1

Historical Introduction

Our immersion in the present state of physics makes it hard for us to understand the difficulties of physicists even a few years ago, or to profit from their experience. At the same time, a knowledge of our history is a mixed blessing — it can stand in the way of the logical reconstruction of physical theory that seems to be continually necessary.

I have tried in this book to present the quantum theory of fields in a logical manner, emphasizing the deductive trail that ascends from the physical principles of special relativity and quantum mechanics. This approach necessarily draws me away from the order in which the subject in fact developed. To take one example, it is historically correct that quantum field theory grew in part out of a study of relativistic wave equations, including the Maxwell, Klein–Gordon, and Dirac equations. For this reason it is natural that courses and treatises on quantum field theory introduce these wave equations early, and give them great weight. Nevertheless, it has long seemed to me that a much better starting point is Wigner's definition of particles as representations of the inhomogeneous Lorentz group, even though this work was not published until 1939 and did not have a great impact for many years after. In this book we start with particles and get to the wave equations later.

This is not to say that particles are necessarily more fundamental than fields. For many years after 1950 it was generally assumed that the laws of nature take the form of a quantum theory of fields. I start with particles in this book, not because they are more fundamental, but because what we know about particles is more *certain*, more directly derivable from the principles of quantum mechanics and relativity. If it turned out that some physical system could not be described by a quantum field theory, it would be a sensation; if it turned out that the system did not obey the rules of quantum mechanics and relativity, it would be a cataclysm.

In fact, lately there has been a reaction against looking at quantum field theory as fundamental. The underlying theory might not be a theory of fields *or* particles, but perhaps of something quite different, like strings.

From this point of view, quantum electrodynamics and the other quantum field theories of which we are so proud are mere 'effective field theories,' low-energy approximations to a more fundamental theory. The reason that our field theories work so well is not that they are fundamental truths, but that any relativistic quantum theory will look like a field theory when applied to particles at sufficiently low energy. On this basis, if we want to know why quantum field theories are the way they are, we have to start with particles.

But we do not want to pay the price of altogether forgetting our past. This chapter will therefore present the history of quantum field theory from earliest times to 1949, when it finally assumed its modern form. In the remainder of the book I will try to keep history from intruding on physics.

One problem that I found in writing this chapter is that the history of quantum field theory is from the beginning inextricably entangled with the history of quantum mechanics itself. Thus, the reader who is familiar with the history of quantum mechanics may find some material here that he or she already knows, especially in the first section, where I discuss the early attempts to put together quantum mechanics with special relativity. In this case I can only suggest that the reader should skip on to the less familiar parts.

On the other hand, readers who have no prior familiarity with quantum field theory may find parts of this chapter too brief to be altogether clear. I urge such readers not to worry. This chapter is not intended as a self-contained introduction to quantum field theory, and is not needed as a basis for the rest of the book. Some readers may even prefer to start with the next chapter, and come back to the history later. However, for many readers the history of quantum field theory should serve as a good introduction to quantum field theory itself.

I should add that this chapter is not intended as an original work of historical scholarship. I have based it on books and articles by real historians, plus some historical reminiscences and original physics articles that I have read. Most of these are listed in the bibliography given at the end of this chapter, and in the list of references. The reader who wants to go more deeply into historical matters is urged to consult these listed works.

A word about notation. In order to keep some of the flavor of past times, in this chapter I will show explicit factors of $\hbar$ and c (and even h), but in order to facilitate comparison with modern physics literature, I will use the more modern *rationalized* electrostatic units for charge, so that the fine structure constant $\alpha \simeq 1/137$ is $e^2/4\pi\hbar c$. In subsequent chapters I will mostly use the 'natural' system of units, simply setting $\hbar = c = 1$.

1.1 Relativistic Wave Mechanics

Wave mechanics started out as relativistic wave mechanics. Indeed, as we shall see, the founders of wave mechanics, Louis de Broglie and Erwin Schrödinger, took a good deal of their inspiration from special relativity. It was only later that it became generally clear that relativistic wave mechanics, in the sense of a relativistic quantum theory of a fixed number of particles, is an impossibility. Thus, despite its many successes, relativistic wave mechanics was ultimately to give way to quantum field theory. Nevertheless, relativistic wave mechanics survived as an important element in the formal apparatus of quantum field theory, and it posed a challenge to field theory, to reproduce its successes.

The possibility that material particles can like photons be described in terms of waves was first suggested[1] in 1923 by Louis de Broglie. Apart from the analogy with radiation, the chief clue was Lorentz invariance: if particles are described by a wave whose phase at position $\mathbf{x}$ and time t is of the form $2\pi(\boldsymbol{\kappa} \cdot \mathbf{x} - \nu t)$, and if this phase is to be Lorentz invariant, then the vector $\boldsymbol{\kappa}$ and the frequency ν must transform like $\mathbf{x}$ and t, and hence like $\mathbf{p}$ and E. In order for this to be possible $\boldsymbol{\kappa}$ and ν must have the same velocity dependence as $\mathbf{p}$ and E, and therefore must be proportional to them, with the same constant of proportionality. For photons, one had the Einstein relation $E = h\nu$, so it was natural to assume that, for material particles,

$$\boldsymbol{\kappa} = \mathbf{p}/h \ , \qquad \nu = E/h \ , \tag{1.1.1}$$

just as for photons. The group velocity $\partial\nu/\partial\boldsymbol{\kappa}$ of the wave then turns out to equal the particle velocity, so wave packets just keep up with the particle they represent.

By assuming that any closed orbit contains an integral number of particle wavelengths $\lambda = 1/|\boldsymbol{\kappa}|$, de Broglie was able to derive the old quantization conditions of Niels Bohr and Arnold Sommerfeld, which though quite mysterious had worked well in accounting for atomic spectra. Also, both de Broglie and Walter Elsasser[2] suggested that de Broglie's wave theory could be tested by looking for interference effects in the scattering of electrons from crystals; such effects were established a few years later by Clinton Joseph Davisson and Lester H. Germer.[3] However, it was still unclear how the de Broglie relations (1.1.1) should be modified for non-free particles, as for instance for an electron in a general Coulomb field.

Wave mechanics was by-passed in the next step in the history of quantum mechanics, the development of matrix mechanics[4] by Werner Heisenberg, Max Born, Pascual Jordan and Wolfgang Pauli in the years 1925–1926. At least part of the inspiration for matrix mechanics was the insistence that the theory should involve only observables, such as the energy levels, or emission and absorption rates. Heisenberg's 1925 paper opens with the manifesto: 'The present paper seeks to establish a basis for theoretical quantum mechanics founded exclusively upon relationships between quantities that in principle are observable.' This sort of positivism was to reemerge at various times in the history of quantum field theory, as for instance in the introduction of the S-matrix by John Wheeler and Heisenberg (see Chapter 3) and in the revival of dispersion theory in the 1950s (see Chapter 10), though modern quantum field theory is very far from this ideal. It would take us too far from our subject to describe matrix mechanics in any detail here.

As everyone knows, wave mechanics was revived by Erwin Schrödinger. In his 1926 series of papers,[5] the familiar non-relativistic wave equation is suggested first, and then used to rederive the results of matrix mechanics. Only later, in the sixth section of the fourth paper, is a relativistic wave equation offered. According to Dirac,[6] the history is actually quite different: Schrödinger first derived the relativistic equation, then became discouraged because it gave the wrong fine structure for hydrogen, and then some months later realized that the non-relativistic approximation to his relativistic equation was of value even if the relativistic equation itself was incorrect! By the time that Schrödinger came to publish his relativistic wave equation, it had already been independently rediscovered by Oskar Klein[7] and Walter Gordon,[8] and for this reason it is usually called the 'Klein–Gordon equation.'

Schrödinger's relativistic wave equation was derived by noting first that, for a 'Lorentz electron' of mass m and charge e in an external vector potential $\mathbf{A}$ and Coulomb potential ϕ, the Hamiltonian H and momentum $\mathbf{p}$ are related by*

$$0 = (H + e\phi)^2 - c^2(\mathbf{p} + e\mathbf{A}/c)^2 - m^2c^4 \, . \tag{1.1.2}$$

For a *free* particle described by a plane wave $\exp\left\{2\pi i(\boldsymbol{\kappa}\cdot\mathbf{x} - \nu t)\right\}$, the de Broglie relations (1.1.1) can be obtained by the identifications

$$\mathbf{p} = h\boldsymbol{\kappa} \to -i\hbar\nabla \ , \qquad E = h\nu \to i\hbar\frac{\partial}{\partial t} \ , \tag{1.1.3}$$

where $\hbar$ is the convenient symbol (introduced later by Dirac) for $h/2\pi$. By an admittedly formal analogy, Schrödinger guessed that an electron in the external fields $\mathbf{A}, \phi$ would be described by a wave function $\psi(\mathbf{x}, t)$ satisfying the equation obtained by making the same replacements in

* This is Lorentz invariant, because the quantities $\mathbf{A}$ and ϕ have the same Lorentz transformation property as $c\mathbf{p}$ and E. Schrödinger actually wrote H and $\mathbf{p}$ in terms of partial derivatives of an action function, but this makes no difference to our present discussion.

(1.1.2):

$$0 = \left[\left(i\hbar \frac{\partial}{\partial t} + e\phi\right)^2 - c^2\left(-i\hbar\nabla + \frac{e\mathbf{A}}{c}\right)^2 - m^2c^4\right]\psi(\mathbf{x},t) . \tag{1.1.4}$$

In particular, for the stationary states of hydrogen we have $\mathbf{A} = 0$ and $\phi = e/4\pi r$, and ψ has the time-dependence $\exp(-iEt/\hbar)$, so (1.1.4) becomes

$$0 = \left[\left(E + \frac{e^2}{4\pi r}\right)^2 - c^2\hbar^2\nabla^2 - m^2c^4\right]\psi(\mathbf{x}) . \tag{1.1.5}$$

Solutions satisfying reasonable boundary conditions can be found for the energy values[9]

$$E = mc^2\left[1 - \frac{\alpha^2}{2n^2} - \frac{\alpha^4}{2n^4}\left(\frac{n}{\ell + \frac{1}{2}} - \frac{3}{4}\right) + \cdots\right] , \tag{1.1.6}$$

where $\alpha \equiv e^2/4\pi\hbar c$ is the 'fine structure constant,' roughly 1/137; n is a positive-definite integer, and ℓ, the orbital angular momentum in units of $\hbar$, is an integer with $0 \leq \ell \leq n - 1$. The α^2 term gave good agreement with the gross features of the hydrogen spectrum (the Lyman, Balmer, etc. series) and, according to Dirac,[6] it was this agreement that led Schrödinger eventually to develop his non-relativistic wave equation. On the other hand, the α^4 term gave a fine structure in disagreement with existing accurate measurements of Friedrich Paschen.[10]

It is instructive here to compare Schrödinger's result with that of Arnold Sommerfeld,[10] obtained using the rules of the old quantum theory:

$$E = mc^2\left[1 - \frac{\alpha^2}{2n^2} - \frac{\alpha^4}{2n^4}\left(\frac{n}{k} - \frac{3}{4}\right) + \cdots\right] . \tag{1.1.7}$$

where m is the electron mass. Here k is an integer between 1 and n, which in Sommerfeld's theory is given in terms of the orbital angular momentum $\ell\hbar$ as $k = \ell + 1$. This gave a fine structure splitting in agreement with experiment: for instance, for $n = 2$ Eq. (1.1.7) gives two levels ($k = 1$ and $k = 2$), split by the observed amount $\alpha^4 mc^2/32$, or 4.53×10^{-5} eV. In contrast, Schrödinger's result (1.1.6) gives an $n = 2$ fine structure splitting $\alpha^4 mc^2/12$, considerably larger than observed.

Schrödinger correctly recognized that the source of this discrepancy was his neglect of the spin of the electron. The splitting of atomic energy levels by non-inverse-square electric fields in alkali atoms and by weak external magnetic fields (the so-called anomalous Zeeman effect) had revealed a multiplicity of states larger than could be accounted for by the Bohr–Sommerfeld theory; this led George Uhlenbeck and Samuel Goudsmit[11] in 1925 to suggest that the electron has an intrinsic angular momentum $\hbar/2$. Also, the magnitude of the Zeeman splitting[12] allowed them to estimate further that the electron has a magnetic moment

$$\mu = \frac{e\hbar}{2mc} . \tag{1.1.8}$$

It was clear that the electron's spin would be coupled to its orbital angular momentum, so that Schrödinger's relativistic equation should not be expected to give the correct fine structure splitting.

Indeed, by 1927 several authors[13] had been able to show that the spin–orbit coupling was able to account for the discrepancy between Schrödinger's result (1.1.6) and experiment. There are really two effects here: one is a direct coupling between the magnetic moment (1.1.8) and the magnetic field felt by the electron as it moves through the electrostatic field of the atom; the other is the relativistic 'Thomas precession' caused (even in the absence of a magnetic moment) by the circular motion of the spinning electron.[14] Together, these two effects were found to lift the level with total angular momentum $j = \ell + \frac{1}{2}$ to the energy (1.1.7) given by Sommerfeld for $k = \ell + 1 = j + \frac{1}{2}$, while the level with $j = \ell - \frac{1}{2}$ was lowered to the value given by Sommerfeld for $k = \ell = j + \frac{1}{2}$. Thus the energy was found to depend only on n and j, but not separately on ℓ:

$$E = mc^2 \left[1 - \frac{\alpha^2}{2n^2} - \frac{\alpha^4}{2n^4} \left(\frac{n}{j + \frac{1}{2}} - \frac{3}{4} \right) + \cdots \right] . \tag{1.1.9}$$

By accident Sommerfeld's theory had given the correct magnitude of the splitting in hydrogen ($j + \frac{1}{2}$ like k runs over integer values from 1 to n) though it was wrong as to the assignment of orbital angular momentum values ℓ to these various levels. In addition, the multiplicity of the fine structure levels in hydrogen was now predicted to be 2 for $j = \frac{1}{2}$ and $2(2j + 1)$ for $j > \frac{1}{2}$ (corresponding to ℓ values $j \pm \frac{1}{2}$), in agreement with experiment.

Despite these successes, there still was not a thorough relativistic theory which incorporated the electron's spin from the beginning. Such a theory was discovered in 1928 by Paul Dirac. However, he did not set out simply to make a relativistic theory of the spinning electron; instead, he approached the problem by posing a question that would today seem very strange. At the beginning of his 1928 paper,[15] he asks 'why Nature should have chosen this particular model for the electron, instead of being satisfied with the point charge.' To us today, this question is like asking why bacteria have only one cell: having spin $\hbar/2$ is just one of the properties that define a particle as an electron, rather than one of the many other types of particles with various spins that are known today. However, in 1928 it was possible to believe that all matter consisted of electrons, and perhaps something similar with positive charge in the atomic nucleus. Thus, in the spirit of the times in which it was asked, Dirac's question can be restated: 'Why do the fundamental constituents of matter have to have spin $\hbar/2$?'

For Dirac, the key to this question was the requirement that probabilities must be positive. It was known[16] that the probability density for the non-relativistic Schrödinger equation is $|\psi|^2$, and that this satisfies a continuity equation of the form

$$\frac{\partial}{\partial t}(|\psi|^2) - \frac{i\hbar}{2m}\nabla \cdot (\psi^*\nabla\psi - \psi\nabla\psi^*) = 0$$

so the space-integral of $|\psi|^2$ is time-independent. On the other hand, the only probability density ρ and current $\mathbf{J}$, which can be formed from solutions of the relativistic Schrödinger equation and which satisfy a conservation law,

$$\frac{\partial\rho}{\partial t} + \nabla \cdot \mathbf{J} = 0\,, \qquad (1.1.10)$$

are of the form

$$\rho = N\,\mathrm{Im}\,\psi^*\left(\frac{\partial}{\partial t} - \frac{ie\phi}{\hbar}\right)\psi\,, \qquad (1.1.11)$$

$$\mathbf{J} = N\,c^2\,\mathrm{Im}\,\psi^*\left(\nabla + \frac{ie\mathbf{A}}{\hbar c}\right)\psi\,, \qquad (1.1.12)$$

with N an arbitrary constant. It is not possible to identify ρ as the probability density, because (with or without an external potential ϕ) ρ does not have definite sign. To quote Dirac's reminiscences[17] about this problem

> I remember once when I was in Copenhagen, that Bohr asked me what I was working on and I told him I was trying to get a satisfactory relativistic theory of the electron, and Bohr said 'But Klein and Gordon have already done that!' That answer first rather disturbed me. Bohr seemed quite satisfied by Klein's solution, but I was not because of the negative probabilities that it led to. I just kept on with it, worrying about getting a theory which would have only positive probabilities.

According to George Gamow,[18] Dirac found the answer to this problem on an evening in 1928 while staring into a fireplace at St John's College, Cambridge. He realized that the reason that the Klein–Gordon (or relativistic Schrödinger) equation can give negative probabilities is that the ρ in the conservation equation (1.1.10) involves a time-derivative of the wave function. This in turn happens because the wave function satisfies a differential equation of *second* order in the time. The problem therefore was to replace this wave equation with another one of first order in time derivatives, like the non-relativistic Schrödinger equation.

Suppose the electron wave function is a multi-component quantity $\psi_n(x)$, which satisfies a wave equation of the form,

$$i\hbar \frac{\partial \psi}{\partial t} = \mathscr{H} \psi \,, \tag{1.1.13}$$

where $\mathscr{H}$ is some matrix function of space derivatives. In order to have a chance at a Lorentz-invariant theory, we must suppose that because the equation is linear in time-derivatives, it is also linear in space-derivatives, so that $\mathscr{H}$ takes the form:

$$\mathscr{H} = -i\hbar c \boldsymbol{\alpha} \cdot \nabla + \alpha_4 mc^2 \,, \tag{1.1.14}$$

where α_1, α_2, α_3, and α_4 are constant matrices. From (1.1.13) we can derive the second-order equation

$$-\hbar^2 \frac{\partial^2 \psi}{\partial t^2} = \mathscr{H}^2 \psi = -\hbar^2 c^2 \alpha_i \alpha_j \frac{\partial^2 \psi}{\partial x_i \partial x_j}$$
$$-i\hbar mc^3(\alpha_i \alpha_4 + \alpha_4 \alpha_i) \frac{\partial \psi}{\partial x_i} + m^2 c^4 \alpha_4^2 \psi \,.$$

(The summation convention is in force here; i and j run over the values 1, 2, 3, or x, y, z.) But this must agree with the free-field form of the relativistic Schrödinger equation (1.1.4), which just expresses the relativistic relation between momentum and energy. Therefore, the matrices $\boldsymbol{\alpha}$ and α_4 must satisfy the relations

$$\alpha_i \alpha_j + \alpha_j \alpha_i = 2\delta_{ij} 1 \,, \tag{1.1.15}$$

$$\alpha_i \alpha_4 + \alpha_4 \alpha_i = 0 \,, \tag{1.1.16}$$

$$\alpha_4^2 = 1 \,, \tag{1.1.17}$$

where δ_{ij} is the Kronecker delta (unity for $i = j$; zero for $i \neq j$) and 1 is the unit matrix. Dirac found a set of 4×4 matrices which satisfy these relations

$$\alpha_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad \alpha_2 = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix},$$
$$\alpha_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \qquad \alpha_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{1.1.18}$$

To show that this formalism is Lorentz-invariant, Dirac multiplied Eq. (1.1.13) on the left with α_4, so that it could be put in the form

$$\left[\hbar c \gamma^\mu \frac{\partial}{\partial x^\mu} + mc^2\right] \psi = 0 \,, \tag{1.1.19}$$

where

$$\boldsymbol{\gamma} \equiv -i\alpha_4 \boldsymbol{\alpha} \,, \qquad \gamma^0 \equiv -i\alpha_4 \,. \tag{1.1.20}$$

(The Greek indices μ, ν, etc. will now run over the values 1, 2, 3, 0, with $x^0 = ct$. Dirac used $x_4 = ict$, and correspondingly $\gamma_4 = \alpha_4$.) The matrices γ^μ satisfy the anticommutation relations

$$\frac{1}{2}(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu) = \eta^{\mu\nu} \equiv \begin{cases} +1 & \mu = \nu = 1, 2, 3 \\ -1 & \mu = \nu = 0 \\ 0 & \mu \neq \nu \end{cases} \,. \tag{1.1.21}$$

Dirac noted that these anticommutation relations are Lorentz-invariant, in the sense that they are also satisfied by the matrices $\Lambda^\mu{}_\nu \gamma^\nu$, where Λ is any Lorentz transformation. He concluded from this that $\Lambda^\mu{}_\nu \gamma^\nu$ must be related to γ^μ by a similarity transformation:

$$\Lambda^\mu{}_\nu \gamma^\nu = S^{-1}(\Lambda) \gamma^\mu S(\Lambda) \,.$$

It follows that the wave equation is invariant if, under a Lorentz transformation $x^\mu \to \Lambda^\mu{}_\nu x^\nu$, the wave function undergoes the matrix transformation $\psi \to S(\Lambda)\psi$. (These matters are discussed more fully, from a rather different point of view, in Chapter 5.)

To study the behavior of electrons in an arbitrary external electromagnetic field, Dirac followed the 'usual procedure' of making the replacements

$$i\hbar \frac{\partial}{\partial t} \to i\hbar \frac{\partial}{\partial t} + e\phi \qquad -i\hbar \nabla \to -i\hbar \nabla + \frac{e}{c}\mathbf{A} \tag{1.1.22}$$

as in Eq. (1.1.4). The wave equation (1.1.13) then takes the form

$$\left(i\hbar \frac{\partial}{\partial t} + e\phi\right) \psi = (-i\hbar c \nabla + e\mathbf{A}) \cdot \boldsymbol{\alpha} \psi + mc^2 \alpha_4 \psi \,. \tag{1.1.23}$$

Dirac used this equation to show that in a central field, the conservation of angular momentum takes the form

$$[\mathscr{H}, \; -i\hbar \mathbf{r} \times \nabla + \hbar \boldsymbol{\sigma}/2] = 0 \,, \tag{1.1.24}$$

where $\mathscr{H}$ is the matrix differential operator (1.1.14) and $\boldsymbol{\sigma}$ is the 4×4 version of the spin matrix introduced earlier by Pauli[19]

$$\boldsymbol{\sigma} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \boldsymbol{\alpha} \,. \tag{1.1.25}$$

Since each component of $\boldsymbol{\sigma}$ has eigenvalues ± 1, the presence of the extra term in (1.1.24) shows that the electron has intrinsic angular momentum $\hbar/2$.

Dirac also iterated Eq. (1.1.23), obtaining a second-order equation, which turned out to have just the same form as the Klein–Gordon equation (1.1.4) except for the presence on the right-hand-side of two additional terms

$$[-e\hbar c\boldsymbol{\sigma}\cdot\mathbf{B} - ie\hbar c\boldsymbol{\alpha}\cdot\mathbf{E}]\,\psi\,. \tag{1.1.26}$$

For a slowly moving electron, the first term dominates, and represents a magnetic moment in agreement with the value (1.1.8) found by Goudsmit and Uhlenbeck.[11] As Dirac recognized, this magnetic moment, together with the relativistic nature of the theory, guaranteed that this theory would give a fine structure splitting in agreement (to order $\alpha^4 mc^2$) with that found by Heisenberg, Jordan, and Charles G. Darwin.[13] A little later, an 'exact' formula for the hydrogen energy levels in Dirac's theory was derived by Darwin[20] and Gordon[21]

$$E = mc^2\left(1 + \frac{\alpha^2}{\left\{n - j - \frac{1}{2} + \left[\left(j+\frac{1}{2}\right)^2 - \alpha^2\right]^{\frac{1}{2}}\right\}^2}\right)^{-1/2}\,. \tag{1.1.27}$$

The first three terms of a power series expansion in α^2 agree with the approximate result (1.1.9).

This theory achieved Dirac's primary aim: a relativistic formalism with positive probabilities. From (1.1.13) we can derive a continuity equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{J} = 0 \tag{1.1.28}$$

with

$$\rho = |\psi|^2\,, \qquad \mathbf{J} = c\psi^\dagger\boldsymbol{\alpha}\psi\,, \tag{1.1.29}$$

so that the positive quantity $|\psi|^2$ can be interpreted as a probability density, with constant total probability $\int |\psi|^2 d^3x$. However, there was another difficulty which Dirac was not immediately able to resolve.

For a given momentum $\mathbf{p}$, the wave equation (1.1.13) has *four* solutions of the plane wave form

$$\psi \propto \exp\left[\frac{i}{\hbar}(\mathbf{p}\cdot\mathbf{x} - Et)\right]\,. \tag{1.1.30}$$

Two solutions with $E = +\sqrt{\mathbf{p}^2c^2 + m^2c^4}$ correspond to the two spin states of an electron with $J_z = \pm\hbar/2$. The other two solutions have $E =$

$-\sqrt{\mathbf{p}^2c^2 + m^2c^4}$, and no obvious physical interpretation. As Dirac pointed out, this problem arises also for the relativistic Schrödinger equation: for each $\mathbf{p}$, there are two solutions of the form (1.1.30), one with positive E and one with negative E.

Of course, even in classical physics, the relativistic relation $E^2 = \mathbf{p}^2c^2 + m^2c^4$ has two solutions, $E = \pm\sqrt{\mathbf{p}^2c^2 + m^2c^4}$. However, in classical physics we can simply assume that the only physical particles are those with positive E. Since the positive solutions have $E > mc^2$ and the negative ones have $E < -mc^2$, there is a finite gap between them, and no continuous process can take a particle from positive to negative energy.

The problem of negative energies is much more troublesome in relativistic quantum mechanics. As Dirac pointed out in his 1928 paper,[15] the interaction of electrons with radiation can produce transitions in which a positive-energy electron falls into a negative-energy state, with the energy carried off by two or more photons. Why then is matter stable?

In 1930 Dirac offered a remarkable solution.[22] Dirac's proposal was based on the exclusion principle, so a few words about the history of this principle are in order here.

The periodic table of the elements and the systematics of X-ray spectroscopy had together by 1924 revealed a pattern in the population of atomic energy levels by electrons:[23] The maximum number N_n of electrons in a shell characterized by principal quantum number n is given by twice the number of orbital states with that n

$$N_n = 2\sum_{\ell=0}^{n-1}(2\ell + 1) = 2n^2 = 2, 8, 18, \ldots \; . \tag{1.1.31}$$

Wolfgang Pauli[24] in 1925 suggested that this pattern could be understood if N_n is the total number of possible states in the nth shell, and if in addition there is some mysterious 'exclusion principle' which forbids more than one electron from occupying the same state. He explained the puzzling factor 2 in (1.1.31) as due to a 'peculiar, classically non-describable duplexity' of the electron states, and as we have seen this was understood a little later as due to the spin of the electron.[11] The exclusion principle answered a question that had remained obscure in the old atomic theory of Bohr and Sommerfeld: why do not all the electrons in heavy atoms fall down into the shell of lowest energy? Subsequently Pauli's exclusion principle was formalized by a number of authors[25] as the requirement that the wave function of a multi-electron system is antisymmetric in the coordinates, orbital and spin, of all the electrons. This principle was incorporated into statistical mechanics by Enrico Fermi[26] and Dirac,[27] and for this reason particles obeying the exclusion principle are generally called 'fermions,' just as particles like photons for which the wave function is symmetric and which obey the statistics of Bose and Einstein are called 'bosons.' The exclusion principle has played a fundamental role in the theory of metals, white dwarf and neutron stars, etc., as well as in chemistry and atomic physics, but a discussion of these matters would take us too far afield here.

Dirac's proposal was that the positive energy electrons cannot fall down into negative energy states because 'all the states of negative energy are occupied except perhaps a few of small velocity.' The few vacant states, or 'holes,' in the sea of negative energy electrons behave like particles with opposite quantum numbers: positive energy and positive charge. The only particle with positive charge that was known at that time was the proton, and as Dirac later recalled,[27a] 'the whole climate of opinion at that time was against new particles' so Dirac identified his holes as protons; in fact, the title of his 1930 article[22] was 'A Theory of Electrons and Protons.'

The hole theory faced a number of immediate difficulties. One obvious problem was raised by the infinite charge density of the ubiquitous negative-energy electrons: where is their electric field? Dirac proposed to reinterpret the charge density appearing in Maxwell's equations as 'the departure from the normal state of electrification of the world.' Another problem has to do with the huge dissimilarity between the observed masses and interactions of the electrons and protons. Dirac hoped that Coulomb interactions between electrons would somehow account for these differences but Hermann Weyl[28] showed that the hole theory was in fact entirely symmetric between negative and positive charge. Finally, Dirac[22] predicted the existence of an electron–proton annihilation process in which a positive-energy electron meets a hole in the sea of negative-energy electrons and falls down into the unoccupied level, emitting a pair of gamma ray photons. By itself this would not have created difficulties for the hole theory; it was even hoped by some that this would provide an explanation, then lacking, of the energy source of the stars. However, it was soon pointed out[29] by Julius Robert Oppenheimer and Igor Tamm that electron–proton annihilation in atoms would take place at much too fast a rate to be consistent with the observed stability of ordinary matter. For these reasons, by 1931 Dirac had changed his mind, and decided that the holes would have to appear not as protons but as a new sort of positively charged particle, of the same mass as the electron.[29a]

The second and third of these problems were eliminated by the discovery of the positron by Carl D. Anderson,[30] who apparently did not know of this prediction by Dirac. On August 2, 1932, a peculiar cosmic ray track was observed in a Wilson cloud chamber subjected to a 15 kG magnetic field. The track was observed to curve in a direction that would be expected for a *positively* charged particle, and yet its range was at least ten times greater than the expected range of a proton! Both the range and the specific ionization of the track were consistent with the hypothesis that this was a new particle which differs from the electron only in the sign of its charge, as would be expected for one of Dirac's holes. (This discovery had been made earlier by P.M.S. Blackett, but not immediately published by him. Anderson quotes press reports of evidence for light positive particles in cosmic ray tracks, obtained by Blackett and Giuseppe Occhialini.) Thus it appeared that Dirac was wrong only in his original identification of the hole with the proton.

The discovery of the more-or-less predicted positron, together with the earlier successes of the Dirac equation in accounting for the magnetic moment of the electron and the fine structure of hydrogen, gave Dirac's theory a prestige that it has held for over six decades. However, although there seems little doubt that Dirac's theory will survive in some form in any future physical theory, there are serious reasons for being dissatisfied with its original rationale:

(i) Dirac's analysis of the problem of negative probabilities in Schrödinger's relativistic wave equation would seem to rule out the existence of any particle of zero spin. Yet even in the 1920s particles of zero spin were known — for instance, the hydrogen atom in its ground state, and the helium nucleus. Of course, it could be argued that hydrogen atoms and alpha particles are not elementary, and therefore do not need to be described by a relativistic wave equation, but it was not (and still is not) clear how the idea of elementarity is incorporated in the formalism of relativistic quantum mechanics. Today we know of a large number of spin zero particles — π mesons, K mesons, and so on — that are no less elementary than the proton and neutron. We also know of spin one particles — the $W^\pm$ and Z^0 — which seem as elementary as the electron or any other particle. Further, apart from effects of the strong interactions, we would today calculate the fine structure of 'mesonic atoms,' consisting of a spinless negative π or K meson bound to an atomic nucleus, from the stationary solutions of the relativistic Klein–Gordon–Schrödinger equation! Thus, it is difficult to agree that there is anything fundamentally wrong with the relativistic equation for zero spin that *forced* the development of the Dirac equation — the problem simply is that the electron happens to have spin $\hbar/2$, not zero.

(ii) As far as we now know, for *every* kind of particle there is an 'antiparticle' with the same mass and opposite charge. (Some purely neutral particles, such as the photon, are their own antiparticles.) But how can we interpret the antiparticles of charged *bosons*, such as the $\pi^\pm$ mesons or $W^\pm$ particles, as holes in a sea of negative energy states? For particles quantized according to the rules of Bose–Einstein statistics, there is no exclusion principle, and hence nothing to keep positive-energy particles from falling down into the negative-energy states, occupied or not. And if the hole theory does not work for bosonic antiparticles, why should we believe it for fermions? I asked Dirac in 1972 how he then felt about this point; he told me that he did not regard bosons like the pion or $W^\pm$ as 'important.' In a lecture[27a] a few years later, Dirac referred to the fact that for bosons 'we no longer have the picture of a vacuum with negative energy states filled up', and remarked that in this case 'the whole theory becomes more complicated.' The next section will show how the development of quantum field theory made the interpretation of antiparticles as holes unnecessary, even though unfortunately it lingers on in many textbooks. To quote Julian Schwinger,[30a] 'The picture of an infinite sea of negative energy electrons is now best regarded as a historical curiosity, and forgotten.'

(iii) One of the great successes of the Dirac theory was its correct prediction of the magnetic moment of the electron. This was particularly striking, as the magnetic moment (1.1.8) is twice as large as would be expected for the orbital motion of a charged point particle with angular momentum $\hbar/2$; this factor of 2 had remained mysterious until Dirac's theory. However, there is really nothing in Dirac's line of argument that leads unequivocally to this particular value for the magnetic moment. At the point where we brought electric and magnetic fields into the wave equation (1.1.23), we could just as well have added a 'Pauli term'[31]

$$\kappa\alpha_4[\gamma^\mu,\gamma^\nu]\psi F_{\mu\nu} \tag{1.1.32}$$

with arbitrary coefficient κ. (Here $F_{\mu\nu}$ is the usual electromagnetic field strength tensor, with $F^{12} = B_3$, $F^{01} = E_1$, etc.) This term could be obtained by first adding a term to the free-field equations proportional to $[\gamma^\mu,\gamma^\nu](\partial^2/\partial x^\mu \partial x^\nu)\psi$, which of course equals zero, and then making the substitutions (1.1.22) as before. A more modern approach would be simply to remark that the term (1.1.32) is consistent with all accepted invariance principles, including Lorentz invariance and gauge invariance, and so there is no reason why such a term should *not* be included in the field equations. (See Section 12.3.) This term would give an additional contribution proportional to κ to the magnetic moment of the electron, so apart from the possible demand for a purely formal simplicity, there was no reason to expect any particular value for the magnetic moment of the electron in Dirac's theory.

As we shall see in this book, these problems were all eventually to be solved (or at least clarified) through the development of quantum field theory.

1.2 The Birth of Quantum Field Theory

The photon is the only particle that was known as a field before it was detected as a particle. Thus it is natural that the formalism of quantum field theory should have been developed in the first instance in connection with radiation and only later applied to other particles and fields.

In 1926, in one of the central papers on matrix mechanics, Born, Heisenberg, and Jordan[32] applied their new methods to the free radiation field. For simplicity, they ignored the polarization of electromagnetic waves and worked in one space dimension, with coordinate x running from 0 to L; the radiation field $u(x,t)$ if constrained to vanish at these endpoints thus has the same behavior as the displacement of a string with ends fixed at $x = 0$ and $x = L$. By analogy with either the case of a string or the full electromagnetic field, the Hamiltonian was taken to have the form

$$H = \frac{1}{2}\int_0^L \left\{ \left(\frac{\partial u}{\partial t}\right)^2 + c^2\left(\frac{\partial u}{\partial x}\right)^2 \right\} dx . \tag{1.2.1}$$

In order to reduce this expression to a sum of squares, the field u was expressed as a sum of Fourier components with $u = 0$ at both $x = 0$ and $x = L$:

$$u(x,t) = \sum_{k=1}^{\infty} q_k(t) \sin\left(\frac{\omega_k x}{c}\right) , \tag{1.2.2}$$

$$\omega_k \equiv k\pi c/L , \tag{1.2.3}$$

so that

$$H = \frac{L}{4}\sum_{k=1}^{\infty} \left\{ \dot{q}_k^2(t) + \omega_k^2 q_k^2(t) \right\} . \tag{1.2.4}$$

Thus the string or field behaves like sum of independent harmonic oscillators with angular frequencies ω_k, as had been anticipated 20 years earlier by Paul Ehrenfest.[32a]

In particular, the 'momentum' $p_k(t)$ canonically conjugate to $q_k(t)$ is determined, as in particle mechanics, by the condition that if H is expressed as a function of the ps and qs, then

$$\dot{q}_k(t) = \frac{\partial}{\partial p_k(t)} H(p(t), q(t)) .$$

This yields a 'momentum'

$$p_k(t) = \frac{L}{2}\,\dot{q}_k(t) \tag{1.2.5}$$

so the canonical commutation relations may be written

$$\left[\dot{q}_k(t), q_j(t)\right] = \frac{2}{L}\left[p_k(t), q_j(t)\right] = \frac{-2i\hbar}{L}\,\delta_{kj}\,, \tag{1.2.6}$$

$$\left[q_k(t), q_j(t)\right] = 0\,. \tag{1.2.7}$$

Also, the time-dependence of $q_k(t)$ is governed by the Hamiltonian equation of motion

$$\ddot{q}_k(t) = \frac{2}{L}\,\dot{p}_k(t) = -\frac{2}{L}\,\frac{\partial H}{\partial q_k(t)} = -\omega_k^2 q_k(t)\,. \tag{1.2.8}$$

The form of the matrices defined by Eqs. (1.2.6)–(1.2.8) was already known to Born, Heisenberg, and Jordan through previous work on the harmonic oscillator. The q-matrix is given by

$$q_k(t) = \sqrt{\frac{\hbar}{L\omega_k}}\,\left[a_k \exp(-i\omega_k t) + a_k^\dagger \exp(+i\omega_k t)\right] \tag{1.2.9}$$

with a_k a time-independent matrix and $a_k^\dagger$ its Hermitian adjoint, satisfying the commutation relations

$$\left[\,a_k\,,\,a_j^\dagger\,\right] = \delta_{kj}\,, \tag{1.2.10}$$

$$\left[\,a_k\,,\,a_j\,\right] = 0\,. \tag{1.2.11}$$

The rows and columns of these matrices are labelled with a set of positive integers n_1, $n_2, \ldots$, one for each normal mode. The matrix elements are

$$(a_k)_{n_1',n_2',\ldots,n_1,n_2\ldots} = \sqrt{n_k}\,\delta_{n_k',n_k-1}\prod_{j\neq k}\delta_{n_j',n_j}\,, \tag{1.2.12}$$

$$(a_k^\dagger)_{n_1',n_2',\ldots,n_1,n_2\ldots} = \sqrt{n_k+1}\,\delta_{n_k',n_k+1}\prod_{j\neq k}\delta_{n_j',n_j}\,. \tag{1.2.13}$$

For a single normal mode, these matrices may be written explicitly as

$$a = \begin{bmatrix} 0 & \sqrt{1} & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{2} & 0 & \cdots \\ 0 & 0 & 0 & \sqrt{3} & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}, \quad a^\dagger = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ \sqrt{1} & 0 & 0 & 0 & \cdots \\ 0 & \sqrt{2} & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{3} & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}.$$

It is straightforward to check that (1.2.12) and (1.2.13) do satisfy the commutation relations (1.2.10) and (1.2.11).

The physical interpretation of a column vector with integer components $n_1, n_2, \ldots$ is that it represents a state with n_k quanta in each normal mode k. The matrix a_k or $a_k^\dagger$ acting on such a column vector will respectively lower or raise n_k by one unit, leaving all n_ℓ with $\ell \neq k$ unchanged; they may therefore be interpreted as operators which annihilate or create one quantum in the kth normal mode. In particular, the vector with all n_k equal to zero represents the vacuum; it is annihilated by any a_k.

This interpretation is further borne out by inspection of the Hamiltonian. Using (1.2.9) and (1.2.10) in (1.2.4) gives

$$H = \sum_k \hbar\omega_k \left(a_k^\dagger a_k + \tfrac{1}{2}\right) . \tag{1.2.14}$$

The Hamiltonian is then diagonal in the n-representation

$$(H)_{n'_1,n'_2,\ldots,n_1,n_2\ldots} = \sum_k \hbar\omega_k \left(n_k + \tfrac{1}{2}\right) \prod_j \delta_{n'_j n_j} . \tag{1.2.15}$$

We see that the energy of the state is just the sum of energies $\hbar\omega_k$ for each quantum present in the state, plus an infinite zero-point energy $E_0 = \frac{1}{2}\sum_k \hbar\omega_k$. Applied to the radiation field, this formalism justified the Bose method of counting radiation states according to the numbers n_k of quanta in each normal mode.

Born, Heisenberg, and Jordan used this formalism to derive an expression for the r.m.s. energy fluctuations in black-body radiation. (For this purpose they actually only used the commutation relations (1.2.6)–(1.2.7).) However, this approach was soon applied to a more urgent problem, the calculation of the rates for spontaneous emission of radiation.

In order to appreciate the difficulties here, it is necessary to go back in time a bit. In one of the first papers on matrix mechanics, Born and Jordan[33] had assumed in effect that an atom, in dropping from a state β to a lower state α, would emit radiation just like a classical charged oscillator with displacement

$$\mathbf{r}(t) = \mathbf{r}_{\beta\alpha} \exp(-2\pi i\nu t) + \mathbf{r}_{\beta\alpha}{}^* \exp(2\pi i\nu t) , \tag{1.2.16}$$

where

$$h\nu = E_\beta - E_\alpha \tag{1.2.17}$$

and $\mathbf{r}_{\beta\alpha}$ is the β,α element of the matrix associated with the electron position. The energy E of such an oscillator is

$$E = \frac{1}{2}m\left(\dot{\mathbf{r}}^2 + (2\pi\nu)^2\mathbf{r}^2\right) = 8\pi^2 m\nu^2 |\mathbf{r}_{\beta\alpha}|^2 . \tag{1.2.18}$$

A straightforward classical calculation then gives the radiated power, and dividing by the energy $h\nu$ per photon gives the rate of photon emission

$$A(\beta \to \alpha) = \frac{16\pi^3 e^2 \nu^3}{3hc^3} |\mathbf{r}_{\beta\alpha}|^2 . \tag{1.2.19}$$

However, it was not at all clear why the formulas for emission of radiation by a classical dipole should be taken over in this manner in dealing with spontaneous emission.

A little later a more convincing though even less direct derivation was given by Dirac.[34] By considering the behavior of quantized atomic states in an oscillating *classical* electromagnetic field with energy density per frequency interval u at frequency (1.2.17), he was able to derive formulas for the rates $uB(\alpha \to \beta)$ and $uB(\beta \to \alpha)$ for absorption or induced emission:

$$B(\alpha \to \beta) = B(\beta \to \alpha) \simeq \frac{2\pi^2 e^2}{3h^2} |\mathbf{r}_{\beta\alpha}|^2 . \tag{1.2.20}$$

(Note that the expression on the right is symmetric between states α and β, because $\mathbf{r}_{\alpha\beta}$ is just $\mathbf{r}_{\beta\alpha}{}^*$.) Einstein[34a] had already shown in 1917 that the possibility of thermal equilibrium between atoms and black-body radiation imposes a relation between the rate $A(\beta \to \alpha)$ of spontaneous emission and the rates uB for induced emission or absorption:

$$A(\beta \to \alpha) = \left(\frac{8\pi h\nu^3}{c^3}\right) B(\beta \to \alpha) . \tag{1.2.21}$$

Using (1.2.20) in this relation immediately yields the Born–Jordan result (1.2.19) for the rate of spontaneous emission. Nevertheless, it still seemed unsatisfactory that thermodynamic arguments should be needed to derive formulas for processes involving a single atom.

Finally, in 1927 Dirac[35] was able to give a thoroughly quantum mechanical treatment of spontaneous emission. The vector potential $\mathbf{A}(\mathbf{x}, t)$ was expanded in normal modes, as in Eq. (1.2.2), and the coefficients were shown to satisfy commutation relations like (1.2.6). In consequence, each state of the free radiation field was specified by a set of integers n_k, one for each normal mode, and the matrix elements of the electromagnetic interaction $e\dot{\mathbf{r}} \cdot \mathbf{A}$ took the form of a sum over normal modes, with matrix coefficients proportional to the matrices a_k and $a_k^\dagger$ defined in Eqs. (1.2.10)–(1.2.13). The crucial result here is the factor $\sqrt{n_k + 1}$ in Eq. (1.2.13); the probability for a transition in which the number of photons in a normal mode k rises from n_k to $n_k + 1$ is proportional to the square of this factor, or $n_k + 1$. But in a radiation field with n_k photons in a normal mode k, the energy density u per frequency interval is

$$u(\nu_k) = \left(\frac{8\pi\nu_k^2}{c^3}\right) n_k \times h\nu_k ,$$

so the rate for emission of radiation in normal mode k is proportional to

$$n_k + 1 = \frac{c^3 u(\nu_k)}{8\pi h\nu_k^3} + 1 .$$

The first term is interpreted as the contribution of induced emission, and the second term as the contribution of spontaneous emission. Hence, without any appeal to thermodynamics, Dirac could conclude that the ratio of the rates uB for induced emission and A for spontaneous emission is given by the Einstein relation, Eq. (1.2.21). Using his earlier result (1.2.20) for B, Dirac was thus able to rederive the Born–Jordan formula[33] (1.2.19) for spontaneous emission rate A. A little later, similar methods were used by Dirac to give a quantum mechanical treatment of the scattering of radiation and the lifetime of excited atomic states,[36] and by Victor Weisskopf and Eugene Wigner to make a detailed study of spectral line shapes.[36a] Dirac in his work was separating the electromagnetic potential into a radiation field **A** and a static Coulomb potential A^0, in a manner which did not preserve the manifest Lorentz and gauge invariance of classical electrodynamics. These matters were put on a firmer foundation a little later by Enrico Fermi.[36b] Many physicists in the 1930s learned their quantum electrodynamics from Fermi's 1932 review.

The use of canonical commutation relations for q and p or a and $a^\dagger$ also raised a question as to the Lorentz invariance of the quantized theory. Jordan and Pauli[37] in 1928 were able to show that the commutators of fields at different spacetime points were in fact Lorentz-invariant. (These commutators are calculated in Chapter 5.) Somewhat later, Bohr and Leon Rosenfeld[38] used a number of ingenious thought experiments to show that these commutation relations express limitations on our ability to measure fields at spacetime points separated by time-like intervals.

It was not long after the successful quantization of the electromagnetic field that these techniques were applied to other fields. At first this was regarded as a 'second quantization'; the fields to be quantized were the wave functions used in one-particle quantum mechanics, such as the Dirac wave function of the electron. The first step in this direction seems to have been taken in 1927 by Jordan.[39] In 1928 an essential element was supplied by Jordan and Wigner.[40] They recognized that the Pauli exclusion principle prevents the occupation number n_k of electrons in any normal mode k (counting spin as well as position variables) from taking any values other than 0 or 1. The electron field therefore cannot be expanded as a superposition of operators satisfying the commutation relations (1.2.10), (1.2.11), because these relations require n_k to take all integer values from 0 to ∞. Instead, they proposed that the electron field should be expanded in a sum of operators a_k, $a_k^\dagger$ satisfying the *anticommutation* relations

$$a_k a_j^\dagger + a_j^\dagger a_k = \delta_{jk} , \tag{1.2.22}$$

$$a_k a_j + a_j a_k = 0 . \tag{1.2.23}$$

The relations can be satisfied by matrices labelled by a set of integers

$n_1, n_2, \cdots$, one for each normal mode, each integer taking just the values zero and one:

$$(a_k)_{n'_1,n'_2,\ldots,n_1,n_2,\ldots} = \begin{cases} 1 & n'_k = 0,\, n_k = 1,\, n'_j = n_j \text{ for } j \neq k \\ 0 & \text{otherwise}, \end{cases} \tag{1.2.24}$$

$$(a_k^\dagger)_{n'_1,n'_2,\ldots,n_1,n_2,\ldots} = \begin{cases} 1 & n'_k = 1,\, n_k = 0,\, n'_j = n_j \text{ for } j \neq k \\ 0 & \text{otherwise}. \end{cases} \tag{1.2.25}$$

For instance, for a single normal mode we have just two rows and two columns, corresponding to the values unity and zero of n' and n; the a and $a^\dagger$ matrices take the form

$$a = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \qquad a^\dagger = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The reader may check that (1.2.24) and (1.2.25) do satisfy the anticommutation relations (1.2.22) and (1.2.23).

The interpretation of a column vector characterized by integers $n_1, n_2, \ldots$ is that it represents a state with n_k quanta in each normal mode k, just as for bosons. The difference is, of course, that since each n_k takes only the values 0 and 1, there can be at most one quantum in each normal mode, as required by the Pauli exclusion principle. Again, a_k destroys a quantum in normal mode k if there is one there already, and otherwise gives zero; also, $a_k^\dagger$ creates a quantum in normal mode k *unless* there is one there already, in which case it gives zero. Much later it was shown by Fierz and Pauli[40a] that the choice between commutation and anticommutation relations is dictated solely by the particle's spin: commutators must be used for particles with integer spin like the photon, and anticommutators for particles with half-integer spin like the electron. (This will be shown in a different way in Chapter 5.)

The theory of general quantum fields was first laid out in 1929, in a pair of comprehensive articles by Heisenberg and Pauli.[41] The starting point of their work was the application of the canonical formalism to the fields themselves, rather than to the coefficients of the normal modes appearing in the fields. Heisenberg and Pauli took the Lagrangian L as the space-integral of a local function of fields and spacetime derivatives of fields; the field equations were then determined from the principle that the action $\int L\, dt$ should be stationary when the fields are varied; and the commutation relations were determined from the assumption that the variational derivative of the Lagrangian with respect to any field's time-derivative behaves like a 'momentum' conjugate to that field (except that commutation relations become anticommutation relations for fermion fields). They also went on to apply this general formalism to the electromagnetic and Dirac fields, and explored the various invariance and conservation laws, including the conservation of charge, momentum, and energy, and Lorentz and gauge invariance.

The Heisenberg–Pauli formalism is essentially the same as that described in our Chapter 7, and so for the present we can limit ourselves to a single example which will turn out to be useful later in this section. For a free complex scalar field $\phi(x)$ the Lagrangian is taken as

$$L = \int d^3x \left[\dot{\phi}^\dagger \dot{\phi} - c^2 (\nabla\phi)^\dagger \cdot (\nabla\phi) - \left(\frac{mc^2}{\hbar}\right)^2 \phi^\dagger \phi \right] . \tag{1.2.26}$$

If we subject $\phi(x)$ to an infinitesimal variation $\delta\phi(x)$, the Lagrangian is changed by the amount

$$\delta L = \int d^3x \Bigg[\dot{\phi}^\dagger \delta\dot{\phi} + \dot{\phi}\delta\dot{\phi}^\dagger - c^2 \nabla\phi^\dagger \cdot \nabla\delta\phi - c^2 \nabla\phi \cdot \nabla\delta\phi^\dagger - \left(\frac{mc^2}{\hbar}\right)^2 \phi^\dagger \delta\phi - \left(\frac{mc^2}{\hbar}\right)^2 \phi\delta\phi^\dagger \Bigg] . \tag{1.2.27}$$

It is assumed in using the principle of stationary action that the variation in the fields should vanish on the boundaries of the spacetime region of integration. Thus, in computing the change in the action $\int L\,dt$, we can immediately integrate by parts, and write

$$\delta \int L\,dt = c^2 \int d^4x \left[\delta\phi^\dagger \left(\Box - \left(\frac{mc}{\hbar}\right)^2 \right) \phi + \delta\phi \left(\Box - \left(\frac{mc}{\hbar}\right)^2 \right) \phi^\dagger \right] .$$

But this must vanish for any $\delta\phi$ and $\delta\phi^\dagger$, so ϕ must satisfy the familiar relativistic wave equation

$$\left[\Box - \left(\frac{mc}{\hbar}\right)^2 \right] \phi = 0 \tag{1.2.28}$$

and its adjoint. The ‘momenta’ canonically conjugate to the fields ϕ and $\phi^\dagger$ are given by the variational derivatives of L with respect to $\dot{\phi}$ and $\dot{\phi}^\dagger$, which we can read off from (1.2.27) as

$$\pi \equiv \frac{\delta L}{\delta\dot{\phi}} = \dot{\phi}^\dagger , \tag{1.2.29}$$

$$\pi^\dagger \equiv \frac{\delta L}{\delta\dot{\phi}^\dagger} = \dot{\phi} . \tag{1.2.30}$$

These field variables satisfy the usual canonical commutation relations, with a delta function in place of a Kronecker delta

$$\left[\pi(\mathbf{x},t),\phi(\mathbf{y},t)\right] = \left[\pi^\dagger(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = -i\hbar\delta^3(\mathbf{x}-\mathbf{y})\,, \tag{1.2.31}$$

$$\left[\pi(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = \left[\pi^\dagger(\mathbf{x},t),\phi(\mathbf{y},t)\right] = 0\,, \tag{1.2.32}$$

$$\left[\pi(\mathbf{x},t),\pi(\mathbf{y},t)\right] = \left[\pi^\dagger(\mathbf{x},t),\pi^\dagger(\mathbf{y},t)\right] = \left[\pi(\mathbf{x},t),\pi^\dagger(\mathbf{y},t)\right] = 0\,, \tag{1.2.33}$$

$$\left[\phi(\mathbf{x},t),\phi(\mathbf{y},t)\right] = \left[\phi^\dagger(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = \left[\phi(\mathbf{x},t),\phi^\dagger(\mathbf{y},t)\right] = 0\,. \tag{1.2.34}$$

The Hamiltonian here is given (just as in particle mechanics) by the 'sum' of all canonical momenta times the time-derivatives of the corresponding fields, minus the Lagrangian:

$$H = \int d^3x \left[\pi\dot{\phi} + \pi^\dagger\dot{\phi}^\dagger\right] - L \tag{1.2.35}$$

or, using (1.2.26), (1.2.29), and (1.2.30):

$$H = \int d^3x \left[\pi^\dagger\pi + c^2(\nabla\phi)^\dagger\cdot(\nabla\phi) + \left(\frac{m^2c^4}{\hbar^2}\right)\phi^\dagger\phi\right]\,. \tag{1.2.36}$$

After the papers by Heisenberg and Pauli one element was still missing before quantum field theory could reach its final pre-war form: a solution to the problem of the negative-energy states. We saw in the last section that in 1930, at just about the time of the Heisenberg–Pauli papers, Dirac had proposed that the negative-energy states of the electron were all filled, but with only the holes in the negative-energy sea observable, rather than the negative-energy electrons themselves. After Dirac's idea was seemingly confirmed by the discovery of the positron in 1932, his 'hole theory' was used to calculate a number of processes to the lowest order of perturbation theory, including electron–positron pair production and scattering.

At the same time, a great deal of work was put into the development of a formalism whose Lorentz invariance would be explicit. The most influential effort was the 'many-time' formalism of Dirac, Vladimir Fock, and Boris Podolsky,[42] in which the state vector was represented by a wave function depending on the spacetime and spin coordinates of all electrons, positive-energy and negative-energy. In this formalism, the total number of electrons of either positive or negative energy is conserved; for instance, production of an electron–positron pair is described as the excitation of a negative-energy electron to a positive-energy state, and the annihilation of an electron and positron is described as the corresponding deexcitation. This many-time formalism had the advantage of manifest Lorentz invariance, but it had a number of disadvantages: In particular, there was a profound difference between the treatment of the photon, described in terms of a quantized electromagnetic field, and that of the electron and positron. Not all physicists felt this to be a disadvantage;

the electron field unlike the electromagnetic field did not have a classical limit, so there were doubts about its physical significance. Also, Dirac[42a] conceived of fields as the means by which we observe particles, so that he did not expect particles and fields to be described in the same terms. Though I do not know whether it bothered anyone at the time, there was a more practical disadvantage of the many-time formalism: it would have been difficult to use it to describe a process like nuclear beta decay, in which an electron and antineutrino are created without an accompanying positron or neutrino. The successful calculation by Fermi[43] of the electron energy distribution in beta decay deserves to be counted as one of the early triumphs of quantum field theory.

The essential idea that was needed to demonstrate the equivalence of the Dirac hole theory with a quantum field theory of the electron was provided by Fock[43a] and by Wendell Furry and Oppenheimer[44] in 1933–4. To appreciate this idea from a more modern standpoint, suppose we try to construct an electron field in analogy with the electromagnetic field or the Born–Heisenberg–Jordan field (1.2.2). Since electrons carry a charge, we would not like to mix annihilation and creation operators, so we might try to write the field as

$$\psi(x) = \sum_k u_k(\mathbf{x}) e^{-i\omega_k t} a_k \,, \tag{1.2.37}$$

where $u_k(\mathbf{x})e^{-i\omega_k t}$ are a complete set of orthonormal plane-wave solutions of the Dirac equation (1.1.13) (with k now labelling the three-momentum, spin, and sign of the energy):

$$\mathscr{H} u_k = \hbar\omega_k u_k \,, \tag{1.2.38}$$

$$\mathscr{H} \equiv -i\hbar c\boldsymbol{\alpha}\cdot\nabla + \alpha_4 mc^2 \,, \tag{1.2.39}$$

$$\int u_k^\dagger u_\ell \, d^3x = \delta_{k\ell} \,, \tag{1.2.40}$$

and a_k are the corresponding annihilation operators, satisfying the Jordan–Wigner anticommutation relations (1.2.22)–(1.2.23). According to the ideas of 'second quantization' or the canonical quantization procedure of Heisenberg and Pauli,[41] the Hamiltonian is formed by calculating the 'expectation value' of $\mathscr{H}$ with a 'wave function' replaced by the quantized field (1.2.37)

$$H = \int d^3x \, \psi^\dagger \mathscr{H} \psi = \sum_k \hbar\omega_k a_k^\dagger a_k \,. \tag{1.2.41}$$

The trouble is, of course, that this is not a positive operator — half the ω_k are negative while the operators $a_k^\dagger a_k$ take only the positive eigenvalues 1 and 0. (See Eqs. (1.2.24) and (1.2.25).) In order to cure this disease, Furry and Oppenheimer picked up Dirac's idea[42] that the positron is the absence of a negative-energy electron; the anticommutation relations are symmetric between creation and annihilation operators, so they defined the positron creation and annihilation operators as the corresponding annihilation and creation operators for negative-energy electrons

$$b_k^\dagger \equiv a_k\,, \qquad b_k \equiv a_k^\dagger \qquad (\text{for } \omega_k < 0) \tag{1.2.42}$$

where the label k on b denotes a positive-energy positron mode with momenta and spin opposite to those of the electron mode k. The Dirac field (1.2.37) may then be written

$$\psi(x) = \sum_k{}^{(+)} a_k u_k(x) + \sum_k{}^{(-)} b_k^\dagger u_k(x)\,, \tag{1.2.43}$$

where $(+)$ and $(-)$ indicate sums over normal modes k with $\omega_k > 0$ and $\omega_k < 0$, respectively, and $u_k(x) \equiv u_k(\mathbf{x})e^{-i\omega_k t}$. Similarly, using the anticommutation relations for the bs, we can rewrite the energy operator (1.2.41) as

$$H = \sum_k{}^{(+)} \hbar\omega_k a_k^\dagger a_k + \sum_k{}^{(-)} \hbar|\omega_k| b_k^\dagger b_k + E_0\,, \tag{1.2.44}$$

where E_0 is the infinite c-number

$$E_0 = -\sum_k{}^{(-)} \hbar|\omega_k|\,. \tag{1.2.45}$$

In order for this redefinition to be more than a mere formality, it is necessary also to specify that the physical vacuum is a state Ψ_0 containing no positive-energy electrons or positrons:

$$a_k\Psi_0 = 0 \qquad (\omega_k > 0)\,, \tag{1.2.46}$$

$$b_k\Psi_0 = 0 \qquad (\omega_k < 0)\,. \tag{1.2.47}$$

Hence (1.2.44) gives the energy of the vacuum as just E_0. If we measure all energies relative to the vacuum energy E_0, then the physical energy operator is $H - E_0$; and Eq. (1.2.44) shows that this is a *positive* operator.

The problem of negative-energy states for a charged spin zero particle was also resolved in 1934, by Pauli and Weisskopf,[45] in a paper written in part to challenge Dirac's picture of filled negative-energy states. Here the creation and annihilation operators satisfy commutation rather than anticommutation relations, so it is not possible to interchange the roles of these operators freely, as was the case for fermions. Instead we must return to the Heisenberg–Pauli canonical formalism[41] to decide which coefficients of the various normal modes are creation or annihilation operators.

Pauli and Weisskopf expanded the free charged scalar field in plane waves in a cube of spatial volume $V \equiv L^3$:

$$\phi(\mathbf{x}, t) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} q(\mathbf{k}, t) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{1.2.48}$$

with the wave numbers restricted by the periodicity condition, that the quantities $k_jL/2\pi$ for $j = 1,2,3$ should be a set of three positive or negative integers. Similarly the canonically conjugate variable (1.2.29) was expanded as

$$\pi(\mathbf{x},t) \equiv \frac{1}{\sqrt{V}}\sum_{\mathbf{k}} p(\mathbf{k},t)e^{-i\mathbf{k}\cdot\mathbf{x}} . \tag{1.2.49}$$

The minus sign is put into the exponent here so that (1.2.29) now becomes:

$$p(\mathbf{k},t) = \dot{q}^\dagger(\mathbf{k},t) . \tag{1.2.50}$$

The Fourier inversion formula gives

$$q(\mathbf{k},t) = \frac{1}{\sqrt{V}}\int d^3x\ \phi(\mathbf{x},t)e^{-i\mathbf{k}\cdot\mathbf{x}} , \tag{1.2.51}$$

$$p(\mathbf{k},t) = \frac{1}{\sqrt{V}}\int d^3x\ \pi(\mathbf{x},t)e^{+i\mathbf{k}\cdot\mathbf{x}} , \tag{1.2.52}$$

and therefore the canonical commutation relations (1.2.31)–(1.2.34) yield for the qs and ps:

$$\left[p(\mathbf{k},t), q(\mathbf{l},t)\right] = \frac{-i\hbar}{V}\int d^3x\ e^{i\mathbf{k}\cdot\mathbf{x}}e^{-i\mathbf{l}\cdot\mathbf{x}} = -i\hbar\delta_{\mathbf{kl}} \tag{1.2.53}$$

$$\begin{aligned}\left[p(\mathbf{k},t), q^\dagger(\mathbf{l},t)\right] &= \left[p(\mathbf{k},t), p(\mathbf{l},t)\right] = \left[p(\mathbf{k},t), p^\dagger(\mathbf{l},t)\right] \\ &= \left[q(\mathbf{k},t), q(\mathbf{l},t)\right] = \left[q(\mathbf{k},t), q^\dagger(\mathbf{l},t)\right] = 0\end{aligned} \tag{1.2.54}$$

together with other relations that may be derived from these by taking their Hermitian adjoints. By inserting (1.2.48) and (1.2.49) in the formula (1.2.36) for the Hamiltonian, we can also write this operator in terms of ps and qs:

$$H = \sum_{\mathbf{k}}\left[p^\dagger(\mathbf{k},t)p(\mathbf{k},t)\ +\ \omega_{\mathbf{k}}^2\, q^\dagger(\mathbf{k},t)q(\mathbf{k},t)\right] , \tag{1.2.55}$$

where

$$\omega_{\mathbf{k}}^2 \equiv c^2\mathbf{k}^2 + \left(\frac{mc^2}{\hbar}\right)^2 . \tag{1.2.56}$$

The time-derivatives of the ps are then given by the Hamiltonian equation

$$\dot{p}(\mathbf{k},t) = -\frac{\partial H}{\partial q(\mathbf{k},t)} = -\omega_{\mathbf{k}}^2\, q^\dagger(\mathbf{k},t) \tag{1.2.57}$$

(and its adjoint), a result which in the light of Eq. (1.2.50) is just equivalent to the Klein–Gordon–Schrödinger wave equation (1.2.28).

We see that, just as in the case of the 1926 model of Born, Heisenberg, and Jordan,[4] the free field behaves like an infinite number of coupled harmonic oscillators. Pauli and Weisskopf could construct p and q operators which satisfy the commutation relations (1.2.53)–(1.2.54) and the 'equations of motion' (1.2.50) and (1.2.57), by introducing annihilation and creation operators a, b, $a^\dagger$, $b^\dagger$ of *two* different kinds, corresponding to particles and antiparticles:

$$q(\mathbf{k},t) = i\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}}}\left[a(\mathbf{k})\exp(-i\omega_{\mathbf{k}}t) - b^\dagger(\mathbf{k})\exp(i\omega_{\mathbf{k}}t)\right] \tag{1.2.58}$$

$$p(\mathbf{k},t) = \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2}}\left[b(\mathbf{k})\exp(-i\omega_{\mathbf{k}}t) + a^\dagger(\mathbf{k})\exp(+i\omega_{\mathbf{k}}t)\right] \tag{1.2.59}$$

where

$$\left[a(\mathbf{k}), a^\dagger(\mathbf{l})\right] = \left[b(\mathbf{k}), b^\dagger(\mathbf{l})\right] = \delta_{\mathbf{kl}}\,, \tag{1.2.60}$$

$$\left[a(\mathbf{k}), a(\mathbf{l})\right] = \left[b(\mathbf{k}), b(\mathbf{l})\right] = 0\,, \tag{1.2.61}$$

$$\begin{aligned}\left[a(\mathbf{k}), b(\mathbf{l})\right] &= \left[a(\mathbf{k}), b^\dagger(\mathbf{l})\right] = \left[a^\dagger(\mathbf{k}), b(\mathbf{l})\right] \\ &= \left[a^\dagger(\mathbf{k}), b^\dagger(\mathbf{l})\right] = 0\,.\end{aligned} \tag{1.2.62}$$

It is straightforward to check that these operators do satisfy the desired relations (1.2.53), (1.2.54), (1.2.50), and (1.2.57). The field (1.2.48) may be written

$$\begin{aligned}\phi(\mathbf{x},t) = \frac{i}{\sqrt{V}}\sum_{\mathbf{k}}\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}}}\,&[a(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{x} - i\omega_{\mathbf{k}}t) \\ &-b^\dagger(-\mathbf{k})\exp(-i\mathbf{k}\cdot\mathbf{x} + i\omega_{\mathbf{k}}t)]\end{aligned} \tag{1.2.63}$$

and the Hamiltonian (1.2.55) takes the form

$$H = \sum_{\mathbf{k}}\frac{1}{2}\hbar\omega_{\mathbf{k}}\left[b^\dagger(\mathbf{k})b(\mathbf{k}) + b(\mathbf{k})b^\dagger(\mathbf{k}) + a^\dagger(\mathbf{k})a(\mathbf{k}) + a(\mathbf{k})a^\dagger(\mathbf{k})\right]$$

or, using (1.2.60)–(1.2.62)

$$H = \sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}\left[b^\dagger(\mathbf{k})b(\mathbf{k}) + a^\dagger(\mathbf{k})a(\mathbf{k})\right] + E_0\,, \tag{1.2.64}$$

where E_0 is the infinite c-number

$$E_0 \equiv \sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}\,. \tag{1.2.65}$$

The existence of two different kinds of operators a and b, which appear in precisely the same way in the Hamiltonian, shows that this is a theory with *two* kinds of particles with the same mass. As emphasized by Pauli and Weisskopf, these two varieties can be identified as particles and the corresponding antiparticles, and if charged have opposite charges. Thus, as we stressed above, bosons of spin zero as well as fermions of spin 1/2 can have distinct antiparticles, which for bosons cannot be identified as holes in a sea of negative energy particles.

We now can tell whether a and b or $a^\dagger$ and $b^\dagger$ are the annihilation operators by taking the expectation values of commutation relations in the vacuum state Ψ_0. For instance, if $a_{\mathbf{k}}^\dagger$ were an annihilation operator it would give zero when applied to the vacuum state, so the vacuum expectation value of (1.2.60) would give

$$-||a(\mathbf{k})\Psi_0||^2 = \left(\Psi_0, \left[a(\mathbf{k}), a^\dagger(\mathbf{k})\right]\Psi_0\right) = +1 \tag{1.2.66}$$

in conflict with the requirement that the left-hand side must be negative-definite. In this way we can conclude that it is a_k and b_k that are the annihilation operators, and therefore

$$a(\mathbf{k})\Psi_0 = b(\mathbf{k})\Psi_0 = 0\,. \tag{1.2.67}$$

This is consistent with all commutation relations. Thus, the canonical formalism forces the coefficient of the $e^{+i\omega t}$ in the field (1.2.58) to be a creation operator, as it also is in the Furry–Oppenheimer formalism[44] for spin 1/2.

Equations (1.2.64) and (1.2.67) now tell us that E_0 is the energy of the vacuum state. If we measure all energies relative to E_0, then the physical energy operator is $H - E_0$, and (1.2.64) shows that this again is positive.

What about the problem that served Dirac as a starting point, the problem of negative probabilities? As Dirac had recognized, the only probability density ρ, which can be formed from solutions of the Klein–Gordon–Schrödinger free scalar wave equation (1.2.28), and which satisfies a conservation law of the form (1.1.10), must be proportional to the quantity

$$\rho = 2\,\mathrm{Im}\left[\phi^\dagger \frac{\partial\phi}{\partial t}\right] \tag{1.2.68}$$

and therefore is not necessarily a positive quantity. Similarly, in the 'second-quantized' theory, where ϕ is given by Eq. (1.2.63), ρ is not a positive operator. Since $\phi^\dagger(x)$ does not commute with $\dot\phi(x)$ here, we can write (1.2.68) in various forms, which differ by infinite c-numbers; it proves convenient to write it as

$$\rho = \frac{i}{\hbar}\left[\frac{\partial\phi}{\partial t}\phi^\dagger - \frac{\partial\phi^\dagger}{\partial t}\phi\right]\,. \tag{1.2.69}$$

The space-integral of this operator is then easily calculated to be

$$N \equiv \int \rho\, d^3x = \sum_{\mathbf{k}}\left(a^\dagger(\mathbf{k})a(\mathbf{k}) - b^\dagger(\mathbf{k})b(\mathbf{k})\right) \tag{1.2.70}$$

and clearly has eigenvalues of either sign.

However, in a sense this problem appears in quantum field theory for spin 1/2 as well as spin zero. The density operator $\psi^\dagger\psi$ of Dirac is indeed a positive operator, but in order to construct a physical density we ought to subtract the contribution of the filled electron states. In particular, using the plane-wave decomposition (1.2.43), we may write the total number operator as

$$N \equiv \int d^3x\, \psi^\dagger\psi = \sum_{\mathbf{k}}{}^{(+)} a^\dagger(\mathbf{k})a(\mathbf{k}) + \sum_{\mathbf{k}}{}^{(-)} b(\mathbf{k})b^\dagger(\mathbf{k})\,.$$

The anticommutation relations for the bs allow us to rewrite this as

$$N - N_0 = \sum_{\mathbf{k}}{}^{(+)} a^\dagger_{\mathbf{k}} a_{\mathbf{k}} - \sum_{\mathbf{k}}{}^{(-)} b^\dagger_{\mathbf{k}} b_{\mathbf{k}}\,, \tag{1.2.71}$$

where N_0 is the infinite constant

$$N_0 = \sum_{\mathbf{k}}{}^{(-)} 1\,. \tag{1.2.72}$$

According to Eqs. (1.2.46) and (1.2.47), N_0 is the number of particles in the vacuum, so Furry and Oppenheimer reasoned that the physical number operator is $N - N_0$, and this now has both negative and positive eigenvalues, just as for spin zero.

The solution to this problem provided by quantum field theory is that neither the ψ of Furry and Oppenheimer nor the ϕ of Pauli and Weisskopf are probability amplitudes, which would have to define conserved positive probability densities. Instead, the physical Hilbert space is spanned by states defined as containing definite numbers of particles and/or antiparticles in each mode. If Φ_n are a complete orthonormal set of such states, then a measurement of particle numbers in an arbitrary state Ψ will yield a probability for finding the system in state Φ_n, given by

$$P_n = |(\Phi_n, \Psi)|^2\,, \tag{1.2.73}$$

where (Φ_n, Ψ) is the usual Hilbert space scalar product. Hence, no question as to the possibility of negative probabilities will arise for any spin. The wave fields ϕ, ψ, etc, are not probability amplitudes at all, but *operators* which create or destroy particles in the various normal modes. It would be a good thing if the misleading expression 'second quantization' were permanently retired.

In particular, the operators N and $N - N_0$ of Eqs. (1.2.70) and (1.2.71) are not to be interpreted as total probabilities, but as number operators: specifically, the number of particles *minus* the number of antiparticles. For charged particles, the conservation of charge forces the charge operators to be proportional to these number operators, so the minus signs in (1.2.70) and (1.2.71) allow us immediately to conclude that particles and antiparticles have opposite charge. In this field-theoretic formalism, interactions contribute terms to the Hamiltonian which are of third, fourth, or higher order in field variables, and the rates of various processes are given by using these interaction operators in a time-dependent perturbation theory. The conceptual framework described in the above brief remarks will serve as the basis for much of the work in this book.

Despite its apparent advantages, quantum field theory did not immediately supplant hole theory; rather, the two points of view coexisted for a while, and various combinations of field-theoretic and hole-theoretic ideas were used in calculations of physical reaction rates. This period saw a number of calculations of cross sections to lowest order in powers of e^2 for various processes, such as $e^- + \gamma \to e^- + \gamma$ in 1929 by Klein and Nishina;[46] $e^+ + e^- \to 2\gamma$ in 1930 by Dirac;[47] $e^- + e^- \to e^- + e^-$ in 1932 by Møller;[48] $e^- + Z \to e^- + \gamma + Z$ and $\gamma + Z \to e^+ + e^- + Z$ (where Z denotes the Coulomb field of a heavy atom) in 1934 by Bethe and Heitler;[49] and $e^+ + e^- \to e^+ + e^-$ in 1936 by Bhabha.[50] (Rules for the calculation of such processes are given in Chapter 8, and worked out in detail there for the case of electron–photon scattering.) These lowest-order calculations gave finite results, in reasonable agreement with the experimental data.

Nevertheless, a general feeling of dissatisfaction with quantum field theory (whether or not in the form of hole theory) persisted throughout the 1930s. One of the reasons for this was the apparent failure of quantum electrodynamics to account for the penetrating power of the charged particles in cosmic ray showers, noted in 1936 by Oppenheimer and Franklin Carlson.[50a] Another cause of dissatisfaction that turned out to be related to the first was the steady discovery of new kinds of particles and interactions. We have already mentioned the electron, photon, positron, neutrino, and, of course, the nucleus of hydrogen, the proton. Throughout the 1920s it was generally believed that heavier nuclei are composed of protons and electrons, but it was hard to see how a light particle like the electron could be confined in the nucleus. Another severe difficulty with this picture was pointed out in 1931 by Ehrenfest and Oppenheimer:[51] the nucleus of ordinary nitrogen, N^{14}, in order to have atomic number 7 and atomic weight 14, would have to be composed of 14 protons and 7 electrons, and would therefore have to be a fermion, in conflict with the result of molecular spectroscopy[52] that N^{14} is a boson. This problem (and others) were solved in 1932 with the discovery of the neutron,[53] and by Heisenberg's subsequent suggestion[54] that nuclei are composed of protons and neutrons, not protons and electrons. It was clear that a strong non-electromagnetic force of short range would have to operate between neutrons and protons to hold nuclei together.

After the success of the Fermi theory of beta decay, several authors[54a] speculated that nuclear forces might be explained in this theory as due to the exchange of electrons and neutrinos. A few years later, in 1935, Hideki Yukawa proposed a quite different quantum field theory of the nuclear force.[55] In an essentially classical calculation, he found that the interaction of a scalar field with nucleons (protons or neutrons) would produce a nucleon–nucleon potential, with a dependence on the nucleon separation r given by

$$V(r) \propto \frac{1}{r} \exp(-\lambda r) \tag{1.2.74}$$

instead of the $1/r$ Coulomb potential produced by electric fields. The quantity λ was introduced as a parameter in Yukawa's scalar field equation, and when this equation was quantized, Yukawa found that it described particles of mass $\hbar\lambda/c$. The observed range of the strong interactions within nuclei led Yukawa to estimate that $\hbar\lambda/c$ is of the order of 200 electron masses. In 1937 such 'mesons' were discovered in cloud chamber experiments[56] by Seth Neddermeyer and Anderson and by Jabez Curry Street and Edward Carl Stevenson, and it was generally believed that these were the hypothesized particles of Yukawa.

The discovery of mesons revealed that the charged particles in cosmic ray showers are not all electrons, and thus cleared up the problem with these showers that had bothered Oppenheimer and Carlson. At the same time, however, it created new difficulties. Lothar Nordheim[56a] pointed out in 1939 that the same strong interactions by which the mesons are copiously produced at high altitudes (and which are required in Yukawa's theory) should have led to the mesons' absorption in the atmosphere, a result contradicted by their copious appearance at lower altitudes. In 1947 it was shown in an experiment by Marcello Conversi, Ettore Pancini, and Oreste Piccioni[57] that the mesons which predominate in cosmic rays at low altitude actually interact weakly with nucleons, and therefore could not be identified with Yukawa's particle. This puzzle was cleared up by a theoretical suggestion,[58] and its subsequent experimental confirmation[59] by Cesare Lattes, Occhialini, and Cecil Powell — there are two kinds of mesons with slightly different masses: the heavier (now called the π meson or pion) has strong interactions and plays the role in nuclear force envisaged by Yukawa; the lighter (now called the μ meson, or muon) has only weak and electromagnetic interactions, and predominates in cosmic rays at sea level, being produced by the decay of π mesons. In the same year, 1947, entirely new kinds of particles (now known as K mesons and hyperons) were found in cosmic rays by George Rochester and Clifford Butler.[60] From 1947 until the present particles have continued to be discovered in a bewildering variety, but to pursue this story would take us outside the bounds of our present survey. These discoveries showed clearly that any conceptual framework which was limited to photons, electrons, and positrons would be far too narrow to be taken seriously as a fundamental theory. But an even more important obstacle was presented by a purely theoretical problem — the problem of infinities.

1.3 The Problem of Infinities

Quantum field theory deals with fields $\psi(x)$ that destroy and create particles at a spacetime point x. Earlier experience with classical electron theory provided a warning that a point electron will have infinite electromagnetic self-mass; this mass is $e^2/6\pi ac^2$ for a surface distribution of charge with radius a, and therefore blows up for $a \to 0$. Disappointingly this problem appeared with even greater severity in the early days of quantum field theory, and although greatly ameliorated by subsequent improvements in the theory, it remains with us to the present day.

The problem of infinities in quantum field theory was apparently first noted in the 1929–30 papers of Heisenberg and Pauli.[41] Soon after, the presence of infinities was confirmed in calculations of the electromagnetic self-energy of a bound electron by Oppenheimer,[61] and of a free electron by Ivar Waller.[62] They used ordinary second-order perturbation theory, with an intermediate state consisting of an electron and a photon: for instance, the shift of the energy E_n of an electron in the nth energy level of hydrogen is given by

$$\Delta E_n = \sum_{m,\lambda} \int d^3k \, \frac{|<m;\mathbf{k},\lambda|H'|n>|^2}{E_n - E_m - |\mathbf{k}|c} , \qquad (1.3.1)$$

where the sums and integral are over all intermediate electron states m, photon helicities λ, and photon momenta $\mathbf{k}$, and H' is the term in the Hamiltonian representing the interaction of radiation and electrons. This calculation gave a self-energy that is formally infinite; further; if this infinity is removed by discarding all intermediate states with photon wave numbers greater than $1/a$, then the self-energy behaves like $1/a^2$ as $a \to 0$. Infinities of this sort are often called ultraviolet divergences, because they arise from intermediate states containing particles of very short wavelength.

These calculations treated the electron according to the rules of the original Dirac theory, without filled negative-electron states. A few years later Weisskopf repeated the calculation of the electron self-mass in the new hole theory, with all negative-energy states full. In this case another term appears in second-order perturbation theory, which in a non-hole-theory language can be described as arising from processes in which the electron in its final state first appears out of the vacuum together with a photon and a positron which then annihilate along with the initial electron. Initially Weisskopf found a $1/a^2$ dependence on the photon wave-number cutoff $1/a$. The same calculation was being carried out (at the suggestion of Bohr) at that time by Carlson and Furry. After seeing Weisskopf's results, Furry realized that while Weisskopf had included an electrostatic term that he and Carlson had neglected, Weisskopf had made a new mistake in the calculation of the magnetic self-energy. After hearing from Furry and correcting his own error, Weisskopf found that the $1/a^2$ terms in the total mass shift cancelled! However, despite this cancellation, an infinity remained: with a wave-number cutoff $1/a$, the self-mass was found to be[63]

$$m_{em} = \frac{3\alpha}{2\pi} m \ln\left(\frac{\hbar}{mca}\right) , \tag{1.3.2}$$

The weakening of the cut-off dependence, to $\ln a$ as compared with the classical $1/a$ or the early quantum $1/a^2$, was mildly encouraging at the time and turned out to be of great importance later, in the development of renormalization theory.

An infinity of quite a different kind was encountered in 1933, apparently first by Dirac.[64] He considered the effect of an external static nearly uniform charge density $\varepsilon(\mathbf{x})$ on the vacuum, i.e., on the negative-energy electrons in the filled energy levels of hole theory. The Coulomb interaction between $\varepsilon(\mathbf{x})$ and the charge density of the negative-energy electrons produces a 'vacuum polarization,' with induced charge density

$$\delta\varepsilon = A\varepsilon + B\left(\frac{\hbar}{mc}\right)^2 \nabla^2\varepsilon + \cdots . \tag{1.3.3}$$

The constant B is finite, and of order α. On the other hand, A is logarithmically divergent, of order $\alpha \ln a$, where $1/a$ is the wave-number cutoff.

Infinities also seemed to occur in a related problem, the scattering of light by light. Hans Euler, Bernard Kockel, and Heisenberg[65] showed in 1935–6 that these infinities could be eliminated by using a more-or-less arbitrary prescription suggested earlier by Dirac[66] and Heisenberg[67]. They calculated an effective Lagrangian density for the non-linear electrodynamic effects produced by virtual electron–positron pairs:

$$\mathscr{L} = \frac{1}{2}\left(\mathbf{E}^2 - \mathbf{B}^2\right) + \frac{e^4\hbar}{360\pi^2 m_e^4 c^7}\left[\left(\mathbf{E}^2 - \mathbf{B}^2\right)^2 + 7\left(\mathbf{E}\cdot\mathbf{B}\right)^2\right] + \cdots , \tag{1.3.4}$$

valid for frequencies $\nu \ll m_e c^2/h$. Soon after, Nicholas Kemmer and Weisskopf[68] presented arguments that in this case the infinities are spurious, and that Eq. (1.3.4) can be derived without any subtraction prescription.

One bright spot in the struggle with infinities was the successful treatment of *infrared* divergences, those that arise from the low-energy rather than the high-energy part of the range of integration. In 1937 it was shown by Felix Bloch and Arne Nordsieck[68a] that these infinities cancel provided one includes processes in which arbitrary numbers of low-energy photons are produced. This will be discussed in modern terms in Chapter 13.

Yet another infinity turned up in a calculation by Sidney Michael Dancoff[69] in 1939 of the radiative corrections to the scattering of electrons by the static Coulomb field of an atom. The calculation contained a mistake (one of the terms was omitted), but this was not realized until later.[69a]

Throughout the 1930s, these various infinities were seen not merely as failures of specific calculations. Rather, they seemed to indicate a gap in the understanding of relativistic quantum field theory on the most fundamental level, an opinion reinforced by the problems with cosmic rays mentioned in the previous section.

One of the symptoms of this uneasy pessimism was the continued exploration throughout the 1930s and 1940s of alternative formalisms. As Julian Schwinger[69b] later recalled, 'The preoccupation of the majority of involved physicists was not with analyzing and carefully applying the known relativistic theory of coupled electron and electromagnetic fields but with changing it.' Thus in 1938 Heisenberg[70] proposed the existence of a fundamental length L, analogous to the fundamental action h and fundamental velocity c. Field theory was supposed to work only for distances larger than L, so that all divergent integrals would effectively be cut off at distances L, or momenta h/L. Several specific proposals[70a] were made for giving field theory a non-local structure. Some theorists began to suspect that the formalism of state-vectors and quantum fields should be replaced by one based solely on observable quantities, such as the S-matrix introduced by John Archibald Wheeler[71] in 1937 and Heisenberg[72] in 1943, whose elements are the amplitudes for various scattering processes. As we shall see, the concept of the S-matrix has now become a vital part of modern quantum field theory, and for some theorists a pure S-matrix theory became an ideal, especially as a possible solution to the problems of the strong interactions.[73] In yet another direction, Wheeler and Richard Feynman[74] in 1945 attempted to eliminate the electromagnetic field, deriving electromagnetic interactions in terms of an interaction at a distance. They were able to show that a pure retarded (or pure advanced) potential could be obtained by taking into account the interaction not only between source and test charges, but also between these charges and all the other charges in the universe. Perhaps the most radical modification of quantum mechanics suggested during this period was the introduction by Dirac[75] of states of negative probability, as a means of cancelling infinities in sums over states. This idea, of an 'indefinite metric' in Hilbert space, has also flourished in quantum field theory, though not in the form originally suggested.

A more conservative idea for dealing with the infinities was also in the air during the 1930s. Perhaps these infinities could all be absorbed into a redefinition, a 'renormalization' of the parameters of the theory. For instance, it was already known that in any Lorentz-invariant classical theory the electromagnetic self-energy and self-momentum of an electron *must* take the form of corrections to the mass of the electron; hence the infinities in these quantities can be cancelled by a negative infinity in the 'bare' non-electromagnetic mass of the electron, leaving a finite measurable 'renormalized' mass. Also, Eq. (1.3.3) shows that the vacuum polarization changes the charge of the electron, from $e \equiv \int d^3x\, \varepsilon$, to

$$e_{\text{TOTAL}} = \int d^3x(\varepsilon + \delta\varepsilon) = (1 + A)e\,. \tag{1.3.5}$$

Vacuum polarization gives finite results in lowest order if observables like scattering cross-sections are expressed in terms of e_{TOTAL} rather than e. The question was, whether all infinities in quantum field theory could be dealt with in this way. In 1936 Weisskopf[76] suggested that this is the case, and verified that known infinities could be eliminated by renormalization of physical parameters in a variety of sample calculations. However, it was impossible with the calculational techniques then available to show that infinities could always be eliminated in this way, and Dancoff's calculation[69] seemed to show that they could not.

Another effect of the appearance of infinities was a tendency to believe that any effect which turned out to be infinite in quantum field theory was actually not there at all. In particular, the 1928 Dirac theory had predicted complete degeneracy of the $2s_{1/2}$–$2p_{1/2}$ levels of hydrogen to all orders in α; any attempt at a quantum electromagnetic calculation of the splitting of these two levels ran into the problem of the infinite self-energy of a bound electron; therefore the existence of such a splitting was generally not taken seriously. Later Bethe[80] recalled that 'This shift comes out infinite in all existing theories, and has therefore always been ignored.' This attitude persisted even in the late 1930s, when spectroscopic experiments[77] began to indicate the presence of a $2s_{1/2}$–$2p_{1/2}$ splitting of order 1000 MHz. One notable exception was Edwin Albrecht Uehling,[78] who realized that the vacuum polarization effect mentioned earlier would produce a $2s_{1/2}$–$2p_{1/2}$ splitting; unfortunately, as we shall see in Chapter 14, this contribution to the splitting is much smaller than 1000 MHz, and of the wrong sign.

The gloom surrounding quantum field theory began to lift soon after World War II. On June 1–4, 1947, the Conference on the Foundations of

Quantum Mechanics at Shelter Island, NY brought theoretical physicists who had been working on the problems of quantum field theory through the 1930s together with a younger generation of theorists who had started scientific work during the war, and — of crucial importance — a few experimental physicists. The discussion leaders were Hans Kramers, Oppenheimer, and Weisskopf. One of the experimentalists (or rather theorist turned experimentalist), Willis Lamb, described a decisive measurement[79] of the $2s_{1/2}$–$2p_{1/2}$ shift in hydrogen. A beam of hydrogen atoms from an oven, many in $2s$ and $2p$ states, was aimed at a detector sensitive only to atoms in excited states. The atoms in $2p$ states can decay very rapidly to the $1s$ ground state by one-photon (Lyman α) emission, while the $2s$ states decay only very slowly by two-photon emission, so in effect the detector was measuring the number of atoms in the metastable $2s$ state. The beam was passed through a magnetic field, which added a known Zeeman splitting to any $2s_{1/2}$–$2p_{1/2}$ splitting naturally present. The beam was also exposed to a microwave-frequency electromagnetic field, with a fixed frequency $\nu \sim 10$ GHz. At a certain magnetic field strength the detector signal was observed to be quenched, indicating that the microwave field was producing resonant transitions from the metastable $2s$ state to the $2p$ state and thence by a rapid Lyman α emission to the ground state. The total (Zeeman plus intrinsic) $2s$–$2p$ splitting at this value of the magnetic field strength would have to be just $h\nu$, from which the intrinsic splitting could be inferred. A preliminary value of 1000 MHz was announced, in agreement with the earlier spectroscopic measurements.[77] The impact of this discovery can be summarized in a saying that was current in Copenhagen when I was a graduate student there in 1954: 'Just because something is infinite does not mean it is zero!'

The discovery of the Lamb shift aroused intense interest among the theorists at Shelter Island, many of whom had already been working on improved formalisms for calculation in quantum electrodynamics. Kramers described his work on mass renormalization in the classical electrodynamics of an extended electron,[79a] which showed that the difficulties associated with the divergence of the self-energy in the limit of zero radius do not appear explicitly if the theory is reexpressed so that the mass parameter in the formalism is identified with the experimental electron mass. Schwinger and Weisskopf (who had already heard rumors of Lamb's result, and discussed the matter on the trip to Shelter Island) suggested that since the inclusion of intermediate states involving positrons was known to reduce the divergence in energy level shifts from $1/a^2$ to $\ln a$, perhaps the *differences* of the shifts in atomic energy levels might turn out to be finite when these intermediate states were taken into account. (In fact, in 1946, before he learned of Lamb's experiment, Weisskopf had already assigned this problem to a graduate student, Bruce French.) Almost immediately after the conference, during a train ride to Schenectady, Hans Bethe[80] carried out a non-relativistic calculation, still without including the effects of intermediate states containing positrons, but using a simple cutoff at virtual photon momenta of order $m_e c^2$ to eliminate infinities. He obtained the encouraging approximate value of 1040 MHz. Fully relativistic calculations using the renormalization idea to eliminate infinities were soon thereafter carried out by a number of other authors,[81] with excellent agreement with experiment.

Another exciting experimental result was reported at Shelter Island by Isidor I. Rabi. Measurements in his laboratory of the hyperfine structure of hydrogen and deuterium had suggested[82] that the magnetic moment of the electron is larger than the Dirac value $e\hbar/2mc$ by a factor of about 1.0013, and subsequent measurements of the gyromagnetic ratios in sodium and gallium had given a precise value[83]

$$\mu = \frac{e\hbar}{2mc}\,[1.00118 \pm 0.00003]\,.$$

Learning of these results, Gregory Breit suggested[83a] that they arose from an order α radiative correction to the electron magnetic moment. At Shelter Island, both Breit and Schwinger described their efforts to calculate this correction. Shortly after the conference Schwinger completed a successful calculation of the anomalous magnetic moment of the electron[84]

$$\mu = \frac{e\hbar}{2mc}\left[1 + \frac{\alpha}{2\pi}\right] = \frac{e\hbar}{2mc}\,[1.001162]$$

in excellent agreement with observation. This, together with Bethe's calculation of the Lamb shift, at last convinced physicists of the reality of radiative corrections.

The mathematical methods used in this period presented a bewildering variety of concepts and formalisms. One approach developed by Schwinger[85] was based on operator methods and the action principle, and was presented by him at a conference at Pocono Manor in 1948, the successor to the Shelter Island Conference. Another Lorentz-invariant operator formalism had been developed earlier by Sin-Itiro Tomonaga[86] and his co-workers in Japan, but their work was not at first known in the West. Tomonaga had grappled with infinities in Yukawa's meson theory in the 1930s. In 1947 he and his group were still out of the loop of scientific communication; they learned about Lamb's experiment from an article in *Newsweek*.

An apparently quite different approach was invented by Feynman,[87] and described briefly by him at the Pocono Conference. Instead of introducing quantum field operators, Feynman represented the S-matrix as a functional integral of $\exp\left(iW\right)$, where W is the action integral for a set of Dirac particles interacting with a *classical* electromagnetic field, integrated over all Dirac particle trajectories satisfying certain initial and final conditions for $t \to \pm\infty$. One result of great practical importance that came out of Feynman's work was a set of graphical rules for calculating S-matrix elements to any desired order of perturbation theory. Unlike the old perturbation theory of the 1920s and 1930s, these Feynman rules automatically lumped together particle creation and antiparticle annihilation processes, and thereby gave results that were Lorentz-invariant at every stage. We have already seen in Weisskopf's early calculation[63] of the electron self-energy, that it is only in such calculations, including particles and antiparticles on the same footing, that the nature of the infinities becomes transparent.

Finally, in a pair of papers in 1949, Freeman Dyson[88] showed that the operator formalisms of Schwinger and Tomonaga would yield the same graphical rules that had been found by Feynman. Dyson also carried out an analysis of the infinities in general Feynman diagrams, and outlined a proof that these are always precisely the sort which could be removed by renormalization. One of the most striking results that could be inferred from Dyson's analysis was a criterion for deciding which quantum field theories are 'renormalizable', in the sense that all infinities can be absorbed into a redefinition of a *finite* number of coupling constants and masses. In particular, an interaction like the Pauli term (1.1.32), which would have changed the predicted magnetic moment of the electron, would spoil the renormalizability of quantum electrodynamics. With the publication of Dyson's papers, there was at last a general and systematic formalism that physicists could easily learn to use, and that would provide a common language for the subsequent applications of quantum field theory to the problems of physics.

I cannot leave the infinities without taking up a puzzling aspect of this story. Oppenheimer[61] in 1930 had already noticed that most of the ultraviolet divergence in the self-energy of a bound electron cancels when one takes the difference between the shifts of two atomic energy levels, and Weisskopf[63] in 1934 had found that most of the divergence in the self-energy of a free electron cancels when one includes intermediate states containing positrons. It would have been natural even in 1934 to guess that including positron intermediate states *and* subtracting the energy shifts of pairs of atomic states would eliminate the ultraviolet divergence in their relative energy shift.* There was even experimental evidence[77] for

* In fact, this guess would have been wrong. As discussed in Section 14.3, radiative corrections to the electron mass affect atomic energy levels not only through a shift in the electron rest energy, which is the same in all atomic energy levels, but also through a change in the electron kinetic energy, that varies from one level to another.

a $2s_{1/2}$–$2p_{1/2}$ energy difference of order 1000 MHz. So why did no one before 1947 attempt an *numerical* estimate of this energy difference?

Strictly speaking, there was one such attempt[88a] in 1939, but it focused on the wrong part of the problem, the charge radius of the *proton*, which has only a tiny effect on hydrogen energy levels. The calculation gave a result in rough agreement with the early experiments.[77] This was a mistake, as shown in 1939 by Lamb.[88b]

A fully relativistic calculation of the Lamb shift including positrons in intermediate states could have been attempted during the 1930s, using the old non-relativistic perturbation theory. As long as one keeps all terms up to a given order, old-fashioned non-relativistic perturbation theory gives the same results as the manifestly relativistic formalisms of Feynman, Schwinger, and Tomonaga. In fact, after Bethe's work, the first precise calculations[81] of the Lamb shift in the USA by French and Weisskopf and Norman Kroll and Lamb were done in just this way, though Tomonoga's group [81] in Japan was already using covariant methods to solve this and other problems.

The one missing element was confidence in renormalization as a means of dealing with infinities. As we have seen, renormalization was widely discussed in the late 1930s. But it had become accepted wisdom in the 1930s, and a point of view especially urged by Oppenheimer,[89] that quantum electrodynamics could not be taken seriously at energies of more than about 100 MeV, and that the solution to its problems could be found only in really adventurous new ideas.

Several things happened at Shelter Island to change this expectation. One was news that the problems concerning cosmic rays discussed in the previous section were beginning to be resolved; Robert Marshak presented the hypothesis[58] that there were two types of 'meson' with similar masses; the muons that had actually been observed, and the pions responsible for nuclear forces. More important was the fact that now there were reliable experimental values for the Lamb shift and the anomalous magnetic moment that forced physicists to think carefully about radiative corrections. Probably equally important was the fact that the conference brought together theorists who had in their own individual ways been thinking about renormalization as a solution to the problem of infinities. When the revolution came in the late 1940s, it was made by physicists who though mostly young were playing a conservative role, turning away from the search by their predecessors for a radical solution.

Bibliography

- □ S. Aramaki, 'Development of the Renormalization Theory in Quantum Electrodynamics,' *Historia Scientiarum* **36**, 97 (1989); *ibid.* **37**, 91 (1989). [Section 1.3.]
- □ R. T. Beyer, ed., *Foundations of Nuclear Physics* (Dover Publications, Inc., New York, 1949). [Section 1.2.]
- □ L. Brown, 'Yukawa's Prediction of the Meson,' *Centauros* **25**, 71 (1981). [Section 1.2.]
- □ L. M. Brown and L. Hoddeson, eds., *The Birth of Particle Physics* (Cambridge University Press, Cambridge, 1983). [Sections 1.1, 1.2, 1.3.]
- □ T. Y. Cao and S. S. Schweber, 'The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory,' *Synthèse* **97**, 33 (1993). [Section 1.3.]
- □ P. A. M. Dirac, *The Development of Quantum Theory* (Gordon and Breach Science Publishers, New York, 1971). [Section 1.1.]
- □ E. Fermi, 'Quantum Theory of Radiation,' *Rev. Mod. Phys.* **4**, 87 (1932). [Sections 1.2 and 1.3.]
- □ G. Gamow, *Thirty Years that Shook Physics* (Doubleday and Co., Garden City, New York, 1966). [Section 1.1.]
- □ M. Jammer, *The Conceptual Development of Quantum Mechanics* (McGraw-Hill Book Co., New York, 1966). [Section 1.1.]
- □ J. Mehra, 'The Golden Age of Theoretical Physics: P.A.M. Dirac's Scientific Work from 1924 to 1933,' in *Aspects of Quantum Theory*, ed. by A. Salam and E. P. Wigner, (Cambridge University Press, Cambridge, 1972). [Section 1.1.]
- □ A. I. Miller, *Early Quantum Electrodynamics – A Source Book* (Cambridge University Press, Cambridge, UK, 1994). [Sections 1.1, 1.2, 1.3.]
- □ A. Pais, *Inward Bound* (Clarendon Press, Oxford, 1986). [Sections 1.1, 1.2, 1.3.]
- □ S. S. Schweber, 'Feynman and the Visualization of Space-Time Processes,' *Rev. Mod. Phys.* **58**, 449 (1986). [Section 1.3.]
- □ S. S. Schweber, 'Some Chapters for a History of Quantum Field Theory: 1938–1952,' in *Relativity, Groups, and Topology II*, ed. by B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984). [Sections 1.1, 1.2, 1.3.]

□ S. S. Schweber, 'A Short History of Shelter Island I,' in *Shelter Island II*, ed. by R. Jackiw, S. Weinberg, and E. Witten (MIT Press, Cambridge, MA, 1985). [Section 1.3.]

□ S. S. Schweber, *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga* (Princeton University Press, Princeton, 1994). [Section 1.1, 1.2, 1.3.]

□ J. Schwinger, ed., *Selected Papers in Quantum Electrodynamics* (Dover Publications Inc., New York, 1958). [Sections 1.2 and 1.3.]

□ S.-I. Tomonaga, in *The Physicist's Conception of Nature* (Reidel, Dordrecht, 1973). [Sections 1.2 and 1.3.]

□ S. Weinberg, 'The Search for Unity: Notes for a History of Quantum Field Theory,' *Daedalus*, Fall 1977. [Sections 1.1, 1.2, 1.3.]

□ V. F. Weisskopf, 'Growing Up with Field Theory: The Development of Quantum Electrodynamics in Half a Century,' 1979 Bernard Gregory Lecture at CERN, published in L. Brown and L. Hoddeson, *op. cit.*. [Sections 1.1, 1.2, 1.3.]

□ G. Wentzel, 'Quantum Theory of Fields (Until 1947)' in *Theoretical Physics in the Twentieth Century*, ed. by M. Fierz and V. F. Weisskopf (Interscience Publishers Inc., New York, 1960). [Sections 1.2 and 1.3.]

□ E. Whittaker, *A History of the Theories of Aether and Electricity* (Humanities Press, New York, 1973). [Section 1.1.]

References

1. L. de Broglie, *Comptes Rendus* **177**, 507, 548, 630 (1923); *Nature* **112**, 540 (1923); Thèse de doctorat (Masson et Cie, Paris, 1924); *Annales de Physique* **3**, 22 (1925) [reprinted in English in *Wave Mechanics*, ed. by G. Ludwig, (Pergamon Press, New York, 1968)]; *Phil. Mag.* **47**, 446 (1924).

2. W. Elsasser, *Naturwiss.* **13**, 711 (1925).

3. C. J. Davisson and L. H. Germer, *Phys. Rev.* **30**, 705 (1927).

4. W. Heisenberg, *A. Phys.* **33**, 879 (1925); M. Born and P. Jordan, *Z. f. Phys.* **34**, 858 (1925); P. A. M. Dirac, *Proc. Roy. Soc.* **A109**, 642 (1925); M. Born, W. Heisenberg, and P. Jordan, *Z. f. Phys.* **35**, 557 (1925); W. Pauli, *Z. f. Phys.* **36**, 336 (1926). These papers are reprinted in *Sources of Quantum Mechanics*, ed. by B. L. van der Waerden (Dover Publications, Inc., New York, 1968).

5. E. Schrödinger, *Ann. Phys.* **79**, 361, 489; **80**, 437; **81**, 109 (1926). These papers are reprinted in English, unfortunately in a somewhat abridged form, in *Wave Mechanics*, Ref. 1. Also see *Collected Papers on Wave Mechanics*, trans. by J. F. Schearer and W. M. Deans (Blackie and Son, London, 1928).

6. See, e.g., P. A. M. Dirac, *The Development of Quantum Theory* (Gordon and Breach, New York, 1971). Also see Dirac's obituary of Schrödinger, *Nature* **189**, 355 (1961), and his article in *Scientific American* **208**, 45 (1963).

7. O. Klein, *Z. f. Phys.* **37**, 895 (1926). Also see V. Fock, *Z. f. Phys.* **38**, 242 (1926); *ibid*, **39**, 226 (1926).

8. W. Gordon, *Z. f. Phys.* **40**, 117 (1926).

9. For the details of the calculation, see, e.g., L. I. Schiff, *Quantum Mechanics*, 3rd edn, (McGraw-Hill, Inc. New York, 1968): Section 51.

10. F. Paschen, *Ann. Phys.* **50**, 901 (1916). These experiments were actually carried out using He^+ because its fine structure splitting is 16 times larger than for hydrogen. The fine structure of spectral lines was first discovered interferometrically by A. A. Michelson, *Phil. Mag.* **31**, 338 (1891); *ibid.*, **34**, 280 (1892).

10*a*. A. Sommerfeld, *Münchner Berichte* 1915, pp. 425, 429; *Ann. Phys.* **51**, 1, 125 (1916). Also see W. Wilson, *Phil. Mag.* **29**, 795 (1915).

11. G. E. Uhlenbeck and S. Goudsmit, *Naturwiss.* **13**, 953 (1925); *Nature* **117**, 264 (1926). The electron spin had been earlier suggested for other reasons by A. H. Compton, *J. Frank. Inst.* **192**, 145 (1921).

12. The general formula for Zeeman splitting in one-electron atoms had been discovered empirically by A. Landé, *Z. f. Phys.* **5**, 231 (1921); *ibid.*, **7**, 398 (1921); *ibid.*, **15**, 189 (1923); *ibid.*, **19**, 112 (1923). At the time, the extra non-orbital angular momentum appearing in this formula was thought to be the angular momentum of the atomic 'core;' A. Sommerfeld, *Ann. Phys.* **63**, 221 (1920); *ibid.*, **70**, 32 (1923). It was only later that the extra angular momentum was recognized, as in Ref. 11, to be due to the spin of the electron.

13. W. Heisenberg and P. Jordan, *Z. f. Phys.* **37**, 263 (1926); C. G. Darwin, *Proc. Roy. Soc.* **A116**, 227 (1927). Darwin says that several authors did this work at about the same time, while Dirac quotes only Darwin.

14. L. H. Thomas, *Nature* **117**, 514 (1926). Also see S. Weinberg, *Gravitation and Cosmology*, (Wiley, New York, 1972): Section 5.1.

15. P. A. M. Dirac, *Proc. Roy. Soc.* **A117**, 610 (1928). Also see Dirac, *ibid.*, **A118**, 351 (1928), for the application of this theory to the calculation of the Zeeman and Paschen–Back effects and the relative strengths of lines within fine-structure multiplets.

16. For the probabilistic interpretation of non-relativistic quantum mechanics, see M. Born *Z. f. Phys.* **37**, 863 (1926); *ibid*, **38**, 803 (1926) (reprinted in an abridged English version in *Wave Mechanics*, Ref. 41); G. Wentzel, *Z. f. Phys.* **40**, 590 (1926); W. Heisenberg, *Z. f. Phys.* **43**, 172 (1927). N. Bohr. *Nature* **121**, 580 (1928); Naturwissenchaften **17**, 483 (1929); *Electrons et Photons – Rapports et Discussions du V^e^ Conseil de Physique Solvay* (Gauthier-Villars, Paris, 1928).

17. Conversation between Dirac and J. Mehra, March 28, 1969, quoted by Mehra in *Aspects of Quantum Theory*, ed. by A. Salam and E. P. Wigner (Cambridge University Press, Cambridge, 1972).

18. G. Gamow, *Thirty Years that Shook Physics*, (Doubleday and Co., Garden City, NY, 1966): p. 125.

19. W. Pauli, *Z. f. Phys.* **37**, 263 (1926); **43**, 601 (1927).

20. C. G. Darwin, *Proc. Roy. Soc.* **A118**, 654 (1928); *ibid.*, **A120**, 621 (1928).

21. W. Gordon, *Z. f. Phys.* **48**, 11 (1928).

22. P. A. M. Dirac, *Proc. Roy. Soc.* **A126**, 360 (1930); also see Ref. 47.

23. E. C. Stoner, *Phil. Mag.* **48**, 719 (1924).

24. W. Pauli, *Z. f. Phys.* **31**, 765 (1925).

25. W. Heisenberg, *Z. f. Phys.* **38**, 411 (1926); *ibid.*, **39**, 499 (1926); P. A. M. Dirac, *Proc. Roy. Soc.* **A112**, 661 (1926); W. Pauli, *Z. f. Phys.* **41**, 81 (1927); J. C. Slater, *Phys. Rev.* **34**, 1293 (1929).

26. E. Fermi, *Z. f. Phys.* **36**, 902 (1926); *Rend. Accad. Lincei* **3**, 145 (1926).

27. P. A. M. Dirac, Ref. 25.

27*a*. P. A. M. Dirac, First W. R. Crane Lecture at the University of Michigan, April 17, 1989, unpublished.

28. H. Weyl, *The Theory of Groups and Quantum Mechanics*, translated from the second (1931) German edition by H. P. Robertson (Dover Publications, Inc., New York): Chapter IV, Section 12. Also see P. A. M. Dirac, *Proc. Roy. Soc.* **A133**, 61 (1931).

29. J. R. Oppenheimer, *Phys. Rev.* **35**, 562 (1930); I. Tamm, *Z. f. Phys.* **62**, 545 (1930).

29*a*. P. A. M. Dirac, *Proc. Roy. Soc.* **133**, 60 (1931).

30. C. D. Anderson, *Science* **76**, 238 (1932); *Phys. Rev.* **43**, 491 (1933). The latter paper is reprinted in *Foundations of Nuclear Physics*, ed. by R. T. Beyer (Dover Publications, Inc., New York, 1949).

30*a*. J. Schwinger, 'A Report on Quantum Electrodynamics,' in *The Physicist's Conception of Nature* (Reidel, Dordrecht, 1973): p.415.

31. W. Pauli, *Handbuch der Physik* (Julius Springer, Berlin, 1932–1933); *Rev. Mod. Phys.* **13**, 203 (1941).

32. Born, Heisenberg, and Jordan, Ref. 4, Section 3.

32*a*. P. Ehrenfest, *Phys. Z.* **7**, 528 (1906).

33. Born and Jordan, Ref. 4. Unfortunately the relevant parts of this paper are not included in the reprint collection *Sources of Quantum Mechanics*, cited in Ref. 4.

34. P. A. M. Dirac, *Proc. Roy. Soc.* **A112**, 661 (1926): Section 5. For a more accessible derivation, see L. I. Schiff, *Quantum Mechanics*, 3rd edn. (McGraw-Hill Book Company, New York, 1968): Section 44.

34*a*. A. Einstein, *Phys. Z.* **18**, 121 (1917); reprinted in English in van der Waerden, Ref. 4.

35. P. A. M. Dirac, *Proc. Roy. Soc.* **A114**, 243 (1927); reprinted in *Quantum Electrodynamics*, ed. by J. Schwinger (Dover Publications, Inc., New York, 1958).

36. P. A. M. Dirac, *Proc. Roy. Soc.* **A114**, 710 (1927).

36*a*. V. F. Weisskopf and E. Wigner, *Z. f. Phys.* **63**, 54 (1930)

36*b*. E. Fermi, *Lincei Rend.* **9**, 881 (1929); **12**, 431 (1930); *Rev. Mod. Phys.* **4**, 87 (1932).

37. P. Jordan and W. Pauli, *Z. f. Phys.* **47**, 151 (1928).

38. N. Bohr and L. Rosenfeld, *Kon. dansk. vid. Selsk., Mat.-Fys. Medd.* **XII**, No. 8 (1933) (translation in *Selected Papers of Leon Rosenfeld*, ed. by R. S. Cohen and J. Stachel (Reidel, Dordrecht, 1979)); *Phys. Rev.* **78**, 794 (1950).

39. P. Jordan, *Z. f. Phys.* **44**, 473 (1927). Also see P. Jordan and O. Klein, *Z. f. Phys.* **45**, 751 (1929); P. Jordan, *Phys. Zeit.* **30**, 700 (1929).

40. P. Jordan and E. Wigner, *Z. f. Phys.* **47**, 631 (1928). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

40*a*. M. Fierz, *Helv. Phys. Acta* **12**, (1939); W. Pauli, *Phys. Rev.* **58**, 716 (1940); W. Pauli and F. J. Belinfante, *Physica* **7**, 177 (1940).

41. W. Heisenberg and W. Pauli, *Z. f. Phys.* **56**, 1 (1929); *ibid.*, **59**, 168 (1930).

42. P. A. M. Dirac, *Proc. Roy. Soc.* **A136**, 453 (1932); P. A. M. Dirac, V. A. Fock, and B. Podolsky, *Phys. Zeit. der Sowjetunion* **2**, 468 (1932); P. A. M. Dirac, Phys. *Zeit. der Sowjetunion* **3**, 64 (1933). The latter two articles are reprinted in *Quantum Electrodynamics*, Ref. 35, pp. 29 and 312. Also see L. Rosenfeld, *Z. f. Phys.* **76**, 729 (1932).

42*a*. P. A. M. Dirac, *Proc. Roy. Soc.* London **A136**, 453 (1932).

43. E. Fermi, *Z. f. Phys.* **88**, 161 (1934). Fermi quotes unpublished work of Pauli for the proposition that an unobserved neutral particle is emitted along with the electron in beta decay. This particle was called the neutrino to distinguish it from the recently discovered neutron.

43*a*. V. Fock, *C. R. Leningrad* 1933, p. 267.

44. W. H. Furry and J. R. Oppenheimer, *Phys. Rev.* **45**, 245 (1934). This paper uses a density matrix formalism developed by P. A. M. Dirac, Proc. Camb. *Phil. Soc.* **30**, 150 (1934). Also see R. E. Peierls, *Proc. Roy. Soc.* **146**, 420 (1934); W. Heisenberg, *Z. f. Phys.* **90**, 209 (1934); L. Rosenfeld, *Z. f. Phys.* **76**, 729 (1932).

45. W. Pauli and V. Weisskopf, *Helv. Phys. Acta* **7**, 709 (1934), reprinted in English translation in A. I. Miller, *Early Quantum Electrodynamics* (Cambridge University Press, Cambridge, 1994). Also see W. Pauli, *Ann. Inst. Henri Poincaré* **6**, 137 (1936).

46. O. Klein and Y. Nishina, *Z. f. Phys.* **52**, 853 (1929); Y. Nishina, *ibid.*, 869 (1929); also see I. Tamm, *Z. f. Phys.* **62**, 545 (1930).

47. P. A. M. Dirac, *Proc. Camb. Phil. Soc.* **26**, 361 (1930).

48. C. Møller, *Ann. d. Phys.* **14**, 531, 568 (1932).

49. H. Bethe and W. Heitler, *Proc. Roy. Soc.* **A146**, 83 (1934); also see G. Racah, *Nuovo Cimento* **11**, No. 7 (1934); *ibid.*, **13**, 69 (1936).

50. H. J. Bhabha, *Proc. Roy. Soc.* **A154**, 195 (1936).

50*a*. J. F. Carlson and J. R. Oppenheimer, *Phys. Rev.* **51**, 220 (1937).

51. P. Ehrenfest and J. R. Oppenheimer, *Phys. Rev.* **37**, 333 (1931).

52. W. Heitler and G. Herzberg, *Naturwiss.* **17**, 673 (1929); F. Rasetti, *Z. f. Phys.* **61**, 598 (1930).

53. J. Chadwick, *Proc. Roy. Soc.* **A136**, 692 (1932). This article is reprinted in *The Foundations of Nuclear Physics*, Ref. 30.

54. W. Heisenberg, *Z. f. Phys.* **77**, 1 (1932); also see I. Curie-Joliot and F. Joliot, *Compt. Rend.* **194**, 273 (1932).

54*a*. For references, see L. M. Brown and H. Rechenberg, *Hist. Stud. in Phys. and Bio. Science*, **25**, 1 (1994).

55. H. Yukawa, *Proc. Phys.-Math. Soc. (Japan)* (3) **17**, 48 (1935). This article is reprinted in *The Foundations of Nuclear Physics*, Ref. 30.

56. S. H. Neddermeyer and C. D. Anderson, *Phys. Rev.* **51**, 884 (1937); J. C. Street and E. C. Stevenson, *Phys. Rev.* **52**, 1003 (1937).

56*a*. L. Nordheim and N. Webb, *Phys. Rev.* **56**, 494 (1939).

57. M. Conversi, E. Pancini, and O. Piccioni, *Phys. Rev.* **71**, 209L (1947).

58. S. Sakata and T. Inoue, *Prog. Theor. Phys.* **1**, 143 (1946); R. E. Marshak and H. A. Bethe, *Phys. Rev.* **77**, 506 (1947).

59. C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, *Nature* **160**, 453, 486 (1947).

60. G. D. Rochester and C. C. Butler, *Nature* **160**, 855 (1947).

61. J. R. Oppenheimer, *Phys. Rev.* **35**, 461 (1930).

62. I. Waller, *Z. f. Phys.* **59**, 168 (1930); *ibid.*, **61**, 721, 837 (1930); *ibid.*, **62**, 673 (1930).

63. V. F. Weisskopf, *Z. f. Phys.* **89**, 27 (1934), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45; *ibid.*, **90**, 817 (1934). The electromagnetic self energy is calculated only to lowest order in α in these references; the proof that the divergence

is only logarithmic in all orders of perturbation theory was given by Weisskopf; *Phys. Rev.* **56**, 72 (1939). (This last article is reprinted in *Quantum Electrodynamics*, Ref. 35).

64. P. A. M. Dirac, XVII Conseil Solvay de Physique, p. 203 (1933), reprinted in *Early Quantum Electrodynamics*, Ref. 45. For subsequent calculations based on less restrictive assumptions, see W. Heisenberg, *Z. f. Phys.* **90**, 209 (1934); *Sachs. Akad. Wiss.* **86**, 317 (1934); R. Serber, *Phys. Rev.* **43**, 49 (1935); E. A. Uehling. *Phys. Rev.* **48**, 55 (1935); W. Pauli and M. Rose, *Phys. Rev.* **49**, 462 (1936). Also see Furry and Oppenheimer, Ref. 44; Peierls. Ref. 44; Weisskopf, Ref. 63.

65. H. Euler and B. Kockel, *Naturwiss.* **23**, 246 (1935); W. Heisenberg and H. Euler, *Z. f. Phys.* **98**, 714 (1936).

66. P. A. M. Dirac, *Proc. Camb. Phil. Soc.* **30**, 150 (1934).

67. W. Heisenberg, *Z. f. Phys.* **90**, 209 (1934).

68. N. Kemmer and V. F. Weisskopf. *Nature* **137**, 659 (1936).

68*a*. F. Bloch and A. Nordsieck, *Phys. Rev.* **52**, 54 (1937). Also see W. Pauli and M. Fierz, *Nuovo Cimento* **15**, 167 (1938), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45.

69. S. M. Dancoff, *Phys. Rev.* **55**, 959 (1939).

69*a*. H. W. Lewis, *Phys. Rev.* **73**, 173 (1948); S. Epstein, *Phys. Rev.* **73**, 177 (1948). Also see J. Schwinger, Ref. 84; Z. Koba and S. Tomonaga, *Prog. Theor. Phys.* **3**/3, 290 (1948).

69*b*. J. Schwinger, in *The Birth of Particle Physics*, ed. by L. Brown and L. Hoddeson (Cambridge University Press, Cambridge, 1983): p. 336.

70. W. Heisenberg, *Ann. d. Phys.* **32**, 20 (1938), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45.

70*a*. G. Wentzel, *Z. f. Phys.* **86**, 479, 635 (1933); *Z. f. Phys.* **87**, 726 (1034); M. Born and L. Infeld, *Proc. Roy. Soc.* **A150**. 141 (1935); W. Pauli, *Ann. Inst. Henri Poincaré* **6**, 137 (1936).

71. J. A. Wheeler, *Phys. Rev.* **52**, 1107 (1937).

72. W. Heisenberg, *Z. f. Phys.* **120**, 513, 673 (1943); *Z. Naturforsch.* **1**, 608 (1946). Also see C. Møller, *Kon. Dansk. Vid. Sels. Mat.-Fys. Medd.* **23**, No. 1 (1945); *ibid.* **23**, No. 19, (1946).

73. See, e.g., G. Chew, *The S-Matrix Theory of Strong Interactions* (W. A. Benjamin, Inc. New York, 1961).

74. J. A. Wheeler and R. P. Feynman, *Rev. Mod. Phys.* **17**, 157 (1945), *ibid.*, **21**, 425 (1949). For further references and a discussion of the application of action-at-a-distance theories in cosmology, see S. Weinberg *Gravitation and Cosmology*, (Wiley, 1972): Section 16.3.

75. P. A. M. Dirac, *Proc. Roy. Soc.* **A180**, 1 (1942). For a criticism, see W. Pauli, *Rev. Mod. Phys.* **15**, 175 (1943). For a review of classical theories of this type, and of yet other attempts to solve the problem of infinities, see R. E. Peierls in *Rapports du 8^{m}e Conseil de Physique Solvay 1948* (R. Stoops, Brussels, 1950): p. 241.

76. V. F. Weisskopf, *Kon. Dan. Vid. Sel., Mat.-fys. Medd.* **XIV**, No. 6 (1936), especially p. 34 and pp. 5–6. This article is reprinted in *Quantum Electrodynamics*, Ref. 35, and in English translation in *Early Quantum Electrodynamics*, Ref. 45. Also see W. Pauli and M. Fierz, Ref. 68*a*; H. A. Kramers, Ref. 79*a*.

77. S. Pasternack, *Phys. Rev.* **54**, 1113 (1938). This suggestion was based on experiments of W.V. Houston, *Phys. Rev.* **51**, 446 (1937); R. C. Williams, *Phys. Rev.* **54**, 558 (1938). For a report of contrary data, see J. W. Drinkwater, O. Richardson, and W. E. Williams, *Proc. Roy. Soc.* **174**, 164 (1940).

78. E. A. Uehling, Ref. 64.

79. W. E. Lamb, Jr and R. C. Retherford, *Phys. Rev.* **72**, 241 (1947). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

79*a*. H. A. Kramers, *Nuovo Cimento* **15**, 108 (1938), reprinted in English translation in *Early Quantum Electrodynamics*, Ref. 45; *Ned. T. Natwink.* **11**, 134 (1944); *Rapports du 8^{m}e Conseil de Physique Solvay 1948* (R. Stoops, Brussels, 1950).

80. H. A. Bethe, *Phys. Rev.* **72**, 339 (1947). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

81. J. B. French and V. F. Weisskopf; *Phys. Rev.* **75**, 1240 (1949); N. M. Kroll and W. E. Lamb, *ibid.*, **75**, 388 (1949); J. Schwinger, *Phys. Rev.* **75**, 898 (1949); R. P. Feynman, *Rev. Mod. Phys.* **20**, 367 (1948); *Phys. Rev.*, **74**, 939, 1430 (1948); **76**, 749, 769 (1949); **80**, 440 (1950); H. Fukuda, Y. Miyamoto, and S. Tomonaga, *Prog. Theor. Phys. Rev. Mod. Phys.* **4**, 47, 121 (1948). The article by Kroll and Lamb is reprinted in *Quantum Electrodynamics*, Ref. 35.

82. J. E. Nafe, E. B. Nelson, and I. I. Rabi, *Phys. Rev.* **71**, 914 (1947); D. E. Nagel, R. S. Julian, and J. R. Zacharias, *Phys. Rev.* **72**, 973 (1947).

83. P. Kusch and H. M. Foley, *Phys. Rev.* **72**, 1256 (1947).

83*a*. G. Breit, *Phys. Rev.* **71**, 984 (1947). Schwinger in Ref. 84 includes a corrected version of Breit's results.

84. J. Schwinger, *Phys. Rev.* **73**, 416 (1948). This article is reprinted in *Quantum Electrodynamics*, Ref. 35.

85. J. Schwinger, *Phys. Rev.* **74**, 1439 (1948); *ibid.*, **75**, 651 (1949); *ibid.*, **76**, 790 (1949); *ibid.*, **82**, 664, 914 (1951); *ibid.*, **91**, 713 (1953); *Proc. Nat. Acad. Sci.* **37**, 452 (1951). All but the first two of these articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

86. S. Tomonaga, *Prog. Theor. Phys. Rev. Mod. Phys.* **1**, 27 (1946); Z. Koba, T. Tati, and S. Tomonaga, *ibid.*, **2**, 101 (1947); S. Kanesawa and S. Tomonaga, *ibid.*, **3**, 1, 101 (1948); S. Tomonaga, *Phys. Rev.* **74**, 224 (1948); D. Ito, Z. Koba, and S. Tomonaga, *Prog. Theor. Phys.* **3**, 276 (1948); Z. Koba and S. Tomonaga, *ibid.*, **3**, 290 (1948). The first and fourth of these articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

87. R. P. Feynman, *Rev. Mod. Phys.* **20**, 367 (1948); *Phys. Rev.* **74**, 939, 1430 (1948); *ibid.*, **76**, 749, 769 (1949); *ibid* **80**, 440 (1950). All but the second and third of these articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

88. F. J. Dyson, *Phys. Rev.* **75**, 486, 1736 (1949). These articles are reprinted in *Quantum Electrodynamics*, Ref. 35.

88*a*. H. Fröhlich, W. Heitler, and B. Kahn, *Proc. Roy. Soc.* **A171**, 269 (1939); *Phys. Rev.* **56**, 961 (1939).

88*b*. W. E. Lamb, Jr, *Phys. Rev.* **56**, 384 (1939); *Phys. Rev.* **57**, 458 (1940).

89. Quoted by R. Serber, in *The Birth of Particle Physics*, Ref. 69*b*, p. 270.