The Penrose dodecahedron revisited

Jordan E. Massad and P. K. Aravind
Physics Department, Worcester Polytechnic Institute, Worcester, Massachusetts 01609

(Received 9 June 1998; accepted 14 October 1998)

This paper gives an elementary account of the ‘‘Penrose dodecahedron,’’ a set of 40 states of a spin-$\frac{3}{2}$ particle used by Zimba and Penrose [Stud. Hist. Phil. Sci. **24**, 697–720 (1993)] to give a proof of Bell’s nonlocality theorem. The Penrose rays are constructed here from the rotation operator of a spin-$\frac{3}{2}$ particle and the geometry of a dodecahedron, and their orthogonality properties are derived and illustrated from a couple of different viewpoints. After recalling how the proof of Bell’s theorem can be reduced to a coloring problem on the Penrose rays, a ‘‘proof-tree’’ argument is used to establish the noncolorability of the Penrose rays and hence prove Bell’s theorem.

I. INTRODUCTION

Some years ago Zimba and Penrose[1] (ZP) gave an ingenious proof of Bell’s nonlocality theorem using a special set of states of a spin-$\frac{3}{2}$ particle. Because these states have a direct connection with the geometry of a dodecahedron, this argument is sometimes referred to as the ‘‘Penrose dodecahedron.’’ A popular account of the Penrose dodecahedron, presented as a puzzle to whet the appetite of the intelligent layman in the mysteries of the quantum theory, was given by Penrose in his book *Shadows of the Mind*.[2] A more technical account of the Penrose dodecahedron, including its role in proofs of Bell’s nonlocality theorem[3] and the Bell–Kochen–Specker (BKS) theorem,[4] was given in the article by ZP.[1]

The core of the Penrose dodecahedron (Fig. 1) is a constellation of 40 special states of a spin-$\frac{3}{2}$ particle. Twenty of these states, termed ‘‘explicit rays’’ by ZP, are the spin $+\frac{1}{2}$ projections of the particle along the 20 directions from the center of a regular dodecahedron to its vertices. The remaining 20 states, termed ‘‘implicit rays’’ by ZP, are chosen to complete orthogonal tetrads whose other members are triads of explicit rays. Penrose presumably arrived at his rays[5] through an impressive feat of geometrical imagination, aided by a deep understanding of the quantum theory of angular momentum. Whatever the genesis of the Penrose rays, ZP deftly manipulated a geometrical picture of quantum spins due to Majorana[6] to deduce all the properties of the rays they needed in their proofs of the Bell and BKS theorems.

The purpose of this article is to revisit the ZP paper and derive some its central conclusions in other ways. There are two motivations we have for doing this. In the first place, the Majorana picture of spin, while very elegant and also remarkably economical for the problem at hand, is still unfamiliar to the vast majority of physicists. We therefore felt it worthwhile to demonstrate how the Penrose rays could be constructed, and all their essential properties deduced, from standard angular momentum theory. Our second purpose is to present a somewhat different argument from that of ZP for the noncolorability of the Penrose rays. Our argument is presented in the form of a ‘‘proof tree’’ (see Fig. 6) that demonstrates the noncolorability of the rays directly from their tetrad table (Fig. 5). This method, while not as concise as Penrose’s vertex coloring argument,[2] has the virtue of being easy to apply and making no demands on one’s geometrical ingenuity. Further, as we have demonstrated elsewhere,[7] the proof-tree approach is useful in proving noncolorability in more complicated situations where an inspired proof of the Penrose type may prove more elusive.

The succeeding sections of this paper are organized as follows. In Sec. II we show how to construct the 40 Penrose rays from the rotation matrix of a spin-$\frac{3}{2}$ particle and the geometry of a dodecahedron. We show especially how all the orthogonalities among the rays, summarized in the tetrad table of Fig. 5, follow from this treatment. To illuminate the orthogonalities from another perspective, we obtain the components of the Penrose rays in a ‘‘natural’’ basis afforded by three explicit rays and one implicit ray. We show that in this basis the Penrose rays take on an extremely simple form (Fig. 4) from which their orthogonalities may be determined by inspection. In Sec. III we show how the proof of the two Bell theorems can be reduced to the purely mathematical problem of demonstrating the impossibility of coloring the Penrose rays according to a simple set of rules. In Sec. IV we use our proof-tree approach to demonstrate the noncolorability of the 40 Penrose rays, as well as a critical subset of 28, and thus complete the proofs of the Bell theorems. The material in Secs. II and IV is our ‘‘poor man’s version’’ of the ZP paper. The argument in Sec. III is based freely on the ideas of ZP and others and is included to make the paper self-contained, since in its absence the reader may be left wondering why one bothers to undertake the rest of the exercise.

It is appropriate to say a few words about the broader context in which the present problem arose. Beginning in the late 80s, several new proofs of the Bell and BKS theorems were given that were either more compelling than earlier proofs or technically simpler. A landmark paper in this regard was the work by Greenberger, Horne and Zeilinger,[8] in which a proof of Bell’s nonlocality theorem was given without inequalities. Soon after this, Peres[9] gave some ingenious noncoloring proofs of the BKS theorem that were much simpler than any of the earlier proofs. A later review article by Mermin[10] presented some spectacular proofs of the BKS theorem based on simple arithmetic contradictions (rather than noncoloring arguments) and also revealed a close kinship between the Bell and BKS theorems. The Penrose dodecahedron seems to have been stimulated in part by the work of Peres,[9] or so one gathers from an amusing incident recounted in Ref. 11. This brief history has only touched on some highlights and makes no claim to completeness. A more authoritative survey may be found, for example, in the book by Peres.[12]

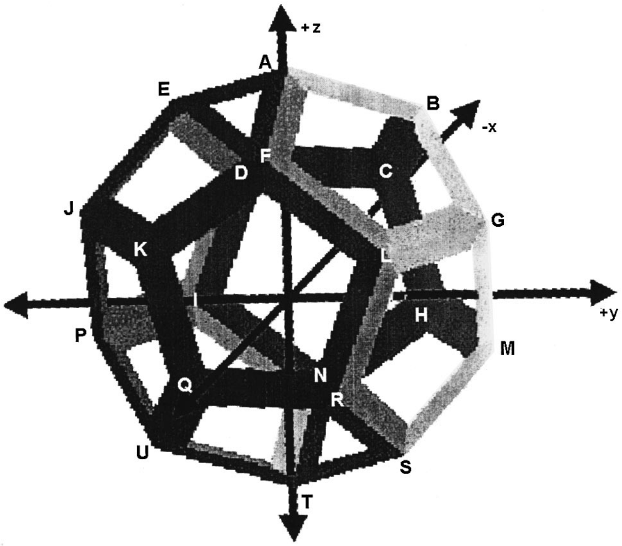


Fig. 1. The Penrose dodecahedron. Shown is a regular dodecahedron with its vertices labelled A through U (with the letter O omitted). The 40 Penrose rays of a spin-$\frac{3}{2}$ particle are associated with the vertices of this dodecahedron as follows. The 20 explicit rays are the eigenstates $|\frac{3}{2}, +\frac{1}{2}\rangle$ whose quantization axes are the 20 directions from the center of the dodecahedron to its vertices; these rays are labelled $|A\rangle, |B\rangle, \ldots$ after the corresponding vertices. The implicit rays are also associated with the vertices and are labelled $|A'\rangle, |B'\rangle, \ldots$. The implicit ray associated with any vertex is the unique ray that is orthogonal to the explicit rays associated with the three neighboring vertices.

II. THE PENROSE RAYS AND THEIR ORTHOGONALITIES

Let $|j,m\rangle_{\hat{\mathbf{n}}}$ denote a spin eigenstate of a spin-j particle referred to the quantization axis $\hat{\mathbf{n}}$ and let $|j,m\rangle_{\hat{\mathbf{n}}'}$ be a similar eigenstate referred to the axis $\hat{\mathbf{n}}$. These eigenstates are related to each other by the rotation operator $\exp(-i\mathbf{J}\cdot\hat{\mathbf{k}}\theta) \equiv R(\hat{\mathbf{k}},\theta)$ according to the equation

$$|j,m\rangle_{\hat{\mathbf{n}}'} = \exp(-i\mathbf{J}\cdot\hat{\mathbf{k}}\theta)|j,m\rangle_{\hat{\mathbf{n}}}. \tag{1}$$

The rotation operator contains the spin operator $\mathbf{J}$ of the particle (in units of $\hbar$) and the axis, $\hat{\mathbf{k}}$, and angle, θ, of any rotation that will take the direction $\hat{\mathbf{n}}$ into the direction $\hat{\mathbf{n}}'$ (see Fig. 2). The evaluation of the rotation operator $\exp(-i\mathbf{J}\cdot\hat{\mathbf{k}}\theta)$, or the equivalent rotation matrix $\langle j,m|\exp(-i\mathbf{J}\cdot\mathbf{k}\theta)|j,m'\rangle$, is discussed in many texts on quantum mechanics.[13] We obtain the rotation operator for a spin-$\frac{3}{2}$ particle in a form that is particularly useful in the calculations below. We begin by observing that the rotation operator for a spin-$\frac{3}{2}$ particle can be written in the form

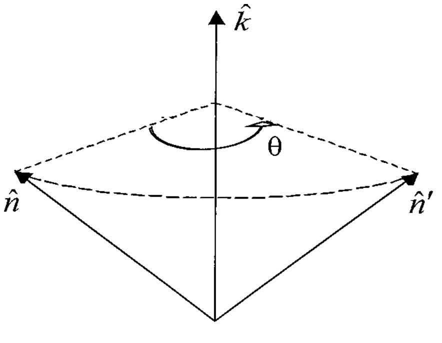


Fig. 2. Illustrating the axis, $\hat{\mathbf{k}}$, and angle, θ, of the rotation operator $\exp(-i\mathbf{J}\cdot\hat{\mathbf{k}}\theta)$ that transforms the eigenstate $|j,m\rangle_{\hat{\mathbf{n}}}$ into the eigenstate $|j,m\rangle_{\hat{\mathbf{n}}'}$ according to Eq. (1) of the text.

$$\exp(-i\mathbf{J}\cdot\hat{\mathbf{k}}\theta) = c_1 I + c_2 \mathbf{J}\cdot\hat{\mathbf{k}} + c_3(\mathbf{J}\cdot\hat{\mathbf{k}})^2 + c_4(\mathbf{J}\cdot\hat{\mathbf{k}})^3, \tag{2}$$

where I is the identity operator and $c_1,\ldots,c_4$ are suitable constants. This form follows if one expands the exponential operator in an infinite power series and notes, from the identity $(\mathbf{J}\cdot\hat{\mathbf{k}}-\frac{3}{2})(\mathbf{J}\cdot\hat{\mathbf{k}}-\frac{1}{2})(\mathbf{J}\cdot\hat{\mathbf{k}}+\frac{1}{2})(\mathbf{J}\cdot\hat{\mathbf{k}}+\frac{3}{2})=0$, that all powers of $\mathbf{J}\cdot\hat{\mathbf{k}}$ higher than the third can be expressed in terms of the third and lower powers. The constants $c_1,\ldots,c_4$ can be determined by noting that (2) continues to be valid if $\mathbf{J}\cdot\hat{\mathbf{k}}$ is replaced by any of its eigenvalues. On successively replacing $\mathbf{J}\cdot\hat{\mathbf{k}}$ by $\frac{3}{2}$, $\frac{1}{2}$, $-\frac{1}{2}$, and $-\frac{3}{2}$ in (2) and solving the resulting simultaneous equations, one finds that

$$\begin{aligned} c_1 &= \tfrac{1}{8}[9\cos(\tfrac{1}{2}\theta) - \cos(\tfrac{3}{2}\theta)], \\ c_2 &= -\tfrac{9}{4}i\sin(\tfrac{1}{2}\theta) + \tfrac{1}{12}i\sin(\tfrac{3}{2}\theta), \\ c_3 &= \tfrac{1}{2}[\cos(\tfrac{3}{2}\theta) - \cos(\tfrac{1}{2}\theta)], \quad c_4 = i\sin(\tfrac{1}{2}\theta) - \tfrac{1}{3}i\sin(\tfrac{3}{2}\theta). \end{aligned} \tag{3}$$

Equations (2) and (3) are the desired form of the spin-$\frac{3}{2}$ rotation operator.

We turn now to the Penrose rays of a spin-$\frac{3}{2}$ particle. The 20 explicit rays are the eigenstates $|\frac{3}{2}, +\frac{1}{2}\rangle$ whose quantization axes lie along the 20 directions from the center of a dodecahedron to its vertices. All the explicit rays may be obtained from any one by the application of suitable rotation operators. To illustrate this in detail, we introduce the two special rotation operators $R_3 \equiv R(\hat{\mathbf{k}}_0, 2\pi/3)$, representing a counterclockwise rotation by $2\pi/3$ about the axis through the vertex A, and $R_2 \equiv R(\hat{\mathbf{k}}_1, \pi)$, representing a half-turn about the axis through the midpoint of the edge AF ($\hat{\mathbf{k}}_0$ and $\hat{\mathbf{k}}_1$ are the unit vectors along these rotation axes). Let $|A\rangle$ denote the explicit ray corresponding to the vertex A. Then 9 of the other explicit rays may be obtained by the action of R_2 and R_3 on $|A\rangle$ as follows:

$$\begin{aligned} &|F\rangle = R_2|A\rangle, \quad |B\rangle = R_3R_2|A\rangle, \quad |E\rangle = R_3^2R_2|A\rangle, \\ &|L\rangle = R_2R_3^2R_2|A\rangle, \quad |C\rangle = R_3R_2R_3^2R_2|A\rangle, \\ &|J\rangle = (R_3^2R_2)^2|A\rangle, \quad |G\rangle = (R_2R_3^2)^2R_2|A\rangle, \\ &|D\rangle = R_3(R_2R_3^2)^2R_2|A\rangle, \quad |K\rangle = R_3^2(R_2R_3^2)^2R_2|A\rangle. \end{aligned} \tag{4}$$

The product of rotations applied to $|A\rangle$ in each case is just that required to send the vertex A into the vertex of the desired explicit ray, while maintaining the overall invariance of the dodecahedron. Replacing the ‘‘seed’’ ray $|A\rangle$ by $|T\rangle$ in (4) yields the remaining 9 explicit rays; for example, the relation $|B\rangle = R_3R_2|A\rangle$ goes over into the relation $|U\rangle = R_3R_2|T\rangle$ between the corresponding antipodal rays.

We now define the implicit ray corresponding to any vertex as the unique ray that is orthogonal to the explicit rays of the three neighboring vertices (this is always possible, since the explicit rays concerned clearly span a three-dimensional Hilbert space). The implicit ray corresponding to vertex A, which we denote $|A'\rangle$, can thus be obtained as the ray orthogonal to the rays $|F\rangle$, $|B\rangle$, and $|E\rangle$. Similarly the implicit

$$|A\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad |F\rangle = \begin{bmatrix} \frac{\tau}{3} q^{3} \\ \frac{\sqrt{3}}{3} q^{-3} \\ \frac{\sqrt{3}}{3} q^{-3} \\ \frac{1}{3\tau} q^{-3} \end{bmatrix} \qquad |B\rangle = \begin{bmatrix} \frac{\tau}{3} q^{-3} \\ \frac{\sqrt{3}}{3} q^{-5} \\ \frac{\sqrt{3}}{3} q^{-1} \\ \frac{1}{3\tau} q^{3} \end{bmatrix} \qquad |E\rangle = \begin{bmatrix} \frac{\tau}{3} q^{3} \\ \frac{\sqrt{3}}{3} q^{5} \\ \frac{\sqrt{3}}{3} q \\ \frac{1}{3\tau} q^{-3} \end{bmatrix}$$

$$|A'\rangle = \begin{bmatrix} -\frac{\sqrt{3}}{3\tau} \\ 0 \\ 0 \\ -\frac{\sqrt{3}}{3}\tau \end{bmatrix} \qquad |T\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad |T'\rangle = \begin{bmatrix} -\frac{\sqrt{3}}{3}\tau \\ 0 \\ 0 \\ \frac{\sqrt{3}}{3\tau} \end{bmatrix} \qquad |Q'\rangle = \begin{bmatrix} \frac{\tau}{3} q^{5} \\ \frac{\sqrt{3}}{3} q^{-1} \\ \frac{\sqrt{3}}{3} q^{3} \\ \frac{1}{3\tau} q^{-1} \end{bmatrix}$$

$$\tau = \frac{1+\sqrt{5}}{2} \qquad q = e^{i\frac{\pi}{6}}$$

Fig. 3. Expressions for some of the explicit and implicit rays.

ray $|T'\rangle$ can be obtained as the ray orthogonal to the rays $|U\rangle$, $|N\rangle$, and $|S\rangle$. The relations in (4), which hold with explicit rays replaced by their implicit counterparts throughout, can then be used to calculate all the remaining implicit rays. We prefer to calculate the implicit rays through rotations rather than orthogonalizations because the latter operation is generally difficult to carry out. The orthogonalizations needed for the ''seed'' rays $|A'\rangle$ and $|T'\rangle$ are, however, very simple.

To obtain numerical values for the components of the rays, we must choose a coordinate system and also a suitable basis for the rays. We choose a right-handed coordinate system with the z axis passing through the vertex A of the dodecahedron and the $x-z$ plane chosen so that the edge AF lies in it (see Fig. 1). We also choose the standard angular momentum basis[13] in which the explicit rays A and T are represented by the column vectors (0,1,0,0) and (0,0,1,0), respectively, and the spin operators J_x, J_y, and J_z by the usual 4×4 matrices. The unit vector $\hat{\mathbf{k}}_0$ occurring in R_3 lies along the z axis and the unit vector $\hat{\mathbf{k}}_1$ occurring in R_2 lies in the $x-z$ plane and at an angle of $\arccos(\tau/\sqrt{3})$ to the z axis, where $\tau = (1+\sqrt{5})/2$ is the golden mean. With these conventions, Eqs. (2)–(4) may be used in the manner indicated above to obtain numerical components for all the 40 Penrose rays.

We do not actually display all the Penrose rays, but only as many as are required to illustrate the 6 orthogonality rules of ZP. Figure 3 shows the components of some selected explicit and implicit rays. These results were obtained with the help of the computer algebra program MAPLE, which allowed us to avoid pages of tedious algebra by means of a few well chosen keystrokes. We now discuss how the rays in Fig. 3 illustrate the 6 orthogonality rules of ZP. These rules, in the numbering of ZP, are the following.

Rule I. Explicit rays corresponding to next-to-adjacent vertices of the dodecahedron are orthogonal. This rule is illustrated by the orthogonalities $\langle F|B\rangle = 0$ and $\langle B|E\rangle = 0$. This rule can actually be deduced directly from the form (2)–(3) of the spin-$\frac{3}{2}$ rotation operator. Consider two eigenstates $|\frac{3}{2}, +\frac{1}{2}\rangle$, the first with its axis along z and the second with its axis in the $x-z$ plane and at an angle θ to z. These two states are connected by the rotation operator $R(\hat{\mathbf{y}}, \theta)$ that performs a rotation through the angle θ about the y axis. Using the standard basis and setting the overlap of these states equal to 0 leads to the condition $R_{22}(\hat{\mathbf{y}}, \theta) = 0$ on the diagonal element of the rotation matrix, which is satisfied only if θ is equal to π or $\arccos(1/3)$. The latter angle is just that subtended by two next-to-adjacent vertices of a dodecahedron at its center.

Rule II. Explicit rays corresponding to antipodal vertices of the dodecahedron are orthogonal. The same is also true of implicit rays. The first part of this rule is illustrated by the (trivial) orthogonality $\langle T|A\rangle = 0$. The second part is illustrated by the less obvious result $\langle T'|A'\rangle = 0$.

Rule III. The explicit and implicit rays corresponding to adjacent vertices of the dodecahedron are orthogonal. This rule, illustrated by $\langle A'|F\rangle = 0$, is actually a direct consequence of the definition of the implicit rays.

Rule IV. The explicit and implicit rays corresponding to the same vertex of the dodecahedron are orthogonal. This is illustrated by the orthogonality $\langle A'|A\rangle = 0$.

Rule V. The explicit and implicit rays corresponding to antipodal vertices of the dodecahedron are orthogonal. This is illustrated by the orthogonality $\langle A'|T\rangle = 0$.

Rule VI. The implicit rays corresponding to next-to-next-to adjacent vertices of the dodecahedron are orthogonal. This is illustrated by the orthogonality $\langle A'|Q'\rangle = 0$. As remarked by ZP, this rule can be deduced from the earlier ones and so a separate proof of it is not really necessary.

We have established the ZP rules above by looking at particular rays. However it is clear, from the symmetry of the dodecahedron, that these rules hold for all similarly related pairs of rays. A systematic application of these rules shows that the Penrose rays form the 40 sets of orthogonal tetrads shown in Fig. 5. This set of tetrads is ''complete'' in the sense that every orthogonality between rays is represented in it. This may be confirmed by noting that each ray (explicit or implicit) is orthogonal to exactly 12 other rays and, further, that each ray occurs in exactly 4 tetrads whose remaining members are all the rays it is orthogonal to. The completeness of the tetrad table implies that it contains all the necessary information (and in fact the only relevant information) for the noncoloring arguments of Sec. IV.

An alternative view of the ray orthogonalities may be obtained by switching to the ''natural basis'' provided by the rays $|F\rangle$, $|B\rangle$, $|E\rangle$, and $|A'\rangle$. In other words, we represent $|F\rangle$, $|B\rangle$, $|E\rangle$, and $|A'\rangle$ by the column vectors (1,0,0,0), (0,1,0,0), (0,0,1,0), and (0,0,0,1), respectively, and work out the components of the remaining rays in this basis. If we recognize that the normalization and overall phase of the rays is unimportant, we find that the rays take on the simplified form shown in Fig. 4: The nonvanishing components of the rays are now pure phases of the form $e^{i\theta}$ where θ is some integer multiple of 60°. The orthogonalities between the rays are now very easy to pick out.

We end this section with a comment. Having constructed the Penrose rays and deduced their properties using conventional angular momentum theory, we must admit that the Majorana method, as exploited by ZP, is definitely superior. Our approach, while elementary, seems to use too much scaffolding to get at essentially geometric properties of the rays that can be made to reveal themselves very clearly through the Majorana approach. An even bigger virtue of the

20 Explicit Rays in the FBEA' Basis

$$|A\rangle=\begin{bmatrix}1\\p\\p^2\\0\end{bmatrix}\quad |F\rangle=\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\quad |B\rangle=\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\quad |E\rangle=\begin{bmatrix}0\\0\\1\\0\end{bmatrix}\quad |L\rangle=\begin{bmatrix}-1\\0\\p^2\\1\end{bmatrix}\quad |G\rangle=\begin{bmatrix}0\\-1\\p\\1\end{bmatrix}\quad |C\rangle=\begin{bmatrix}p^2\\1\\0\\1\end{bmatrix}$$

$$|D\rangle=\begin{bmatrix}p\\0\\1\\1\end{bmatrix}\quad |J\rangle=\begin{bmatrix}0\\p^2\\1\\-1\end{bmatrix}\quad |K\rangle=\begin{bmatrix}1\\p^{-2}\\0\\1\end{bmatrix}\quad |R\rangle=\begin{bmatrix}0\\p\\-1\\1\end{bmatrix}\quad |M\rangle=\begin{bmatrix}p^{-1}\\0\\1\\1\end{bmatrix}\quad |H\rangle=\begin{bmatrix}1\\0\\p\\-1\end{bmatrix}$$

$$|I\rangle=\begin{bmatrix}1\\p^2\\0\\1\end{bmatrix}\quad |P\rangle=\begin{bmatrix}p^{-2}\\1\\0\\1\end{bmatrix}\quad |Q\rangle=\begin{bmatrix}0\\1\\p^2\\-1\end{bmatrix}\quad |S\rangle=\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\quad |N\rangle=\begin{bmatrix}0\\1\\1\\-1\end{bmatrix}\quad |U\rangle=\begin{bmatrix}-1\\0\\1\\1\end{bmatrix}\quad |T\rangle=\begin{bmatrix}p^2\\p\\1\\0\end{bmatrix}$$

(a)

$$\boxed{p=e^{i\frac{\pi}{3}}}$$

20 Implicit Rays in the FBEA' Basis

$$|A'\rangle=\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\quad |F'\rangle=\begin{bmatrix}0\\p^2\\1\\p\end{bmatrix}\quad |B'\rangle=\begin{bmatrix}p\\0\\1\\p^2\end{bmatrix}\quad |E'\rangle=\begin{bmatrix}p^{-2}\\p^2\\0\\1\end{bmatrix}\quad |L'\rangle=\begin{bmatrix}0\\1\\1\\p\end{bmatrix}\quad |G'\rangle=\begin{bmatrix}1\\0\\-1\\p^{-1}\end{bmatrix}\quad |C'\rangle=\begin{bmatrix}1\\0\\-1\\p\end{bmatrix}$$

$$|D'\rangle=\begin{bmatrix}1\\1\\0\\p^2\end{bmatrix}\quad |J'\rangle=\begin{bmatrix}1\\1\\0\\p^{-2}\end{bmatrix}\quad |K'\rangle=\begin{bmatrix}0\\1\\1\\p^{-1}\end{bmatrix}\quad |R'\rangle=\begin{bmatrix}-1\\1\\p^{-1}\\0\end{bmatrix}\quad |M'\rangle=\begin{bmatrix}-1\\1\\p\\0\end{bmatrix}\quad |H'\rangle=\begin{bmatrix}p^2\\-1\\1\\0\end{bmatrix}$$

$$|I'\rangle=\begin{bmatrix}p\\1\\-1\\0\end{bmatrix}\quad |P'\rangle=\begin{bmatrix}1\\p^{-1}\\1\\0\end{bmatrix}\quad |Q'\rangle=\begin{bmatrix}1\\p\\1\\0\end{bmatrix}\quad |S'\rangle=\begin{bmatrix}1\\p^2\\0\\p^{-2}\end{bmatrix}\quad |N'\rangle=\begin{bmatrix}0\\1\\p^2\\p\end{bmatrix}\quad |U'\rangle=\begin{bmatrix}p\\0\\p^2\\1\end{bmatrix}\quad |T'\rangle=\begin{bmatrix}1\\-1\\1\\0\end{bmatrix}$$

(b)

$$\boxed{p=e^{i\frac{\pi}{3}}}$$

Fig. 4. Penrose rays in the ‘‘natural basis’’ provided by the rays $|F\rangle$, $|B\rangle$, $|E\rangle$, and $|A'\rangle$: (a) explicit rays, and (b) implicit rays.

Majorana approach is that it seems to have hinted at the existence of the Penrose dodecahedron in the first place.

III. REDUCTION OF BELL’S THEOREM TO A COLORING ARGUMENT

The object of Bell’s nonlocality theorem is to rule out the existence of (local) hidden variables theories that make definite predictions in situations where quantum mechanics predicts only probabilities.[14] We assume, for the purposes of this discussion, that the reader has some knowledge of the nature and scope of hidden variables theories. A clear discussion of such theories may be found in the article by Greenberger *et al.*[8] that appeared earlier in this journal. Other good references are the review article by Mermin[10] and the book by Peres.[12] We confine ourselves here to demonstrating that a hidden variables model (whether noncontextual or local) of a spin-$\frac{3}{2}$ particle must be able to color each of the Penrose rays either red or green in such a way that there is exactly one green ray (and three red rays) in each of the 40 tetrads formed by the rays; the impossibility of such a coloring (demonstrated in Sec. IV) then constitutes a proof of the two Bell theorems.

We proceed to the desired result in three steps. First, we analyze some aspects of the measurement process according to orthodox quantum mechanics; specifically, we introduce the notion of compatible projection measurements and derive an important constraint on their outcomes. Next, we outline

Natural Tetrads		Antipodal Tetrads	Tetrahedral Tetrads
F B E A'	L Q S R'	A A' T T'	A' I' Q' M'
A L K F'	G H S M'	B B' U U'	A' P' R' H'
A G C B'	C M N H'	C C' Q Q'	B' K' I' S'
A D J E'	D P N I'	D D' R R'	B' J' N' R'
F G R L'	J I U P'	E E' S S'	C' F' P' S'
B L M G'	K R U Q'	F F' N N'	C' L' J' T'
B D H C'	R M T S'	G G' P P'	D' K' G' T'
E C I D'	H I T N'	H H' K K'	D' F' M' U'
E K P J'	P Q T U'	I I' L L'	E' L' H' U'
F J Q K'	S N U T'	J J' M M'	E' G' N' Q'

Fig. 5. The 40 tetrads of mutually orthogonal rays formed by the Penrose rays. Explicit and implicit rays are denoted by unprimed and primed letters, respectively. The 20 ''natural tetrads'' each consist of an implicit ray associated with a vertex and the explicit rays associated with the three neighboring vertices. The 10 ''antipodal tetrads'' consist of the explicit and implicit rays associated with antipodal vertices. And the 10 tetrahedral tetrads consist of implicit rays associated with the 10 tetrahedra that can be inscribed in a dodecahedron (these tetrahedra can be obtained by first inscribing five cubes in a dodecahedron and then picking out alternate vertices in each cube to obtain a pair of mutually reciprocal tetrahedra).

the contrasting attitudes of quantum mechanics (QM) and hidden variables theories (HVTs) to the results of the projection measurements, and particularly to one's ability to predict these results in advance. And finally we show how the position of the HVTs, in the case of a spin-$\frac{3}{2}$ particle, requires them to color the Penrose rays in the manner stated above. Our argument is patterned closely on that of ZP, but differs in some minor details.

We first take up the subject of measurement. Consider an N-state quantum system (such as a spin-j particle, for which $N=2j+1$). A simple measurement that one can carry out on this system is a so-called ''projection measurement.'' A projection measurement amounts to asking the system the question: are you in the state $|\Psi\rangle$? (where $|\Psi\rangle$ is some arbitrary normalized state). The system answers this question with either a yes or a no. From the viewpoint of standard QM, the projection measurement amounts to measuring the projection operator (or ''projector'') $P\equiv|\Psi\rangle\langle\Psi|$ on the system. The projector is Hermitian ($P^{\dagger}=P$) and idempotent ($P^2=P$), so its only eigenvalues are 0 and 1. If the eigenvalue 1 is returned upon measurement the system has answered ''yes'' to our question, and if 0 is returned it is has answered ''no.''

We next consider multiple measurements. According to QM, if the operators corresponding to a set of observables commute, then it is possible to measure these observables in the same context. What this means is that if we are given a single copy of the system in an arbitrary state, then we can conceive of an experimental arrangement that would allow us to determine the eigenvalues of all the (commuting) observables in that state. Now the projectors associated with two or more mutually orthogonal states commute with each other, hence, according to QM, they represent a set of yes–no questions that can be asked in the same context. It turns out that there are some important constraints on the answers the system can give to such questions, which we now proceed to derive.

Consider the projectors associated with N mutually orthogonal states. The sum of these (commuting) projectors is the identity operator, from the completeness relation. The last statement implies that if the N yes–no questions corresponding to these projectors are posed in the same context, exactly one will receive the answer yes and the remaining $N-1$ will receive the answer no. This constraint is a direct consequence of Mermin's principle[15] which states that if a set of commuting observables obeys a certain identity, their measured eigenvalues in an arbitrary state also obey that identity; thus the identity $P_1+P_2+\cdots+P_N=I$ obeyed by the projectors translates into the constraint $\lambda_1+\lambda_2+\cdots+\lambda_N=1$ satisfied by their eigenvalues, from which the desired result follows.

The above constraint is usually stated in the form of two separate constraints: (i) If the projectors corresponding to two orthogonal states are measured in the same context, they will not both be found to have the eigenvalue 1. (ii) If the projectors corresponding to N mutually orthogonal states are measured in the same context, they will not all be found to have the eigenvalue 0. This slightly more flexible formulation covers the case in which less than N commuting projectors are measured in the same context. Although the constraints (i) and (ii) have been derived here from the formalism of orthodox QM, it should be added that they are confirmed by all known experiments.

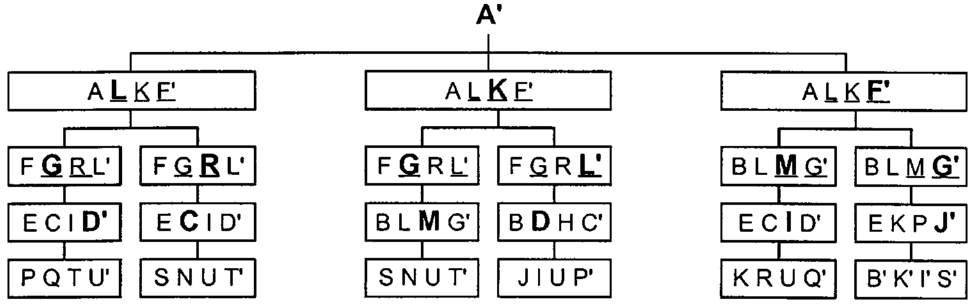


Fig. 6. Noncolorability of Penrose's 40 rays. One begins by assuming that a coloring exists and shows that this leads to a contradiction. The argument is presented in the form of a ''proof tree'' that proceeds from the root, at the top, down through the various branches to the contradictions at the bottom. Rays colored green are shown in boldface and those colored red (all the rest) in ordinary type. Any ray that is underlined has its color assigned to it as a matter of choice while any ray that is not underlined has its color forced. A forced red ray always results from a (forced or unforced) green ray in an earlier step, while a forced green ray is always forced by its three red companions in the same step. The (attempted) coloring begins by assigning the color green to the ray A'. There is no arbitrariness involved in this step since the existence of the tetrahedral tetrads (Fig. 5) implies that at least one of the implicit rays must be green and, from the dodecahedral symmetry of the problem, we can always arrange for this ray to be A'. The argument then proceeds down the different branches of the proof tree, corresponding to the various (mutually exclusive and exhaustive) choices of green rays at the second and third steps. All six branches of the proof tree eventually lead to completely red tetrads, which establishes the noncolorability of the Penrose rays.

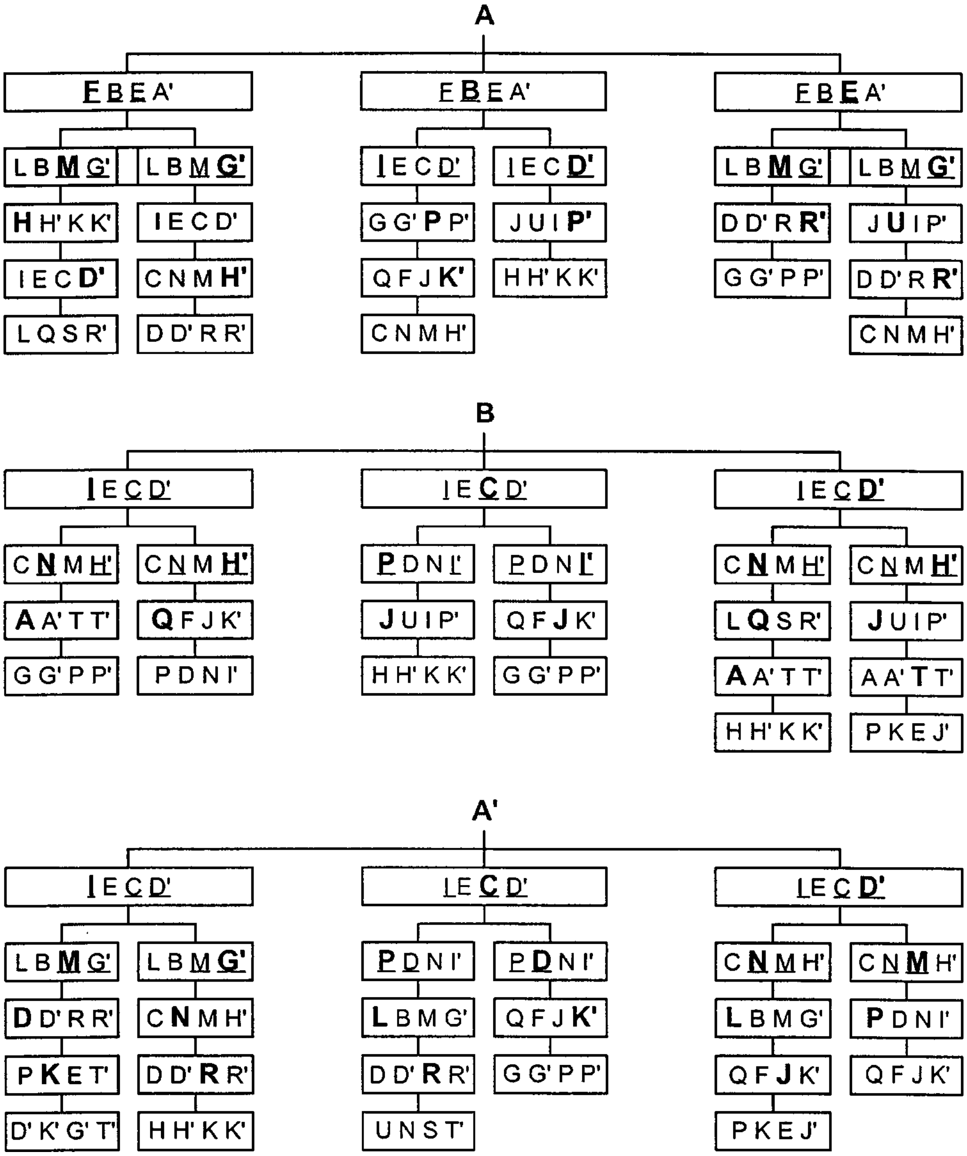


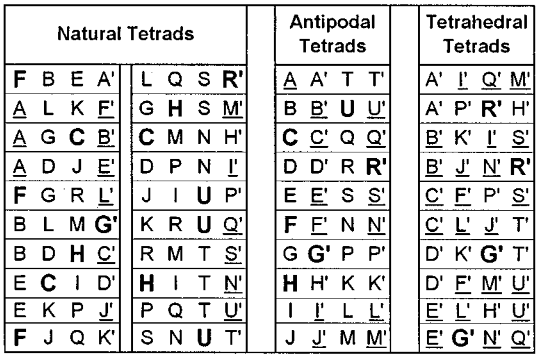
Fig. 7. Noncolorability of the 28 Penrose rays A–U, A', D', G', H', K', P', R', and T'. This set consists of all the explicit rays and the 8 implicit rays corresponding to one of the cubes that can be inscribed in the dodecahedron. The demonstration again proceeds by assuming that a coloring exists and showing that this leads to a contradiction. The initial ray to be colored green can be either (1) an explicit ray whose implicit partner belongs to the set (e.g., A), or (2) an explicit ray whose implicit partner does not belong to the set (e.g., B), or (3) an implicit ray in the set (e.g., A'). From the overall (cubic) symmetry of this configuration it does not matter which ray we pick in each set, so we choose A, B, and A'. The impossibility of a viable coloring resulting from each of these choices is demonstrated above. The notational conventions are identical to those of Fig. 6.

This concludes our brief review of measurement theory. We now contrast the abilities of QM and HVTs to predict the results of the projection measurements in advance, given as complete a description of the state of the system as each deems possible. QM maintains that, for an arbitrary state of the system, it is generally impossible to predict the answers to the yes–no questions in advance. In fact, it goes even further and says that the answers do not exist before the questions are asked and only come into being when the questions are actually asked (i.e., the projection measurements made). The HVTs claim, on the contrary, that the answers given by the system reflect objective properties of its physical state that exist prior to any act of measurement and that merely reveal themselves upon measurement. The HVTs further charge that the inability of QM to give definite answers to the questions stems from the incomplete description of a physical state provided by the wave function.

How can one put these daring (or merely reasonable?) claims of the HVTs to the test? Here is one way, suggested by the previous discussion. We present a HVT with a list of yes–no questions (projectors) pertaining to a quantum system. The HVT is then obligated to give an unambiguous answer to each question (because it represents an objective property of the state under consideration) and to do so in conformity with the requirements (i) and (ii) stated above (since it would otherwise be in conflict with both QM and experiment). If the HVT fails to pass this test, it is untenable and must be rejected.

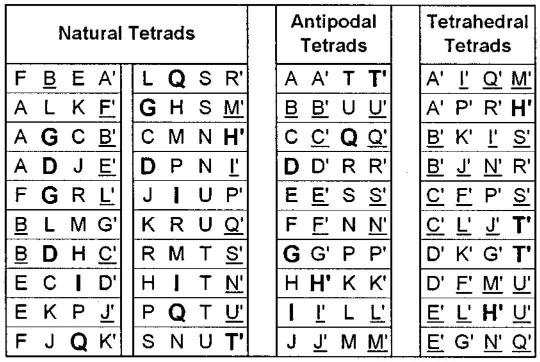
Let us now specialize the discussion to the case of a spin-$\frac{3}{2}$ particle. Suppose we choose the questions to be projectors based on the Penrose rays and present a HVT with the 40 questions based on the 40 Penrose rays. Suppose further that a yes answer to a question is taken to imply that the corresponding ray is colored green and a no answer that it is

(1) Ray A Deleted



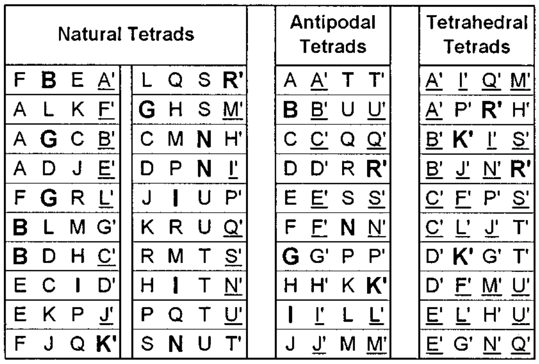
Natural Tetrads		Antipodal Tetrads	Tetrahedral Tetrads
F B E A'	L Q S **R'**	A A' T T'	A' I' Q' M'
A L K F'	G **H** S M'	B B' **U** U'	A' P' **R'** H'
A G **C** B'	**C** M N H'	**C** C' Q Q'	B' K' I' S'
A D J E'	D P N I'	D D' R **R'**	B' J' N' **R'**
F G R L'	J I **U** P'	E E' S S'	C' F' P' S'
B L M **G'**	K R **U** Q'	**F** F' N N'	C' L' J' T'
B D **H** C'	R M T S'	G **G'** P P'	D' K' **G'** T'
E **C** I D'	**H** I T N'	**H** H' K K'	D' F' M' U'
E K P J'	P Q T U'	I I' L L'	E' L' H' U'
F J Q K'	S N **U** T'	J J' M M'	E' **G'** N' Q'

(2) Ray B Deleted



Natural Tetrads		Antipodal Tetrads	Tetrahedral Tetrads
F B E A'	L **Q** S R'	A A' T **T'**	A' I' Q' M'
A L K F'	**G** H S M'	B B' U U'	A' P' R' **H'**
A **G** C B'	C M N **H'**	C C' **Q** Q'	B' K' I' S'
A **D** J E'	**D** P N I'	**D** D' R R'	B' J' N' R'
F **G** R L'	J **I** U P'	E E' S S'	C' F' P' S'
B L M G'	K R U Q'	F F' N N'	C' L' J' **T'**
B **D** H C'	R M T S'	**G** G' P P'	D' K' G' **T'**
E C **I** D'	H **I** T N'	H **H'** K K'	D' F' M' U'
E K P J'	P **Q** T U'	**I** I' L L'	E' L' **H'** U'
F J **Q** K'	S N U **T'**	J J' M M'	E' G' N' Q'

(3) Ray A' Deleted



Natural Tetrads		Antipodal Tetrads	Tetrahedral Tetrads
F **B** E A'	L Q S **R'**	A A' T T'	A' I' Q' M'
A L K F'	**G** H S M'	**B** B' U U'	A' P' **R'** H'
A **G** C B'	C M **N** H'	C C' Q Q'	B' **K'** I' S'
A D J E'	D P **N** I'	D D' R **R'**	B' J' N' **R'**
F **G** R L'	J **I** U P'	E E' S S'	C' F' P' S'
B L M G'	K R U Q'	F F' **N** N'	C' L' J' T'
B D H C'	R M T S'	**G** G' P P'	D' **K'** G' T'
E C **I** D'	H **I** T N'	H H' K **K'**	D' F' M' U'
E K P J'	P Q T U'	**I** I' L L'	E' L' H' U'
F J Q **K'**	S **N** U T'	J J' M M'	E' G' N' Q'

Fig. 8. Criticality of ZP's 28 rays. Deletion of even a single ray from the set of 28 considered in Fig. 7 leads to a colorable set, as demonstrated here. The deleted ray can be (1) an explicit ray whose implicit partner belongs to the set (e.g., A), or (2) an explicit ray whose implicit partner does not belong to the set (e.g., B), or (3) an implicit ray in the set (e.g., A'). From symmetry it suffices to consider just a single example from each of these three classes, and we choose A, B, and A'. The above tables show viable colorings for the remaining 27 rays in each of these cases. Note the three-part code for rays: green rays are indicated in boldface, red rays in ordinary type, and uncolored rays in ordinary type with underscoring. Tetrads containing uncolored rays are shown to confirm (via the absence of more than one green ray in them) that this is a viable coloring.

colored red. Then we arrive at just the requirement stated at the beginning of this section: A viable HVT must be able to color the Penrose rays red or green in such a way that there is exactly one green (and three reds) in each of the 40 tetrads. The requirements (i) and (ii) can be collapsed into a single requirement in this case because there are no orthogonalities between the rays beyond those already contained in the tetrads.

There is, however, an objection to the above argument, first pointed out by Bell.[4] Let us again specialize to a spin-$\frac{3}{2}$ particle and consider the two sets of questions $ALKF'$ and $AGCB'$ that have the question A in common (we take the questions to be synonymous with Penrose rays and the question sets with tetrads of Penrose rays). We have required that a HVT give the same answer to question A whether it is asked as a part of $ALKF'$ or $AGCB'$. But this is an unreasonable requirement because the two question sets can never be posed in the same context, and the sought for consistency can never be tested in practice.[16] HVTs that give the same answer to a question (projector) no matter which question set it is asked as a part of are known as noncontextual HVTs, and the task of demolishing such HVTs is the province of the Bell–Kochen–Specker theorem. Bell himself downplayed the importance of this theorem,[4] but Mermin has recently given a powerful argument[17] in support of noncontextuality that underscores the continuing interest of this theorem.

We now discuss how the above argument can be rescued from Bell's objection, and that brings us to Bell's nonlocality theorem. The strategem suggested by ZP involves a pair of spin-$\frac{3}{2}$ particles in a singlet state that fly off in opposite directions (left and right) from a common source. Each particle is intercepted by an observer who can ask any of the 40 questions (based on the Penrose rays) that we have been discussing. Moreover the two observers have their dodecahedra perfectly aligned and similarly labelled, so that all the axes of one are parallel to the corresponding axes of the other. Now, the singlet state of a pair of spin-$\frac{3}{2}$ particles has the following remarkable anticorrelation property: If the observers ask questions corresponding to antipodal Penrose rays (with both rays being explicit or implicit), they always receive identical answers. We proceed to show how this anticorrelation property can be used to justify the noncontextualist labelling of the Penrose rays associated with either particle.

We begin by supposing that there exists a coloring of the rays on the left that could be context dependent (i.e., the color of a ray might depend on the other rays it is considered together with). If we fix a particular context for any ray on the left and then measure the antipodal ray on the right, the anticorrelation rule mandates a definite color for the latter. However the locality assumption implies that the context on the left can have no influence on the outcome on the right, and so the inescapable conclusion is that there exists a definite coloring of the rays on the right that is quite independent of any conditions on the left. To see that the coloring on the right is noncontextual we note, to quote ZP, that ''the anticorrelation rule is phrased without reference to the measurement context.'' In other words, the color mandated by the anticorrelation rule for any ray on the right is independent of the other rays with which this ray is measured (or colored). This establishes the noncontextuality of the coloring on the right and, by symmetry, the same must be true of the coloring on the left as well.

The argument of the previous paragraph used the anticorrelation property of the singlet state, together with the assumption of locality, to justify the questionable assumption of noncontextuality made earlier. This argument (when completed by the demonstration of noncoloring in Sec. IV) constitutes one example of Bell's nonlocality theorem, whose purpose is to rule out HVTs based on the locality assumption. Bell's nonlocality theorem is generally regarded as more powerful than the BKS theorem because locality is a testable assumption while noncontextuality is not. For a further discussion of the meaning of the Bell and BKS theorems, see Refs. 8, 10, and 12 already cited.

IV. NO-COLORING PROOFS FOR PENROSE RAYS

Figure 6 uses a ''proof tree'' to demonstrate the impossibility of coloring the 40 Penrose rays green or red in such a way that there is only one green ray in each tetrad. The caption to Fig. 6 lays out the argument, so nothing further need be said here. We point out that the only input needed to make this argument is the tetrad table of Fig. 5. To assist in visualizing this argument, we have written a QBASIC program that displays the 40 Penrose tetrads and asks the user to pick and color any ray he wishes. The user can repeat this process as often as needed, and in this way he can follow an entire branch of the proof tree down to the contradiction at the bottom. We have found that such a ''live demonstration'' can be quite helpful in arousing the interest of a general audience in Bell's theorem.

As explained in Sec. III, the noncoloring argument of Fig. 6, when interpreted suitably, serves to prove both the Bell and BKS theorems. However, a purist may find this proof a little unsatisfying because it involves coloring both the explicit and the implicit rays. A projection measurement corresponding to an explicit ray may be readily made by means of a suitably oriented Stern–Gerlach magnet and a detector in the spin $+\frac{1}{2}$ exit channel. However, projection measurements corresponding to implicit rays are not so readily made. If one wishes to avoid measuring (or coloring) implicit rays altogether, the vertex coloring argument of Penrose[2] involving only explicit rays must be invoked. However, if one regards the Penrose argument as a thought experiment and is unconcerned with matters of practical realization, the argument of Fig. 6 certainly suffices.

ZP pointed out that the 40 Penrose rays contain several subsets of 28 rays that cannot be colored in accordance with the constraints (i) and (ii) stated earlier. Each such subset consists of all the explicit rays together with the 8 implicit rays corresponding to one of the cubes that can be inscribed in the dodecahedron (five such subsets can be constructed, corresponding to the five cubes that can be inscribed in a dodecahedron). The proof tree of Fig. 7 demonstrates the noncolorability of one of the subsets of 28 rays. In following the argument of Fig. 7 it is important to keep in mind that although the 28 rays form only 14 tetrads among themselves, all 40 tetrads need to be kept in view during the coloring process in order to ensure that the constraint (i) is never violated. ZP refer to the 28 rays as a ''critical set'' because the deletion of any one ray from this set leads to a colorable set; this fact is demonstrated in Fig. 8.

V. CONCLUSION

The Penrose dodecahedron is an ingenious toy that can be used as the centerpiece of thought experiment that establishes both Bell's nonlocality theorem and the Bell–Kochen–Specker theorem. The charm of the demonstration lies in the fact that the proof of these deep and subtle theorems can be reduced to a coloring problem on a familiar geometrical object. The coloring problem is simple enough to be understood (and perhaps even solved) by a smart high school student, although understanding why it proves Bell's theorem is a much harder task. The Penrose dodecahedron has a wide appeal (to physicists, philosophers, laymen,...) and can be enjoyed at a variety of levels. Inasmuch as one can profit by examining an ingenious device, or a breathtaking magic trick, in more than one way, it is hoped that the present exposé of the dodecahedron will be welcomed by physicists looking for an alternative explanation of its inner machinery. Though we now understand how this trick works (or think we do), we still continue to be amazed by it and would like to applaud the magicians who performed it.

---

[1] J. Zimba and R. Penrose, ''On Bell non-locality without probabilities: More curious geometry,'' Stud. Hist. Phil. Sci. **24**, 697–720 (1993).

[2] R. Penrose, *Shadows of the Mind* (Oxford U.P., New York, 1994), Chap. 5.

[3] J. S. Bell, ''On the Einstein–Podolsky–Rosen paradox,'' Physics (Long Island City, NY) **1**, 195–200 (1964). Reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge U.P., Cambridge, NY, 1987).

[4] J. S. Bell, ''On the problem of hidden variables in quantum mechanics,'' Rev. Mod. Phys. **38**, 447–52 (1966); reprinted in the book quoted in Ref. 3; S. Kochen and E. P. Specker, ''The problem of hidden variables in quantum mechanics,'' J. Math. Mech. **17**, 59–88 (1967).

[5] A ray is a physical state without regard to its normalization or overall phase. Many different states correspond to the same ray. Since we will be concerned only with the orthogonalities between the Penrose states, it suffices for us to deal entirely with the corresponding rays.

[6] E. Majorana, ''Atomi orientati in campo magnetico variabile'' Nuovo Cimento **9**, 43–50 (1932).

[7] P. K. Aravind and Forest Lee-Elkin, ''Two non-colorable configurations in four dimensions illustrating the Kochen-Specker theorem,'' J. Phys. A **31**, 9829–9834 (1998).

[8] D. M. Greenberger, M. A. Horne and A. Zeilinger, ''Going beyond Bell's theorem,'' in *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafatos (Kluwer, Dordrecht, 1989), p. 73; D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, ''Bell's theorem without inequalities,'' Am. J. Phys. **58**, 1131 (1990); N. D. Mermin, ''Quantum Mysteries Revisited,'' *ibid.* **58**, 731 (1990).

[9] A. Peres, ''Two simple proofs of the Kochen–Specker theorem,'' J. Phys. A **24**, 174–178 (1991).

[10] N. D. Mermin, ''Hidden variables and the two theorems of John Bell,'' Rev. Mod. Phys. **65**, 803–815 (1993).

[11] Sci. Am. (Int. Ed.) SCAMAC, p. 12, Feb. 1993.

[12] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Boston, 1995), Chaps. 5–7.

[13] See, for example, J. J. Sakurai, *Modern Quantum Mechanics* (Addison–Wesley, New York, 1994) or A. R. Edmonds, *Angular momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1960).

[14] This statement, while convenient for our purposes, is perhaps too narrow. In a recent article in Physics Today (April '98), Sheldon Goldstein writes, ''The implications of his (Bell's) work have been widely misunderstood as demonstrating the impossibility of hidden variables rather than the inevitability of nonlocality.'' Even Bell declared that what was proved by impossibility proofs is ''lack of imagination.''

[15] N. D. Mermin, ''Simple Unified Form for the Major No-Hidden Variables Theorems,'' Phys. Rev. Lett. **65**, 3373–3376 (1990). ''Mermin's principle'' is laid out in Eqs. (1) and (2) of this paper. The justification of this principle is direct: If a set of commuting observables obeys a certain functional identity, the expectation value of this identity in a simultaneous eigenstate of the observables implies a similar identity among their simultaneous eigenvalues; but a measurement of these observables in an arbitrary state of the system always collapses it into one of the simultaneous eigenstates. This principle is evidently very old and well known.

[16] A. Peres (Ref. 12) characterizes the requirement that a HVT give the same answer to a yes–no question (projector) no matter which compatible set it is a part of as ''counterfactual compatibility.''

[17] See Ref. 10, Sec. VII. The first sentence of the fourth paragraph of this section has a rather obvious typographical error: the word ''contextual'' is used where the word ''noncontextual'' was intended.