

# Real Spinor Fields

David Hestenes

The Dirac equation is expressed entirely in terms of geometrical quantities by providing a geometrical interpretation for the  $(-1)^{1/2}$  which appears explicitly in the Dirac equation. In the modification of the Dirac electron theory which ensues, the  $(-1)^{1/2}$  appears as the generator of rotations in the spacelike plane orthogonal to the plane containing the electron current and spin vectors. This amounts to a further “relativistic” constraint on the spinor theory and so may be expected to have physical consequences. It does not, however, conflict with well-substantiated features of the Dirac theory.

## 1. Introduction

In 1928, Dirac proposed a relativistically invariant first-order wave equation for the electron.<sup>1</sup> Dirac’s theory has led to a complex of physical explanations and predictions at once so surprising and convincing that it has gained general acceptance among physicists today. The previously perplexing phenomena of electron spin was not only accounted for, but fine details of the hydrogen spectra and an accurate value of the electron magnetic moment were calculated without arbitrary assumptions. Moreover, after some theoretical trauma, it was realized that Dirac’s equation entails the existence of a positively charged electron—at just about the same time that such a particle was discovered experimentally. We do well to understand precisely what features of the Dirac equation entail these remarkable results.

To produce a wave equation which is both first order in time and relativistically invariant, Dirac constructed the matrix algebra which bears his name. As Dirac averred, this leads to an explanation of electron spin “without arbitrary assumptions.” But one mysterious feature of the electron wavefunction seems to be left unexplained. Why is it a complex function? In Dirac’s equation, which is largely determined by the requirement of relativistic invariance, why does an uninterpreted  $(-1)^{1/2}$  appear explicitly? As the increasing theoretical importance of antiparticle conjugation tends to show, the appearance of this  $(-1)^{1/2}$  is no triviality.

We submit that the  $(-1)^{1/2}$  in Dirac’s equation can be interpreted geometrically and that the reason for appearance in physics is inseparable from that spin. To appreciate this viewpoint it is necessary understand the full geometrical significance of the Dirac algebra.

The metrical properties of space-time can be represented by introducing appropriate rules for the multiplication of space-time vectors. The result is a Clifford algebra which can be thought of as an algebra of directions in space-time. It is conveniently called the real Dirac algebra, because it is isomorphic to the algebra of Dirac matrices over the real numbers. The requirements of relativity can be satisfied by writing all physical equations in terms of the real Dirac algebra. Relativity provides no justification for the use of additional “complex” numbers such as those in the “complex Dirac algebra.”

---

<sup>1</sup> P. A. M. Dirac, Proc. Roy. Soc. (London) 117, 610 (1928).

In contrast to the “complex matrix algebra of Dirac,” the real Dirac algebra has a thoroughly geometrical significance. By re-expressing Dirac’s theory in terms of the real Dirac algebra, we find that the  $(-1)^{1/2}$  which appears explicitly in Dirac’s equation acquires a geometrical interpretation as the generator of rotations in a spacelike plane. Moreover, the orientation of this plane is described by the electron “spin.” In this way spin and “complex” numbers are combined in a single geometric entity.

The theory presented in this paper is algebraically isomorphic to Dirac’s; it can be provided with an equivalent physical interpretation as well. It differs from Dirac’s in that all its algebraic ingredients have a geometrical significance determined by properties we attribute to space-time. This produces a clarification of the Dirac theory, especially as regards the role of complex numbers. It also introduces new possibilities for modifying the theory. But to avoid prejudging the issues, a fairly conservative approach is adopted in this paper. Only a relatively minor amendment to the physical interpretation of the Dirac theory is suggested. In subsequent papers, upon the firm base of the Dirac theory, more general theories will be constructed which incorporate additional physical facts and exploit the geometrical and physical interpretations of  $(-1)^{1/2}$  given here.

This paper is divided into three sections. In Sec. A the real Dirac algebra is briefly described. In Sec. B the algebraic expression for a real spinor field is given. For a Dirac particle, the wavefunction and its symmetries are provided with both physical and geometrical interpretations. The only departure from Dirac theory is to be found in a new and simple distinction between positive and negative energy states. This innovation requires a new expression for the its charge-current density. However, since in the one-particle approximation this current is the same as Dirac’s current, any consequences which this alteration may entail is manifested only in quantum electrodynamical calculations—somewhat beyond the scope of this paper. Finally, connection with the Dirac theory is made in Sec. C. Readers who are convinced that Sec. B is correct will not find it necessary to read Sec. C.

## A. Space-Time Algebra

We use the geometric algebra developed by Hestenes.<sup>2</sup> However, without assuming prior familiarity with space-time algebra (STA), this section tries to give an explanation of the algebra which will be sufficient for the purposes of this paper. To do it succinctly, it is assumed that the reader is well acquainted with the matrix form of the Dirac algebra. He is then asked to re-interpret this algebra geometrically in a prescribed way. If the reader finds the treatment here excessively difficult to follow or too sparse to be satisfying, he is referred to the more detailed discussion given in Ref. 2.

### 1. Vectors

Physicists are accustomed to thinking of the  $\gamma_\mu$  ( $\mu = 0, 1, 2, 3$ ) as four by four matrices which are the four components of single world vector in space-time. Instead, the reader is asked to think of the  $\gamma_\mu$  as a frame of four orthonormal vectors in space-time. Think of  $\gamma_0$  as a unit vector in the forward light cone and the  $\gamma_i$  ( $i = 1, 2, 3$ ) as a right-handed set of

---

<sup>2</sup> D. Hestenes, Space-Time Algebra (Gordon and Breach Science Publishers, Inc., New York, 1966).

spacelike vectors. Thus,  $\gamma_0$  is the unit normal to the (3-dimensional) spacelike hyperplane spanned by the  $\gamma_i$ . Moreover, any world vector  $A$  can be written as a linear combination of the  $\gamma_\mu$ ,

$$A = A^\mu \gamma_\mu. \quad (1.1)$$

The  $A^\mu$  are the components of the vector  $A$  with respect to the set of base vectors  $\{\gamma_\mu\}$ .

Although the reader is asked to think of the  $\gamma_\mu$  as vectors instead of matrices, they still retain the multiplication rules possessed by the Dirac matrices. This makes new demands on our understanding of geometry. Presumably, the matrix multiplication is clear enough, but, in view of the geometrical significance of vectors, it would seem that, in order to justify multiplication of vectors, we must supply it with a geometric interpretation. This can, in fact, be done. Assuming that two vectors  $A$  and  $B$  can be multiplied “like matrices,” the product  $AB$  can be understood geometrically by separating it into symmetric and antisymmetric parts:

$$AB = (A \cdot B + B \wedge A) \quad (1.2a)$$

$$A \cdot B = \frac{1}{2}(AB + BA) \quad (1.2b)$$

$$A \wedge B = \frac{1}{2}(AB - BA) \quad (1.2c)$$

A special notation has been introduced for the symmetric and antisymmetric products because each has a geometrical interpretation independent of the other. The quantity  $A \cdot B$  is a scalar; the dot signifies the familiar *inner product* of vectors. The quantity  $A \wedge B$  is commonly called a *bivector* or *2-vector*; the wedge signifies the *outer product* of vectors. Although the outer product is not so familiar to physicists, it was invented and interpreted geometrically many years before the invention of matrix algebra. For the purposes of this paper, it is not so important that the reader appreciate the geometrical interpretation of the outer product<sup>3</sup>; he need only believe that such an interpretation can be supplied, in order to appreciate that by (1.1) the different geometrical products  $A \cdot B$  and  $A \wedge B$  are united in a single product  $AB$ . Thus, while the product  $AB$  of two vectors is not itself a vector, it is nevertheless composed of quantities with geometrical significance every bit as definite as that of vectors. Let the reader be assured that, in a similar way, the geometric character of the product of any number of vectors can be divined.

It may help the reader to see (1.2) written out for the base vectors  $\{\gamma_\mu\}$ . Then Eq. (1.2a) becomes

$$\gamma_\mu \gamma_\nu = \gamma_\mu \cdot \gamma_\nu + \gamma_\mu \wedge \gamma_\nu. \quad (1.3a)$$

Because of the orthogonality of the  $\gamma_\mu$ ,

$$\gamma_\mu \gamma_\nu = \gamma_\mu \cdot \gamma_\nu, \quad \text{if } \mu = \nu \quad (1.3b)$$

$$\gamma_\mu \gamma_\nu = \gamma_\mu \wedge \gamma_\nu, \quad \text{if } \mu \neq \nu \quad (1.3c)$$

The set of inner products of the base vectors  $\{g_{\mu\nu} \equiv \gamma_\mu \cdot \gamma_\nu\}$  is the so-called “metric tensor” expressed in the frame  $\{\gamma_\mu\}$ . Using the metric tensor, (1.2b) allows us to write down an equation which is familiar to everyone acquainted with the Dirac algebra,

$$g_{\mu\nu} = \frac{1}{2}(\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu) \quad (1.4)$$

---

<sup>3</sup> The geometrical significance of the outer product is discussed in Pt. I of Ref. 2.

To display explicitly the signature we attribute to the metric of space-time, we write out the nonvanishing components of (1.4),

$$\gamma_0^2 = 1, \quad \gamma_i^2 = -1 \quad (i = 1, 2, 3). \quad (1.5)$$

A word of warning is in order to help the reader guard against misunderstanding. It is evident that the word “vector” is used in the algebraic sense (as an element of a linear space) rather than in the tensor sense (as a set of elements with certain transformation properties). The word “scalar” is also used in the algebraic sense (as an element of the number field over which a linear space is defined) rather than in the tensor sense (as an invariant of some group of transformations). For instance, we would call a component of the vector in (1.1) a scalar; it is the inner product of two vectors.

$$A_\mu = A \cdot \gamma_\mu = A^\nu g_{\mu\nu}. \quad (1.6)$$

(Note that we use the usual convention of tensor analysis to raise and lower indices.) It is also worth saying that  $A_\mu$  must be a real number (or function). We do not allow the field of complex numbers, for this would introduce a  $(-1)^{1/2}$  without geometrical significance. Soon it should be clear that we already have many algebraic objects with a well-defined geometrical significance which can play the role of  $(-1)^{1/2}$ .

By taking all linear combinations (over the real numbers) of all linearly independent polynomials of the  $\gamma_\mu$ , we obtain the elements of a geometric algebra with 16 linearly independent elements. We can call it the “real Dirac algebra” to emphasize its similarity, to Dirac’s matrix algebra except as regards the number field. Or, we can call it a “vector algebra” to emphasize that all the elements can be constructed out of vectors. Perhaps it is best to call it a “space-time algebra” to emphasize its use as an algebraic representation of the primitive geometrical properties of space-time.

Once we have the space-time algebra, we need only define differentiation to have a complete geometric calculus for space-time. Let  $\{x^\mu; \mu = 0, 1, 2, 3\}$  be a set of inertial coordinates<sup>4</sup> in space-time. The fundamental vector differential operator, written as  $\square$ , can be introduced by the formula

$$\square = \gamma^\mu \partial_\mu, \quad \text{where} \quad \partial_\mu \equiv \partial/\partial x^\mu. \quad (1.7)$$

This operator was first introduced into physics by Dirac. In his effort to construct a “relativistically invariant” first-order differential equation for the electron wavefunction, he was led, in effect, to take the square root of the d’Alembertian  $\square^2$ . The appropriate square root is  $\square$ , as our notation indicates. But, by now the reader needs hardly be reminded of the difference between our interpretation of the  $\gamma^\mu$  and Dirac’s.

It is preferable to call  $\square$  the *gradient* operator, because the name agrees with common parlance when  $\square$  operates on scalars. For example, by (1.7), the gradient of the coordinate function  $x^\mu$  is

$$\square x^\mu = \gamma^\mu \quad (1.8)$$

---

<sup>4</sup> Among other names, these coordinates are sometimes called Minkowski, pseudo-Euclidean, Euclidean, or Cartesian coordinates. We prefer the physical name, since in these coordinates the graph of the world line of a point particle with no forces acting on it is a straight line.

This shows that  $\gamma^\mu$  is a vector pointing in the direction of maximum increase in  $x^\mu$ ; the magnitude of  $\gamma^\mu$  indicates the rate of change of  $x^\mu$ . It may be noted that the meaning of  $\gamma^\mu$  is surreptitiously altered to the point where  $\gamma^\mu$  is to be thought of as a vector field, assigning a tangent vector  $\gamma^\mu(x)$  to each point  $x$  in space-time. However, to be consistent with the definition of inertial coordinates,  $\gamma^\mu$  must be a constant vector field, i.e.,

$$\partial_\nu \gamma^\mu = 0. \quad (1.9)$$

So, in a sense, the  $\gamma^\mu$  at different points are equivalent.

We are accustomed to think of the gradient of a scalar field as the normal to an “equipotential” surface. The gradient of a vector field  $A$  is interpreted differently. To get at this interpretation, we follow (1.2) and decompose  $\square A$  into symmetric and antisymmetric parts.

$$\square A = \square \cdot A + \square \wedge A. \quad (1.10)$$

The scalar part  $\square \cdot A$  is the usual *divergence* of a vector, and  $\square \wedge A$  is the *curl*. If  $A$  is the electromagnetic vector potential, then  $F = \square \wedge A$  is the electromagnetic field strength.<sup>5</sup>

## 2. Conjugation

We are familiar with the operations of transpose, complex conjugation, and Hermitian conjugation in matrix algebra. Analogous “conjugation” operations can be defined for the real Dirac algebra. But first we must get more insight into the structure of the algebra.

Consider the product  $A_1 A_2 \cdots A_r$  of  $r$ -vectors  $A_1, \dots, A_r$ . The part of  $A_1 A_2 \cdots A_r$  which is antisymmetric under interchange of any two vectors is called an  $r$ -vector; it is a generalization of (1.2c) and is written  $A_1 \wedge \cdots \wedge A_r$ . This quantity vanishes if the vectors  $A_1, A_2, \dots, A_r$  are linearly dependent. Since in space-time we can find at most four linearly independent vectors, we must have  $r \leq 4$ .

Let us call an element of the real Dirac algebra a  $d$ -number. It can be proved that any  $d$ -number is expressible as a linear combination of  $r$ -vectors simply by using linearly independent  $r$ -vectors constructed from the  $\gamma_\mu$  and the convention that scalars be called 0-vectors. Thus, any  $d$ -number  $\psi$  can be written

$$\psi = \psi_S + \psi_V + \psi_B + \psi_T + \psi_P, \quad (2.1)$$

where the subscripts  $S, V, B, T, P$  mean, respectively, scalar (0-vector) part, vector (1-vector) part, bivector (2-vector) part, trivector or pseudovector (3-vector) part, pseudoscalar (4-vector) part. The decomposition (2.1) is analogous to the separation of a complex number into real and imaginary parts, which partly explains why  $\psi$  is sometimes called a “hypercomplex” number.

By reversing the products of all vectors in the Dirac algebra, we obtain from  $\psi$  a new  $d$ -number  $\tilde{\psi}$

$$\tilde{\psi} = \psi_S + \psi_V - \psi_B - \psi_T + \psi_P, \quad (2.2)$$

---

<sup>5</sup> The equations of electrodynamics are formulated in terms of the real Dirac algebra in Pt. II of Ref. 2.

This is an invariant kind of “conjugation” in the sense that it is independent of any basis in the algebra. The unit pseudoscalar is so important that we represent it by the special symbol  $i$ . We can express  $i$  in terms of the  $\gamma_\mu$ ,

$$i = \gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3 = \gamma_0 \gamma_1 \gamma_2 \gamma_3. \quad (2.3)$$

Reasons for using the symbol  $i$  for this quantity are apparent in the discussion that follows.

With the help of  $i$  we can define an operation which reverses the direction of all vectors in space-time. We call space-time conjugation<sup>6</sup> the operation which maps  $\psi$  into  $\bar{\psi}$ , where

$$\bar{\psi} = -i\psi i = \psi_S - \psi_V + \psi_B - \psi_T + \psi_P. \quad (2.4)$$

A  $d$ -number  $\psi$  is called *even* if  $\bar{\psi} = \psi$  and *odd* if  $\bar{\psi} = -\psi$ .

The set of all even  $d$ -numbers  $\psi$  forms a subalgebra which can be identified as the Pauli algebra. An inertial frame determines a right-handed set of base vectors  $\sigma_i$  ( $i = 1, 2, 3$ ) in the Pauli algebra by the definitions

$$\sigma_i = \gamma_i \gamma_0. \quad (2.5)$$

From (2.3), it is clear that

$$\sigma_1 \sigma_2 \sigma_3 = i. \quad (2.6)$$

It is useful to introduce special symbols for two other kinds of conjugation,

$$\psi^* = \gamma_0 \psi \gamma_0 \quad (2.7)$$

$$\psi^\dagger = \gamma_0 \tilde{\psi} \gamma_0 \quad (2.8)$$

Clearly  $\psi^*$  and  $\psi^\dagger$  depend on the choice of  $\gamma_0$ . The dagger symbol in (2.8) is particularly appropriate because  $\psi^\dagger$  corresponds to the Hermitian conjugate of  $\psi$  in the usual matrix representations of the Dirac algebra.

Notice that the operation (2.7) changes  $\{\sigma_i\}$  and also  $\{\gamma_i\}$  into left-handed frames without affecting  $\gamma_0$ . For this reason, the operation which takes  $\psi$  into  $\psi^*$  is called space conjugation.<sup>6</sup>

### 3. Frames and Lorentz Rotations

It can be proved that every field of orthonormal frames  $\{e_\mu(x)\}$  with the same orientation as  $\{\gamma\}$  can be obtained from  $\{\gamma_\mu\}$  by a field of *Lorentz rotations*<sup>7</sup> expressed as follows:

$$e_\mu(x) = R(x) \gamma_\mu \tilde{R}(x),$$

where  $R$  is a  $d$ -number field with the properties<sup>8</sup>

$$R \tilde{R} = 1, \quad \text{and} \quad R = \bar{R}.$$

---

<sup>6</sup> The word “reflection” is reserved for an operation which inverts points of space-time. The more general term “conjugate” is used for this operation which reverses the direction of tangent vectors but leaves space-time points unchanged.

<sup>7</sup> Lorentz rotations are discussed in detail in Part IV of Ref. 2.

<sup>8</sup> We will usually suppress the argument  $x$ .

It may help the reader to observe that (3.1) has the same form as a similarity transformation, which in Dirac's matrix algebra is used to transform from one matrix representation of the  $\{\gamma_\mu\}$  to another. Here, of course, the interpretation is different, and the operator  $R$  is allowed to be a function of the space-time point  $x$ . Except for a sign,  $R$  is uniquely determined by the  $e_\mu$  and the  $\gamma_\mu$ .<sup>9</sup> Because, by (3.1),  $R$  determines the transformation of the  $\gamma_\mu$  into the  $e_\mu$ , *it is sometimes convenient to refer to  $R$  itself as a Lorentz rotation.*

The conditions (3.2) imply that  $R$  can be written in the form

$$R = \pm e^{-\frac{1}{2}\phi}, \quad (3.3)$$

where  $\phi$  is a bivector field. Conversely, every  $d$ -number of the form (3.3) represents a unique Lorentz rotation.

We call (3.1) a *spatial rotation* if  $R = R^*$  and a *special timelike rotation* (or a *boost*) if  $R = R^\dagger$ . Any Lorentz rotation can be expressed as a spatial rotation followed by a special timelike rotation. This means that  $R$  can always be written in the form

$$R = e^{-\frac{1}{2}\chi} e^{-\frac{1}{2}i\theta}, \quad (3.4)$$

where  $\chi$  and  $\theta$  are bivectors satisfying  $\chi^* = \chi$  and  $\theta^* = -\theta$ .

Any Lorentz transformation can be expressed as some combination of space-time conjugation (2.4) and space conjugation (2.7), and a Lorentz rotation (3.1).

## B. Spinor Fields

At this point, it is hoped that the reader is sufficiently prepared to grapple with a reformulation of Dirac's electron theory in terms of our "space-time calculus." This theory can, in a certain sense, be derived from Dirac's theory. But, the derivation is tedious. Moreover, the end result is relatively simple and be set forth and understood as a physical theory in its own right. Therefore, we present and discuss the reformulated theory directly. The derivation has been relegated to the end of this paper where it can be studied separately.

The electron wavefunction has certain suggestive features which may have a physical significance transcending the Dirac equation, so we begin by discussing the wavefunction and its interpretation without reference to a "wave equation." Then we write down the reformulated Dirac equation and discuss some of the implications it has for interpretation of the wavefunction. The theory given here is algebraically isomorphic to Dirac's theory. However, it differs in that all the algebraic quantities involved have a definite geometrical significance determined by the properties already ascribe to space-time in our construction of the space-time algebra. This, in turn, suggests certain modification of the physical theory, although, except for a relatively minor exception, we resist the temptation to pursue them here and are content with showing how the physical features of the Dirac theory are to be formulated in our new language.

## 4. The Wavefunction

We begin simply by writing down the most succinct expression for a spinor field,

$$\psi(x) = e^{\frac{1}{2}\mu(x)}, \quad (4.1)$$

---

<sup>9</sup> This is shown explicitly by Eq. (17.24) of Ref. 2.

where  $\mu(x)$  is an *even  $d$ -number field*. Let us call  $\psi(x)$  the *real* spinor field to emphasize that it is expressed in terms of the real Dirac algebra and so is susceptible to a geometrical interpretation. “Real spinors” are equivalent to “Dirac spinors” in the sense that two can be put into one-to-one correspondence. This is proved in Sec. C, but a simple check on the asserertion can be made immediately by noting that since  $\mu$  is an even  $d$ -number it has eight linearly independent components. This agrees with the fact that a spinor is a column matrix with four complex components.

While (4.1) may well be the most appropriate expression of a spinor field for certain kinds of mathematical analysis, to explain its geometrical and physical significance we must decompose a spinor into its “geometrically distinct” parts. Therefore, write  $\mu$  in the form

$$\mu(x) = \alpha(x) + i\beta(x) + \phi(x), \quad (4.2)$$

where  $\alpha$  and  $\beta$  are scalars and  $\phi$  is a bivector in the Dirac algebra. Also define

$$\rho(x) = e^{\alpha(x)} > 0 \quad (4.3)$$

and use (3.3) to write  $\psi$  in the form

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R. \quad (4.4)$$

Now observe that  $\psi$  determines four “current vectors”  $J_\mu$  as follows:

$$J_\mu \equiv \psi \gamma_\mu \tilde{\psi} = \rho e_\mu. \quad (4.5)$$

The  $e_\mu$  are the orthonormal vectors defined by (3.1);  $J_0$  is a “positive” timelike vector,<sup>10</sup> equivalent to the *probability current* density of Dirac theory. Following Dirac, we must interpret  $\rho(x)$  as the probability density in the instantaneous rest frame at the point  $x$ . More briefly,  $\rho(x)$  is the *proper probability density*. To conserve probability, the divergence of  $J_0$  must vanish, i.e.,

$$\square \cdot J_0 = 0. \quad (4.6)$$

This provides a restriction on the wave equation for  $\psi$ .

The vector  $J_0$  describes the orientation of the electron “spin.” In commonly used language, it is “the expectation value of the electron spin.” The essential property of the “spin vector” is that it is orthogonal to  $J_0$ . The labeling of the “spin vector” with the subscript three, rather than with one or two is merely a convention.

The vectors  $J_1$  and  $J_2$  do not inherit independent physical interpretations from the Dirac theory. However, the plane containing  $J_1$  and  $J_2$ , which can be described by the bivector  $J_1 \wedge J_2$ , does have physical significance. The abstract complex plane in which the phase factor of the Dirac wavefunction is defined corresponds to this plane. The  $(-1)^{\frac{1}{2}}$  in the electron wavefunction can now be interpreted geometrically as the generator of rotations in the plane  $J_1 \wedge J_2$ . Therefore, we propose that this plane be called the *phase plane* of the wavefunction. It can also fairly be called the *spin plane*, because it is completely determined

---

<sup>10</sup> “Positive” means that  $J_0$  is in the forward light cone. This then can be conveniently expressed by  $J_0 \cdot \gamma_0 > 0$ .

by the spin and current vectors. This can be seen easily by noting that, by virtue of (2.3), the generator  $e_2e_1$  of rotations in the phase plane can be written

$$e_2e_1 = ie_0e_3. \quad (4.7)$$

The notions of spin and phase are now inextricably united. The “phase” describes the magnitude of a rotation in the phase plane and the “spin” describes the orientation of the phase plane.

So far we understand the factors  $\rho$  and  $R$  in the wavefunction (4.4). The scalar  $\rho$  describes the probability density, and the Lorentz rotation  $R$  describes in a single unit the orientation of the probability current vector and the spin and the phase of the wavefunction.

But what can be the meaning of the factor  $e^{i\beta}$ ? Evidently, it does not play a role in determining the observables just mentioned because it does not appear in the expression (4.5) for the current vectors. Inversely, given the “observables”  $J_\mu$ , we can find  $\rho$  and  $R$ , but to determine  $\psi$  uniquely we need one more bilinear function of  $\psi$ , namely,

$$\psi\tilde{\psi} = \rho e^{i\beta}. \quad (4.8)$$

Now there is one “observable” appearing in the Dirac theory of which we have given no account, namely, the sign of the energy. This “observable” is described by the quantity  $\rho e^{i\beta}$ . We suppose that  $\rho e^{i\beta}$  describes the relative admixture of positive and negative energy components of  $\psi$ . More precisely, we take  $\rho e^{i\beta(x)}$  to be the relative probability of observing a particle at the point  $x$ ; so that if  $\rho e^{i\beta} = 1$  everywhere,  $\psi$  describes a pure *one-particle* state, and if  $\rho e^{i\beta} = -1$  everywhere  $\psi$  describes a pure *antiparticle* state. In view of the interpretation we have given to  $\psi$  the scalar part of  $\psi\tilde{\psi}$ ,

$$(\psi\tilde{\psi})_S = \rho \cos \beta, \quad (4.9)$$

can be interpreted as the proper particle density of the spinor field  $\psi$ . Thus, a negative value of this quantity indicates the likelihood of observing an antiparticle. Of course, in a quantized version of our theory all the physical observables we have discussed are vacuum expectation values of quantum field operators.

This is a good place to summarize once again the relation of our theory to Dirac’s. The theories are equivalent algebraically, but our theory admits a geometrical interpretation of physical quantities which is absent in Dirac’s theory. We can, if we wish, give a physical interpretation of our theory which is precisely equivalent to Dirac’s. However, the geometric significance which our theory endows to physical quantities suggests other possibilities. Just the same, in this paper we proffer only a relatively minor amendment to Dirac’s interpretation by identifying a particular factor in the wavefunction as responsible for the admixture of positive and negative energy states. In all honesty, it must be admitted that this idea seems to produce some difficulty with the “superposition principle” which is not fully understood. On the other hand, we encounter more to be said in its favor when we study the Dirac equation in the next section.

With the notions we already have at our disposal, we can discuss the nonrelativistic limit of the electron wavefunction without reference to a specific wave equation. Earlier we identified the  $\gamma_\mu$  with an inertial frame. To “fix the phase” of  $\psi$ , we can identify the  $\gamma_\mu$  in

(4.5) with the “laboratory frame.” Now suppose the electron is in a pure positive energy state, so  $e^{i\beta} = 1$ , and use (3.3) to write the wavefunction in the form

$$\psi = \rho^{\frac{1}{2}} e^{-\frac{1}{2}i\phi}. \quad (4.10)$$

The factor  $\rho^{\frac{1}{2}}$  is the amplitude of the wavefunction and the bivector  $\frac{1}{2}\phi$  may be thought of as a “relativistic phase.” The “relativistic phase factor” can be decomposed by using (3.4) to express it as a spatial rotation  $e^{-\frac{1}{2}i\theta(x)}$  in the laboratory frame followed by a Lorentz transformation  $e^{-\frac{1}{2}\chi(x)}$  which, without rotation, takes the laboratory frame into the instantaneous rest frame of the electron:

$$\psi = \rho^{\frac{1}{2}} e^{-\frac{1}{2}\chi} e^{-\frac{1}{2}i\theta}. \quad (4.11)$$

In the nonrelativistic approximation, the factor  $e^{-\frac{1}{2}\chi}$  is negligible, so the state of the electron can be described by the effective wavefunction

$$\psi' = \rho^{\frac{1}{2}} e^{-\frac{1}{2}i\theta}. \quad (4.12)$$

This is equivalent to the Pauli wavefunction. The bivector  $\frac{1}{2}i\theta(x)$  has an orientation which describes the orientation of the electron “spin” and magnitude which is just the “scalar phase” of the wavefunction. If, as a further approximation, it is supposed that the precession of the spin plane is negligible, then we can write

$$\frac{1}{2}i\theta(x) = i\sigma_3\delta(x) = \gamma_2\gamma_1\delta(x), \quad (4.13)$$

where  $\delta(x)$  is a “scalar phase.” The effective wavefunction can be written as

$$\psi''(x) = \rho^{\frac{1}{2}} \exp[-\gamma_2\gamma_1\delta(x)]. \quad (4.14)$$

This is equivalent to the Schrödinger wavefunction. However, in (4.14) the phase factor is to be thought of as a rotation in a spacelike plane (the spin plane) rather than as a function in some abstract complex space. In the “Schrödinger approximation” the orientation of the spin plane is a constant.

## 5. Symmetries of the Wavefunction

For  $\psi$  the equivalent of the *Dirac equation* for an electron is

$$\square\psi = (m\psi\gamma_0 + eA\psi)i\sigma_3. \quad (5.1)$$

Here,  $m$  is the mass,  $e$  is the charge of the electron, and  $A$  is the electromagnetic vector potential. In this section, we discuss certain symmetries of  $\psi$  associated with the Dirac equation. A *symmetry* of a spinor field  $\psi(x)$  is a mapping of the field  $\psi$  onto itself which preserves the wave equation for  $\psi$  or changes it in a definite and physically meaningful way.

We have already seen that  $\psi$  determines a frame of tangent vectors  $J_\mu(x) = \psi\gamma_\mu\tilde{\psi}$  at each point  $x$  of space-time, and inversely, except for a factor  $e^{i\beta}$ , the  $J_\mu$  determine  $\psi$ . A transformation on  $\psi$  induces a corresponding transformation on the  $J_\mu$ . Because we are familiar with the representation of directions by vectors, a transformation of the  $J_\mu$

can be interpreted geometrically. This enables us to give a geometric interpretation to transformations of  $\psi$ .

There are two distinct kinds of geometrical transformations. First, there is a *transformation of the tangent vectors*  $J_\mu(x)$  at a point  $x$  of space-time into a new set of vectors  $J'_\mu(x)$  at the same point  $x$ . As we see, charge conjugation is a transformation of this kind. Second, there is a *transformation of points* in space-time, wherein the tangent vectors at a point of space-time are mapped into “equivalent” vectors at a different point  $x'$ . Displacements are the simplest transformations of this kind. Symmetries of a spinor field can be interpreted geometrically as some combination of these two kinds of transformations.

First note that (5.1) is invariant under

$$\psi \rightarrow \psi\gamma_0, \quad (5.2)$$

which induces

$$J_0 \rightarrow J_0, \quad J_i \rightarrow -J_i. \quad (5.3)$$

This transformation tells us that the Dirac equation does not distinguish (or couple) even and odd spinor fields—a fact which is not discovered and so is not interpreted in the usual form of the Dirac theory. Because of this equivalence of even and odd fields, we may, without further comment, confine the rest of our discussion to transformations which leave  $\psi$  even.

An operation which changes the sign of the electromagnetic coupling while leaving the rest of the Dirac equation invariant is called *charge conjugation*. By inspection, this can be accomplished only by multiplying (5.1) on the right by some linear combination of  $\sigma_1$  and  $\sigma_2$ . Therefore, charge conjugation has the general form

$$\begin{aligned} C : \psi &\rightarrow \psi_C = \psi\sigma_2 \exp(i\sigma_3\phi_C) \\ &= \psi \exp(-i\sigma_2\frac{1}{2}\pi) \exp(i\sigma_3\phi_C i), \end{aligned} \quad (5.4)$$

where  $\phi_C$  is a constant scalar. This induces a rotation of  $\pi$  of the  $J_i$  about some axis orthogonal to  $J_3$ . For instance, if  $\phi_C = 0, \pi$ ,

$$J_0 \rightarrow J_0, \quad J_1 \rightarrow -J_1, \quad J_2 \rightarrow J_2, \quad J_3 \rightarrow -J_3. \quad (5.5)$$

We can interpret (5.4) itself as a rotation of the  $J_\mu$  but this is not the whole truth, because (5.4) does not leave  $\psi\tilde{\psi}$  invariant,

$$\psi_C\tilde{\psi}_C = -\psi\tilde{\psi}. \quad (5.6)$$

This is an important point, for it shows that charge conjugation changes the sign of the particle density (4.7). To be more specific, note that the factor

$$\exp(i\sigma_2\frac{1}{2}\pi) = -i\sigma_2$$

in (5.4) reverses the direction of the spin vector  $J_3$  in (5.5) but does not affect the expression  $\psi\tilde{\psi}$ , while the factor  $i$  changes the sign of the particle density but does not affect  $J_3$ .

In Minkowski space-time, a point  $x$  with coordinates  $x^\mu$  can be represented by a “position vector”  $x = x^\mu \gamma_\mu$ . “Space reflection” of points in the hyperplane orthogonal to  $\gamma_0$  can be represented by the mapping

$$P : \quad x \rightarrow x^* = \gamma_0 x \gamma_0 . \quad (5.7)$$

The form of the Dirac equation will be preserved if the point transformation (5.7) induces the following transformation of fields:

$$P : \quad A(x) \rightarrow A^*(x^*) , \quad (5.8)$$

$$P : \quad \psi(x) \rightarrow \psi_P(x) = \psi^*(x^*) e^{i\sigma_3 \phi_P} . \quad (5.9)$$

We must have  $\phi_P = 0$  or  $\pi$ , if we require that successive applications of (5.9) map  $\psi(x)$  back to itself. The transformation  $P$  is commonly called parity conjugation. It induces the transformations,

$$P : \quad J_0(x) \rightarrow J_0^*(x^*) , \quad (5.10)$$

$$P : \quad J_3(x) \rightarrow -J_3^*(x^*) . \quad (5.11)$$

Because the components of the tangent vector  $J_0$  in (5.10) transform in the same way as the components of the position vector  $x$  in (5.7),  $J_0$  is usually said to “transform as a vector” under parity conjugation. Because  $J_3$  transforms with opposite sign it is said to “transform as an axial vector.” It is not necessary for us to subscribe to this distinction between “vector” and “axial vector,” first, because it is not needed to distinguish  $J_0$  from  $J_3$ , and second, because it is purely gratuitous, for by using (5.2) we can keep  $J_0$  fixed and change the sign of  $J_3$  at will without affecting the Dirac equation.

A time-reversal transformation  $T$  can be constructed in analogy to parity conjugation. The transformation

$$T : \quad x \rightarrow \tilde{x}^* = -x^* . \quad (5.12)$$

The corresponding transformation of the wavefunction is

$$T : \quad \psi(x) \rightarrow \psi_T(x) = i\psi^*(-x^*) \exp(i\sigma_3 \phi_T) . \quad (5.13)$$

As long as  $A$  is unique, the Dirac equation determines  $\psi$  to within a factor of a constant Lorentz rotation which can be fixed by a choice of the  $\gamma_\mu$ , so that (4.5) yields definite  $J_\mu$ . But Maxwell’s equation determines  $A$  only up to a gauge transformation,

$$A \rightarrow A' = A + \square \chi . \quad (5.14)$$

If the Dirac equation is to be left invariant, then (5.14) must be accompanied by the transformation

$$\psi \rightarrow \psi' = \psi \exp(i\sigma_3 \chi) . \quad (5.15)$$

This induces the transformation

$$\begin{aligned} J_0 &\rightarrow J_0 , & J_3 &\rightarrow J_3 , \\ J_1 &\rightarrow J_1 \cos 2\chi - J_2 \sin 2\chi , \end{aligned}$$

$$J_2 \rightarrow J_2 \cos 2\chi + J_1 \sin 2\chi,$$

which shows that electromagnetic interactions do not distinguish  $J_1$  from  $J_2$ .

Closely related to the electromagnetic gauge transformations are the conservation laws found by calculating the divergence of the  $J_\mu$  using (5.1) to evaluate

$$[(\square\psi)\gamma_\mu\tilde{\psi} + \psi\gamma_\mu(\square\psi)\tilde{\psi}]_S = (\square\psi\gamma_\mu\tilde{\psi})_S = \square \cdot J_\mu,$$

we find that

$$\square \cdot J_\mu = eiA \wedge e_3 \wedge e_0 \wedge J_\mu \quad (5.17)$$

The expression on the right vanishes identically if  $\mu = 0$  or  $3$ . Therefore, the ‘‘probability current’’  $J_0$  and the ‘‘spin current’’  $J_3$  are conserved.

If, as has been suggested, the expression (4.9) is interpreted as the proper particle density, then the charge-current density  $J(x)$  must be

$$J(x) = eJ_0(x) \cos \beta(x). \quad (5.18)$$

If the total charge is to be conserved, then the divergence of  $J$  must vanish, i.e.,

$$\square \cdot J = 0. \quad (5.19)$$

Since, as (5.17) shows, the Dirac equation implies that the divergence of  $J_0$  also vanishes, (5.19) implies that

$$J_0 \cdot \square\beta = 0. \quad (5.20)$$

Let us see what this condition means.

Imagine that we begin with a charge distribution  $e\rho(x)$  in the rest frame at the point  $x$ . As the electromagnetic interaction is turned on, there is a redistribution of charge due to the creation of particle-antiparticle pairs. The new charge density is  $e\rho(x) \cos \beta(x)$ . Thus, insofar as the problem of renormalization is to calculate the redistribution of charge due to interaction, it is the problem of calculating  $\beta(x)$ . Equation (5.20) says that the gradient of  $\beta$  is orthogonal to  $J_0$ , which is just the reasonable condition that the redistribution of charge is orthogonal to the direction of the current.

Heretofore  $J_0$  has been the only invariant of the Dirac field used in the definition of the charge-current density. Naturally, difficulties were encountered because  $J_0$  is a positive timelike vector and does not change sign under charge conjugation. In the past these difficulties have been surmounted by reinterpreting the wavefunction as a quantized field operator. By contrast, the current vector  $J$  defined in (5.18) achieves the proper behavior under charge conjugation without resorting to the strenuous expedient of second quantization. Whether or not the added condition on the wavefunction which this approach entails will conspire to give a correct account of renormalization is a matter ultimately to be determined by calculation. Unfortunately, such a calculation is a many-particle problem which is beyond the scope of this paper.

## 6. Expansion in Plane Waves

The Dirac equation for a free field is

$$\square\psi = m\psi\gamma_0 i\sigma_3. \quad (6.1)$$

The two positive energy solutions to this equation can be written

$$\psi_i^{(+)} = u_i e^{i\sigma_3 p \cdot x}, \quad (6.2)$$

where the  $u_i$  ( $i = 1, 2$ ) are constant Lorentz rotations. Substituting (6.2) into (6.1) we find

$$p u_i = m u_i \gamma_0,$$

or, since  $u_i \tilde{u}_i = 1$ ,

$$p = m u_i \gamma_0 \tilde{u}_i. \quad (6.3)$$

The solutions  $u_1$  and  $u_2$  described as “spin up” and “spin down,” respectively, can be written in the explicit form

$$u_1 = e^{-\frac{1}{2}\chi}, \quad (6.4)$$

$$u_2 = e^{-\frac{1}{2}\chi} i \sigma_2. \quad (6.5)$$

Corresponding to these solutions, the spin vector is

$$e_3 = u_1 \gamma_3 \tilde{u}_1 = -u_2 \gamma_3 \tilde{u}_2. \quad (6.6)$$

The negative energy solutions to (6.1) can be written

$$\psi_i^{(-)} = v_i e^{i\sigma_3 p \cdot x}, \quad (6.7)$$

where  $i = 1, 2$ , and

$$v_1 = i u_1^* = i e^{\frac{1}{2}\chi}, \quad (6.8)$$

$$v_2 = i u_2^* = i e^{-\frac{1}{2}\chi} i \sigma_2. \quad (6.9)$$

Substituting (6.7) into (6.1), we get

$$p v_i = m v_i \gamma_0, \quad p u_i^* = -m u_i \gamma_0,$$

or

$$p = -m (u_i \gamma_0 \tilde{u}_i)^*. \quad (6.10)$$

Since  $u_i \gamma_0 \tilde{u}_i$  is a positive timelike vector, the energy  $p_0 = p \cdot \gamma_0$  is negative in (6.10).

A general “free-particle” solution  $\psi(x)$  of the Dirac equation can be expanded in terms of the plane-wave solutions,

$$\psi(x) = \int \frac{d^3 p}{(2\pi)^{3/2}} \left( \frac{m}{p_0} \right)^{\frac{1}{2}} \sum_{i=1}^2 (u_i a_i e^{i\sigma_3 p \cdot x} + v_i b_i e^{-i\sigma_3 p \cdot x}). \quad (6.11)$$

In this expression  $p_0 = p \cdot \gamma_0 > 0$ , and the  $a_i$  (and  $b_i$ ) are “complex numbers” of the form

$$a_i = \alpha_i + i \sigma_3 \beta_i,$$

where  $\alpha_i$  and  $\beta_i$  are scalars. It is clear that here role of  $(-1)^{\frac{1}{2}}$  in the usual form of the Dirac theory taken over by  $i \sigma_3 = \gamma_1 \gamma_2$ , the generator of rotations in the 12-plane. The

plane-wave “phase factor”  $\exp(i\sigma_3 p \cdot x)$  is literally a rotation in the plane containing  $\gamma_1$  and  $\gamma_2$ . It induces a rotation in the plane of  $J_1$  and  $J_2$ . The expansion (6.11) is basically an expansion of  $\psi(x)$  in terms of rotations in the 12-plane.

## C. Connection with Dirac Theory

### 7. Bilinear Forms

In STA the Dirac equation for an electron interacting with an electromagnetic field was written

$$\square \Psi = (m + eA)\Psi i, \quad (7.1)$$

where

$$\Psi = \Psi \frac{1}{2}(m + \sigma_3). \quad (7.2)$$

Equation (7.1) is algebraically equivalent to the Dirac equation in its matrix form.<sup>11</sup> But there is a significant difference in that the uninterpreted  $(-1)^{\frac{1}{2}}$  which appears in the Dirac theory has been replaced by a geometrically significant “root of minus one,” namely the pseudoscalar  $i$ . This substitution in no way abrogates the well-known and verified features of the Dirac equation. However, it does give all elements of Dirac equation a well-defined geometrical meaning, and it is just this feature which enables us to replace  $\Psi$  by the simple wavefunction  $\psi$  which we have already discussed in detail. Actually, as we have seen, it is  $i\sigma_3$  rather than  $i$  which is properly identified as the  $(-1)^{\frac{1}{2}}$  which appears in the Dirac and Schrödinger equations. In the form (7.1),  $i\sigma_3$  and  $i$  seem indistinguishable for, according to (7.2),  $\Psi i = \Psi i\sigma_3$ . But as we take  $\Psi$  apart, the difference between these two roots of minus one will become important.

We begin our study of  $\Psi$  by analyzing the structure of certain bilinear forms of  $\Psi$  which are given physical interpretation in the Dirac theory. The vector part of  $\Psi\gamma_0\tilde{\Psi}$  is just that vector which was identified by Dirac as the probability current  $J_0$  of the electron field. Note that  $(\Psi\gamma_0\tilde{\Psi}) = (\Psi\gamma_0\tilde{\Psi})$ , so by virtue of (2.2)  $\Psi\gamma_0\tilde{\Psi}$  has only scalar, vector, and pseudoscalar parts. Hence, we can write

$$\Psi\gamma_0\tilde{\Psi} = \Psi\Psi^\dagger\gamma_0 = J_0 + \rho e^{i\beta}. \quad (7.3)$$

Here the scalar part is  $\rho \cos\beta$  and the pseudoscalar  $i\rho \sin\beta$ .

Now note that  $(\tilde{\Psi}\Psi) = \tilde{\Psi}\Psi$  so that  $\tilde{\Psi}\Psi$  can have only scalar, vector, and pseudoscalar parts. But, because of (7.2),

$$\tilde{\Psi}\Psi = \frac{1}{2}(1 - \sigma_3)\tilde{\Psi}\Psi\frac{1}{2}(1 + \sigma_3),$$

---

<sup>11</sup> A spinor can be defined as a solution to the equations  $\Psi\tilde{\Psi} = \Psi i\tilde{\Psi} = 0$ . For every such solution, there is a bivector  $\sigma_3$  (with  $\sigma_3^2 = 1$ ) so that (7.2) holds. Equation (7.1) implies that  $\sigma_3$  must be constant. The definition of spinor given here is equivalent to the definition (given in Ref. 2) of a spinor as an element of a minimal ideal in the Dirac algebra. Actually, there are an infinite number of minimal ideals. The electron wavefunction is an element of a particular minimal ideal characterized by  $\sigma_3$ , as (7.2) shows. As our analysis in this paper shows, the significance of  $\sigma_3$  is to be found in its relation to the spin of the electron.

which implies that the scalar and pseudoscalar parts of  $\tilde{\Psi}\Psi$  must vanish. By taking the inner product of  $\tilde{\Psi}\Psi$  with  $\gamma_0$  and comparing with (7.3), we find

$$\gamma_0 \cdot (\tilde{\Psi}\Psi) = (\gamma_0 \tilde{\Psi}\Psi)_S = (\Psi\gamma_0 \tilde{\Psi})_S = \rho \cos \beta.$$

It follows that

$$\tilde{\Psi}\Psi = \rho \cos \beta \gamma_0 (1 + \sigma_3) = \rho(\gamma_0 - \gamma_3) \cos \beta. \quad (7.4)$$

Obviously,  $\tilde{\Psi}\Psi$  is a null vector.

Now, using (7.3) and (7.4), we can evaluate  $\Psi\gamma_0 \tilde{\Psi}\Psi$  in two different ways,

$$\Psi\gamma_0 \tilde{\Psi}\Psi = (J_0 + \rho e^{i\beta})\Psi = \Psi 2\rho \cos \beta.$$

Hence

$$(J_0 - \rho e^{-i\beta})\Psi = 0. \quad (7.5)$$

Multiplying (7.5) on the left by  $(J_0 + \rho e^{i\beta})$ , we find

$$J_0^2 = \rho^2. \quad (7.6)$$

If  $J_0^2 \neq 0$  then  $\rho^2 \neq 0$  and we can write

$$J_0 = \rho v. \quad (7.7)$$

where  $v$  is a unit timelike vector. Let us follow out the implications of (8.7) and afterwards, at the end of Sec. 8, discuss the case when  $J_0^2 = 0$  but  $J_0 \neq 0$ . If  $\rho$  is identified with the proper probability of the field, then  $\rho > 0$ . It follows  $J_0 > 0$  (i.e., when  $J_0$  does not vanish, it is a positive timelike vector).<sup>12</sup> We now write (7.3) in the form

$$\Psi\gamma_0 \tilde{\Psi} = \rho(e^{i\beta} + v), \quad (7.8a)$$

where

$$\rho > 0, \quad v > 0, \quad v^2 = 1. \quad (7.8b)$$

We can also construct bivector and trivector quantities which are bilinear in  $\Psi$ . Observe that, by virtue of (2.2),  $(\Psi i\gamma_0 \tilde{\Psi})^\sim = -\Psi\gamma_0 \tilde{\Psi}$  shows that  $\Psi i\gamma_0 \tilde{\Psi}$  can have only bivector and trivector parts. By using (7.4) and (7.8) to reduce  $\Psi i\gamma_0 \tilde{\Psi}\Psi\gamma_0 \tilde{\Psi}$  in two different ways, we find that  $\Psi i\gamma_0 \tilde{\Psi}$  can be written in the following form:

$$\Psi i\gamma_0 \tilde{\Psi} = \rho i s (v e^{i\beta} + 1), \quad (7.9a)$$

where the only new quantity  $s$  is a spacelike vector orthogonal to  $v$ , i.e.,

$$s^2 = -1, \quad s \cdot v = 0. \quad (7.9b)$$

The vector  $s$  describes the orientation of the electron spin.

<sup>12</sup> See Theorem 6 of Appendix B in Ref. 2.

## 8. Structure of Dirac Spinors

We can take advantage of the fact that, because  $v$  and  $s$  are orthogonal unit vectors, there exists a field of Lorentz rotations which enables us to identify  $v$  with the  $e_0$  and  $s$  with the  $e_3$  in Eq. (3.1). We write

$$v = e_0 = R\gamma_0\tilde{R}, \quad (8.1)$$

$$s = e_3 = R\gamma_3\tilde{R}. \quad (8.2)$$

So far, since we have not specified  $e_1$  and  $e_2$ ,  $R$  is not uniquely determined; Eqs. (8.1) and (8.2) are unchanged if  $R$  is replaced by  $R' = Re^{i\sigma_2\chi}$ , where  $\chi$  is any scalar function.

Now define  $\psi$  by

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}}R \quad (8.3)$$

and write

$$\Psi = \psi U. \quad (8.4)$$

Using (7.2) and (7.5), we can write

$$\frac{1}{2}(1+v)e^{-\frac{1}{2}i\beta}\Psi\frac{1}{2}(1+\sigma_3) \quad (8.5)$$

which, when translated into a condition on  $U$ , becomes

$$U = \frac{1}{2}(1+\gamma_0)U\frac{1}{2}(1+\sigma_3). \quad (8.6)$$

By separating  $U$  into an even part  $U_E = \frac{1}{2}(U + \bar{U})$  and an odd part  $U_O = \frac{1}{2}(U - \bar{U})$ , we see that (8.6) can be written

$$U = \frac{1}{2}(1+\gamma_0)M\frac{1}{2}(1+\sigma_3), \quad (8.7)$$

where  $M = U_E + \gamma_0 U_O$  is even. By separating  $M$  into a part  $M_1 = \frac{1}{2}(M + M^*)$  which commutes with  $\gamma_0$ , and a part  $M_2 = \frac{1}{2}(M - M^*)$  which anticommutes with  $\gamma_0$ , we see that

$$\begin{aligned} U &= \frac{1}{2}(1+\gamma_0)(M_1 + M_2\sigma_3)\frac{1}{2}(1+\sigma_3) \\ &= \frac{1}{4}(M_1 + M_2\sigma_3)(1+\gamma_0)(1+\sigma_3). \end{aligned} \quad (8.8)$$

By writing  $N = (M_1 + M_2\sigma_3)$ , substituting (8.8) into (8.4), and using the condition that (7.3) must be satisfied, we find that

$$N\tilde{N} = NN^\dagger = 1 \quad (8.9)$$

which means that  $N$  is a spatial rotation. Therefore, we can absorb  $N$  in the Lorentz rotation  $R$  of (8.3) and write  $\Psi$  in the canonical form

$$\Psi = \psi\frac{1}{4}(1+\gamma_0)(1+\sigma_3). \quad (8.10)$$

It should be emphasized that (8.10) has been derived from the definition of  $\Psi$  as a spinor and independent of any particular form for the Dirac equation.

To replace the Dirac equation for  $\Psi$  by an equation for the simpler wavefunction  $\psi$ , substitute (8.10) into (7.1), separate the equation into even and odd parts, multiply the odd part on the right by  $\gamma_0$ , and add it again to the even part. The result is Eq. (5.1).

We have one piece of unfinished business. To analyze the case where  $J_0$  is a null vector we return to Eq. (7.6) and argue as in the non-null case. We find that if

$$J_0^2 = 0, \quad J_0 \neq 0 \quad (8.11)$$

then  $\Psi$  can be written in the canonical form

$$\Psi = \rho^{\frac{1}{2}} R \gamma_0 2^{-\frac{1}{2}} (1 + \sigma_3) = \rho^{\frac{1}{2}} R 2^{-\frac{1}{2}} (\gamma_0 - \gamma_3), \quad (8.12)$$

where  $R$  is a Lorentz rotation and  $\rho$  is a positive scalar. It is important to study the difference between (8.10) and (8.12). By the algebraic operations described in the last paragraph, (8.10) can be reduced to (8.3). But an analogous reduction of (8.12) is not possible. We can do no more than rewrite it in the equivalent even form

$$\psi = \rho^{\frac{1}{2}} R 2^{-\frac{1}{2}} (1 - \sigma_3). \quad (8.13)$$

This quantity has perhaps as much right to be called a spinor as the quantity (8.3). In fact it can be written as the sum of two spinors of the type (8.3)

$$\rho^{\frac{1}{2}} R 2^{-\frac{1}{2}} (1 - \sigma_3) = \left(\frac{1}{2}\rho\right)^{\frac{1}{2}} R + \left(\frac{1}{2}\rho e^{i\pi}\right)^{\frac{1}{2}} R e^{-i\sigma_2 \frac{1}{2}\pi}.$$

Therefore, if the sum of any two spinors is again a spinor, then both (8.3) and (8.13) must be spinors. Still, we have not discussed spinors of the type (8.13) because they cannot be solutions of the Dirac equation for an electron. The Dirac equation does allow (8.11) to be satisfied. Spinors of the type (8.13) are equivalent to the ‘‘two-component spinors’’ commonly used to describe neutrinos.

#### D. Prospects

To bring the program initiated here to its logical conclusion, it is necessary to carry out a general reassessment of the role of complex numbers in elementary particle theory. The fact that the  $(-1)^{\frac{1}{2}}$  has a geometrical interpretation in electron theory strongly suggests that a similar interpretation can be given to every appearance of  $(-1)^{\frac{1}{2}}$  in the basic equations of physics, though it is by no means evident that the  $(-1)^{\frac{1}{2}}$  will have the same meaning in each instance. We have learned that the  $(-1)^{\frac{1}{2}}$  in the electron wavefunction is inextricably tied up with the spin. Yet the  $(-1)^{\frac{1}{2}}$  which seems to play an essential role in the quantization of fermion fields has no evident connection with spin. It will take a careful analysis of quantum electrodynamics to resolve this apparent conflict. Again, the tie up of  $(-1)^{\frac{1}{2}}$  and spin suggests that analytic continuation of scattering amplitudes cannot be properly understood without taking spin into account.<sup>13</sup> On the other hand, there may be geometrically different kinds of analytic continuation, for there are several different geometical roots of minus one in the Dirac algebra.

---

<sup>13</sup> A formulation of the ‘‘complex’’ Lorentz group in terms of the real Dirac algebra is given in Sec. 19 of Ref. 2.

Anyone who plays the game of theoretical physics with the rules suggested here will not allow a  $(-1)^{\frac{1}{2}}$  in his theory unless it has a geometrical significance grounded in space-time. This may be regarded as another constraint imposed by space-time on permissible physical equations which had already been restricted by the requirement of relativistic covariance or invariance.

We have exhibited several bilinear “observables” of a spinor field  $\psi$ . The Dirac theory supplies a physical interpretation of  $J_0 = \psi\gamma_0\tilde{\psi}$  and of the general orientation of the frame  $\{J_i = \psi\gamma_i\tilde{\psi}\}$ . We have suggested an interpretation of  $\psi\tilde{\psi}$  which awaits final justification. To supply a more detailed interpretation of the  $J_\mu$ , we must go well beyond the Dirac theory. This possibility will be explored in another paper.