PERGAMON

Studies in History and Philosophy of Modern Physics 34 (2003) 621–650

Studies in History and Philosophy of Modern Physics

www.elsevier.com/locate/shpsb

Weinberg's proof of the spin-statistics theorem

Michela Massimi*, Michael Redhead

Centre for Philosophy of Natural and Social Sciences, London School of Economics and Political Science, London WC2A 2AE, UK

Abstract

The aim of this paper is to offer a conceptual analysis of Weinberg's proof of the spin-statistics theorem by comparing it with Pauli's original proof and with the subsequent textbook tradition, which typically resorts to the dichotomy positive energy for half-integral spin particles/microcausality for integral-spin particles. In contrast to this tradition, Weinberg's proof does not directly invoke the positivity of the energy, but derives the theorem from the single relativistic requirement of microcausality. This seemingly innocuous difference marks an important change in the conceptual basis of quantum physics. Its historical, theoretical, and conceptual roots are here reconstructed. The link between Weinberg's proof and Pauli's original is highlighted: Weinberg's proof turns out to do justice to Pauli's anti-Dirac lines of thought. The work of Furry and Oppenheimer is also surveyed as a "third way" between the textbook tradition established by Pauli and Weinberg's approach.

Keywords: Spin-statistics theorem; Microcausality; Pauli's Exclusion Principle; *S* matrix theory

1. Introduction

The spin-statistics theorem has not ceased to represent a challenge to human understanding since Pauli (1940) originally presented it. As Richard Feynman (1963) once said,

> Why is it that particles with half-integral spin are Fermi particles (…), whereas particles with integral spin are Bose particles (…)? We apologize for the fact that we cannot give you an elementary explanation. An explanation has been worked

*Corresponding author. Present address: Girton College, Cambridge CB3 0JG, UK.
E-mail addresses: mm483@cam.ac.uk (M. Massimi), mlr1000@cam.ac.uk (M. Redhead).

1355-2198/$ - see front matter
doi:10.1016/S1355-2198(03)00066-2

out by Pauli from complicated arguments of quantum field theory and relativity (Chapter 4, Section 1).

The spin-statistics theorem called for a simple proof, and several attempts have been made in this direction in the past sixty years as the increasing literature on the subject testifies. For example, Feynman (1949a, b) derived the theorem from the Feynman Rules; Schwinger (1951) from strong-reflection symmetry; Lüders and Zumino (1958) and Burgoyne (1958) resorted to the axioms for quantum field theory; more recently, Berry (in Berry & Robbins, 1997) and Sudarshan (in Duck & Sudarshan, 1998) have offered a proof based on topological arguments and rotational invariance, respectively. A detailed and well-documented account of many of the alternative proofs of the spin-statistics theorem can be found in Duck and Sudarshan's (1997) source book.

The aim of this paper is to focus on a relatively neglected proof, which is not even discussed in Duck and Sudarshan's source book: Weinberg's proof. As we will argue in this paper, Weinberg's proof seems to us a good candidate for the "elementary explanation" of the spin-statistics theorem that Feynman was hoping for. In contrast to a well-established textbook tradition that typically resorts to the postulate of the positive energy to get the spin-statistics connection for half-integral spin particles, Weinberg derived the spin-statistics connection for both bosons and fermions from the single requirement of microcausality suitably interpreted.

Why has Weinberg's proof been so far overlooked? Maybe because it was originally presented within Weinberg's research programme on a Lorentz-invariant S matrix: the spin-statistics connection was disparagingly presented as a side result of this more general programme. Since then, the theoretical elegance and the conceptual implications of Weinberg's proof have been widely neglected in the literature on the spin-statistics theorem: in this paper we hope to do justice to it, and indirectly to Pauli himself. In fact, the upshot of our paper is to reconcile Weinberg's proof with Pauli's by showing that Weinberg has vindicated Pauli's original intentions: a derivation of the spin-statistics connection solely from relativity theory, in particular from the microcausality condition. As we will see, Pauli's project failed. He had to make recourse to the additional postulate of the positivity of the energy, and accordingly, to his old enemy, Dirac's hole theory, which warranted this postulate. In this respect, Weinberg's proof serves as an admirable foil for Pauli's proof and for the subsequent textbook tradition.

In Section 2, we briefly reconstruct the history of Pauli's proof as a part of his long-lasting polemic against Dirac's hole theory from the Pauli–Weisskopf "anti-Dirac" paper in 1934 to Pauli's first incomplete proof in 1936 until the final one in 1940. At the same time, we reconstruct the conceptual origins of the postulate of positive energy as grounded in Dirac's hole theory.

In Section 3, the persistence of the hole picture behind the postulate of the positive energy will appear in a recent version of the theorem developed by Greiner and Reinhardt (1996). Greiner and Reinhardt's proof will be presented as a paradigmatic example of the textbook tradition originally established by Pauli. In Section 4, we finally present Weinberg's proof, while in Section 5 the requirement of microcausality, on which the proof hinges, will be more closely analysed. However, before any analysis can be undertaken, it is necessary to briefly recall the theorem.

As is well known, in the original semi-classical model (Uhlenbeck & Goudsmit, 1925, 1926), the spin was regarded as the intrinsic angular momentum of a spinning electron. Later it was realized that the spin has no classical analogue; it is simply the eigenvalue s of the spin operator S acting in the spin state space E_S, and it can be either *integral* ($s = 1, 2, \ldots$), in the case for instance of photons and mesons, or *half-integral* ($s = \frac{1}{2}, \frac{3}{2}, \ldots$) in the case of electrons, protons, positrons, neutrons, and muons, among others.

Quantum statistics date back to 1924 and 1926, respectively. First Bose (1924) and then Einstein (1924, 1925a, b) elaborated the statistics for a photon gas and the Bose–Einstein distribution for the corresponding ideal gas. In 1926, it was first Fermi's (1926) turn, and Dirac's (1926) shortly afterwards, to formulate the quantum statistics for an ideal gas of identical particles obeying Pauli's Exclusion Principle. According to the *Bose–Einstein statistics*, indistinguishable particles are allowed to be in only symmetric states. The *Fermi–Dirac statistics*, on the other hand, allows indistinguishable particles to be in only antisymmetric states. Just to recall briefly, symmetric states for—say—two indistinguishable particles are such that the state vector of the composite system does not change sign under permutation of the space and spin coordinates of the two particles:

$$1/\sqrt{2}\left(|a_1^r\rangle \otimes |a_2^s\rangle + |a_1^s\rangle \otimes |a_2^r\rangle\right), \tag{1}$$

where $|a_q^r\rangle$ is the eigenvector of a 1-particle operator A associated with the eigenvalue a^r for the qth particle (with $q = 1, 2$).

In antisymmetric states, on the other hand, the state vector *does* change sign under permutation of the space and spin coordinates of the two particles:

$$1/\sqrt{2}\left(|a_1^r\rangle \otimes |a_2^s\rangle - |a_1^s\rangle \otimes |a_2^r\rangle\right). \tag{2}$$

To say that the state vector of the composite system is antisymmetric is mathematically equivalent to saying that the entangled states of the two particles are different (otherwise the antisymmetric state vector would vanish). This is nothing but Pauli's Exclusion Principle forbidding the possibility of any two indistinguishable particles being in the same dynamic state (Pauli, 1925).

The two quantum statistics define two different kinds of particles:

- *fermions* are particles obeying Pauli's Exclusion Principle and the corresponding Fermi–Dirac statistics;
- *bosons* are particles not obeying the Exclusion Principle and following the Bose–Einstein statistics, which allows more than one particle per state.

The theoretical breakthrough of the spin-statistics theorem consists in proving the existence of a strict connection between spin and quantum statistics, namely:

- half-integral spin particles follow the Fermi–Dirac statistics (they are *fermions*);
- integral spin particles follow the Bose–Einstein statistics (they are *bosons*).

Prima facie, the theorem seems to imply the following dichotomy: all elementary particles divide into fermions or bosons. But as a matter of fact, the spin-statistics theorem does not rely on this dichotomy, which has attained an undeserved status through the so-called symmetrization postulate[1] and has indeed been questioned by the development of parastatistics since the 1960s (see Messiah & Greenberg, 1964). The aim and scope of the spin-statistics theorem is only to forge a link between the kind of spin a particle has and the kind of quantum statistics the particle does not follow, without ruling out *a priori* the possibility of intermediate (neither symmetric nor antisymmetric) states.

But it is now time to turn our attention to the proof of the theorem. In the following historical interlude, we present the way Pauli arrived at the spin-statistics theorem through his long-lasting and severe criticism of Dirac's hole theory.

2. Pauli and the spin-statistics theorem: historical interlude

The spin-statistics connection was originally pursued in the attempt to find a relativistic theory for particle creation and annihilation alternative to the Dirac hole theory. The theorem was the conclusion of Pauli's theoretical fight against Dirac's theory; hence, we have to begin our historical reconstruction from it.

In 1928, Dirac announced his equation for the electron (Dirac, 1928), in modern notation:

$$\left[\mathrm{i}\sum_{\mu}\gamma^{\mu}p_{\mu}+mc\right]\psi=0, \tag{3}$$

where γ^{μ} are the four Dirac matrices with $\mu = 0, 1, 2, 3$ satisfying $\gamma^{\mu}\gamma^{\nu}+\gamma^{\nu}\gamma^{\mu}=2g^{\mu\nu}$ where $g^{\mu\nu}$ is the contravariant Lorentz metric diag $(1, -1, -1, -1)$, and ψ is a four-component vector, the so-called Dirac spinor.

The Dirac equation (see Wightman, 1972) was meant to overcome some difficulties affecting another relativistic wave equation that Gordon (1926), and independently Klein (1927), had earlier introduced. The dreamt-of relativistic wave equation was in fact expected to separate the negative energy solutions from the positive energy ones, but the Klein–Gordon equation allowed a superposition of both. Despite its theoretical and empirical successes,[2] the Dirac equation fared no better on the score of the negative energy solutions. As Klein quickly retorted against Dirac, in the case of time-dependent external fields, transitions from positive to negative energy states also affected the Dirac equation (the Klein Paradox). The

[1] The symmetrization postulate asserts that only certain kets of the state space of a system including several indistinguishable particles can describe its physical states. Depending on the nature of the particles (bosons or fermions) physical kets are either completely symmetric or antisymmetric with respect to permutation of these particles.

[2] The Dirac equation allowed the derivation of the value 2 of Landé's g factor as the electron spin magnetic moment. Sommerfeld's formula for the fine structure of the hydrogen spectrum was also derived from the Dirac equation, and so was the Klein–Nishina formula for the scattering of light by electrons.

existence of negative energy states was a conundrum that beset Dirac for the next few years.

Shortly after the breakthrough of 1928, Dirac returned to this problem (Dirac, 1930). Remarking that no hard-and-fast distinction between positive and negative energy solutions was available in quantum theory, Dirac identified the wave packet constituted by the superposition of the negative energy solutions as describing the motion of

> an electron of charge $+e$ (and positive energy) moving in the original electromagnetic field. Thus *an electron with negative energy moves* in an *external field as though it carries a positive charge* (1930, p. 361; emphasis in the original).

Dirac was very close to the introduction of the antiparticle of the electron: the negative energy solutions of his equation were associated with allegedly positive-charged electrons. Antiparticles would have been easily available at this point, if Dirac had not followed Weyl's (1929) mistaken identification of the negative energy electrons with protons. This identification violated the conservation of the electric charge, and—most importantly—it was incompatible with the different masses of electrons and protons. But as Dirac himself later recalled,[3] at that time he was much more concerned with getting a satisfactory theory of the electron than with bringing in protons; and indeed, the real upshot of Dirac's (1930) paper was to bring in the Fermi–Dirac statistics via the Exclusion Principle. In fact, in the second section of the paper, Dirac introduced what is known in the literature as the Dirac "negative energy sea" (Dirac, 1930):

> The most stable states for an electron (i.e., the states of lowest energy) are those with negative energy and very high velocity. All the electrons in the world will tend to fall into these states with emission of radiation. The Pauli's Exclusion Principle, however, will come into play and prevent more than one electron going into any one state. Let us assume there are so many electrons in the world that all the most stable states are occupied, or, more accurately, that *all the states of negative energy are occupied except perhaps a few of small velocity.* (…) *Only the small departures from exact uniformity, brought about by some of the negative energy states being unoccupied, can we hope to observe* (p. 362; emphasis in the original).

The few vacant states or "holes"—as Dirac called them—in the negative energy sea were introduced by analogy with X-rays, emitted when an internal vacancy in the electronic configuration occurs. However, while in the X-ray, the holes count as states of negative energy because ordinary positive energy electrons are required to fill them up; in the Dirac negative energy sea, the holes counted as states of positive energy because negative energy electrons were required to fill them up. Thus, the holes were supposed to behave like ordinary particles with a positive charge $+e$. Whenever an electron jumped from a negative energy state to a positive energy one,

[3] See Kuhn's interview with Dirac (14 May 1963) in Kuhn and Heilbron (1967).

it left behind a hole (electron–hole creation); *vice versa*, when it jumped from a positive energy state to a negative energy one, there was a process of electron–hole annihilation. The processes of pair creation/annihilation resembled the similar processes of emission/absorption of radiation that in 1927 Dirac had described in terms of transitions from an unobservable "zero state."[4]

The hole theory was a useful heuristic tool to deal with the problem of negative energy states. However, it faced several difficulties. First, the infinite number of negative energy states was expected to produce an electric field of infinite divergence according to Maxwell's equation

$$\mathrm{div}\,\mathbf{E} = 4\pi\rho. \tag{4}$$

Second, the different masses of electrons and protons had to be explained. A further difficulty concerned the infinite annihilation probability of electrons and protons. Oppenheimer (1930) pointed out that electrons and protons would annihilate each other so that the mean lifetime for matter would be of the order of 10^{-10} s. When Weyl (1930) finally proved that the masses of the positive- and negative-charged electrons had to be identical, Dirac published a paper in which the identification of the holes with protons was abandoned, and antiparticles were for the first time explicitly introduced (Dirac, 1931):

> A hole, if there were one, would be a new kind of particle unknown to experimental physics, having the same mass and opposite charge to an electron. We may call such a particle an anti-electron (p. 61).

With almost prophetic words, Dirac anticipated—via the hole theory—the discovery of the positron, detected two years later by Anderson (1933) in the photographs of cosmic-ray tracks in a Wilson cloud chamber. The discovery of the positron vindicated the hole theory, which in 1931 Dirac himself was about to give up as a sick idea because of the impossibility of making any further progress with it.[5]

Antiparticles were then conceptually introduced as "holes in a negative energy sea." Indeed, they were entities *negatively* defined as absence (holes) with respect to a plenitude of being (negative energy sea). They were the necessary—until then, missing—link of a sort of great chain of being (echoing Lovejoy's famous expression) going from a lower to an upper bound of a uniformly distributed and filled continuum of energy states: they were introduced by the *principle of plenitude* and *continuity* (see Lovejoy, 1936) to fill up the few unoccupied links of an infinite chain.

But the principles of plenitude and continuity were not the only conceptual path to antiparticles. Physicists who did not have a penchant for Dirac's hole theory undertook a different conceptual route to antiparticles; through this same route, the

[4] "When a light-quantum is absorbed it can be considered to jump into a zero state (i.e., stationary state of an atom), and when one is emitted it can be considered to jump from the zero state to one in which it is physically in evidence, so that it appears to have been created." (Dirac, 1927, pp. 260–261).

[5] See Kuhn's interview with Dirac (14 May 1963) in Kuhn and Heilbron (1967).

spin-statistics theorem was introduced. Before reconstructing Pauli's way to the theorem, two important theoretical steps against the hole theory must be mentioned: the first was taken by Fermi, and the second by Furry and Oppenheimer.

Despite its success in dealing with negative energy states, the hole theory soon turned out not to be an entertainable method to account for the theory of β-decay. As Pauli wrote to Heisenberg in 1933, "the β-decay weighs on the entire [hole] theory as a big X."[6] If the hole theory had been correct, one would expect a neutron to decay into a proton, an electron, and a positron. But this would violate conservation of charge. Moreover, the decay $n \rightarrow p, e^-$ was forbidden by the conservation of angular momentum. It was for this latter reason, and also to accommodate the continuous energy spectrum for the electron, that Pauli postulated a new kind of particle with mass equal or inferior to the mass of the electron and with null electric charge: the neutrino.[7] But where did the electron and the neutrino originate? Indeed, as Pauli soon realized, "the hole-question and the neutrino-question let themselves be solved only together."[8] Fermi's theory of β-decay (1934) provided an answer to both questions. The hole picture could not be maintained in the physical scenario of Fermi's theory. The electrons emitted in β-decay could not be taken as particles jumping from the negative energy sea to a positive energy state; i.e., their creation process could not be interpreted as the creation of an electron–hole pair. Electrons and, indeed, neutrinos were created in β-decay "out of nothing" so to speak (or, more strictly speaking, out of the energy released in the nuclear transmutation); i.e., without assuming a hypothetical negative energy sea from which they would jump by leaving behind a hole. In fact, if the hole picture had been right, one would have expected the total number of electrons and neutrinos to remain constant. But this was not the case in Fermi's theory.[9]

[6] Pauli to Heisenberg, 14 July 1933 (Pauli, 1985, p. 187).

[7] Pauli originally postulated the existence of this new particle in 1931 to interpret the spectrum of β-rays. At that time two rival hypotheses were available to explain this phenomenon: Bohr claimed that in β-decay, the energy was not strictly conserved; Pauli, on the other hand, contended that the energy was strictly conserved, and the apparent violation was rather due to the emission of another type of particle, not yet observed, having null electric charge and spin-$\frac{1}{2}$ so that the sum of the energies of this new particle and the electron was constant. The new particle that Pauli originally called "neutron" was later called "neutrino" by Fermi, after Chadwick's experimental discovery of neutron in 1932.

[8] Pauli to Heisenberg, 11 November 1933 (Pauli, 1985, p. 226).

[9] Fermi arrived at this result by following an analogy with a theory of emission of light-quanta very different from Dirac's (1927) radiation theory, which—as mentioned in fn. 4—had anticipated the hole picture in the idea of a zero state from which light-quanta were emitted. And since in this other theory the total number of light-quanta was not constant because "light-quanta emerge when they are emitted from an atom, and disappear when they are absorbed," Fermi claimed that "analogously, we want here to ground the theory of the emission of β-rays on the following assumptions: (a) the total number of electrons, as well as that of neutrinos, is not necessarily constant. Electrons (or neutrinos) can emerge and disappear. However, this possibility has *no analogy with the creation and annihilation of an electron–positron pair*; indeed, if the positron is interpreted as Dirac's hole, the latter process can easily be interpreted as an electron jumping from a state with negative energy to one with positive energy with conservation of the total number of the electrons" (Fermi, 1934, pp. 161–162; emphasis added).

The next theoretical step against the hole theory implied a radical revision of the picture of positrons as holes. More or less at the same time as Fermi was elaborating his theory of β-decay, Furry and Oppenheimer (1934) offered a reinterpretation of the hole theory "without holes,"[10] which anticipated Weinberg's proof of the spin-statistics theorem in some important formal aspects. In an article published on 15 February 1934, Furry and Oppenheimer announced a development of Dirac's theory into a perfectly symmetric theory of electrons and positrons:

> The formal changes that are required in the theory are simple, and correspond closely to Dirac's most recent suggestion (1931) for interpreting the negative kinetic energy states (p. 245).

The formal changes at issue resulted in the second quantization of Dirac's theory in its (almost) current standard formalism, in which any reference to the hole picture has been lost. The second quantization of Dirac's theory began with the introduction of the occupation number N_r to describe a system of N indistinguishable particles satisfying the Exclusion Principle, whose eigenvalues were $N_r = 1$ if a particle was present in the state r, or $N_r = 0$ if no particle was present. Accordingly, the wave function of the system was given by

$$\psi(0\cdots 1_{r1}\cdots 0\cdots 1_{r2}\cdots 1_{rN}\cdots) = (N!)^{1/2}\psi(r_1\cdots r_N), \tag{5}$$

where the LHS of (5) refers to the occupation number representation (with the occupied states labelled as $r_1, r_2, \ldots, r_N$), while the RHS $\psi(r_1\cdots r_N)$ specifies the probability amplitude for particle 1 to be in state r, particle 2 in state r_2, etc. (the states were assumed to be denumerable).

Any dynamical variable of the system expressible as a symmetric sum of 1-particle operators was introduced as an operator on this wave function:

$$\sum_{r,r'} \Omega_{r'r} a^\dagger_{r'} a_r, \tag{6}$$

where $\Omega_{r'r}$ is the 1-particle matrix element connecting the rth and r'th states and with the creation and annihilation operators $a^\dagger_{r'}, a_r$, satisfying the following anticommutation relations:

$$a_r a_s + a_s a_r = 0; \quad a^\dagger_r a^\dagger_s + a^\dagger_s a^\dagger_r = 0; \quad a^\dagger_r a_s + a_s a^\dagger_r = \delta_{rs}. \tag{7}$$

The occupation numbers were expressed as

$$N_r = a^\dagger_r a_r; \quad 1 - N_r = a_r a^\dagger_r. \tag{8}$$

The crucial formal change, whose physical meaning corresponded to Dirac's holes without however assuming them any longer, consisted in a radical conceptual reinterpretation of the unoccupied negative energy states as *positive particles in their*

[10] For a similar theoretical move, see Fock (1933).

own right. Furry and Oppenheimer restricted the states r into three possible categories: wholly positive (denoted by the Latin letter r), wholly negative (denoted by the Greek letter ρ) and superpositions of positive and negative states (denoted by the brackets (r)). As a consequence of this restriction, the troublesome transitions of positive energy states into negative ones, which typically affected Dirac's theory, were avoided. Most importantly, any dynamical variable of a system containing N_r electrons and M_ρ positrons of diagonal form should now be associated with the corresponding operator

$$\sum_r \Omega_{rr} N_r + \sum_\rho (-\Omega_{\rho\rho}) M_\rho. \tag{9}$$

Formally, by subtracting $\sum_-(\Omega) = \sum_\rho \Omega_{\rho\rho}$ from the operator $\sum_{r,r'} \Omega_{r'r} a^\dagger_{r'} a_r$, a new operator $\tilde{\omega}$ was introduced

$$\tilde{\omega} = \sum_{(r)(s)} \Omega_{(r)(s)} \omega_{(r)(s)}$$

with

$$\omega_{r(s)} = a^\dagger_r a_{(s)}; \quad \omega_{(r)s} = a^\dagger_{(r)} a_s; \quad \omega_{\rho\sigma} = a^\dagger_\rho a_\sigma - \delta_{\rho\sigma} = -a_\sigma a^\dagger_\rho. \tag{10}$$

Note the switch of the creation and annihilation operators $a^\dagger_\rho a_\sigma$ in the field operator $\omega_{\rho\sigma}$ of the negative energy states. This switch brought along with it a redefinition of the Hamiltonian as the sum of the positive kinetic energies of the electrons and the positive kinetic energies $(-T_{\rho\rho})$ of the *positrons*.

$$\sum_r N_r T_{rr} - \sum_\rho M_\rho T_{\rho\rho}. \tag{11}$$

As a result, the Hamiltonian turned out to be positive semi-definite without resorting to Dirac's picture of a fully occupied negative energy sea. From a more modern perspective, the Furry–Oppenheimer prescription is equivalent to normal ordering in which creation operators are placed to the left and annihilation operators to the right, and for Dirac fields the sign is reversed on switching the order of the creation and annihilation operators.[11]

The conceptual reinterpretation of antiparticles as positive particles in their own right and the consequent redefinition of the Hamiltonian with positive energy put in "by hand"—so to speak—anticipates in some important aspects Weinberg's proof of the spin-statistics theorem. Furry and Oppenheimer unwittingly opened up the possibility of a uniform proof of the spin-statistics theorem from microcausality

[11] Normal ordering was introduced by Wick (1950) (see also Houriet & Kind, 1949; Dyson, 1949). Normal ordering for the free Dirac Hamiltonian leads from the not positive semi-definite $\hat{H} = \int d^3p\omega_p(\hat{b}^\dagger\hat{b} - \hat{d}\hat{d}^\dagger)$ to the following positive semi-definite $\hat{H} = \int d^3p\omega_p(\hat{b}^\dagger\hat{b} + \hat{d}^\dagger\hat{d})$ (for notation, see Section 3). Normal ordering is sometimes justified by the fact that it is provably equivalent to symmetrizing the theory with respect to the Dirac field and its charge conjugate. The Furry–Oppenheimer prescription was originally justified by their reinterpretation of the hole picture; i.e., by reinterpreting antiparticles as positive particles in their own right. But the modern point of view regards the Furry–Oppenheimer prescription as freestanding, so to speak.

alone, *formally* analogous to the quite different *S*-matrix approach that Weinberg adopted thirty years later.[12] Furry and Oppenheimer then occupy an intermediate position between Dirac's hole theory and Weinberg's proof of the spin-statistics theorem. They used the hole picture as a ladder that can be thrown away once one has climbed it. Their work constitutes a potential (quantum-field theoretical) "third way" to the spin-statistics theorem, an alternative to the textbook tradition set forth by Pauli as well as to Weinberg's innovative proof.

More or less at the same time, during the winter of 1933–34, Pauli, too, was working on the hole theory. Despite Anderson's recent discovery of the positron, Pauli did not believe in the "holes." As he wrote to Heisenberg on 6 February 1934 (in Pauli, 1985),

> Thus is Dirac's construal of laws of nature set upon Mount Sinai. All is expressed mathematically very elegantly. But physically I am not at all convinced (p. 276).

The main problem of Dirac's hole theory—according to Pauli—was the electron's infinite self-energy; i.e., the energy of the electromagnetic field generated by the electron (electrostatic self-energy) plus the energy of the interaction of the electron with this field (electrodynamic self-energy). A further problem was the vacuum polarization, which Dirac (1934) addressed at the Solvay Congress in October 1933, and whose solution Pauli judged unsatisfactory.

In March, Weisskopf (1934) submitted his contribution on the self-energy of the electron, while in July, the Pauli–Weisskopf "anti-Dirac" paper was submitted to *Helvetica Physica Acta* (Pauli & Weisskopf, 1934).[13] Pauli and Weisskopf tried to force the spin into the quantization of the Klein–Gordon equation on pain of

[12] In fact, the question of the bosonic quantization of the free Dirac field can now be posed as follows: should the commutator properties be used to control the spectrum of the antiparticle number operator $M_\rho = a_\rho^\dagger a_\rho$, or equivalently, in Greiner's notation, $N = \hat{d}^\dagger \hat{d}$ (see Section 3)? If we apply the anticommutation relations for the antiparticle operators $[\hat{d}, \hat{d}^\dagger]_+ = 1$, we get the spectrum $\{0, 1\}$ in agreement with the Exclusion Principle. On the other hand, if we apply the commutation relations $[\hat{d}, \hat{d}^\dagger]_- = 1$, the spectrum is expanded to $\{0, 1, 2, 3, \dots\}$ in agreement with the Bose–Einstein statistics that permits more than one particle per state. As we will see in Section 3, Greiner and Reinhardt use the commutation relations $[\hat{d}, \hat{d}^\dagger]_- = -1$, justified by their roundabout route *via* the hole theory, but not by the Furry–Oppenheimer route, in which positive energy is built into the formalism without any appeal to the hole picture. However, as we will see, Greiner–Reinhardt's commutation relations do not formally rule out bosonic quantization for the Dirac fields, and prevent Greiner and Reinhardt from getting a uniform proof of the spin-statistics theorem from microcausality alone. By contrast, in Section 4 we will see that Weinberg adopts different commutation relations for the antiparticle operators $[\hat{d}, \hat{d}^\dagger]_- = 1$ which reflect the formal change of putting positive energy in "by hand" without harking back to the hole theory, and which will play a major role in the derivation of the spin-statistics theorem from the single requirement of microcausality.

[13] Weisskopf recalled this crucial episode a few years later as follows (Weisskopf, 1983): "At that time (…) the hole theory of the filled vacuum was still the accepted way of dealing with positrons. Pauli called our work the 'anti-Dirac paper.' He considered it a weapon in the fight against the filled vacuum that he never liked. We thought that this theory only served the purpose of a non-realistic example of a theory that contained all the advantages of the hole theory without the necessity of filling the vacuum. We had no idea that the world of particles would abound with spin-zero entities a quarter of a century later. That was the reason why we published it in the venerable but not widely read *Helvetica Physica Acta*" (p. 70).

revoking the gauge and relativistic invariance of the theory as well as the positive semi-definite energy density. This difficulty shed light on the impossibility of applying the Exclusion Principle to spin-0 particles; as Pauli wrote to Heisenberg on 7 November 1934 (in Pauli, 1985),

> Our theory can be implemented only with the Bose–Einstein statistics, because here a necessary connection begins to dawn between spin and statistics (p. 361).

The Pauli–Weisskopf "anti-Dirac" paper smoothed the path to the spin-statistics theorem and yielded an important side result: without resorting to the hole picture, Pauli and Weisskopf introduced antiparticles as positive particles in their own right, like Furry and Oppenheimer did. Antiparticles were the symmetric counterparts of particles, having only a bundle of properties opposite to those of the particles. They were no longer "holes" with respect to a plenitude of being (negative energy sea): their existence was not dictated by the principles of plenitude and continuity, but by the formal symmetry of the quantized Klein–Gordon theory with respect to positive and negative charges (for details, see Massimi, 2002). As we will see below, this apparently innocuous difference about the concept of antiparticle marks one of the most profound divisions in the history of modern physics, and it is particularly relevant to our analysis of Weinberg's proof in Section 5.

The spin-statistics connection for particles obeying the Exclusion Principle was the main obstacle towards a full-blown proof of the spin-statistics theorem. In 1936, Pauli announced a first—incomplete—proof: following up on the remarkable results of the Pauli–Weisskopf paper, in this new paper Pauli (1936) explored "the possibility of a formal development of a relativistic scalar theory for particles without spin obeying the Exclusion Principle" (p. 110). But the fermionic quantization of scalar fields turned out to be impossible as a closer analysis of the formal apparatus already developed in 1934 easily revealed. Following the same procedure of second quantization of the Klein–Gordon equation adopted in the Pauli–Weisskopf paper, Pauli derived the spin-statistics connection for scalar fields from a crucial property of the electric charge: the charge density ρ at two space-like separated points (x and x') was required to commute

$$[\rho(x), \rho(x')]_- = 0. \tag{12}$$

This fundamental property of the charge density was nothing but the microcausality condition imposed by special relativity, which will play a main role in the rest of our story. As a classical locality condition, microcausality forbids that measurements of a physical quantity at a space-like separation (i.e., measurements that cannot get into contact through light signals) can influence each other. Pauli then showed that it was impossible to retain microcausality for ρ together with the relativistic invariance of the theory when spin-0 particles were quantized according to the Exclusion Principle (i.e., using anticommutators). In this way, by appealing to microcausality, Pauli proved that spin-0 particles could be quantized only with Bose statistics. However, the corresponding case for half-integral spin particles still had to be proved.

Pauli's (1936) proof set forth the guidelines for Markus Fierz's proof (Fierz, 1939). In fact, Fierz's proof for integral spin particles was based on the same requirement of relativistically invariant and local commutation relations (microcausality) that Pauli had already required for the charge density. The real novelty consisted in the proof for half-integral spin particles, which Fierz derived by introducing a further postulate: the positivity of the energy. The fields corresponding to half-integral spin particles were associated with spinors. Fierz introduced anti-commutation relations for the quantized Fourier coefficients of the half-integral spin fields with the consequence that "the particles described satisfy the Pauli's Exclusion Principle; a circumstance which makes it possible, *through a 'hole theory*,' to make the energy positive" (Fierz, 1939, p. 17; emphasis added). The roundabout route via Dirac's hole theory seemed to Fierz an inevitable step in order to make the Hamiltonian positive semi-definite. Thus, the anticommutator turned out to be necessarily connected with half-integral spin particles because only the Exclusion Principle—*via* the hole theory—warranted a positive semi-definite Hamiltonian. The perennial enemy (Dirac's hole theory) could not be avoided and indeed played an essential role in proving the spin-statistics connection for half-integral spin particles.

Pauli's final proof of the theorem essentially repeated Fierz's point. Despite the emphasis that (Pauli, 1940) "the connection between spin and statistics is one of the most important applications of the special relativity theory" (p. 722), Pauli was not able to prove the whole theorem only on the basis of the relativistic requirement of microcausality, and he fell back on the dichotomy microcausality/positive energy:

> Hence we come to the result: For integral spin, the quantization according to the Exclusion Principle is not possible. On the other hand, it is formally possible to quantize the theory for half-integral spins according to Bose–Einstein statistics, but (…) the energy of the system would not be positive. Since for physical reasons it is necessary to postulate this, we must apply the Exclusion Principle in connection with Dirac's hole theory (1940, p. 722).

By postulating microcausality for the spinors, Pauli concluded that for integral-spin fields, the commutator between the field and its Hermitian conjugate vanished at space-like $x - x'$, giving the right (bosonic) spin-statistics connection. But for *half-integral spin* fields, there was no *a priori* restriction of this type: both the commutator and the anticommutator were formally permitted. To rule out the bosonic commutator, Pauli, like Fierz, had to resort to the requirement of positive energy, which was in turn grounded on the hole theory. Ironically enough, Pauli ended up with relying on his old enemy to get the right spin-statistics connection for fermions.

A unified treatment of the two cases (integral and half-integral spin) under the single requirement of microcausality seemed to be unavailable to Pauli's generation. And it has continued to remain such in the eyes of recent generations, if we consider the deeply instilled tendency—typical of many current quantum-field theory textbooks—to present the spin-statistics connection by resorting to the dichotomy microcausality/positive energy.[14] In the next section, we present Greiner and Reinhardt's proof as a paradigmatic example of this ongoing textbook tradition.

3. The textbook tradition: the Greiner–Reinhardt proof and the hole picture

Greiner and his colleagues at the Johann Wolfgang Goethe University in Frankfurt are famous for producing a series of textbooks covering the whole field of theoretical physics. In this section, we focus on Greiner and Reinhardt's (1996) proof of the spin-statistics theorem as it appears in their book, *Field Quantization*. We have chosen it because it provides a paradigmatic example of an ongoing textbook tradition in relativistic quantum-field theory that follows Pauli's original proof in requiring positive energy to get the right spin-statistics connection for half-integral spin particles. Like Pauli, Greiner and Reinhardt also appeal to two different postulates:

- Microcausality is used to derive the right (Bose–Einstein) statistics for the Klein–Gordon fields (spin-0 particles).
- Positive energy is introduced to get the right (Fermi–Dirac) statistics for Dirac fields (spin-$\frac{1}{2}$ particles).

As we will see below, the requirement of positive energy is here regarded as essential to get the spin-statistics connection for Dirac fields, and it is conceptually rooted in the hole picture, as in Fierz and Pauli. Of course, the hole picture is nowadays regarded as devoid of any physical meaning and remains only as a heuristic device. However, interestingly enough, Greiner and Reinhardt ground the existence of antiparticles precisely in the hole picture, as Dirac originally did.

Let us flesh out Greiner–Reinhard's proof, starting with some preliminary technical remarks. Greiner and Reinhardt take the Dirac wave function $\psi(\mathbf{x}, t)$ as a classical field having four components and satisfying the transformation laws of a relativistic spinor. They proceed then to quantize the Dirac field by replacing the spinors $\psi(\mathbf{x}, t)$ and $\psi^\dagger(\mathbf{x}, t)$ by field operators $\hat{\psi}(\mathbf{x}, t)$ and $\hat{\psi}^\dagger(\mathbf{x}, t)$. The plane-wave expansion of the field operator $\hat{\psi}(\mathbf{x}, t)$ is expressed in terms of the states (1996, Section 5.3, pp. 124–130):

$$\psi_p^{(r)}(\mathbf{x}, t) = (2\pi)^{-3/2}\sqrt{\frac{m}{\omega_p}}w_r(\mathbf{p})\,\mathrm{e}^{-\mathrm{i}\varepsilon_r(\omega_p t - \mathbf{p}\cdot\mathbf{x})}. \tag{13}$$

[14] The axiomatic field theory proofs (e.g., Burgoyne, 1958) do give a unified treatment for bosons and fermions, but both microcausality and positive energy are included among the axioms, making it impossible to disentangle their individual roles.

Here $w_r(\mathbf{p})$ is the Dirac unit spinor, where the index r enumerates the four solutions:

$r = \{1,2\}$ denotes positive energy solutions ($E = +\omega_p = +\sqrt{\mathbf{p}^2 + m^2}$).

$r = \{3,4\}$ denotes negative energy solutions ($E = -\omega_p = -\sqrt{\mathbf{p}^2 + m^2}$).

These are expressed by the sign function ε_r, which takes two values: $+1$ if $r = \{1,2\}$ and -1 if $r = \{3,4\}$.

The plane-wave expansion of the field operator $\hat{\psi}\ (\mathbf{x}, t)$ is then written as

$$\hat{\psi}(\mathbf{x}, t) = \sum_{r=1}^{4} \int \mathrm{d}^3 p \hat{a}(\mathbf{p}, r) \psi_p^r(\mathbf{x}, t) \tag{14}$$

and the corresponding one for the Hermitian conjugate field operator $\hat{\psi}^\dagger(\mathbf{x}, t)$ as

$$\hat{\psi}^\dagger(\mathbf{x}, t) = \sum_{r=1}^{4} \int \mathrm{d}^3 p \hat{a}^\dagger(\mathbf{p}, r) \psi_p^r(\mathbf{x}, t). \tag{15}$$

The operators $\hat{a}(\mathbf{p}, r)$ and $\hat{a}^\dagger(\mathbf{p}, r)$ satisfy the anticommutation relations:

$$\begin{aligned} &[\hat{a}(\mathbf{p}, r), \hat{a}^\dagger(\mathbf{p}', r')]_+ = \delta^3(\mathbf{p} - \mathbf{p}')\delta_{rr'}, \\ &[\hat{a}(\mathbf{p}, r), \hat{a}(\mathbf{p}', r')]_+ = [\hat{a}^\dagger(\mathbf{p}, r), \hat{a}^\dagger(\mathbf{p}', r')]_+ = 0. \end{aligned} \tag{16}$$

The operator $\hat{a}(\mathbf{p}, r)$ is the *annihilation operator for particles* ($r = 1,2$) and the *creation operator for antiparticles* ($r = 3,4$); similarly, the operator $\hat{a}^\dagger(\mathbf{p}, r)$ is the *creation operator for particles* ($r = 1,2$) and the *annihilation operator for antiparticles* ($r = 3,4$). But, as Greiner and Reinhardt (1996) point out,

> The double role played by the operators $\hat{a}(\mathbf{p}, r)$ and $\hat{a}^\dagger(\mathbf{p}, r)$ is a bit confusing. Therefore, it is customary to introduce separate notations for the particle and the *hole operators*. We take this opportunity also to change the notation for the wave functions. The names $u(\mathbf{p}, s)$ and $v(\mathbf{p}, s)$ will be introduced for the unit Dirac spinors of the upper and lower continuum [where s denotes the spin projection along the Z-axis in units of $\hbar/2$ in the rest-frame of the particle]. They are related to the spinors $w_r(\mathbf{p})$ as follows:
>
> $$\begin{aligned} w_1(\mathbf{p}) &= u(\mathbf{p}, +1), \\ w_2(\mathbf{p}) &= u(\mathbf{p}, -1), \\ w_3(\mathbf{p}) &= v(\mathbf{p}, -1), \\ w_4(\mathbf{p}) &= v(\mathbf{p}, +1), \end{aligned}$$
>
> (…) [Associated with these new] spinors, the following new operators are introduced:
>
> $$\begin{aligned} \hat{a}(\mathbf{p}, 1) &= \hat{b}(\mathbf{p}, +1), \\ \hat{a}(\mathbf{p}, 2) &= \hat{b}(\mathbf{p}, -1), \\ \hat{a}(\mathbf{p}, 3) &= \hat{d}^\dagger(\mathbf{p}, -1), \\ \hat{a}(\mathbf{p}, 4) &= \hat{d}^\dagger(\mathbf{p}, +1) \end{aligned}$$ (pp. 127−128; emphasis added).

If the total energy is represented as a continuum with a positive upper bound ($r = 1, 2$) and a negative lower bound ($r = 3, 4$), the operators $\hat{b}$ correspond to the solutions of the Dirac equation for the upper continuum and the operators $\hat{d}^{\dagger}$ to the solutions for the lower continuum. The new operators $\hat{b}^{\dagger}$ and $\hat{b}$ are, respectively, the creation and annihilation operators for particles, while $\hat{d}^{\dagger}$ and $\hat{d}$ are the creation and annihilation operators for antiparticles or *holes*. The mathematical construction of the operators—as Greiner and Reinhardt[15] explicitly remark—is rooted in Dirac's hole picture. The idea of a lower continuum of negative energy is nothing but Dirac's idea of a "negative energy sea" in which all the states are filled up by negative energy particles (one per state as required by the Exclusion Principle), and an unoccupied state is a hole corresponding to an antiparticle.

The conceptual link with Dirac's hole picture here is deeper than it might seem at first sight. It is not just a question of historical curiosity. Nor does it merely provide a heuristic tool that a time-honoured agenda dictates to quantum-field theory textbooks.[16] Rather, the reference to the hole picture turns out to be crucial to prove the spin-statistics theorem for spin-$\frac{1}{2}$ particles via the postulate of positive energy. In fact, since the hole picture provides the motivation for the aforementioned mathematical construction of the operators, Greiner and Reinhardt (1996) can claim,

> as the number of particles in the "lower continuum" ($r = 3, 4$) grows, the total energy of the system can drop to negative values beyond any bound. To circumvent this unacceptable conclusion Dirac's hole picture comes to the rescue (p. 126).

Dirac's hole picture would guarantee that the lower continuum is filled up and would consequently prevent a surplus of negative energy particles leading to a negative Hamiltonian. In turn, positive energy safeguards the derivation of the spin-

[15] "According to [Dirac's hole picture] in the vacuum state all the levels of the lower continuum (energy $E < -m$) are occupied by particles (…). These particles, which fill up the 'Dirac sea', are always present and are distributed homogeneously (in the absence of electromagnetic field) all over space. Therefore, they cannot be observed in any experiment. Their energy and charge can be removed by a simple subtraction." Greiner and Reinhardt (1996, p. 126). And again at p. 130: "Dirac's concept of a fully occupied lower continuum […] finds its mathematical expression in [the definition of the new operators $\hat{b}$ and $\hat{d}^{\dagger}$]".

[16] Greiner and Reinhardt (1996) seem to regard the hole picture as providing QFT with clear-cut heuristics: "This mathematical construction is endowed with physical meaning if the holes are interpreted as antiparticles, e.g., as positrons" (p. 127). Saunders (1991), too, regards the hole picture as providing a simple and direct physical meaning for Dirac field theory (the standard formalism), but he complains that "nowadays no one would regard the use of the Fourier expansion or of the normal-ordering as logically dependent on the Dirac hole theory; they are supposed to stand in their own right" (p. 86). Saunders investigates the logical independence of the standard formalism from Dirac's hole theory within the context of a justification for the plane-wave expansion. He considers such a logical independence as a drawback of the standard formalism precisely because it robs it of a clear-cut physical meaning. And in order to restore it, he opts for Irving Segal's geometric quantization (Segal, 1964; see also Woodhouse, 1980), in which Dirac's "negative-energy sea is encoded into the mathematical description of the antiparticle states" (Saunders, 1991, p. 66; for details see pp. 91–106). As we will see in the next section, Weinberg's proof marks the definitive detachment from the hole picture not only as a heuristic device, but also as a conceptual framework playing a crucial role in the proof of the spin-statistics theorem.

statistics connection for spin-$\frac{1}{2}$ particles:

$$\begin{gathered}\text{Dirac's hole picture}\\ \Downarrow \\ \text{Positivity of the energy}\\ \Downarrow \\ \text{Spin-}\tfrac{1}{2}\text{ particles obey Fermi}-\text{Dirac statistics.}\end{gathered}$$

Let us see, then, how Greiner and Reinhardt derive the spin-statistics connection for bosons and fermions. To do this, let us go back to the operators $\hat{b}$ and $\hat{d}$. These new operators obey the same anticommutation relations as the previous operator $\hat{a}(\mathbf{p}, r)$:

$$\begin{aligned}&[\hat{b}(\mathbf{p}, s), \hat{b}^{\dagger}(\mathbf{p}', s')]_{+} = \delta^3(\mathbf{p} - \mathbf{p}')\delta_{ss'},\\ &[\hat{d}(\mathbf{p}, s), \hat{d}^{\dagger}(\mathbf{p}', s')]_{+} = \delta^3(\mathbf{p} - \mathbf{p}')\delta_{ss'}.\end{aligned} \tag{17}$$

The Dirac field operator is now expanded as

$$\hat{\psi}(\mathbf{x}, t) = \sum_s \int \frac{\mathrm{d}^3 p}{(2\pi)^{3/2}} \sqrt{\frac{m}{\omega_p}} (\hat{b}(\mathbf{p}, s) u(\mathbf{p}, s) \mathrm{e}^{-\mathrm{i}p\cdot x} + \hat{d}^{\dagger}(\mathbf{p}, s) v(\mathbf{p}, s) \mathrm{e}^{+\mathrm{i}p\cdot x}), \tag{18}$$

where $p \cdot x = w_p t - \mathbf{p} \cdot \mathbf{x}$ is the 4-dimensional scalar product.

Using the Lorentz boost operator the following explicit form of the unit spinors can be derived:

$$u(\mathbf{p}, s) = \frac{\not{p} + m}{\sqrt{2m(\omega_p + m)}} u(0, s), \quad v(\mathbf{p}, s) = \frac{-\not{p} + m}{\sqrt{2m(\omega_p + m)}} v(0, s), \tag{19}$$

where the slash notation $\not{p}$ denotes $p^{\mu}\gamma_{\mu}$.

Greiner and Reinhardt then compute the anticommutator for the free Dirac field operator $\hat{\psi}(x)$ and its adjoint $\hat{\bar{\psi}}(y)$, where $\hat{\bar{\psi}} = \hat{\psi}^{\dagger}\gamma^0$, at arbitrary space-time points x and y, which in the plane-wave expansion becomes

$$\begin{aligned}[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{+} = &\int \frac{\mathrm{d}^3 p}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3 p'}{(2\pi)^{3/2}} \sqrt{\frac{m}{\omega_p}} \sqrt{\frac{m}{\omega_{p'}}} \sum_{s,s'} [\hat{b}(\mathbf{p}, s) u(\mathbf{p}, s) \mathrm{e}^{-\mathrm{i}p\cdot x}\\ &+ \hat{d}^{\dagger}(\mathbf{p}, s) v(\mathbf{p}, s) \mathrm{e}^{+\mathrm{i}p\cdot x}, \hat{b}^{\dagger}(\mathbf{p}', s') \bar{u}(\mathbf{p}', s') \mathrm{e}^{+\mathrm{i}p'\cdot y}\\ &+ \hat{d}(\mathbf{p}', s') \bar{v}(\mathbf{p}', s') \mathrm{e}^{-\mathrm{i}p'\cdot y}]_{+}.\end{aligned} \tag{20}$$

Only the anticommutators for $[\hat{b}, \hat{b}^{\dagger}]_{+}$ and $[\hat{d}^{\dagger}, \hat{d}]_{+}$ contribute:

$$\begin{aligned}[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{+} = &\int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{m}{\omega_p} \sum_s (u(\mathbf{p}, s) \bar{u}(\mathbf{p}, s)\, \mathrm{e}^{-\mathrm{i}p\cdot(x-y)}\\ &+ v(\mathbf{p}, s) \bar{v}(\mathbf{p}, s) \mathrm{e}^{+\mathrm{i}p\cdot(x-y)})\end{aligned} \tag{21}$$

The spin summation over the unit spinors u and v leads to the projection operators

$$\sum_s u(\mathbf{p}, s) \bar{u}(\mathbf{p}, s) = \left(\frac{\not{p} + m}{2m}\right) \quad \text{and} \quad \sum_s v(\mathbf{p}, s) \bar{v}(\mathbf{p}, s) = \left(\frac{\not{p} - m}{2m}\right) \tag{22}$$

so that the anticommutator becomes

$$
\begin{aligned}
[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{+} &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} ((\not{p} + m) e^{-ip\cdot(x-y)} - (-\not{p} + m) e^{+ip\cdot(x-y)}) \\
&= (i\not{\nabla} + m) \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} (e^{-ip\cdot(x-y)} - e^{+ip\cdot(x-y)}) \\
&= (i\not{\nabla} + m)\, i\Delta(x-y), \qquad (23)
\end{aligned}
$$

where $\Delta(x-y)$ is the commutation function that Jordan and Pauli (1928) originally introduced for the special case of the (massless) electromagnetic field. This function is Lorentz invariant (LI). A fundamental property of this commutation function is

$$\Delta(x-y) = 0 \quad \text{for} \quad (x-y)^2 < 0.$$

In other words, if the argument is a space-like four-vector, the Pauli–Jordan function vanishes outside the light cone; i.e., for space-like separations $(x-y)^2 < 0$. As Greiner and Reinhardt (1996) point out,

> In our case this implies that measurements at two points that have a space-like separation (…) *do not influence each other*. To put it into another way: disturbances cannot propagate with superluminal velocity. This is one of the most fundamental demands to be imposed on a physical theory. It is also known as the condition of *microcausality* (pp. 102–103. Emphasis in the original).

When some observable quantities like current and charge density are constructed out of products of field operators, the microcausality condition requires that either the commutator or the anticommutator (or both) of the field operators vanish at space-like separations. This is satisfied because physical observables are bilinear forms in the fields and their adjoints.

Thus, the Pauli–Jordan function obeys microcausality, and using this property, Greiner and Reinhardt prove that the Klein–Gordon fields (spin-0 particles) do not satisfy the Fermi–Dirac statistics (because microcausality would otherwise be violated) (see 1996, p. 104). However, and this is the main point we are concerned with, Greiner and Reinhardt deny that microcausality can be deployed to rule out the wrong (bosonic) quantization for Dirac fields (spin-$\frac{1}{2}$ particles). They contend rather that microcausality is not a sufficient condition to establish the right spin-statistics connection for spin-$\frac{1}{2}$ particles. Their reasoning is *by reductio* (see 1996, pp. 139–140).

Suppose—they say—we quantize Dirac fields according to Bose–Einstein. The bosonic commutator for the free Dirac field operators $\hat{\psi}(x)$ and $\hat{\bar{\psi}}(y)$ would read

$$
\begin{aligned}
[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{-} = \int \frac{d^3p}{(2\pi)^{3/2}} \int \frac{d^3p'}{(2\pi)^{3/2}} \sqrt{\frac{m}{\omega_p}} \sqrt{\frac{m}{\omega_{p'}}} \sum_{s,s'} ([\hat{b}(\mathbf{p},s), \hat{b}^{\dagger}(\mathbf{p}',s')]_{-} u(\mathbf{p},s) \\
\times \bar{u}(\mathbf{p}',s')\, e^{-ip\cdot x + ip'\cdot y} + [\hat{d}^{\dagger}(\mathbf{p},s), \hat{d}(\mathbf{p}',s')]_{-} v(\mathbf{p},s) \bar{v}(\mathbf{p}',s')\, e^{+ip\cdot x - ip'\cdot y} \\
+ \text{mixed terms}). \qquad (24)
\end{aligned}
$$

The commutation relations would be $[\hat{a}(\mathbf{p},r),\hat{a}^{\dagger}(\mathbf{p}',r')]_{-}=\delta^3(\mathbf{p}-\mathbf{p}')\delta_{rr'}$ with $r,r'=1,\ldots,4$, or, reverting to the notation in terms of $\hat{b}$ and $\hat{d}$ operators, they would be

$$[\hat{b}(\mathbf{p},s),\hat{b}^{\dagger}(\mathbf{p}',s')]_{-}=[\hat{d}^{\dagger}(\mathbf{p},s),\hat{d}(\mathbf{p}',s')]_{-}=\delta^3(\mathbf{p}-\mathbf{p}')\delta_{ss'}. \tag{25}$$

This implies that the expansion coefficients of the above commutator are both equal to 1, and then the terms of the commutator reduce to those of the anticommutator, with the result that the commutator vanishes for space-like separations $x-y$ just as the anticommutator does.

The moral that Greiner and Reinhardt draw is that microcausality cannot be invoked to distinguish between fermionic and bosonic quantization of Dirac fields, and therefore it does not lead to the right spin-statistics connection for spin-$\frac{1}{2}$ particles: "The derivation of the spin-statistics theorem for Dirac fields therefore can be only based on the positivity condition of the energy" (1996, p. 140).

In fact, the Hamiltonian of the hypothetically bosonic-quantized Dirac field would be

$$\hat{H}=\sum_{s}\int \mathrm{d}^3p\,\omega_p(\hat{b}^{\dagger}(\mathbf{p},s)\hat{b}(\mathbf{p},s)-\hat{d}(\mathbf{p},s)\hat{d}^{\dagger}(\mathbf{p},s)) \tag{26}$$

with the first term corresponding to the positive energy solution and the second term corresponding to the negative energy solution. Since the commutation relations for the hole operators are

$$[\hat{d}^{\dagger}(\mathbf{p},s),\hat{d}(\mathbf{p}',s')]_{-}=\delta^3(\mathbf{p}-\mathbf{p}')\delta_{ss'},$$

the reordering of the operators would not invert the sign of the second term of the Hamiltonian, which would not be positive semi-definite in every case. Greiner and Reinhardt conclude, "the Hamiltonian of the Dirac field thus cannot be made a positive-definite operator if the 'wrong' (i.e., bosonic) quantization prescription is employed" (1996, p. 129). But it is at this point that the Dirac hole picture comes to the rescue: it sets everything right by granting an overall positive Hamiltonian so that "the condition of positive energies [can be] used to derive another special case of the spin-statistics theorem (spin$-\frac{1}{2}\rightarrow$Fermi–Dirac statistics)" (1996, p. 130).

To summarize, Greiner and Reinhardt fall back on the Fierz–Pauli agenda in blending different postulates to prove the spin-statistics theorem: microcausality for the Klein–Gordon fields, and positive energy for Dirac fields. Once again, no unified proof seems to be available under the single relativistic requirement of microcausality. Weinberg's proof overturns this generally accepted textbook tradition. As we will see in the next section, by dispensing with Dirac's hole picture in the interpretation of the antiparticle operators, Weinberg is doing more than simply abandoning a deeply entrenched heuristic picture. He is elaborating a new proof of the theorem that no longer needs to invoke positive energy because microcausality alone can do the entire job.

4. Weinberg's proof of the spin-statistics theorem: overturning the textbook tradition

Steven Weinberg has offered a remarkably simple proof of the spin-statistics theorem (at any rate, for the case of non-interacting fields), whose conceptual innovation with respect to the textbook tradition has been widely ignored in the literature. We aim at doing justice to Weinberg's proof by highlighting its points of divergence from the traditional proof and, at the same time, by showing its unexpected conceptual connection with Pauli's original intent.

Weinberg announced the proof for the first time in an article (1964, B1322–B1323) and then again in his voluminous monograph on quantum-field theory (1995, pp. 219–224). Henceforth, we will focus mainly on the 1964 proof (the later one differs only slightly from it). The declared purpose of Weinberg's (1964) article was to calculate Feynman rules for particles of any spin, where "our calculation uses field theory, but only as a convenient instrument for the construction of a Lorentz-invariant S matrix" (Weinberg, 1964, B1318). The spin-statistics theorem is presented as a side result of the S matrix approach: the advantage—in Weinberg's words—is that while "Pauli's proof of the connection between spin and statistics is straightforward for integer j, but rather indirect for half-integer j," the microcausality requirement that Weinberg invokes yields the spin-statistics connection straightforwardly also for half-integer j (1964, B1319).

To get causal Dirac fields, Weinberg introduces particle annihilation and antiparticle creation fields, respectively (see Weinberg, 1995, p. 219)[17]

$$\hat{\psi}^{+}(x) = (2\pi)^{-3/2} \sum_{s} \int \mathrm{d}^3 p \sqrt{\frac{m}{\omega_p}} u(\mathbf{p}, s) \mathrm{e}^{-\mathrm{i}p\cdot x} \hat{b}(\mathbf{p}, s), \tag{27}$$

$$\hat{\psi}^{-}(x) = (2\pi)^{-3/2} \sum_{s} \int \mathrm{d}^3 p \sqrt{\frac{m}{\omega_p}} v(\mathbf{p}, s) \mathrm{e}^{\mathrm{i}p\cdot x} \hat{d}^{\dagger}(\mathbf{p}, s). \tag{28}$$

In order to make the comparison with Greiner and Reinhardt as close as possible, we shall assume that the wave functions u and v satisfy the Dirac equations $(\not{p} - m)u = 0$ and $(\not{p} + m)v = 0$ (cf. Eq. (19)). In Weinberg's own treatment, these equations are actually derived from the group-theoretic analysis of LI supplemented by the microcausality condition. There is in effect a double use of microcausality in deriving the spin summation formulae of Eq. (22) as well as the spin-statistics theorem itself. Our treatment is closer to the "third way" of Furry and Oppenheimer referred to in Section 2: it has the advantage of avoiding Weinberg's group-theoretic analysis and yet bringing out the essential step in Weinberg's proof involving the choice of commutation/anticommutation relations for the annihilation/creation operators.

[17] We have adapted Weinberg's notation to correspond to Greiner–Reinhardt.

Following Weinberg, we now combine $\hat{\psi}^+$ and $\hat{\psi}^-$ in a linear combination:[18]

$$\hat{\psi}(x) = k\hat{\psi}^+(x) + \lambda\hat{\psi}^-(x) \tag{29}$$

with the constants k and λ to be determined so that for $x - y$ space-like separations, the Dirac field anticommutes or commutes with itself and its adjoint (depending on whether the particles destroyed and the antiparticles created by the components $\hat{\psi}^+(x)$ and $\hat{\psi}^-(x)$ are fermions or bosons, respectively):

$$[\hat{\psi}(x), \hat{\psi}(y)]_\pm = [\hat{\psi}(x), \hat{\bar{\psi}}(y)]_\pm = 0. \tag{30}$$

This is the microcausality requirement, which is to be satisfied by Dirac fields as well as by the Klein–Gordon fields (Weinberg, 1964, B1318; 1995, p. 222) and on which Weinberg relies to get the right spin-statistics connection for both cases. Let us see how.

To see the connection with the Greiner–Reinhardt notation, let $\hat{b}$ and $\hat{b}^\dagger$ be the annihilation and creation operators for particles and $\hat{d}$ and $\hat{d}^\dagger$ those for antiparticles. These operators satisfy Fermi anticommutation or Bose commutation rules:

$$\begin{aligned} [\hat{b}(\mathbf{p}, s), \hat{b}^\dagger(\mathbf{p}', s')]_\pm &= \delta^3(\mathbf{p} - \mathbf{p}')\delta_{ss'}, \\ [\hat{d}(\mathbf{p}, s), \hat{d}^\dagger(\mathbf{p}', s')]_\pm &= \delta^3(\mathbf{p} - \mathbf{p}')\delta_{ss'}. \end{aligned} \tag{31}$$

The microcausality condition for Dirac fields then reads

$$\begin{aligned} [\hat{\psi}(x), \hat{\bar{\psi}}(y)]_\pm = \int \frac{d^3p}{(2\pi)^3} \frac{m}{\omega_p} \Bigg[&|k|^2 \sum_s u(\mathbf{p}, s)\bar{u}(\mathbf{p}, s) e^{-ip\cdot(x-y)} \\ &\pm |\lambda|^2 \sum_s v(\mathbf{p}, s)\bar{v}(\mathbf{p}, s) e^{ip\cdot(x-y)} \Bigg]. \end{aligned} \tag{32}$$

Given the projection operators

$$\sum_s u(\mathbf{p}, s)\bar{u}(\mathbf{p}, s) = \left(\frac{\not{p} + m}{2m}\right) \quad \text{and} \quad \sum_s v(\mathbf{p}, s)\bar{v}(\mathbf{p}, s) = \left(\frac{\not{p} - m}{2m}\right) \tag{22}$$

and putting them outside the squared brackets as $(i\not{\nabla} + m)$, the above microcausality condition reduces to

$$[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_\pm = (i\not{\nabla} + m) \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} [|k|^2 e^{-ip\cdot(x-y)} \pm (-)|\lambda|^2 e^{ip\cdot(x-y)}]. \tag{33}$$

But the quantity

$$\int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} [e^{-ip\cdot(x-y)} - e^{ip\cdot(x-y)}] = i\Delta(x - y)$$

[18] Following Novozhilov's reconstruction of Weinberg's argument (Novozhilov, 1975, p. 77), Saunders (1991) points out that "if we want to have a covariant causal field that is complex, but transforms simply under gauge transformations, then we must introduce a new Fock space (the antiparticle space), and the linear combination of annihilation operators on the particle space and creation operators on the antiparticle space (together with its adjoint) is the only possible operator expansion" (p. 88).

is nothing but the Lorentz-invariant Pauli–Jordan function, which vanishes if the argument is a space-like $x - y$. Now, in order to meet microcausality, it must be the case that

$$[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = 0 \text{ for } (x-y)^2 < 0, \tag{30}$$

and, in turn, this can be obtained if the integrand in (33) vanishes. The necessary and *sufficient condition* for the integrand to vanish (given the vanishing Pauli–Jordan function) reduces then to the coefficients of the exponential functions being equal in magnitude and opposite in sign; i.e.,

$$-|k|^2 = \pm(-)|\lambda|^2. \tag{34}$$

This is possible if and only if

- $|k| = |\lambda|$, and
- between $\pm$, we choose the top sign $(+)$. But this means that we must choose the Fermi anticommutation rule $(+)$. Hence Dirac fields must obey the Fermi–Dirac statistics. QED

In sum, in order to meet microcausality, Weinberg has got the right (Fermi–Dirac) statistics for spin-$\frac{1}{2}$ particles, *contra Greiner–Reinhardt* and more generally counter an entire textbook tradition that contends that microcausality is not sufficient to arrive at this result. How could Weinberg succeed where everyone else failed?

At a closer look, the discrepancy between Weinberg's proof and Greiner–Reinhardt's can be traced back to the use of different anticommutation/commutation relations. In fact, for Greiner and Reinhardt, microcausality is not sufficient because

$$[\hat{b}(\mathbf{p}, s), \hat{b}^{\dagger}(\mathbf{p}', s')]_{\pm} = [\hat{d}^{\dagger}(\mathbf{p}, s), \hat{d}(\mathbf{p}', s')]_{\pm} = \delta^3(\mathbf{p} - \mathbf{p}')\delta_{ss'}. \tag{25}$$

If we plugged these anticommutation/commutation relations for particle and antiparticle operators into Weinberg's microcausality formula, we would get

$$[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = (\mathrm{i}\nabla + m) \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{2\omega_p} [|k|^2 \mathrm{e}^{-\mathrm{i}p\cdot(x-y)} - |\lambda|^2 \mathrm{e}^{\mathrm{i}p\cdot(x-y)}], \tag{35}$$

that clearly vanishes in both cases (i.e., the anticommutator as well as the commutator) whenever $|k|^2 = |\lambda|^2$, *by contrast with Weinberg's corresponding formula*

$$[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = (\mathrm{i}\nabla + m) \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{2\omega_p} [|k|^2 \mathrm{e}^{-\mathrm{i}p\cdot(x-y)} \pm (-)|\lambda|^2 \mathrm{e}^{\mathrm{i}p\cdot(x-y)}], \tag{33}$$

where the crucial ($\pm$) sign comes from different anticommutation/commutation relations for the operators

$$[\hat{b}(\mathbf{p},s),\hat{b}^{\dagger}(\mathbf{p}',s')]_{\pm} = \delta^3(\mathbf{p}-\mathbf{p}')\delta_{ss'},$$
$$[\hat{d}(\mathbf{p},s),\hat{d}^{\dagger}(\mathbf{p}',s')]_{\pm} = \delta^3(\mathbf{p}-\mathbf{p}')\delta_{ss'}. \tag{31}$$

To summarize, while from Weinberg's formula it is possible to get the right (Fermi–Dirac) statistics for spin-$\frac{1}{2}$ particles, it is not from Greiner–Reinhardt's because the terms of the commutator coincide with those of the anticommutator with the result that the commutator vanishes for space-like separations as does the anticommutator (see the synoptic schema below).

	Greiner-Reinhardt	***Weinberg***
Anticommutation/commutation relations	$[\hat{b}(\mathbf{p},s),\hat{b}^{\dagger}(\mathbf{p}',s')]_{\pm} =$ $[\hat{d}^{\dagger}(\mathbf{p},s),\hat{d}(\mathbf{p}',s')]_{\pm} =$ $\delta^3(\mathbf{p}-\mathbf{p}')\delta_{ss'}$	$[\hat{b}(\mathbf{p},s),\hat{b}^{\dagger}(\mathbf{p}',s')]_{\pm} =$ $[\hat{d}(\mathbf{p},s),\hat{d}^{\dagger}(\mathbf{p}',s')]_{\pm} =$ $\delta^3(\mathbf{p}-\mathbf{p}')\delta_{ss'}$
Consequences for Dirac fields	Both the commutator and the anticommutator satisfy microcausality: **Hence microcausality is not a sufficient condition to prove the spin-statistics theorem for Dirac fields.** ⇓ Positivity of energy is required	Only the anticommutator satisfies microcausality: **Hence microcausality is a sufficient condition to prove the spin-statistics theorem for Dirac fields.**

Furthermore, Greiner–Reinhardt's commutation relations for antiparticles,

$$[\hat{d}^{\dagger}(\mathbf{p},s),\hat{d}(\mathbf{p}',s')]_{-} = \delta^3(\mathbf{p}-\mathbf{p}')\delta_{ss'}$$

were responsible for the fact that the reordering of the antiparticle operators left the hypothetically bosonic-quantized Hamiltonian for Dirac fields not positive semi-definite: it strengthened the need for introducing positive energy as a necessary additional assumption. But, as we have just seen, this is no longer required in Weinberg's proof.

What makes the comparison interesting is the fact that the discrepancy between Weinberg and the textbook tradition springs from the deeper level of the interpretation of microcausality and, consequently, of antiparticles.

5. Conceptual roots of Weinberg's proof: the requirement of microcausality

The seemingly innocuous difference between Weinberg's proof and the traditional one (paradigmatically exemplified by Greiner and Reinhardt) marks in fact a profound divergence in the history and philosophy of physics as far as the interpretation of the classical relativistic requirement of microcausality is concerned. As we have seen in Section 3, according to Greiner and Reinhardt, microcausality is a classical locality condition imposed by special relativity and designed to avoid the risk of superluminal signals. As such, it is a fundamental requirement of any physical theory and "nothing indicates that the causality principle might break down at the scale of atoms or elementary particles" (Greiner & Reinhardt, 1996, p. 103).

But Weinberg's concept of microcausality does not have to do with locality, properly speaking (Weinberg, 1995):

> The condition $[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = 0$ is often described as a causality condition, because if $x-y$ is space-like then no signal can reach x from y, so that a measurement of $\hat{\psi}$ at point x should not be able to interfere with a measurement of $\hat{\bar{\psi}}$ at y. Such considerations of causality are plausible for the electromagnetic field, anyone of whose components may be measured at a given space-time point. However, we will be dealing here with fields like the Dirac field of the electron that do not seem in any sense measurable. The point of view taken here is that $[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = 0$ *is needed for the Lorentz invariance of the S matrix*, without any ancillary assumptions about measurability or causality (p. 198; emphasis added).

Thus, *contra* Greiner and Reinhardt, $[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = 0$ for $(x-y)^2 < 0$ is no longer interpreted as a real causality condition, nor is the quantum field regarded as a real field but as "a mere artifice to be used in the construction of an invariant S matrix" (Weinberg, 1964, B1319). For Weinberg, microcausality is a mere heuristic device—devoid of any causal or local meaning—useful to guarantee the LI of the S matrix. This discrepancy sheds light on two different theoretical attitudes:

(a) Greiner and Reinhardt, as *quantum-field theorists*, have a robust attitude toward the quantum field: they are committed to Dirac's picture of a *continuum of positive and possibly negative energy states* in the Hilbert space with operators indexed for all space-time points between any initial and final state.

(b) Weinberg, as a hybrid *field theorist/S-matrix theorist*, has a weaker attitude toward the quantum field: the S matrix describes which in-state evolves into which out-state, without being committed to the space-time facts intervening between the in- and the out-state (the S-matrix theorists are the "behaviourists"—so to speak—of the quantum field).[19]

[19] "The hope of S matrix theory was that, by using the principles of unitarity, analyticity, LI and other symmetries, it would be possible to calculate the S matrix, and you would never have to think about a quantum field. In a way, this hope reflected a kind of positivistic Puritanism: we cannot measure the field of a pion or a nucleon, so we should not talk about it, while we do measure S-matrix elements, so this is what we should stick to as ingredients of our theories" Weinberg (1999, p. 248).

The spin-statistics theorem is just a corollary of Weinberg's research programme on a Lorentz-invariant S matrix. Let us explain this point with a bit more clarity.

As is well known (see Teller, 1995, pp. 117–120), the S-matrix operator allows us to deal with *interacting* quantum fields because interactions are regarded as perturbative refinements of the free field motions. Given the Schrödinger equation in the so-called interaction representation

$$i\frac{\partial}{\partial t}|\psi(t)\rangle = \hat{H}_I|\psi(t)\rangle$$

where $\hat{H}_I$ is the interaction Hamiltonian, the time evolution of an interacting field state is described by a unitary operator $\hat{U}^I$

$$\left|\psi^I(t)\right\rangle = \hat{U}^I(t, t_0)\left|\psi^I(t_0)\right\rangle \tag{36}$$

so that the time derivative of the interaction state is equivalent to

$$i\frac{\partial}{\partial t}\hat{U}^I(t, t_0) = \hat{H}_I(t)\hat{U}(t, t_0). \tag{37}$$

This can be integrated with the boundary condition $\hat{U}(t_0, t_0) = \hat{I}$:

$$\hat{U}(t, t_0) = \hat{I} - i\int_{t_0}^{t} dt_1 \hat{H}_I(t_1)\hat{U}(t_1, t_0). \tag{38}$$

Since the unitary operator appears in both sides of this equation, a series of approximations are used to solve this difficulty:

- The zeroth approximation $\hat{U}^{(0)}(t, t_0) = \hat{I}$ substituted for $\hat{U}(t, t_0)$ gives the first approximation

$$\hat{U}^{(1)}(t, t_0) = \hat{I} - i\int_{t_0}^{t} dt_1 \hat{H}_I(t_1). \tag{39}$$

- The first approximation $\hat{U}^{(1)}(t, t_0)$ substituted for $\hat{U}(t, t_0)$ gives the second approximation

$$\begin{aligned}\hat{U}^{(2)}(t, t_0) &= \hat{I} - i\int_{t_0}^{t} dt_1 \hat{H}_I(t_1)\left[\hat{I} - i\int_{t_0}^{t_1} dt_2 \hat{H}_I(t_2)\right] \\ &= \hat{I} - i\int_{t_0}^{t} dt_1 \hat{H}_I(t_1) + (-i)^2\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, \hat{H}_I(t_1)\hat{H}_I(t_2).\end{aligned} \tag{40}$$

Continuing in this way, by taking $t_0 \to -\infty$ and $t \to +\infty$, we get the *S-matrix operator expansion*

$$\hat{S} = \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{-\infty}^{+\infty} dt_1 \cdots \int_{-\infty}^{+\infty} dt_n T\{\hat{H}_I(t_1)\cdots\hat{H}_I(t_n)\}, \tag{41}$$

where $T\{\hat{H}_I(t_1)\cdots\hat{H}_I(t_n)\}$ means that the operators $\hat{H}_I(t_i)$ are ordered with time indexes decreasing from left to right.

For the S matrix to be invariant under proper orthochronous Lorentz transformations, the interaction Hamiltonian $\hat{H}_I(t)$ must be rewritten as an interaction Hamiltonian density (Weinberg, 1964, B1318)

$$\hat{H}_I(t) = \int \mathrm{d}^3 x \hat{\mathbf{H}}(\mathbf{x}, t), \tag{42}$$

where $\hat{\mathbf{H}}(x)$ is a scalar and the variable x denotes a vector in four-dimensional space time. The S matrix can accordingly be rewritten as

$$\hat{S} = \sum_{n=0}^{\infty} \frac{(-\mathrm{i})^n}{n!} \int_{-\infty}^{+\infty} \mathrm{d}^4 x_1 \cdots \mathrm{d}^4 x_n T\{\hat{\mathbf{H}}(x_1) \cdots \hat{\mathbf{H}}(x_n)\}. \tag{43}$$

But in order to guarantee the LI of the S matrix a further condition is necessary. In fact, the ϑ-functions $\vartheta(x_i - x_j)$ implicit in the definition of the time-ordered product $T\{\hat{\mathbf{H}}(x_1) \cdots \hat{\mathbf{H}}(x_n)\}$ are not scalars unless their argument $(x_i - x_j)$ is time-like.[20] Thus, it must be forbidden for any ϑ-function to have a space-like argument. And this veto is equivalent to imposing the condition that for any space-like $(x - y)$

$$[\hat{\mathbf{H}}(x), \hat{\mathbf{H}}(y)]_- = 0. \tag{44}$$

But as Weinberg (1964, B1318) points out, "the only known way of making sure that such an $\hat{\mathbf{H}}(x)$ will satisfy the restrictions" of (a) being a scalar and (b) vanishing at space-like separations "is to form it as a function of one or more fields $\hat{\psi}(x)$, which are linear combinations of the creation and annihilation operators, and which have the properties: (…) (b) for $(x - y)$ space-like $[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = 0$."

We can now better understand why microcausality is here invoked not as a causality condition but as a mere prescription on the construction of the Dirac fields so as to guarantee the LI of the S matrix. As Bain (1998, pp. 7–9) has pointed out, Weinberg introduces local quantum-field theory as the result of a demonstrative induction,[21] whose phenomenal premise includes empirical evidence about scattering experiments, and whose major premise says that, "A physically satisfactory S matrix satisfies the principles of LI and cluster decomposition (CD)." Now for the S matrix to be LI, it is sufficient that:

(i) The interaction Hamiltonian density $\hat{\mathbf{H}}(x)$ is a Lorentz scalar
(ii) $[\hat{\mathbf{H}}(x), \hat{\mathbf{H}}(y)]_- = 0$ for space-like $(x-y)$.

On the other hand, for the S matrix to satisfy CD, which is another locality condition requiring that scattering experiments at great distances do not interfere, the full Hamiltonian $\hat{H}$ must be built up as the sum of products of creation and annihilation operators with coefficients which are smooth functions of the momenta. On the whole, for the S matrix to satisfy LI and CD, it is sufficient that $\hat{\mathbf{H}}(x)$ is a sum

[20] $\vartheta(x_i - x_j) = 1$ or 0, if $t_i > t_j$ or $t_i < t_j$, respectively.

[21] Demonstrative induction is a non-ampliative inference in which a conclusion of intermediate generality is drawn from two premises: a so-called phenomenal premise including experimental evidence obtained by inductive generalization, and a so-called major premise including theoretical principles of the background knowledge. On demonstrative induction as a scientific method see Norton (1993; 1994) and Massimi (forthcoming).

of products of local quantum fields $\hat{\psi}(x)$, which satisfy microcausality $[\hat{\psi}(x), \hat{\bar{\psi}}(y)]_{\pm} = 0$ for space-like $(x-y)$ and are linear combinations of creation and annihilation operators with coefficients that are smooth functions of momenta. Schematically,

$$\text{(LI \& CD of the } S \text{ matrix)} \Rightarrow \text{(field decomposition of } \hat{\mathbf{H}}(x) \text{ with microcausality for fields)}$$

There is however an obstacle to the construction of fields like $\hat{\psi}(x) = k\hat{\psi}^{+}(x) + \lambda\hat{\psi}^{-}(x)$ satisfying microcausality.[22] In Weinberg's words (1995),

> ...it may be that the particles that are destroyed and created by these fields carry non-zero values of one or more conserved quantum numbers like the electric charge. For instance, if particles of species n carry a value $q(n)$ for the electric charge $\hat{Q}$, then
>
> $$[\hat{Q}, \hat{a}(\mathbf{p}, s, n)]_{-} = -q(n)\hat{a}(\mathbf{p}, s, n),$$
> $$[\hat{Q}, \hat{a}^{\dagger}(\mathbf{p}, s, n)]_{-} = +q(n)\hat{a}^{\dagger}(\mathbf{p}, s, n) \text{ (p. 199)},$$

where $\hat{a}$ and $\hat{a}^{\dagger}$ denote generic annihilation and creation operators for particles of momentum $\mathbf{p}$, spin projection s and species n.

Thus, for the scalar $\hat{\mathbf{H}}(x)$ to commute with the charge operator $\hat{Q}$, $\hat{\mathbf{H}}(x)$ must be constructed as a sum of products of fields $\hat{\psi}_n(x)$ and their Hermitian conjugates $\hat{\psi}^{\dagger}_m(x)$ satisfying simple commutation relations with $\hat{Q}$ (e.g., $[\hat{Q}, \hat{\psi}_n(x)]_{-} = -q_n\hat{\psi}_n(x)$) so that

$$q_{n1} + q_{n2} + \cdots - q_{m1} - q_{m2} - \cdots = 0.$$

But the above condition is satisfied if and only if

(1) All particle species n that are destroyed by the annihilation field $\hat{\psi}^{+}(x)$ carry the same charge $q(n) = q$.
(2) All particle species $\bar{n}$ that are created by the creation field $\hat{\psi}^{-}(x)$ carry the charge $q(\bar{n}) = -q$.

Hence, in order to conserve the electric charge, there must be a doubling of particle species carrying non-zero values of $\hat{Q}$ (Weinberg, 1995):

> If a particular component of the annihilation field destroys a particle of species n, then the same component of the creation field must create particles of species $\bar{n}$ known as the *antiparticles* of the particles of species n, which have opposite values

[22] Weinberg (1999) has emphasized that his derivation of field theory from the idea of an S matrix satisfying LI and CD is really only a "folk theorem." To begin with, it relies on perturbation theory, a notoriously unreliable tool in rigorous quantum field theory, but more importantly it achieves necessary rather than sufficient conditions for an acceptable field theory. A counterexample to sufficiency is the fact that string theory also satisfies the conditions of the theorem. Weinberg concludes that what he has really shown is that his conditions are sufficient for an *effective field theory*; i.e., a field theory that works at low energies, where examples such as string theories degenerate into field theories anyway.

of all conserved quantum numbers. *This is the reason for antiparticles* (p. 199; emphasis in the original).

So, Weinberg does not introduce antiparticles as holes in a negative energy sea *á la* Dirac, as Greiner and Reinhardt do: the reason for antiparticles does not lie in the mathematical construction of the Hamiltonian as a continuum of energy states with a lower and an upper bound. Rather, antiparticles are for Weinberg the necessary consequence of the conservation of quantum numbers such as the electric charge in the construction of a Lorentz-invariant S matrix.

The way Weinberg introduces antiparticles reminds us—*mutatis mutandis*—of Furry–Oppenheimer's similar conceptual route. Furry and Oppenheimer introduced antiparticles from the mathematical constraints of their theory, which was perfectly symmetric with respect to positive and negative charges. Similarly, Weinberg justifies antiparticles in the light of the formal constraints of his Lorentz-invariant S matrix, which also requires symmetry with respect to positive and negative charges. Most importantly, as in Furry and Oppenheimer, antiparticles are taken here as positive particles in their own right introduced as symmetric counterparts of particles and having only a bundle of properties opposite to the properties of particles. No concession is made to the hole picture.

Through this conceptual shift about antiparticles, not only is Weinberg dispensing with the idea of holes in a negative energy sea; he is dispensing more generally with a pictorial representation of the quantum field like the one offered by the hole theory. The quantum field is only an artifice in the construction of the S matrix: and all that we know about the quantum field is what the S matrix tells us. No wonder we then have a fictitious microcausality condition, which does not refer to any real "causal" relation between space-time points x and y, but is only a formal constraint for the construction of a Lorentz-invariant S matrix. It is no wonder, either, that Weinberg can rely on microcausality alone to get the spin-statistics theorem: positive energy needs no longer to be postulated apart and to be warranted by the hole picture. Positive energy has been built into the formalism, and it stands on its own, as Furry and Oppenheimer had anticipated in 1934.

Pauli seems to agree with Greiner and Reinhardt in resorting to the hole picture to grant the requirement of positive energy. But he is surely closer to Weinberg as far as the concept of antiparticle and the programmatic intent of applying relativity alone to get the spin-statistics connection is concerned. This contradictory position testifies to Pauli's difficulties in finding his own theoretical way independently of his old enemy: Pauli's proof of the spin-statistics theorem was only a *Pyrrhic* victory over Dirac.

6. Conclusions

The aim of this paper was a clarification of the theoretical and conceptual roots of Weinberg's proof of the spin-statistics theorem. In the light of the historical reconstruction of Pauli's original proof, and of the subsequent textbook tradition, we have argued that Weinberg's proof constitutes a turning point in this tradition: the requirement of positive energy is finally dispensed with together with the underpinning hole picture. Microcausality turns out to be the only postulate for the proof.

In this respect, Weinberg's proof vindicates Pauli's anti-Dirac line of thought and his intention of getting the spin-statistics connection from relativity theory alone: he has achieved what Pauli merely declared as a programme.

We can now go back to Feynman's comment, quoted at the beginning of this paper (Feynman, 1963).

> An explanation [of the spin-statistics connection] has been worked out by Pauli from complicated arguments of quantum field theory and relativity. He has shown that the two must necessarily go together, but we have not been able to find a way of reproducing his arguments on an elementary level…This probably means that we do not have a complete understanding of the fundamental principle involved… (Chapter 4, Section 1).

We think that Weinberg's proof provides an advance in this direction by combining quantum-field theory and relativity in a long-sought elementary proof, whose only fundamental principle involved is microcausality.

Acknowledgements

We are grateful to an anonymous referee for helpful comments on an earlier draft of this paper. M.M. gratefully acknowledges the London School of Economics and Political Science and the Arts and Humanities Research Board for financial support. Much of the material presented here formed part of a Ph.D. thesis (see Massimi, 2002) submitted to the University of London.

References

Anderson, C. D. (1933). The positive electron. *Physical Review*, *43*, 491–494.

Bain, J. (1998). Weinberg on QFT: Demonstrative induction and underdetermination. *Synthese*, *117*, 1–30.

Berry, M. V., & Robbins, J. M. (1997). Indistinguishability for quantum particles: Spin, statistics and the geometric phase. *Proceedings of the Royal Society of London*, *A453*, 1771–1790.

Bose, S. N. (1924). Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift für Physik*, *26*, 178–181.

Burgoyne, N. (1958). On the connection of spin and statistics. *Nuovo Cimento*, *VIII*, 607–609.

Dirac, P. A. M. (1926). On the theory of quantum mechanics. *Proceedings of the Royal Society of London*, *A112*, 661–677.

Dirac, P. A. M. (1927). On the theory of the emission and absorption of radiation. *Proceedings of the Royal Society of London*, *A114*, 243–265.

Dirac, P. A. M. (1928). The quantum theory of the electron. *Proceedings of the Royal Society of London*, *A117*, 610–624.

Dirac, P. A. M. (1930). A theory of electrons and protons. *Proceedings of the Royal Society of London*, *A126*, 360–365.

Dirac, P. A. M. (1931). Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London*, *A133*, 60–72.

Dirac, P. A. M. (1934). Theory of the positron. In *Structure et Propriétés des Noyaux Atomiques. Rapports et discussions du septiéme conseil de physique tenue á Bruxelles du 22 au 29 Octobre 1933 sous les auspices de l'Institut International de Physique Solvay* (pp. 203–230). Paris: Gauthier-Villars.

Duck, I., & Sudarshan, E. C. G. (1997). *Pauli and the spin-statistics theorem*. Singapore: World Scientific.

Duck, I., & Sudarshan, E. C. G. (1998). Toward an understanding of the spin-statistics theorem. *American Journal of Physics*, *66*, 284–303.

Dyson, F. J. (1949). The *S*-matrix in quantum electrodynamics. *Physical Review*, *75*, 1736–1755.

Einstein, A. (1924). Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte Preussische Akademie der Wissenschaften, Phys.-Math. Klasse* (pp. 261–267).

Einstein, A. (1925a). Quantentheorie des einatomigen idealen Gases. 2. Abhandlung. *Sitzungsberichte Preussische Akademie der Wissenschaften, Phys.-Math. Klasse* (pp. 3–14).

Einstein, A. (1925b). Quantentheorie des idealen Gases. *Sitzungsberichte Preussische Akademie der Wissenschaften, Phys.-Math. Klasse* (pp. 18–25).

Fermi, E. (1926). Sulla quantizazione del Gas Perfetto Monoatomico. *Rendiconti della Reale Accademia dei Lincei*, *3*, 145–149. German trans. Zur Quantelung des idealen einatomigen Gases. *Zeitschrift für Physik*, *36*, 902–912.

Fermi, E. (1934). Versuch einer Theorie der β-Strahlen. *Zeitschrift für Physik*, *88*, 161–171.

Feynman, R. P. (1949a). The theory of positrons. *Physical Review*, *76*, 749–759.

Feynman, R. P. (1949b). Space-time approaches to quantum electrodynamics. *Physical Review*, *76*, 769–789.

Feynman, R. P. (1963). *The Feynman lectures on physics*, *Vol. 3*. Reading, MA: Addison-Wesley.

Fierz, M. (1939). Über die relativistische Theorie kräftefreier Teilchen mit beliebigem Spin. *Helvetica Physica Acta*, *12*, 3–37.

Fock, V. (1933). Zur Theorie des Positrons. *Doklady Akademii Nauk USSR*, *1*, 267–271.

Furry, W. H., & Oppenheimer, J. R. (1934). On the theory of the electron and positive. *Physical Review*, *45*, 245–262.

Gordon, W. (1926). Der Compton Effekt nach der Schrödingerschen Theorie. *Zeitschrift für Physik*, *40*, 117–133.

Greiner, W., & Reinhardt, J. (1996). *Field quantization*. Heidelberg-New York: Springer.

Houriet, A., & Kind, A. (1949). Classification invariante des termes de la matrix S. *Helvetica Physica Acta*, *22*, 319–330.

Jordan, P., & Pauli, W. (1928). Zur Quantenelektrodynamik ladungsfreier Felder. *Zeitschrift für Physik*, *47*, 151–173.

Klein, O. (1927). Elektrodynamik und Wellenmechanik von Standpunkt des Korrespondenzprinzips. *Zeitschrift für Physik*, *41*, 407–442.

Kuhn, T. S., & Heilbron, J. L. (Eds.). (1967). *Sources for history of quantum physics*. Philadelphia: American Philosophical Society (Contains inventory of the holdings of the Archive for the History of Quantum Physics).

Lovejoy, A. O. (1936). *The great chain of being: A study in the history of an idea*. Cambridge, MA: Harvard University Press.

Lüders, G., & Zumino, B. (1958). Connection between spin and statistics. *Physical Review*, *110*, 1450–1453.

Massimi, M. (2002). *Pauli's Exclusion Principle: A philosophical perspective*. Ph.D. thesis, London: University of London.

Massimi, M. (forthcoming). What demonstrative induction can do against the threat of underdetermination: Bohr, Heisenberg, and Pauli on spectroscopic anomalies 1921–1924. *Synthese*, forthcoming.

Messiah, A. M. L., & Greenberg, O. W. (1964). Symmetrization postulate and its experimental foundation. *Physical Review*, *136B*, 248–267.

Norton, J. D. (1993). The determination of theory by evidence: The case for quantum discontinuity. *Synthese*, *97*, 1–31.

Norton, J. D. (1994). Science and certainty. *Synthese*, *99*, 3–22.
Novozhilov, Y. V. (1975). *Introduction to elementary particle theory*. Oxford: Pergamon Press.
Oppenheimer, J. R. (1930). Two notes on the probability of radiative transitions. *Physical Review*, *35*, 939–947.
Pauli, W. (1925). Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik*, *31*, 765–783.
Pauli, W. (1936). Théorie quantique relativiste des particules obeisant á la statistique de Einstein–Bose. *Annals de Institut Henry Poincaré*, *6*, 109–136.
Pauli, W. (1940). The connection between spin and statistics. *Physical Review*, *58*, 716–722.
Pauli, W. (1985). In K. von Meyenn (Ed.), *Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u.a., Band 2: 1930–1939*. Berlin: Springer.
Pauli, W., & Weisskopf, V. (1934). Über die Quantisierung der skalaren relativistischen Wellengleichung. *Helvetica Physica Acta*, *7*, 709–731.
Saunders, S. (1991). The negative-energy sea. In S. Saunders, & H. R. Brown (Eds.), *The philosophy of vacuum* (pp. 65–109). Oxford: Clarendon Press.
Schwinger, J. (1951). The theory of quantized fields. I. *Physical Review*, *82*, 914–927.
Segal, I. (1964). Quantum field and analysis in the solution manifolds of differential equations. In W. Martin, & I. Segal (Eds.), *Analysis in function space* (pp. 129–153). Cambridge, MA: MIT Press.
Teller, P. (1995). *An interpretive introduction to quantum field theory*. Princeton: Princeton University Press.
Uhlenbeck, G. E., & Goudsmit, S. (1925). Ersetzung der Hypothese vom unmechanischer Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons. *Naturwissenschaften*, *13*, 953–954.
Uhlenbeck, G. E., & Goudsmit, S. (1926). Spinning electron and the structure of spectra. *Nature*, *117*, 264–265.
Weinberg, S. (1964). Feynman rules for any spin. *Physical Review*, *133*, B1318–B1332.
Weinberg, S. (1995). *The quantum theory of fields*, *Vol. I*. Cambridge: Cambridge University Press.
Weinberg, S. (1999). What is quantum field theory, and what did we think it was? In Tian Yu Cao (Ed.), *Conceptual foundations of quantum field theory*. Cambridge: Cambridge University Press.
Weisskopf, V. (1934). Über die Selbstenergie des Elektrons. *Zeitschrift für Physik*, *89*, 27–39.
Weisskopf, V. F. (1983). Growing up with field theory: The development of quantum electrodynamics. In L. M. Brown, & L. Hoddeson (Eds.), *The birth of particle physics* (pp. 56–81). Cambridge: Cambridge University Press.
Weyl, H. (1929). Elektrons und Gravitation. I. *Zeitschrift für Physik*, *56*, 330–352.
Weyl, H. (1930). *The theory of groups and quantum mechanics*. New York: Dover.
Wick, G. C. (1950). Evaluation of the collision matrix. *Physical Review*, *80*, 268–273.
Wightman, A. S. (1972). The Dirac equation. In A. Salam, & E. P. Wigner (Eds.), *Aspects of quantum theory* (pp. 95–115). Cambridge: Cambridge University Press.
Woodhouse, N. (1980). *Geometric quantization*. Oxford: Clarendon Press.