14): You did not – to confess immediately – come very far in the discussion of the possible meaning of this fact. But in this paper you not only have put the hopeless confusion of the Weyl theory to an end, but also even had the resonance character of the quantum postulate in your hands long before de Broglie, and also thought about whether you should take $\gamma = h/2\pi i$ or e^2/c! (p. 23) – Will you now immediately confess that, like a priest, you kept secret the truth which you held in your hands, and give notice to your contemporaries of all you know! The

* According to Raman and Forman, who reprinted this letter in their article in *Historical Studies in the Physical Sciences*, Vol. 1, pp. 291–314, the letter was written around December 10, 1926.

most important thing is yet to be done, that remark made in 1922 being a theorem of the old quantum mechanics. One can with certainty expect that it will show its whole significance when it is brought into meaningful connection with wave mechanics. (I have not done this yet.) I think that it is your duty, after you have mystified the world in such a manner, now to clarify everything.

Now, that is enough. Thank you very much for spending so much time on my stupid letter*. For the moment I have discontinued my study on this matter. I think on the whole the Kaluza–Klein Space Theory has to be considered as a set-back (*Rückschritt*) since the existence of the beautiful Weyl Space Theory, and I would like to look at this more closely. I have different clues (*Anhalt*) which show that it will not be difficult to make Weyl's and Kaluza's theory consistent with each other (plot for yourself for every world-point the scale unit (*Eicheinheit*) as the 5th dimension, one immediately sees a lot of beautiful things!) I am eagerly looking forward to reading your manuscript† (until now it is still not here) especially after the hints given by Fues. Even if it would be just for one day, I would very, very much like to be able to see it.

By the way, the Rockefeller is granted; the telegram arrived yesterday. I am very happy that it is now certain that I may work with you.

I wish you a good journey. I am looking forward to your return.

With hearty greetings, I am

Yours very faithfully

Fritz London

References

Born, M. (1926*a*) *Z. f. Phys.* 37, 863 (received June 25)
Born, M. (1926*b*) *Z. f. Phys.* 38, 803 (received July 21)
Born, M. and Jordan, P. (1925) *Z. f. Phys.* 34, 858 (received September 27)
Dirac, P. A. M. (1925) *Proc. Roy. Soc.* A109, 642 (received November 7)
Dirac, P. A. M. (1972) *Fields & Quanta* 3, 139
Freed, D. and Uhlenbeck, K. (1984) *Instantons and Four-manifolds*. Springer
Fock, V. (1927) *Z. f. Phys.* 39, 226
Hanle, P. (1977) *Isis* 68, 606
Hanle, P. (1979) *Am. J. Phys.* 47, 644
Heisenberg, W. (1925) *Z. f. Phys.* 33, 879 (received July 29)
Lawson, Jr, H. B. (1985) *The Theory of Gauge Fields in Four Dimensions*. American Mathematics Society 58, Providence, RI
London, F. (1927*a*) *Z. f. Phys.* 42, 375
London, F. (1927*b*) *Naturwiss.* 15, 187
Pais, A. (1986) *Inward Bound*, p. 257. Oxford University Press

* Raman and Forman comment that this refers to a letter of December 1, 1926, to which Schrödinger replied on December 7, 1926.

† Raman and Forman comment that this refers to a manuscript on relativistic wave equations mentioned in Schrödinger's letter of December 7, 1926.

Raman, V. V. and Forman, P. (1969) *Hist. Studies Phys. Sci.* 1, 291
Schrödinger, E. (1922) *Z. f. Phys.* 12, 13
Schrödinger, E. (1926*a*) *Ann. d. Phys.* 79, 361 (received January 27)
Schrödinger, E. (1926*b*) *Ann. d. Phys.* 79, 489 (received February 23)
Schrödinger, E. (1926*c*) *Ann. d. Phys.* 79, 734 (received March 18)
Schrödinger, E. (1926*d*) *Ann. d. Phys.* 80, 437 (received May 10)
Schrödinger, E. (1926*e*) *Die Naturw.* 28, 664
Schrödinger, E. (1926*f*) *Ann. d. Phys.* 81, 109 (received June 23)
Wessels, L. (1977) *Studies Hist. & Phil. Sci.* 10, 311
Weyl, H. (1929) *Z. f. Phys.* 56, 330
Weyl, H. (1968) in (Chandrasekharan, K., ed.) *Gesammelte Abhandlungen*, vol. II, p. 261. Springer
Wu, T. T. and Yang, C. N. (1975) *Phys. Rev. D* 12, 3845
Yang, C. N. (1983) *Selected Papers 1945–1980 with Commentary*, p. 564. Freeman
Yang, C. N. (1987) (Chandrasekharan, K., ed.). Springer (in press)