Evaluations of Dirac's electron, 1928–1932

Donald Franklin Moyer

Citation: American Journal of Physics **49**, 1055 (1981); doi: 10.1119/1.12643
View online: http://dx.doi.org/10.1119/1.12643
View Table of Contents: http://scitation.aip.org/content/aapt/journal/ajp/49/11?ver=pdfcov
Published by the American Association of Physics Teachers

Evaluations of Dirac's electron, 1928–1932

Donald Franklin Moyer
2025 Sherman Ave, Evanston, Illinois 60201
(Received 18 August 1980; accepted 14 April 1981)

This the second of a three-part series. This essay surveys evaluations of Dirac's theory by other physicists, especially by Bohr who used Dirac's speculations about negative energy electrons as evidence for the failure of quantum mechanics at nuclear dimensions.

I. BOUND STATES

In April of 1928 Charles Galton Darwin and Walter Gordon both published detailed studies of the bound-state solutions of Dirac's equation.[1] Since Darwin himself had been trying to construct a relativistic quantum theory of the electron his surprise at the way Dirac succeeded is especially interesting. Darwin had been seeking to represent the electron as a vector wave "by empirically constructing a pair of equations to represent the fine structure of the hydrogen spectrum." To do this in Hamiltonian form "suitable for the relativity transformation" he used the Klein–Gordon operator. When Darwin saw how Dirac "has brilliantly removed the defects before existing in the mechanics of the electron, and has shown how the phenomena usually called the 'spinning electron' fit into place in the complete theory" he was surprised. He was surprised because "the whole theory of general relativity is based on the idea of invariance of form, and here is a system invariant in fact but not in form." Darwin noted that "Dirac's guiding principle is that the 'Hamiltonian equation' must be linear" and was surprised that this principle rather than the empirical correspondence worked, showing "the great superiority of principle over the previous empirical method."

Darwin's chief interest in this paper was to study the bound-state solutions for a central field in detail to show that Dirac's equation gives the observed energy levels, etc. Darwin's work, which remains the standard presentation, showed that Dirac's equation passed this test with only one difficulty. The difficulty was that to get a complete set of solutions it is necessary to admit "negative values of the energy, and we have at present little idea of what this means."

This difficulty did not prevent successful use of Dirac's equation for calculation of bound states of many-electron atoms using Hartree's self-consistent field approximation and even for calculation of hyperfine structure.[2] The negative energy difficulty became unavoidably disasterous for bound states of electrons within the nucleus—which were believed to be necessary to account for nuclear systematics—when Klein showed that the steep potential required to bind an electron within the nucleus would necessarily produce transformations to the negative energy states entailed by Dirac's equation.[3]

II. SCATTERING

Problems of physical interpretation such as those that led Dirac to reject the Klein–Gordon electron theory and seek a linear equation are not so difficult in the case of periodic, bound-state phenomena. Aperiodic, scattering phenomena provide the testing ground for differences in physical interpretation. Bohr recognized this explicitly in his Como address, as did Darwin and others elsewhere.[4] Dirac's own worries about physical interpretation became acute with his attempt to use his early formulation of quantum mechanics to calculate intensities and polarizations of Compton scattering in 1926.

In September of 1928 Klein and Nishina reported their use of Dirac's "more rational relativistic dynamics" to calculate the intensity and polarization of Compton scattering. Their results agreed with experiments "rather better" than the previous calculations of Dirac and Gordon, which used the Klein–Gordon Hamiltonian.[5] Reanalysis of existing data by L. H. Gray indicated that the Klein–Nishina result "agrees best with the somewhat meagre data which are available concerning the absorption of x-rays of known frequency."[6] Rutherford reported this in his presidential address to the Royal Society in November of 1928 and again in his Friday evening discourse at the Royal Institution in March of 1929.[7] He was especially interested in this result because the Klein–Nishina formula could be used as a tool for analyzing the energies of cosmic rays (as indeed Klein and Nishina had suggested in their first report). Analysis of the energies of cosmic rays was important to test the hypotheses of Jeans that cosmic rays came from the annihilation of electrons and protons and of Millikan that cosmic rays come from the creation of helium nuclei from electrons and protons. The Klein–Nishina formula did very quickly become a trustworthy tool reflecting favorably on Dirac's theory from which it was obtained.

Scattering of electrons would be another application of Dirac's theory of the electron but also involved basic issues not directly related to the Dirac equation. The general formulation of collision problems was not yet clear and there were problems normalizing aperiodic wave functions. Scattering of electrons by electrons, for example, depended on the specific interraction that was not clear. Work by Oppenheimer and Mott[8] brought some clarity to the general problem and yielded an especially interesting result not specifically related to Dirac's equation but generally related to his physical interpretation of quantum mechanics. Oppenheimer noted that for electron scattering it would be necessary to use antisymmetric wave functions consistent with Pauli's exclusion principle. Mott did this for electron–electron scattering and, as an aside, for alpha–alpha scattering where symmetric wave functions are required. The second case produced what has become known as "Mott scattering." Thus the purely quantum-mechanical property of spin, manifest theoretically in the symmetry or antisymmetry of the wave functions, yields a "scattering law differing considerably from the classical." Mott had shown that quantum-mechanical calculation of Rutherford scattering for nuclei would give the classical result obtained by Rutherford. Now he found theoretically a purely quantum-mechanical variation from the classical result. Thus several tests of Mott scattering were generally regarded to be tests of the quantum theory, and these successful confir-

 0002-9505/81/111055-08$00.50

mations raised confidence in the quantum theory of scattering.[9] The more troublesome specific electron–electron interaction was clarified by Møller and confirmed by Champion.[10]

Because so many other theoretical and experimental issues were involved, Compton scattering and electron scattering were not sensitive tests of Dirac's theory. Nonetheless Dirac's theory of the electron did pass the test of trustworthy use for scattering as it had for bound states. The problem of negative energies was lurking here also, however, since scattering interactions must necessarily induce transitions to the negative energies. If Dirac's theory is used to calculate scattering intensities what is to be made of the negative energy states necessarily involved?

III. SPIN

Without building in the electron magnetic moment in any way and without even any adjustment of parameters Dirac's equation automatically yielded exactly the magnetic moment of the electron needed to account for observations of atomic spectra. Inasmuch as the spin is a necessary consequence of Dirac's general transformation theory and the principle of relativity Dirac's theory answers his olympian question and explains why nature chose this particular form for the electron. This was the most impressive success of Dirac's theory but again there were some puzzles.

Gregory Breit noticed that Dirac's α matrices "may be thought of as operational matrix representation of the velocity vector $\dot{x}/c$, $\dot{y}/c$, $\dot{z}/c$" with eigenvalues ± 1: "The only possible values of $\dot{x}$, $\dot{y}$, $\dot{z}$, are therefore $\pm c$. At first sight this looks absurd because it implies that the only possible measured velocity of an electron is the velocity of light." This puzzle shows the often troublesome relationship between theoretical representations and experimental observations in quantum mechanics. Darwin followed with a study of the electrical density and current calculated from Dirac's theory that actually manifests the motions and magnetic moment of electrons. In a later study of the Dirac electric density current Schrödinger discovered the Zitterbewungen bringing back the puzzle noticed by Breit.[11]

The beta decay of nuclei meant, all agreed, that nuclei were constructed of protons and electrons. This presented several problems: the problem of confining the electron within the nucleus, the statistics problem, and the problem of a continuous distribution of energies of electrons emerging from nuclei themselves changing discontinuously between discreet energy states. Bohr suggested that spin may not be a real property of the electron and that the magnetic moment of the electron is not observable at all. Mott answered that in Dirac's theory "formally at least, ... the electron has a magnetic moment." "But, [he continued] when the electron is in an atom we can not observe this magnetic moment directly .... The question arises has the *free* electron 'really' got a magnetic moment, a magnetic moment that we can by any conceivable experiment observe?" Mott published Bohr's agrument that because of the uncertainty principle it was impossible to observe the magnetic moment of a free electron. However, Moss was "unwilling to give up altogether the idea of the direction of the spin axis of the free electron" so in the body of the paper he calculated the polarization of an electron beam scattered by a potential. He determined that an unpolarized beam would be partly polarized upon scattering by nuclei and that this polarization could be observed by detecting the asymmetry of the scattering from a second target. The effect is small—about 10% in excellent conditions, Mott suggested—and very difficult to observe because so many other effects—including nonconservation of parity—intervene.[12]

By the Volta Conference in Rome in October of 1932 even Mott believed "that the spin of the electron is still not properly understood," "that no meaning can be attached to the statement that a free electron has a magnetic moment ... . In fact it is best nowadays to abandon the assumption that the electron has itself a magnetic moment," and that Dirac's theory "is only an approximation." At the Volta Conference, and elsewhere, Bohr argued that the failure of Dirac's theory—failure because it entailed a property not interpretable by the correspondence principle and because it predicted absurd negative energies—indicated the breakdown of quantum mechanics at nuclear dimensions. Ironically, measurements of spin—gyromagnetic ratio, actually—have provided "the cleanest and most precise verification of the theoretical structure and calculational procedure of quantum electrodynamics yet devised."[13]

IV. MAGIC AND SICKNESS

Two metaphors—magic and sickness—recur in recollections of early reactions to Dirac's theory. Oppenheimer, for example, used both in close juxtaposition.[14] The theory magically yielded properties of the electron from general formal considerations. But the therory was also afflicted with the sickness of predicting properties not observed.

The magic came with Dirac's relativistic generalization. When Darwin noted[15] with surprise that Dirac's equation was "invariant in fact but not in form" he did recognize that it is "possible to give it formal invariance as well." But, he continued, this would require 16 equations each with a real and imaginary part and added, "It seems quite preposterous to think that a single electron should require 32 equations to express its behavior." Others, especially Eddington, found new magic just here. Eddington answered Darwin and sought to show[16] how a symmetrical treatment of Dirac's equation "appears advantageous in deducing general properties." He argued that the inverse fine-structure constant, hc/e^2, should be 136 and is just the number of degrees of freedom assigned to the electron by the elements of the space of the symmetric 16×16 matrix required to make Dirac's theory invariant in form. Later Eddington obtained "fuller insight into the more obscure parts of the theory" and was able "to bring the theory into an improved form" in which another degree of freedom was found to increase the inverse fine-structure constant to 137 closer to Millikan's experimental value of 137.1. Even though some poked fun at Eddington arguing, for example, that the inverse fine-structure constant was actually related to the absolute zero of temperature by $-273\,°\mathrm{C} = -(2/137 - 1)$, Eddington's theory did attract attention and "much prominence has been given to the paper in the public press." Public excitement was fed by news that Einstein was "about to publish the results of a protracted investigation into the possibility of generalizing the theory of relativity so as to include the phenomena of electromagnetism." Shortly after Einstein's new theory appeared Norbert Wiener and M. S. Vallarta reported that in discussion with Dirik J. Struik they found that "Einstein's 4-Bein's" allowed a "harmonizing with the quantum theory of the spinning electron ... [which] enables us to carry over the Dirac theory into general relativity almost without alteration."[17] Excitement about this bit of magic soon abated and the relationships between relativity and Dirac's theory were clarified by the spinor analysis developed by Bruce Van der Waerden and used by Otto Laport and George Uhlenbeck. Work on the purely formal aspects of Dirac's theory continued; for example, searches for equations invariant with respect to coordinates, spin, and gauge transformations yielded Dirac's equation, among several others with no apparent physical meaning Dirac studied versions of his equation in five-dimensional deSitter space and six-dimensional conformal space but saw no promise in this direction. Work on more physical unified field theories such as the Born–Infeld theory also continued, also without much pay off. Here there were basic unsolved problems concerning the interactions of matter and radiation. As usual Darwin put the problem drolly: "It seems not out of place to fit the electromagnetic equations into the general scheme; if they are wrong, it is still interesting to know why Maxwell made the mistake of inventing them."[18]

By mid-1929 the mainstream of research was concerned with the nucleus. The theory of quantum mechanics applied to the atom was generally completed. Dirac's remark—"the general theory of quantum mechanics is now almost complete"—was often cited. There was new vigor in Ernest Rutherford's group studying the nucleus at the Cavendish laboratory. Enrico Fermi's group in Rome was preparing for an experimental assault on the nucleus.[19] Bohr was raising philosophical questions about the use of quantum mechanics in the nucleus. He argued that theoretically Dirac's equation showed that quantum mechanics could not account for the electrons in the nucleus since the potential required to confine an electron would necessarily, as Klein showed, induce transitions to the negative energy states entailed by Dirac's equation. Furthermore, he argued that the experimental evidence of the continuous beta-decay energy spectrum suggested that energy was not conserved in the nucleus.

Bohr wrote to Dirac about these difficulties in 1929. Dirac's response deserves to be quoted in full:[20]

St Johns College, Cambridge
26-11-29

Dear Professor Bohr,

Many thanks for your letter. The question of the origin of the continuous β-ray spectrum is a very interesting one and may prove to be a serious difficulty in the theory of the atom. I had previously heard Gamow give an account of your views at Kapitza's club. My own opinion of this question is that I should prefer to keep rigorous conservation of energy at all costs and would rather abandon even the concept of matter consisting of separate atoms and electrons than the conservation of energy.

There is a simple way of avoiding the difficulty of electrons having negative kinetic energy. Let us suppose the wave equation $[w/c + e/cA_0 + \rho_1(\sigma\cdot\gamma + e/cA + \rho_0 me]\psi' = 0$ does accurately describe the motion of a single electron. This means that if the electron is started off with a + ve energy, there will be a finite probability of its suddenly changing into a state of negative energy and emitting the surplus energy in the form of high-frequency radiation. It cannot then very well change back into a state of + ve energy, since to do so it would have to absorb high-frequency radiation and there is not very much of this radiation actually existing in nature. It would still be possible, however, for the electron to increase its velocity (provided it can get the momentum from somewhere) as by so doing its energy would be still further reduced and it would emit more radiation. Thus the most stable states for electron are those of negative energy with very high velocity.

Let us now suppose there are so many electrons in the world that all these most stable states are occupied. The Pauli principle will then compel some electrons to remain in less stable states. For example if all the states of − ve energy are occupied and also a few of + ve energy, those electrons with + ve energy will be unable to make transitions to states of − ve energy and will therefore have to behave quite properly. The distribution of − ve electrons will, of course, be of infinite density, but it will be quite uniform so that it will not produce any electromagnetic field and one would not expect to be able to observe it.

It seems reasonable to assume that not all the states of negative energy are occupied, but that there are a few vacancies or "holes." Such a hole which can be described by a wave function, like an x-ray orbit would appear experimentally as a thing with + ve energy, since to make a hole disappear (i.e., to fill it up), one would have to put − ve energy into it. Further, one can easily see that such a hole would move in an elecromagnetic field as though it had a + ve charge. These holes I believe to be the protons. When an electron of + ve energy drops into a hole and fills it up, we have an electron and proton disappearing simultaneously and emitting their energy in the form of radiation.

I think one can understand in this way why all the things one actually observes in nature have a positive energy. One might also hope to be able to account for the dissymmetry between electrons and protons. So long as one neglects interaction one has complete symmetry between electrons and protons; one could regard the protons as the real particles and the electrons as the holes in the distribution of protons of − ve energy. However, when the interaction between the electrons is taken into account this symmetry is spoilt. I have have not yet worked out mathematically the consequences of the interaction. It is the "Austausch" effect that is important and I have not yet been able to get a relativisitic formulation of this. One can hope, however, that a proper theory of this will enable one to calculate the ratio of the masses of proton and electron.

I was very glad to hear that you will visit Cambridge in the spring and I am looking forward to your visit. With kind regards from

Yours sincerely,
P. A. M. Dirac

Bohr answered within a few days. Again his arguments, which he, and others, elaborated in the next three years, deserve full quotation[21]:

December 5th, 1929

Dear Dirac,

Many thanks for your most interesting letter, which has givenKlein and me cause to much thinking and discussion. Your idea is indeed very fascinating, but I must confess that we do not see how it works out in detail. Before all we do not understand, how you avoid the effect of the infinite electric density in space. According to the principles of electrostatics it would seen that even a finite uniform electrification should give rise to a considerable, if not infinite, field of force. In the difficulties of your old theory I still feel inclined to see a limit of the fundamental concepts on which atomic theory hitherto rests rather than a problem of interpreting the experimental evidence in a proper way by means of these concepts. Indeed according to my view the fatal transition from positive to negative energy should not be regarded as an indication of what may happen under certain conditions but rather as a limitation in the applicability of the energy concept.

In the case of electrons impinging on a potential barrier examined by Klein we have, on the one hand, a striking example of the difficulties involved in an unlimited use of the concept of potentials in relativistic quantum mechanics. On the other hand, we have just in this case an example of the actual limit of applying the idea of potentials in connection with possible experimental arrangements. In fact, due to the existence of an elementary unit of electrical charge we cannot build up a potential barrier of any height and steepness desired without facing a definite atomic problem. In Klein's example the critical height of the barrier is of order mc^2, and the rise of potential shall take place within a distance of the order h/mc which is the order of magnitude of the wavelength of the electrons concerned. But if the dimensions of the barrier perpendicular to the electric force shall be large compared with this wavelength λ_0, it claims the presence of a double layer of electricity of such a strength that a surface element of size λ_0^2 of the negative layer contains at least hc/e^2 electrons. It is therefore clear that the problem in question cannot legitimately be treated as that of one electron moving in a given potential field, but is essentially a many electron problem which falls outside the range of present quantum mechanics.

On the the whole it appears that the circumstance that hc/e^2 is large compared with unity does not only indicate the actual limit of the applicability of the quantum theory in its present form, but at the same time ensures its consistency within these limits. In fact the radius r_0 of the electron estimated on classical theory is $e^2/mc^2 = (h/me)(e^2/hc)$, and we can therefore never determine the position of an electron within an accuracy comparable with r_0 without allowing an uncertainty in its momentum larger than mc, thus entailing an uncertainty with energy surpassing the critical value mc^2. The idea that the reach of quantum mechanics is bound up with the actual existence of the electron would also seem to be in harmony with fact that the symbols e and m appear in the fundamental equations of the present theory.

In the problem of β-decay spectra we may now be outside the natural limit for the consistent applicability of the concepts of energy and momentum, and in this sense we may regard the expulsion of a β ray from a nucleus as the birth of an electron as a dynamical individual. In the fact that the total charge of the nucleus can be measured before and after the β-ray disintegration and that the results are in conformity with conservation of electricity I see a support for upholding the conservation of the elementary charges even at the risk of abandoning the conservation of energy, and I do not quite understand your reasons for taking the opposite view. Of course, I do not wish to advocate any of the scepsis of old and new as to the strict conservation of energy in ordinary quantum theory. On the contrary my view is that the legitimate field of application of the conservation theorems may be just the same as that of a consistant application of quantum mechanics in its present form, where the problems arising in classical electrodynamics in connection with the constitution of the electron are neglected.

As regards the transitions from positive to negative energy accompanied by radiation I am not sure that they present as serious a difficulty for your wave equation as it might appear. The question is, how much those features of the theory which claim the transitions in question are involved in the problems, where your theory has been found in so wonderful agreement with experiments. In this connection I must correct the statement in my former letter regarding the probability of these transitions which is not nearly so large as I believed. In discussing the problem more closely with Klein we convinced ourselves that the estimation of this probability did not take sufficient regard to the smallness of the wavelength of the radiation concerned compared with atomic dimensions. We have not made an actual calculation of any such probability, and if you have considered the problem in detail I should be very thankful for any information regarding this point. My hope is that it should be possible to defend all the successful applications of your wave equation, but I suspect that the natural limitation of these applications prevents an extrapolation of the kind you describe in your letter. As regards the problem of annihilation of electrons and protons which you mention in this connection it appears to me that the astrophysical evidence is of a very conflicting nature. Thus Eddington's theory of the equilibrium of stars seems to indicate that the rate of energy production per unit mass ascribed to such annihilation is larger in the earlier stages of stellar evolution where the density in the interior is smaller than in the later stages where the interior density is larger. As far as I can see any views like yours would claim a variation with the density in the opposite direction. On the whole it seems to me that an understanding of the laws of stellar evolution claims some new radical departure from our present view regarding energy balance.

With kindest regards from Klein and myself,

Yours sincerely,

Dirac submitted this hole interpretation to the Royal Society 6 December 1929 and in mid-December presented it at l'Institute Henri Poincaré.[22] The Royal Society paper—"A Theory of Electrons and Protons"—began with a discussion "of unwanted solutions with negative kinetic energy for the electron, which appear to have no physical meaning." They arise because the corresponding classical relativistic Hamiltonian is quadratic and cannot be suppressed because perturbations of external fields must cause transitions to these states and "furthermore, in the accurate quantum theory in which the electromagnetic field also is subjected to quantum laws, transitions can take place in which the energy of the electron changes from a positive to a negative value even in the absence of any external field." Dirac noted in his original theory of the electron than an electron in a negative kinetic energy state would move as a positive charge would, and Weyl had suggested that it might thus be a proton. Dirac now noted that "one cannot simply assert that a negative-energy electron *is* a proton as that would lead to ... paradoxes" regarding conservaton of charge, momentum, and energy. Dirac's resolution of the negative-energy problem had three parts. First, he used Pauli's exclusion principle to argue that "*all the states of negative energy are occupied* except perhaps a few of small velocity." Second, he asserted that these states would not themselves be observable. Even though there would be an "infinite number per unit volume all over the world ... if their distribution is exactly uniform we should expect them to be completely unobservable. *Only the small departure from exact uniformity, brought about by some negative-energy states being unoccupied, can we hope to observe*." Only departures from the otherwise uniformly infinite volume charge density of the vacuum are observable and used in the Maxwell equation, $\nabla \cdot \mathbf{E} = -4\pi\rho$. Third, Dirac recognized that unoccupied negative states, or holes, would move like positively charged particles with positive kinetic energies and argued: "We are therefore led to the assumption that *the holes ... are the protons*." The great beauty of this assumption was than now "we require to postulate only one fundamental kind of particle, instead of the two, electron and proton, that were previously necessary." Dirac expected that the dissymmetry between the masses of the electron and proton would be accounted for by "some more prefect theory of the interactions, based perhaps on Eddington's calculation of the fine-structure constant." Because it was known that the negative energy states had to play a role in scattering—as Waller informed Dirac "in some important practical cases nearly all the scattering comes from intermediate states with negative energy for the electron"—and to give the newly interpreted theory some more tangible plausibility Dirac used it to provide "a justification for the scattering formula of Klein and Nishina which was deduced by these authors with the help of classical analogies not rigorously proved to be consequences of general quantum mechanics."

Dirac's new magic received quick notice,[23] in the 1 February issue of *Nature*, for example: "The existence of positive electricity can be predicted by a fairly direct line of argument." However, in the 1 March issue of *Physical Review* Robert Oppenheimer noted a new sickness. He calculated that if Dirac's holes were protons they would annihilate with electrons with a mean lifetime of 10^{-10} sec which is inconsistent with the observed stability of matter: "We should hardly expect any states of negative energy to remain empty ... [and should] return to the assumption of two independent elementary particles of opposite charges." *Nature* then noted that filling *all* the negative energy states "involves the rejection of the fundamental similarity of positive and negative electricity, which was perhaps the most attractive feature of Dr. Dirac's theory." Oppenheimer had also noted that the mass dissymmetry befouled Dirac's scattering calculations. Dirac noted this as well along with his own calculation that the annihilation rate "is much too large to agree with the known stability of electrons and protons." Nonetheless the new magic was attractive and it was suggested that Dirac's holes could be used to make a theory of beta decay, used to show "why Pauli's principle governs the world," and used to show why the ether is unobservable. Bohr used his visit to England in the spirng of 1930 to criticize Dirac's hole interpretation and to continue his argument about the limit of quantum mechanics. Gamow was to telegraph to Bohr "Wunderbar" if Dirac had renounced his heresy at the BAAS meeting in the fall or "Quatsch" if Dirac persisted. He sent "quatsch" since Dirac was not so easily detered from the formal unity he had discovered. *Nature* also poked some fun but also published Dirac's full BAAS paper. In the first part of this paper Dirac stressed the olympian unity afforded by the hole interpretation: "It has always been the dream of philosophers to have all matter built up from one fundamental kind of particle, so that it is not altogether satisfactory to have two in our theory, the electron and proton. There are, however, reasons for believing that the electron and proton are really not independent, but are just two manifestations of one elementary kind of particle." In the middle part of the paper Dirac argued that a "connection between the electron and the proton is, in fact, rather forced upon us by general considerations about the ... tracks in space-time" associated with the relativistic Hamiltonian and by the Pauli exclusion principle. In the last part of the paper he acknowledged that the theory "involves certain difficulties" with the infinite charge density of filled negative energy states, with the annihilation rate that is "much too large to be true," and with the dissimilar masses of electrons and protons, but in the end he hoped "further advances in the theory of quantum electrodynamics...will settle the difficulties."

Magnetism was the topic of the sixth Solvay Conference in the fall of 1930 and Dirac's electron was featured in the papers and discussions.[24] For example, Fermi gave a paper on nuclear magnetic moments in which he constructed configurations and motions of electrons confined within the nucleus such that their contribution to the nuclear magnetic moment would be consistent with measurements of hyperfine structure. In the discussion the general difficulties of testing a theory based on specific unobserved machinery such as Fermi's were emphasized. Dirac remarked that the problem of the magnetic moment of confined electrons was not so serious to require a special explanation because it could be explained generally by the great speed of change of orientation of electrons subject to the great forces within the nucleus. Fermi responded that his theory was based on the relationship between the small volume of confinement and the energy of the electron. But, Dirac rebutted, the magnetic moment and position of the electron commute and according to the relativistic theory the magnetic moment itself is independent of speed. Heisenberg criticized Dirac's general argument that the electron magnetic moments could disappear on the average in the nucleus because this could not explain the details of special cases. Heisenberg emphasized that the problem of nuclear statistics, of $_7N^{14}$, for example, was serious. Pauli gave a major paper on the magnetic electron with a major section on Dirac's relativistic theory in which he repeated Bohr's argument that the magnetic moment of free electrons was unobservable because of the uncertainty principle. In the discussion Fermi suggested an indirect method of polarizing electrons by thermal diffusion but Bohr repeated his general philosophical argument that experiments could not disclose any intrinsic magnetic moment of the electron. O. W. Richardson asked if Bohr denied the reality of electron spin in the face of spectroscopic evidence. Van Vleck noted the usefulness of this property in explaining paramagnetism and Richardson added that the whole theory of metals uses the idea of free eelctrons with a property called spin. Bohr replied that the property called spin could not be interpreted because there was no corresponding classical property.

Bohr interpreted quantum-mechanical properties constructively by experimental correspondence with classical properties. Dirac interpreted quantum-mechanical properties generally with his transformation theory and used the general transformation theory to prove formal analogies with classical mechanics such as "the true analogue of Llouville's theorem," $\dot{\rho} = [\rho, H]$, which he often used to illustrate his formal appraoch to quantum mechanics. However, his approach was too formal for most of his colleagues.[25]

Even though the Copenhagen interpretation of quantum mechanics was becoming standard, Bohr's use of the uncertainty principle to show limitations of his correspondence principle was causing problems. Darwin, for example, had to apologize for "rather careless remarks that I had made" that misrepresented Bohr's argument that the uncertainty principle asserts itself to make the magnetic moment of the free electron unmeasurable. Bohr was also unhappy with a paper by Landau and Peierls that concluded than "in the correct relativistic quantum theory (which does not yet exist), there will therefore be no physical quantities and no measurements." They justified this odd result by incorporating relativity into the uncertainty principle to show "that all the physical quantities occuring in wave mechanics can in general no longer be defined in the relativistic range." They argued that since field strengths, positions, and momenta all become unmeasurable, "it is therefore not surprising that the formalism leads to various infinities; it would be surprising if the formalism bore any resemblance to reality." This result is confirmed, they added, by "the physically meaningless solutions [of Dirac's equation] with negative energy" and by the fact of the beta-decay spectrum that "means that the law of conservation is probably invalid for nucelar electrons." Though Bohr objected because Landau and Peierls suggested there could be no correspondence between relativistic quantum mechanics and classical mechanics. Landau and Peierl's argument gained wide notice.[26]

In April of 1931 Dirac wrote to Van Vleck that "Bohr is at present trying to convince everyone that the places where relativistic quantum theory fails are just those where one would expect it to fail from general philosophical consideration." Dirac was at that time composing a strong and clear description of his own policy that was published in September: "The most powerful method of advance that can be suggested at present is to employ all the resources of pure mathematics in attempts to perfect and generalize the mathematical formalism that forms the existing basis of theoretical physics, and *after* each success in this direction, to try to interpret the new mathematical features in terms of physical entities." Dirac's relativistic quantum theory of the electron and hole interpretation were examples of this "method of advance." Dirac now modified his hole interpretation because he recognized that the hole "necessarily has the same mass an electron" and because in the original interpretation electrons and protons could not be stable. Since his new interpretation has so often been misrepresented, it should be quoted in full:

> It thus appears that we must abandon the identification of the holes with protons and must find some other interpretation for them. Following Oppenheimer, we can assume that in the world as we know it, all, and not merely nearly all, of the negative-energy states for electrons are occupied. A hole, if there were one, would be a new kind of particle, unknown to experimental physics, having the same mass and opposite charge to an electron. We may call such a particle an antielectron. We should not expect to find any of them in nature, on account of their rapid rate of recombination with electrons, but if they could be produced experimentally in high vacuum they would be quite stable and amenable to observation. An encounter between two hard γ rays (of energy at least half a million volts) could lead to the creation simultaneously of an electron and antielectron, the probability of occurrence of this process being of the same order of magnitude as that of the collision of the two γ rays on the assumption that they are spheres of the same size as classical electrons. This probability is negligible, however, with the intensities of γ rays at present available.
>
> The protons on the above view are quite unconnected with electrons. Presumably the protons will have their own negative-energy states, all of which normally are occupied, an unoccupied one appearing as an antiproton. Theory at present is quite unable to suggest a reason why there should be any differences between electrons and protons.

The main business of Dirac's paper was to give another example of his "general scheme of advance" and predict the possible existence of a magnetic monopole. Dirac did not predict that antielectrons, or magnetic monopoles, would be observed. Quite the contrary, he gave several reasons why they could not be observed, at least not easily.[27]

In 1930 Pauli had suggested the existence of another new particle—which we now call the neutrino, but he called the neutron—to resolve problems of nuclear spins and beta-decay energetics. There is an interesting distinction here between asserting the existence of a new particle to resolve some difficulty and discovering that the formalism allows some new particle to exist. In 1931 Pauli discovered that the Dirac equation allows the existence of a neutral magnetic dipole that he called the magnetic neutron and lectured on in Berkeley, Ann Arbor, and Princeton. Dirac wrote to Van Vleck on October 2nd that "Pauli came to Princeton last night and he and I gave a colloquium on neutrons and magnetic poles." Chatting with Rabi about his neutron hypothesis in a Chinese restaurant in New York during this visit Pauli remarked, "I think I'll be cleverer than Dirac. I don't think I'll publish it." Pauli was cleverer for if we regard his multifaceted hypothesis to be false, we respect his wisdom for not publishing; but, if we identify his hypothesis with Fermi's neutrino we respect Pauli's wisdom for making it known to him.

While Dirac and Pauli were having fun with magnetic monopoles and dipoles at Princeton, at the BAAS meeting and Volta Conference Bohr was arguing that general philosophical consideration and failure of Dirac's relativistic quantum theory and specific problems of nuclear electrons and beta-decay energetics showed the limits of quantum mechanics. Although Fermi designed the Volta Conference to give younger physicists the opportunity to give their papers, Bohr's view was clearly apparent in Mott's report "On the Present Status of the Theory of the Electron." He began with a discussion of how Schrödinger's equation could be deduced "from the minimum of experimental evidence" and how potential could be interpreted only by use of the correspondence principle and concluded "that it is not possible to use the Dirac equation to describe the behavior of electrons in the nucleus." Bohr's views were most fully expressed in the published versions of his Faraday Lecture and of his remarks at the Volta Conference. Bohr first argued "that the very stability of atomic structures, which is essential for our analysis of natural phenomena, imposes unavoidable limitations on the use of space-time pictures in accounting for atomic reactions." He supported this with an account of the historical development of the crisis and revolution in atomic theory and his familiar philosophical argument that the quantum of action introduces a basic uncertainty limiting the use of space-time coordination and energy-momentum conservation to "two *complementary aspects of ordinary causality* which ... exclude one another to a certain extent, although neither of them has lost its intrinsic validity." He developed also a new argument that when properties with unambiguous classical meanings are used together with properties that have no unambiguous classical meaning such as the quantum of action or the spin then "the whole attack on atomic problems leaning on the correspondence argument is an *essentially approximative procedure* made possible by the smallness of the ratio" e^2/hc. From the thesis that the very stability of atoms imposed limits on classical mechanics he extrapolated to assert that the stability of elementary particles imposes limits on quantum mechanics. He supported this with his argument that "the theory of Dirac ... has disclosed new aspects of the fundamental difficulties involved in the reconciliation of the intrinsic stability of the electron with the existence of the quantum of action." The difficulties included Dirac's unobservable negative energies, the ambiguity of force indicated by the Klein paradox, the unmeasurable magnetic moment of the electron, the uninterpretable spin, and the unresolved infinities abounding in relativistic quantum mechanics. Furthermore, Bohr argued, "the present formulation of quantum mechanics fails essentially" when applied to even the simplest nuclei because, for example, "the idea of spin is found not to be applicable to intranuclear electrons" which is "a very direct indication, indeed, of the essential limitation of the idea of separate dynamical entities when applied to electrons" and means that "the expulsion of a β ray from a nucleus may be regarded as the creation of an electron as a mechanical entity." Bohr suspected that resolution of these problems "would appear to be out of reach of the present formulation of the quantum theory." Just as resolution of problems of atomic stability "implies a renunciation of the classical idea of causality" resolution of the new problems, Bohr suggested, "may force us to renounce the very idea of energy balance", adding that "we have no argument either experimental or theoretical for upholding the energy principle in the case of β-ray disintegrations, and are even led to complications and difficulties in trying to do so."[28]

Fermi, in his great review of the quantum theory of radiation,[29] shared Bohr's judgment that because of unresolved infinities "we may therefore say that practically all the problems in radiation theory that do not involve the structure of the electron have their satisfactory explanation; while the problems connected with the internal properties of the electron are still very far from their solution." Fermi noted that Dirac's anamolous negative energy values" play a necessary role in Compton scattering but since they "have no physical significance" transitions to them "in reality certainly do not take place" and a "correct theory should find some way of preventing them."

Oppenheimer remembered that Pauli's judgment of Dirac's relativistic quantum mechanics was that "any theory which had such a sickness must agree with experience only be accident." Indeed, both the tone and substance of Pauli's great review of relativistic quantum mechanics bear this out. Here he argued that because of the anamolous infinites and negative energies it was not even possible to establish the limits of usefullness of the theory.[30]

Pauli's magnetic neutron, Dirac's magnetic monopole, Dirac's "donkey electrons," and the infinities of QED were the butt of jokes at Copenhagen, especially in the Blegdamsvej Faust" in April of 1932.[31]

The magic and sickness of Dirac's electron were discussed at the Paris International Electrical Congress in the late summer of 1932. There was no suspicion at all that Dirac's antielectron would turn out to be real even though two phenomena that would shortly vindicate Dirac's magic were presented in Fermi's discussion of the "Meitner–Hupfeld effect" and in Millikan's discussion of "positives" observed by Carl Anderson in cosmic rays.[32]

[1]C. G. Darwin, Proc. Soc. London **118**, 654 (1928); W. Gordon, Z. Phys. **48**, 11 (1928).

[2]D. R. Hartree, Proc. Cambridge Philos. Soc. **25**, 225 (1929); J. A. Gaunt, Proc. R. Soc. London **124**, 163 (1929); P. A. M. Dirac, *ibid.* **123**, 714 (1929); J. Hargreaves, *ibid.* **124**, 568 (1929); E. Fermi, *Collected Papers* (University of Chicago, Chicago, 1926), p. 328.

[3]O. Klein, Z. Phys. **53**, 157 (1929).

[4]N. Bohr, Nature **121**, 580 (1928); C. G. Darwin, Proc. R. Soc. London **117**, 258 (1927).

[5]O. Klein and Y. Nishina, Nature **122**, 398 (1928); Z. Phys. **52**, 853 (1929); Y. Nishina, Nature **122**, 843 (1928); Z. Phys. **52**, 869 (1929); Nature **123**, 349 (1929).

[6]Nature **123**, 174 (1929); J. A. Gray, *ibid.* **123**, 241 (1929).

[7]E. Rutherford, Proc. R. Soc. London **122**, 1 (1929); Nature **123**, (1929).

[8]J. R. Oppenheimer, Phys. Rev. **31**, 66 (1928); **32**, 361 (1928); Proc. Natl. Acad. Sci. U.S. **14**, 261 (1928); N. F. Mott, Proc. R. Soc. London **118**, 542 (1928); **124**, 422 (1929); **125**, 222 (1929); **126**, 79 (1929); **126**, 259 (1930); **127**, 658 (1930); Nature **123**, 717 (1929); Proc. Cambridge Philos. Soc. **25**, 304 (1929).

[9]Nature **126**, 257 (1930); **127**, 179 (1931); J. A. Chadwick, Proc. R. Soc. London **128**, 114 (1930); P. M. S. Blackett and F. C. Champion, Proc. R. Soc. London **130**, 380 (1930–31).

[10]H. C. Wolfe, Phys. Rev. **37**, 591 (1931); C. Møller, Z. Phys. **70**, 786 (1931); F. C. Champion, Proc. R. Soc. London **136**, 630 (1932); **137**, 688 (1932).

[11]G. Breit, Proc. Natl. Acad. Sci. U.S. **14**, 553 (1928); C. G. Darwin, Proc. R. Soc. London **120**, 621 (1928); **120**, 631 (1928); Nature **122**, 980 (1928); E. Schrödinger, Preuss. Abad. Wiss. Berlin **24**, 418 (1930).

[12]N. F. Mott, Proc. R. Soc. London **124**, 425 (1929); A. Franklin, Studies in History and Philosophy of Science (to be published).

[13]N. F. Mott, R. Acad. Ital. Roma, 22 (1932); A. Rich and J. C. Wesley, Rev. Mod. Phys. **44**, 250 (1972).

[14]J. R. Oppenheimer, interview in *Sources for History of Quantum Physics* (American Philosophical Society, Philadelphia, 1967); T. S. Kuhn, J. L. Heilbron, P. Forman, and L. Allen, *ibid.*

[15]C. G. Darwin, Proc. R. Soc. London **118**, 654 (1928).

[16]A. S. Eddington, Proc. R. Soc. London **121**, 524 (1928); **122**, 358 (1929); Nature **124**, 840 (1929).

[17]G. Beck, H. Bethe, and W. Riezler, Naturwiss. **19**, 39 (1931); Nature **123**, 174 (1929); N. Weiner and M. S. Vallarta, Nature **123**, 317 (1929).

[18]O. Laporte and G. Uhlenbeck, Phys. Rev. **37**, 1380 (1931); O. Veblen, J. Wash. Acad. Sci. **24**, 281 (1934); A. H. Taub, O. Veblen, and J. von Neuman, Proc. Natl. Acad. Sci. U.S. **20**, 383 (1934; P. A. M. Dirac, Ann. Math. **30**, 657 (1935); **37**, 492 (1936); M. Born, Proc. R. Soc. London **143**, 4101 (1934); M. Born and L. Infeld, **144**, 425 (1934); **147**, 522 (1934); W. Heisenberg and W. Pauli, Z. Phys. **56**, 1 (1929); **59**, 168 (1930); J. R. Oppenheimer, Phys. Rev. **35**, 562 (1930); **38**, 725 (1931); C. G. Darwin, Nature **123**, 203 (1928).

[19]P. A. M. Dirac, Proc. R Soc. London **123**, 714 (1929); E. Rutherford, Proc. R. Soc. London **123**, 373 (1929); O. M. Corbino, Minerva **9**, 528 (1971).

[20]P. A. M. Dirac, letter to Bohr, Ref. 14.

[21]P. A. M. Dirac, letter from Bohr, Ref. 14.

[22]P.A. M. Dirac, Am. Inst. Henri Poincaré **1**, 357 (1931); Proc. R. Soc. London **126**, 360 (1930); H. Weyl Z. Phys. **31**, 1006 (1930); I Waller, *ibid.* **61**, 837 (1930).
[23]Nature **125**, 182 (1930); J. R. Oppenheimer, Phys. Rev. **35**, 562 (1930); Nature **125**, 616 (1930); P. A. M. Dirac, Proc. Cambridge Philos. Soc. **26**, 361 (1930); V. Ambarzumian and D. Iwanenko, C. R. Acad. Sci. **190**, 582 (1930); E. U. Condon and J. E. Mack, Phys. Rev. **35**, 579 (1930); Nature **126**, 455 (1930); **125**, 610 (1930); G. Beck, interview deposited in Library of American Philosophical Society (I thank Laurie Brown for this reference) (unpublished); P. A. M. Dirac, BAAS Report, 303 (1930); P. A. M. Dirac, Nature **126**, 605 (1930); **126**, 449 (1930).
[24]*Le Magnetisme...*, 6th Solvay Conference (Gauthier-Villars, Paris, 1932).
[25]P. A. M. Dirac, Proc. Cambridge Philos. Soc. **25**, 62 (1929); Nature **127**, 699 (1931).
[26]W. Heisenberg, *The Physical Principles of the Quantum Theory* (Dover, New York, 1949); C. G. Darwin, Proc. R. Soc. London **130**, 632 (1931); L. Landau and R. Peierls, Z. Phys. **69**, 56 (1931); N. Bohr and L. Rosenfeld, Det. Kgl. Dansk. Videnskabernes Selskab., Matematisk-fysiske Meadelelser **12**, 1 (1933); L. deBroglie, *Exposes de Physique Theorique* (Herman, Paris, 1932); D. Iwanenbo, Z. Phys. **72**, 621 (1931).
[27]P. A. M. Dirac, letter to Van Vleck, Ref. 14; Proc. R. Soc. London, **133**, 60 (1931).
[28]L. M. Brown, Phys. Today **31**, 23 (September 1978); P. A. M. Dirac, letter to Van Vleck, Ref. 14; I. Rabi, interview in Ref. 14.
[29]Volta Conference, R. Acad. Ital. Rome (1932); N. Bohr. BAAS Report, 333 (1931); J. Chem. Soc., 349 (1932).
[30]E. Fermi, Rev. Mod. Phys. **4**, 87 (1932).
[31]J. R. Oppenheimer, interview in Ref. 14; W. Pauli, Handbuch der Physik, 2nd ed. **24**, II (1933).
[32]G. Gamow, *Thirty Years That Shook Physics* (Doubleday, Garden City, NY, 1966).
[33]*Comptes Rendus du Congress International d'Electricit* (Gauthier-Villars, Paris, 1932).

Generalized Lagrangian and conservation law for the damped harmonic oscillator

Leon Y. Bahar and Harry G. Kwatny
Department of Mechanical Engineering and Mechanics, Drexel University, Philadelphia, Pennsylvania 19104
(Received 19 May 1980; accepted 21 January 1981)

An explicit construction of the time-dependent Lagrangian for a damped harmonic oscillator is given. The connection between this Lagrangian with the time-independent one developed by Morse and Feshbach is established. Various methods of deriving a first integral for the equation of motion are given. It is shown that the results developed in this paper by direct means are in agreement with those obtained through the use of group-theoretic methods, such as invariance principles and Noether's theorem, which are mathematically more advanced approaches.

I. INTRODUCTION

In recent papers by Yan,[1] Edwards,[2] Ray,[3] and Lemos[4] the Lagrangian

$$L = \exp[(c/m)t\,][(m\dot{x}^2/2) - (kx^2/2)] \tag{1}$$

is used to generate the equation of motion for the linearly damped harmonic oscillator in the form

$$\exp[(c/m)t\,](m\ddot{x} + c\dot{x} + kx) = 0, \tag{2}$$

which reduces to the standard equation of motion

$$m\ddot{x} + c\dot{x} + kx = 0 \tag{3}$$

in view of the fact that the exponential multiplier in Eq. (2) does not introduce any extraneous solutions to the equation of motion described by Eq. (3).

It is of interest to note that none of the articles cited above attempt to derive the Lagrangian given by Eq. (1) directly from the equation of motion (3) through the use of the methodology of the "inverse problem of Newtonian mechanics," an exhaustive account of which has been recently given by Santilli.[5] Even recent texts such as Logan[6] follow the same pattern, merely stating that "It is not difficult to see that the Lagrangian that leads to Eq. (3) is the one given by Eq. (1)."

Because of the frequent use of Eq. (1) in the literature, it is of importance to construct that expression explicitly, and to investigate its implications. First, however, it is important to discuss, in general, the motivation for seeking an arbitrary $L(x, \dot{x}, t)$ (not necessarily of the form T-V), which leads to the correct equation of motion up to a trivial multiplier through the prescription

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0. \tag{4}$$

It should be emphasized that in Eq. (4) L is no longer the conventional Lagrangian defined by

$$L = T - V, \tag{5}$$

where V is at most a generalized potential of the form $V(x, \dot{x}, t)$ from which the corresponding generalized force is derivable through the defining relation

$$Q = \frac{d}{dt}\left(\frac{\partial V}{\partial \dot{x}}\right) - \frac{\partial V}{\partial x}. \tag{6}$$

Instead, L in Eq. (4) can be called a generalized Lagrangian, which reduces to the classical Lagrangian only when it coincides with Eq. (5).

Obviously, Eq. (3) can be obtained by other Lagrangian formalisms, such as using the damping term as a generalized force, or using the Rayleigh dissipation function. It is therefore important to emphasize that the motivation for using the Lagrangian characterized by Eq. (1) is not to generate the corresponding equation of motion (3). Rather, it is in order to find a Lagrangian whose structure is such that it will produce the equation of motion solely through the prescription defined by Eq. (4). The search for a generalized Lagrangian of this form is dictated by the necessity to ex-