Textbooks[4] usually employ the conservation of angular momentum to establish that the orbit stays in a plane (perpendicular to $\mathbf{r}\times\mathbf{p}$, where $\mathbf{r}$ and $\mathbf{p}$ are not collinear). The present treatment is strictly two-dimensional and does not address the direction of angular momentum. However, the persistence of an orbital plane can be seen by noting that no component of a central force exists perpendicular to the plane.

V. INVERSE-SQUARE FORCE

Motion under an inverse-square force law,

$$f(r) = -k/r^2, \tag{22}$$

conserves a third quantity, a complex number A, defined in the orbital plane,

$$A = -(i\alpha P + R_0), \tag{23}$$

where

$$\alpha = l/km. \tag{24}$$

It is demonstrated below that A is constant.

The vector equivalent of A is the well-known Lenz[5,6] vector $\mathbf{A}$,

$$\mathbf{A} = (1/mk)\mathbf{p}\times\mathbf{L} - \hat{\mathbf{r}}, \tag{25}$$

where $\mathbf{L}$ is the angular momentum vector and $\hat{\mathbf{r}}$ is a unit vector in the radial direction. The equivalence follows by first noting that $\mathbf{p}\times\mathbf{L}$ is the vector $l\mathbf{p}$ rotated by 90° in the $-\mathbf{L}$ direction. Converting to complex number notation, this is P multiplied by (rotated by) $\exp(-i\pi/2)$ or $-i$. Replacing $\mathbf{p}\times\mathbf{L}$ in Eq. (25) by $-ilP$ and writing $\mathbf{r}$ as its complex counterpart R_0 establishes the equivalence with A. Of course, the aforementioned attributes of A were first realized for the Lenz vector, but the treatment requires handling the cross product and triple product.[7]

Differentiating Eq. (23) with respect to time and substituting from Newton's second law and conservation of angular momentum gives

$$-\frac{dA}{dt} = i\alpha\dot{P} + iR_0\dot{\theta}$$

$$= -i\alpha R_0 k/r^2 + ilR_0/(mr^2) = 0. \tag{26}$$

This establishes that A is a constant of the motion.

The polar equation of the orbit is found by taking the dot product of R with Eq. (23),

$$\mathrm{Re}(R^*A) = -\alpha\,\mathrm{Re}(iR^*P) - \mathrm{Re}(R^*R_0). \tag{27}$$

Choosing the orientation[8] of A to be along $\theta = 0$, the left-hand side can be written as $ra\cos\theta$, where a is the magnitude of A. The equation for a conic section follows,

$$\alpha l/r = 1 + a\cos\theta, \tag{28}$$

and a is identified as the eccentricity.

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), pp. 349–351.

[2] See, for example, Morse and Feshbach, pp. 73–76. A mathematician's viewpoint is expressed in S. MacLane and G. Birkhoff, *Algebra* (Macmillan, New York, 1968), pp. 253–255.

[3] Magnitudes are denoted by the corresponding lowercase letters, $f = |F|$. These conventions are convenient but not essential.

[4] See, for example, G. R. Fowles, *Analytical Mechanics* (Saunders, New York, 1986), 4th ed., pp. 141–142.

[5] W. Lenz "The evolution of the motions and quantum condition of disturbed Keplerian motion," Z. Phys. **24**, 197–207 (1924).

[6] The A vector appears to have been introduced first by Laplace. See H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980), 2nd ed., p. 153.

[7] S. Borowitz, *Fundamentals of Quantum Mechanics* (Benjamin, New York, 1967), pp. 305–306.

[8] The overall negative sign in Eq. (23) orients the A vector from the coordinate origin to the pericenter of the orbit. This orientation corresponds to the convention for $\theta = 0$ used in astronomy.

Quantum mechanics made transparent

Richard C. Henry
Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, Maryland 21218 and Earth and Space Science Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 6 September 1988; accepted for publication 14 May 1990)

This article is a "sampler," which shows how quantum mechanics may be presented to students in a way that makes apparent how natural quantum mechanics is as a description of the world. The mathematical machinery of Hilbert space, the idea of representing observables by operators, the Schrödinger equation, and the position-momentum uncertainty relation all follow from natural assumptions that students can readily accept. The basic ideas of quantum mechanics are developed from intuitive first principles to the point where one can connect with more traditional treatments of quantum mechanics.

SIMPLICIO: Salviati, why is our world quantum mechanical, instead of being classical? Quantum mechanics seems unduly complicated!

SALVIATI: Does that question worry you too, Sagredo?

SAGREDO: It does, and I also worry about the abstract and apparently arbitrary nature of the postulates from which quantum mechanics is developed. In Dicke and Wittke's[1] famous book on quantum mechanics, there are seven postulates, including one that has a footnote that contains a reference to a book in German. And, in Shan-

kar[2] (which is one of the best books for learning quantum mechanics), there are four postulates, but two of these contain two parts each, so really six postulates are involved. It seems a lot of postulates!

SIMP: Furthermore, all those postulates are utterly incomprehensible to the beginning student, and never seem in the least degree natural.

SALV: Gentlemen! I can obtain quantum mechanics so naturally it will seem inevitable.

SAGR: What!

SIMP: Do it.

SALV: I'll just sketch it, as it is perfectly ordinary quantum mechanics, just seen from a slightly, but crucially, different perspective. You'll follow it easily, since you have a good knowledge of Dirac Hilbert-space quantum mechanics.

SAGR: (Gulp!) Of course.

SIMP: Hold it, Salviati. I know that $\Delta x\ \Delta p \geqslant \hbar/2$, which got me through my Ph.D. comprehensive, but I'll confess I don't know much more. If you want me to understand, you'll have to fill in the details.

SALV: Fine; the only price we pay for that is length. I will have to cut off at some point, however, let's say at where we easily connect with various points in Shankar's book.

SIMP: Good; proceed.

SALV: Let's stick to one dimension, for simplicity. A typical problem in *classical* mechanics is "given the position of a particle, and given the velocity of the particle, and given that $F = ma$, and given the force F, predict what the position of the particle will be in so many seconds." That's a lot of "givens!" Let's, instead, tackle this more basic problem: "given *nothing*, predict the result of our impending position measurement." Sagredo?

SAGR: Given *nothing*? I can make *no* prediction.

SALV: I am not asking for a perfect result, just do the best you can.

SAGR: Given no information, I cannot make any prediction at all.

SALV: Not so, Sagredo. Here is our labeled x axis. Look at it, and tell me one possible result of the measurement that you are about to make.

SAGR: Well, 2.5 is a possible result.

SALV: Excellent. Now, tell me, what is the probability that you will obtain that result, when you make your measurement?

SAGR: Given no more information?

SALV: Yes.

SAGR: I cannot estimate the probability at all.

SALV: Well, can the probability be 1.7? Can it be -0.2?

SAGR: No; the probability must lie somewhere between 0 and 1.

SALV: And the same is true for every one of the possible values, which of course are the real numbers $-\infty$ to $+\infty$. Let me sketch the probability curve, using a dashed line, since we know nothing of its shape (Fig. 1).

SIMP: You have sketched it incorrectly, Salviati—you don't have it decreasing toward plus and minus infinity, whereas we know that it must decrease to zero, so that the integral under it is unity.

SALV: Excellent, Simplicio! I will redraw it (Fig. 2). So, Sagredo, you were wrong; you were able to make a prediction after all!

SAGR: It's not much of a prediction, Salviati! It seems to me to be nothing more than a concise summary of our ignorance.

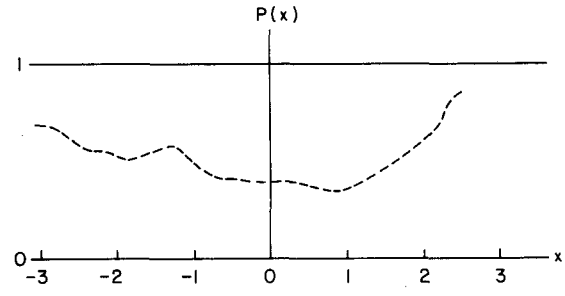

Fig. 1. Simplicio points out the defect in this probability curve, that the area under the dashed (that is, unknown) curve is not unity, or even finite. The corrected version appears in Fig. 2.

SALV: Yes, but it has three great virtues: (1) Poor as it may be, it is a prediction of the result of our impending position measurement; (2) it is true; (3) it is based on nothing—no "laws of physics," or results of previous measurements of anything, or *anything*.

SIMP: But Salviati, what if this whole approach is wrong, and there do exist laws, and a better prediction than Fig. 2 is possible?

SALV: Any "better" theory can only produce a more constrained curve in Fig. 2. Everything we develop in the discussion that follows must be true, regardless of what else may be true as well.

So let me continue. The curve in Fig. 2, I shall now show, still does not contain, in itself, all of the information that we in fact possess. For suppose I were to push part of the curve down (reduce the probabilities, for a certain range of possible results). What would happen?

SIMP: Some *other* part of the curve must come *up*!

SALV: That is what *should* happen, but it does *not* happen automatically. It is for that reason that we now need to develop rather elaborate machinery for discussing curves of the type shown in Fig. 2. To do so, let's start with a

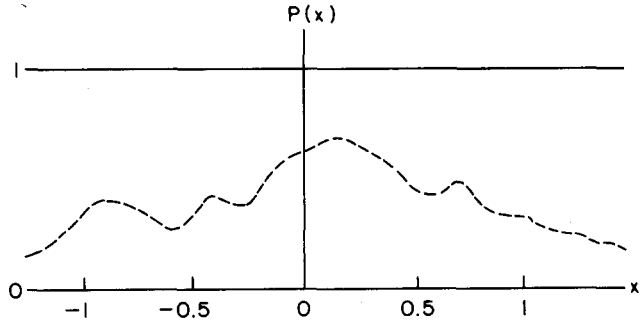

Fig. 2. The possible values for a position measurement are any one of the real numbers, which stretch from $-\infty$ to $+\infty$. The *probability* of any particular result is some unknown number having a value lying between 0 and 1. *Some* result for the position measurement will be obtained, upon measurement, so the area under the curve is unity.

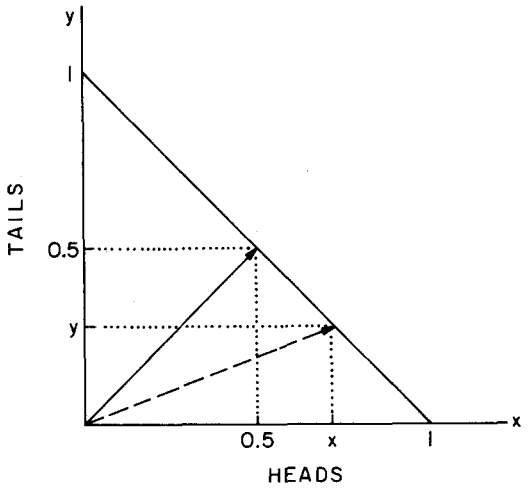

Fig. 3. Our first Hilbert space, that for a tossed coin. The two perpendicular lines are not coordinate axes, they are the two orthogonal eigenvectors of the "coin toss" operator. There are *two* eigenvectors because there are *two* possible results for the "measurement," heads and tails. The solid arrow is the state vector for an honest coin (that is, one having probability 0.5 landing "heads," and 0.5 of landing "tails"). The dashed arrow is the state vector for a biased coin. The probability of heads for the biased coin is x; of tails y. Of course, $x + y = 1$.

situation that is much simpler than that shown in Fig. 2 (where there is an infinite number of possible outcomes). Let's, instead, start with a situation where there are only two possible outcomes, a tossed coin.

Our machinery, called Hilbert space, is shown in Fig. 3. It consists of two lines at right angles to each other; we associate an outward *direction* with these lines, and call them "eigenvectors of the coin-toss operator." (Don't try to remember what the terminology means from previous exposure to quantum mechanics; we are *defining* these terms here.) The spinning coin is represented by the solid arrow, called the "state vector." The probability of the two possible results of the "measurement"—"heads" or "tails"—is given by the coordinates of the tip of the arrow, that is, by the projection of the state vector on the x and y axes, respectively. You will see that we are dealing with an honest coin—the probability of heads is 0.5, and the probability of tails is also 0.5. The state vector for a coin that is biased in favor of heads is shown by a dashed arrow. For that coin, the probability of heads is x, and the probability of tails is y. Of course, $x + y = 1$, since you must get a result, and you will inevitably get either heads or tails (our coin is assumed not to land on edge!). So we draw into Fig. 2 the line that is defined by the equation $x + y = 1$; it slopes up to the left.

Figure 4 shows Hilbert space after I have made my measurement. It just so happens that the result was heads. The new position of the state vector is that it lies on the x axis. This means, of course, that the probability of heads is now 1, and the probability of tails is now zero, which is clearly so.

A most important point, now! Suppose that our two eigenvectors had not been orthogonal (i.e., at right angles to each other), but instead had been at, say, 45°. Our present situation would be that shown in Fig. 5, namely, that the probability of heads is 1, and the probability of tails $\neq 0$, since the state vector has a nonzero projection onto the y axis, despite lying on the x axis. Well, this is just not so—the coin is sitting there, heads upward, and there is no probability that it is tails upward. So we conclude that *the eigenvectors of our operator must be orthogonal*, for our machinery (Hilbert space) to work properly.

Now suppose that I have a three-sided die, so that there are three possible outcomes. This Hilbert space is shown in Fig. 6, with the state vector for an honest (and still-spin-

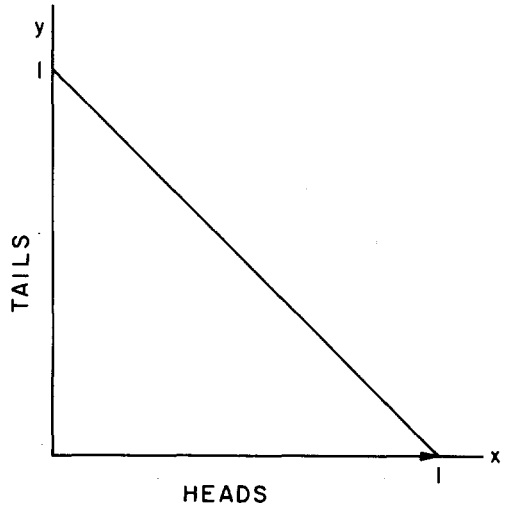

Fig. 4. The coin of Fig. 3 has landed, and landed heads. Note the new position of the state vector.

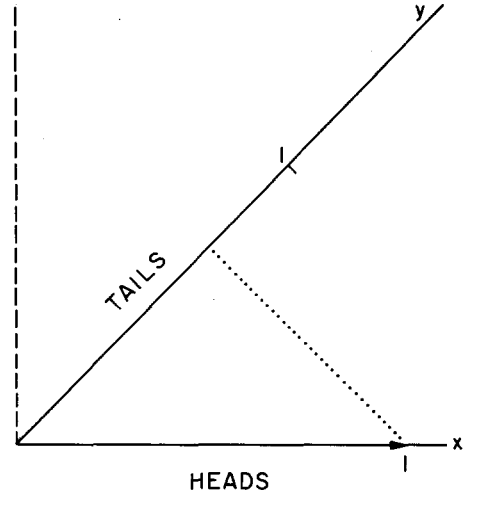

Fig. 5. Our second Hilbert space, a defective one for a tossed coin. Even though the coin has landed heads, this Hilbert space implies, incorrectly, that there is still some probability that the coin is in fact tails! The defect, of course, is that the two eigenvectors are not orthogonal. We learn that all operators, where a definite result is obtained in a measurement, must have mutually orthogonal eigenvectors.

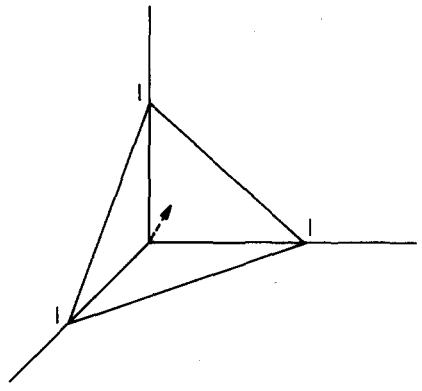

Fig. 6. Hilbert space is "3-D" for a three-sided die. The state vector for an honest die is shown as a dashed arrow. This operator has three eigenvectors, mutually perpendicular of course.

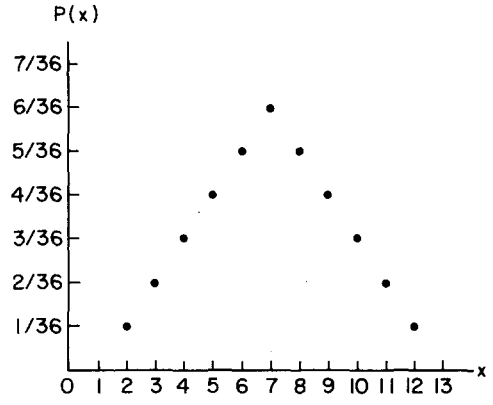

Fig. 8. Here is the probability curve for a pair of dice thrown together. This probability curve looks a bit more like that in Fig. 2. The numbers on the x axis are the eigenvalues.

ning) die included in the diagram. Also shown is the plane $x + y + z = 1$. With this figure, you can see *why* we are developing our present machinery: If I were to move the state vector in Fig. 6 so that the probability of one particular outcome is, say *reduced* (that is, so that the state vector has a *smaller* projection on one particular axis), the probabilities of the *other* possible outcomes will be, on average, *increased*, and this happens *automatically*.

Actually, I don't know how to construct a three-sided die!

You can almost imagine the Hilbert space for an ordinary six-sided die (even though I can't draw the space, of course): the space is six-dimensional, with six mutually orthogonal axes and contains a "six-plane" having equation $a + b + c + d + e + f = 1$, corresponding to the equation $x + y = 1$ for our coin. You can even almost visualize the state vector for an honest die! The possible results of your measurements are, of course, 1, and 2, and 3, and 4, and 5, and 6; these are called the *eigenvalues* of the die-toss operator. (The eigenvalues of the coin-toss operator were "heads" and "tails," or if you wish, you can put a numerical label on each side of the coin. A physicist would likely choose labels $+\frac{1}{2}$ and $-\frac{1}{2}$.)

The probability curve for an ordinary (honest) die is given in Fig. 7.

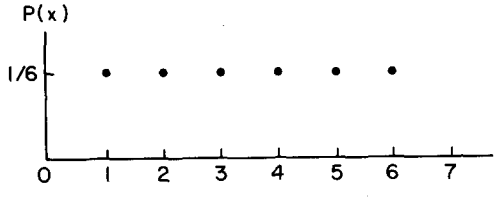

Fig. 7. Here is the probability curve (like that of Fig. 2) for a normal six-sided die. We cannot draw the corresponding Hilbert space, because it is six-dimensional.

Suppose you were to throw a pair of identical dice. The Hilbert space is eleven-dimensional, and the probability curve for honest dice is given in Fig. 8.

Comparing Fig. 8 with Fig. 2 makes clear the relationship between this work that we have been doing on dice/coins, and our original problem. Clearly, for our original problem, Hilbert space is infinite-dimensional. There are infinitely many mutually orthogonal eigenvectors of the "position" operator, and the eigenvalues of the position operator are all the real numbers between $-\infty$ and $+\infty$. While I cannot draw an infinite-dimensional Hilbert space, I can extract a chunk of it for examination, and I have done so in Fig. 9. Three typical eigenvalues (possible results of our impending position measurement) are 2.5, 4.7, and -12.8; these numbers are used to label their eigenvectors.

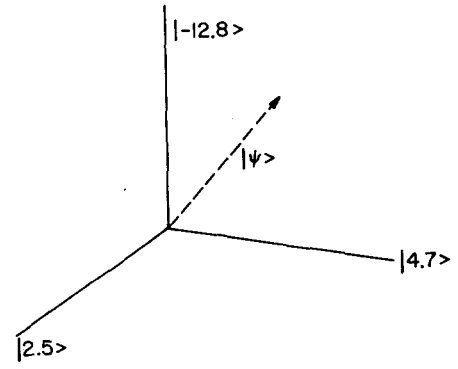

Fig. 9. Previous Hilbert spaces were fakes, for learning purposes. Here is a genuine Hilbert space, or rather, part of one, because this space, the "position" Hilbert space, is infinite-dimensional. I have extracted a small piece of it for display. The dashed arrow is the state vector (its direction is unknown). Three eigenvectors, labeled with their corresponding eigenvalues, are shown.

The state vector is shown dashed, since we have no idea (see Fig. 2) where it is located, that is, what the relative probabilities are of the various outcomes (such as 4.7). Notice also that in Fig. 9 I have used Professor Dirac's symbols for designating vectors in Hilbert space: $|4.7\rangle$ means "the eigenvector of the position operator that has eigenvalue 4.7"; $|\psi\rangle$ is the state vector.

Suppose now that I make my position measurement, and that the result happens to be 4.7. Then the Hilbert space looks as in Fig. 10. The state vector is no longer dashed, as we now know precisely where it is—the probability of the position being anything other than 4.7 (the value that has just been actually measured!) is clearly zero, so the state vector has no projection on any other of the infinite number of position eigenvectors and has length unity along $|4.7\rangle$.

SAGR: Suppose we now repeat our position measurement. Will we get the *same* result, from a *second* position measurement?

SALV: Our machinery as constructed *says* we will, and experiment (always the ultimate arbiter) agrees.

I told you that our machinery was rather elaborate! But now we have it fairly well defined. Let's use it to obtain an important insight!

We've been talking about a position measurement. Now let's talk about a measurement of any other quantity (e.g., humidity, momentum, spin of an electron, ...). Always a definite result is obtained, so the Hilbert space for, say, the humidity operator, or the operator corresponding to any other physical quantity, must be made up of orthogonal eigenvectors. And, let's insist on having an infinite number of dimensions, so that identical machinery will accommodate all possible physical quantities. (For example, you know that in the case of the spin of the electron there are observed to be only two possibilities, spin $+\frac{1}{2}$ and spin $-\frac{1}{2}$, but if we label half of our infinite number of eigenvectors $|1/2\rangle$, and half of them $|-1/2\rangle$, then we successfully accommodate the observed values of spin in our general purpose machinery). One of our great aims, of course, is to discover exactly what the eigenvalues are for various operators (that is, physical quantities). At the present point, we only know the eigenvalues for position, and we only know those because Pythagoras taught us that space is real numbers.[3]

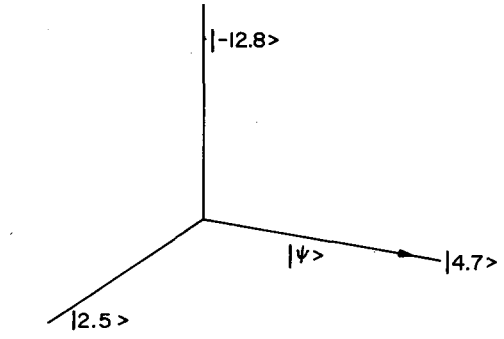

Fig. 10. Following Fig. 9, a position measurement has now been actually made, and the result is $x = 4.7$. The new position of the state vector is no longer dashed, because we know where it is now.

Finally, we will insist that eigenvalues turn out to be real numbers, because every measured quantity is a real number.

SIMP: How can I be sure *all* measurable quantities are real numbers?

SALV: Consider humidity; voltage; air pressure; and dozens of others. You make your measurement, in the end, by reading the *position* of a pointer in a *scale* that is laid out along one dimension of *space*: Pythogoras tells us that our reading is a real number!

SIMP: Yes! And for electron spin, I hold a ruler at the output of the Stern–Gerlach experiment, and I get a real number as well—but not just any real numbers; just two values, the two possible values of electron spin.

SIMP: So we have our Hilbert spaces for each and every quantity. Now let's discuss the relationship between the Hilbert space corresponding to the position of a particle and that corresponding to its momentum. In each case, the Hilbert space is infinite-dimensional, the eigenvectors are mutually orthogonal, and a state vector (one in each space) represents the physical situation whatever it may be.

First, actually *make* a *position* measurement: The position state vector of course snaps to being some position eigenvector.

Now place the origins of the two Hilbert spaces on top of each other, and rotate the two Hilbert spaces so that the two state vectors coincide. This is obviously always possible, and clearly simplifies matters. This does not yet completely determine the relative orientations of the position and momentum Hilbert spaces, however; any rotation of the Hilbert spaces around the common state vector is possible. Let us rotate these two Hilbert spaces around the common state vector until as many eigenvectors as possible coincide.[4]

(Carry out the same procedure for the operators corresponding to every physical variable, while you are at it. Now all the state vectors coincide.)

What is the result? We don't know, because at this point we know nothing at all about the momentum operator. What we can do, however, is describe the *possible* results, and discuss what they would mean.

(1) One possibility is that *all* the momentum eigenvectors lie exactly along position eigenvectors. In particular, this would mean that a momentum eigenvector would lie along the state vector; that is, the particle is in a momentum eigenstate (as well as being in a position eigenstate). This, in turn, would mean that if you were now to make a momentum measurement, you could be certain that your measured value would be the eigenvalue (whatever it is) of the momentum eigenvector that is coincident with the position eigenvector that is the state vector.

That's one possibility.

(2) A second possibility is that *some* of the momentum eigenvectors lie along position eigenvectors, but not all.

(3) The final possibility is that the particle is not in a momentum eigenstate and that no amount of rotation about the common state vector (which is a position eigenvector) brings even a *single* momentum eigenvector into coincidence with a *single* position eigenvector. What this possibility would mean is that when the state vector is a position eigenstate "the coin is still spinning" as far as momentum is concerned; that is, the probability of obtaining any specific value of the momentum is less than unity.

These are the *only* possibilities. Now, you know perfectly well, from previous knowledge of quantum mechanics, that we are going to find, eventually, that we are driven to "possibility 3" in the case of momentum. The main thing to realize at this point is that if indeed we find a logical reason why possibility 3 must represent the situation (in the case of momentum), then there is no mystery as to why the uncertainty principle holds or why uncertainty is fundamentally involved in measurements: It simply must be, logically.

SIMP: So how do we get driven to possibility 3?

SALV: Patience, Simplicio, we have a long way to go yet. But the great importance of the results fully justifies the investment of your time.

Let's now go back to Fig. 3, where we first developed the machinery of Hilbert space and ask whether our design is unique. Are there other (perhaps better) ways to build Hilbert space?

Yes, an esthetically perhaps more satisfying version of Hilbert space is shown in Fig. 11. In place of the line $x + y = 1$ we have the circle $x^2 + y^2 = 1$. The probabilities are no longer x and y, of course; they are x^2 and y^2, the quantities having sum unity. The quantities x and y themselves we will give a name; we will call them the probability *amplitudes.*

Our new version of Hilbert space has only one (minor) advantage over the old, and that is, that the state vector is of constant length.

Figure 11 does not exhaust the possibilities. Try plotting up $x^3 + y^2 = 1$. This example loses the advantage of constant length in the state vector, but otherwise there is nothing wrong with it. In this case, we would again refer to x and y as the probability amplitudes, and x^3 and y^2 would be the two probabilities, the sum of which must of course be unity.

So we see that, in general, the two probabilities can be any (I suppose I should say, any "within reason") functions of x and y, and also the two functions need not be the same function.

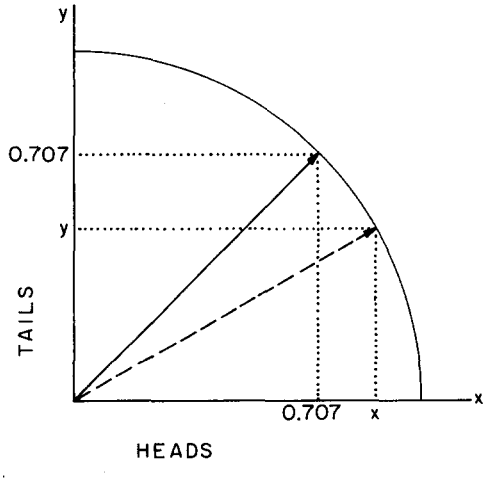

Fig. 11. A "new improved" version of Hilbert space, for a tossed coin. The probabilities are now $x^2 + y^2$ instead of x and y; the latter are renamed the "probability amplitudes." The sum of the probabilities is, of course, unity, so $x^2 + y^2 = 1$ (this curve, a circle, is illustrated). Now the state vector is of constant length!

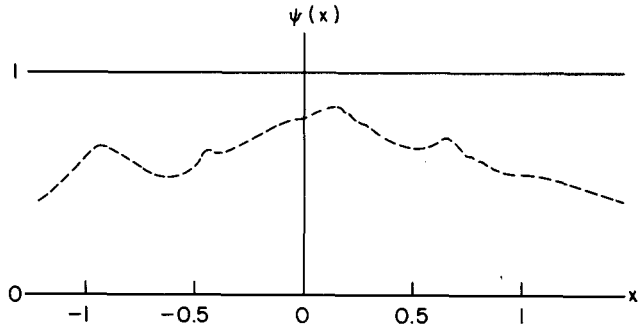

Fig. 12. Since, with our new version of Hilbert space, the probability curve is the square of the probability amplitude, our probability curve of Fig. 2 is now joined by a corresponding curve, shown here, giving the probability amplitudes. This curve is called $\psi(x)$, the wave function.

So what have we learned? That we have some freedom left in our construction of Hilbert space!. We will need that freedom before too long. For the time being, let us adopt the $x^2 + y^2 = 1$ version as *our* Hilbert space, because of its nice feature that the state vector is of constant length.

This choice means that each plot of the kind shown in Fig. 2 is now joined by an associated plot (Fig. 12) of the probability amplitudes, $\psi(x)$. This curve, $\psi(x)$, is called "the wave function." In the $x^2 + y^2 = 1$ case that we have chosen, one simply squares the wave function to obtain the probability curve (Fig. 2).

Now I want you to recall some elementary mathematics. Forget what we are doing and turn your attention to Cartesian coordinate systems in "plane ordinary" two-dimensional space (Fig. 13). I have no doubt that you remember

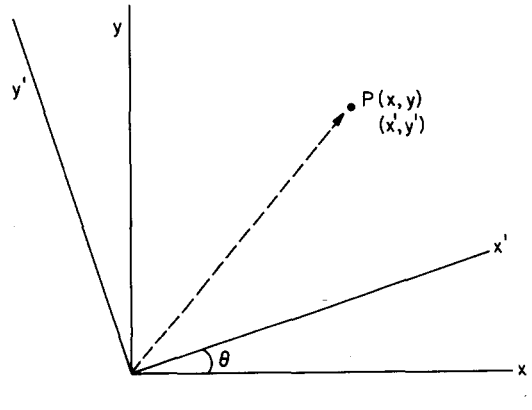

Fig. 13. First we consider this as two sets of ordinary coordinate axes in the two-dimensional plane, and we review the mathematical connection between the coordinates (x,y) and (x',y') of the point P, referred to the two coordinate systems. Second, however, we consider this as a 2-D chunk of Hilbert space, and consider the lines x,y and x',y' (which formerly were axes) to be eigenvectors of two operators, which for convenience we designate the xy operator and the $x'y'$ operator.

the formulas connecting the coordinates (x,y) of point P in one coordinate system with the values of the coordinates (x',y') of the same point in the other coordinate system; the equations are

$$x' = x\cos\theta + y\sin\theta,$$
$$y' = y\cos\theta - x\sin\theta,$$

where θ is the angle of rotation between the two coordinate systems. No doubt you also recall that this pair of equations is isomorphic to the following matrix equation:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

For every value of θ, there is a unique matrix

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Now let's return to Hilbert space! Imagine Fig. 13, now, as a chunk of Hilbert space. The lines labeled x and y, being orthogonal, and having an outward direction associated with them, represent eigenvectors of some operator. Merely because these two lines happen to be horizontal and vertical, we will call this particular operator (whichever one it is) our "basis operator."

Similarly, the lines labeled x' and y', being orthogonal, represent eigenvectors of some other operator. (Notice that these two operators stand in a "possibility 3" relationship to each other. For suppose the state vector lies on the line labeled x. Then, certainly, no rotation of the two operators, or either of them, around the line labeled x, will cause any two lines to coincide. The same is true for each eigenvector of each of the two operators.)

But now to the point. In the x, y basis, we have discovered that there is a unique 2×2 matrix that specifies the orientation of the set of eigenvectors of any operator with respect to our reference set of eigenvectors. This means that the operator is, in some sense, "represented" by such a matrix.

This result generalizes easily to any number of dimensions, although it requires more and more independent parameters (one less than the number of dimensions) to specify the matrix.

Notice that the values of the matrix elements depend on the basis chosen: If the basis were different in Fig. 13 (say, tipped slightly) the values of θ would be different, and the matrix elements would be different.

So, from now on, when we have chosen a specific basis to work in, and we are thinking of an operator corresponding to some physical variable, we are also thinking of a certain specific matrix.

The reason for making this switch, away from geometry, and instead to (completely equivalent) matrices, is that matrices can be manipulated easily, using algebraic methods.

In the case of the operator having eigenvectors x, y in Fig. 13, what is its matrix in its own basis? Clearly it is the general matrix, but with θ set equal to zero (so that $x' = x$ and $y' = y$), or

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos 0 & \sin 0 \\ -\sin 0 & \cos 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

So we see that an operator is represented, in it own basis (which is called its "eigenbasis"), by the unit matrix. (The unit matrix is also called the identity operator; we will use it frequently.)

Now the column matrix $\binom{x}{y}$ corresponds to any vector, that is, any point P. Consider a point on the x axis, that is, consider the vector $\binom{x}{0}$. This vector represents the x axis itself, which is one of the eigenvectors of the basis operator. Normalize it to unit length [the length of the vector $\binom{x}{y}$ is of course $\sqrt{x^2+y^2}$], and the eigenvector is represented by $\binom{1}{0}$. Similarly, the other eigenvector of this operator is represented by $\binom{0}{1}$, in the eigenbasis.

So, returning to our planned measurement of position in one dimension, where we have an infinite-dimensional Hilbert space, we know that in the position basis, the position operator can be represented by an infinite unit matrix, and a typical eigenvector is

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

But now let's be clever and elegant! We know that with each position eigenvector is associated an eigenvalue, and furthermore, thanks to Pythagoras, we know what these (position) eigenvalues are: They are the real numbers. Let's invert a method of tacking the eigenvalue onto the eigenvector, without doing any damage! We can do this by representing our operator, not by the unit matrix, but by a diagonal matrix having the eigenvalues as its diagonal elements (all other elements remain zero). Thus the particular tiny chunk of the infinite Hilbert space consisting of the eigenvectors $|5.0\rangle$ and $|5.1\rangle$ would have its position operator represented, in the position basis, like this:

$$\begin{pmatrix} 5.0 & 0 \\ 0 & 5.1 \end{pmatrix}.$$

Now let's decide to multiply the eigenvector $|5.0\rangle$ by the matrix representing the position operator, just to see what happens:

$$\begin{pmatrix} 5.0 & 0 \\ 0 & 5.1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 5.0 \\ 0 \end{pmatrix} = 5.0 \begin{pmatrix} 1.0 \\ 0 \end{pmatrix}$$

We have discovered that when we choose to "let the position operator act on" a given position eigenvector, the result is a vector that, clearly, lies in the same direction as the eigenvector in question, but that is stretched, so that its length...is the eigenvalue! [Try it now for the other eigenvector, $\binom{0}{1}$.]

So we have changed our operator into a "stapling machine" that staples the eigenvalues onto the eigenvectors. The order of the eigenvalues down the diagonal doesn't matter, because the eigenvectors are "all the same"—each perpendicular to all the others. For convenience, one naturally tends to have the labels increase prettily down the diagonal.

Suppose we act on an arbitrary vector, with our position operator:

$$\begin{pmatrix} 5.0 & 0 \\ 0 & 5.1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5.0x \\ 5.1y \end{pmatrix}.$$

The result is a vector that clearly lies in a different direction than the original vector. It is obvious that we will obtain, as our result, a vector that lies in the same direction as the original vector, if and only if the original vector is an eigenvector of the position operator; that is,

$$X|x\rangle = x|x\rangle$$

is true if and only if $|x\rangle$ is a position eigenvector, having eigenvalue x. (X designates the position operator or matrix.)

The same is true for any operator:

$$\Omega|\omega\rangle = \omega|\omega\rangle$$

is an infinite set of equations which, if you are given the operator Ω (that is, given the matrix that represents Ω in some basis), you can in fact solve to obtain the eigenvalues ω; and the eigenvectors $|\omega\rangle$ (that is, the column matrices representing the eigenvectors in that basis).

Since we still know nothing about the momentum operator, we have no idea what its representation is in the position basis (that is what we are trying to find out!). But we do know what its representation is in the momentum basis, its "own eigenbasis" (if I may be permitted the tautology): It is diagonal, and the diagonal elements are its eigenvalues, whatever they may eventually turn out to be.

Now, finally, let us turn away from our discussion of the position operator and seek to actually find another operator. What operator shall we pick?

SIMP: Momentum!

SALV: Why?

SIMP: What do you mean, why! Isn't that what you've been leading up to? Are you...

SALV: Calm down, Simplicio! I'm just asking why it is that we care so much about momentum.

SIMP: Well, it is conserved.

SALV: Right. Around the time of Newton, there was great confusion about the notion "quantity of motion" of a body. What was the important quantity? In the end, it was decided that there are *two* important quantities, mv and $\frac{1}{2}mv^2$, and that other possible quantities ($\sqrt{2}mv^5$, $3mv^3$, etc.) are unimportant. They are unimportant because, although we can measure them if we like, they are not conserved. And *why* are mv and $\frac{1}{2}mv^2$ conserved?

SAGR: I remember that from classical mechanics, Salviati. Momentum is conserved because nature is symmetrical regarding translation through space, and energy is conserved because nature is symmetrical regarding translation through time. Because there are three space dimensions, momentum has three components; energy is a scalar because time is one-dimensional. It is all explained nicely by Marion.[5]

SALV: Right, Sagredo. Also, because nature is symmetrical regarding rotation in space, angular momentum is conserved, and because in weak interactions nature is *not* symmetrical with regard to mirror-image experiments, parity is *not* conserved. Well, we need none of this; we just bring it out in order to provide us with a hint as to how to find, or rather how to define, the momentum operator. Clearly, translation through space is the path we should follow (and *conservation* under translation through space is the definition we are seeking).

SIMP: But if we are starting from first principles *how do we know* that there will be symmetry with regard to translation through space (and hence an associated conserved quantity, which it would be natural to name "linear momentum")?

SALV: We don't. (We can seek such a quantity in our measurements; and of course we do find it, and that does indicate that space is symmetrical with regard to translations.) All that we are saying is, suppose such a symmetry exists and such a conserved quantity is found, *what must its operator be*? That is all that we are aiming to accomplish.

So we are motivated to consider the spatial translations operator, T. We are now broadening our definition of "operator" to include any matrix, regardless of whether it happens to correspond to a physical observable. T is to be that matrix which, when we let it act on (multiply) the state vector, causes the state vector to move (rotate) in such a way that its corresponding wave function is displaced (translated) in Fig. 12; that is, the particle has moved in the x direction.

We can consider *active* translation, or *passive* translation: In Fig. 12, active translation is defined as moving the dashed curve (the wave function of the particle) a certain amount, a, to (say) the right.

Now doing this looks a bit dangerous. It requires mucking around with the state vector in Hilbert space. We are trying to avoid making any unnecessary physical assumptions. So instead, let's play it safe and choose a passive translation, which is defined to be merely a relabeling of our position axis so as to move the scale an amount a to the left. Since only a relabeling of the axis is involved here, no physics is involved at all. But! Miracle! The result is exactly the same as if we have used the active approach. The proof is apparent in Fig. 14, which has been produced from Fig. 12 by either an active or a passive translation of $a = 1$ unit. You can't tell which! (This occurs because Hilbert space has no absolute orientation.)

Now the result of this translation is clearly that the state vector has moved in Hilbert space (since it now has a different projection on each of the infinite number of axes). We have defined T to be the operator that produces this movement, when it "acts on" the state vector:

$$T|\psi\rangle = |\psi'\rangle.$$

where $|\psi\rangle$ is the original state vector, and $|\psi'\rangle$ is the state vector after our relabeling of the x axis.

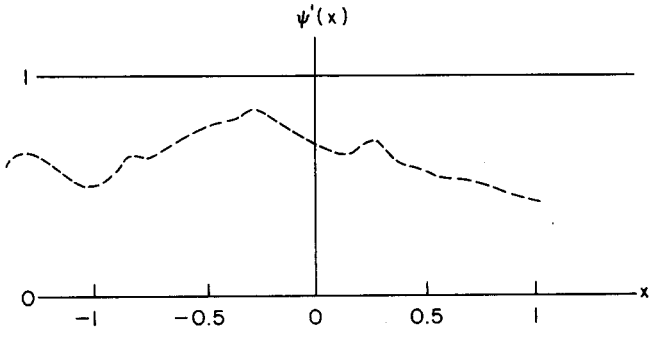

Fig. 14. The result of the translation operator T acting on the state vector is to shift the wave function of Fig. 12 to the left by one unit. That is, whatever it is that you are measuring the position of has shifted by 1 unit to the left (or we have moved our scale 1 unit to the right).

Now, recall that we carefully lined up the state vectors associated with *every* other physical variable to coincide with the particular state vector that is involved with position. Relabeling the x axis clearly cannot affect the relative probabilities of obtaining various values for other physical quantities, so if we want to keep all the state vectors together (as we do, so that the complete physical state of a particle will be always represented by a *single* state vector in Hilbert space), we must hope that if we now act on all the vectors, in all these superimposed Hilbert spaces, with T, the result will be that all the other state vectors are moved so that they stay with the position state vector; and, more importantly, that T moves all the eigenvectors of all the other operators in just such a way that they maintain precisely their original relationship with their state vector; in particular, each state vector's projection on its eigenvectors, we hope, will remain the same.

Now, the translation operator T (some matrix) rotates the position state vector so that it points in some new direction in Hilbert space. Since all the other state vectors have exactly the same components as the position state vector (being coincident with it), they will be rotated exactly the way the position state vector is rotated, and so they *will* stay with it, as desired. So far so good! We still need to show that the translation operator shifts all the eigenvectors of the various operators appropriately. Let's leave this as a "loose end" for now; we will return to it!

How do we go about connecting Hilbert space (e.g., Fig. 9) with its corresponding wave function (Fig. 12)? The wave function is, of course, a display of the components of the state vector. How do we obtain those components, from Hilbert space, in order to construct a probability curve? Consider Fig. 13. The individual components of the vector P in (for example) the $x'y'$ basis can be obtained as follows:

$$(1,0)\begin{pmatrix}x'\\y'\end{pmatrix}=x' \quad (0,1)\begin{pmatrix}x'\\y'\end{pmatrix}=y',$$

where (1, 0) and (0, 1) are the eigenvectors of the $x'y'$ operator, in its eigenbasis: Just carry out the matrix multiplication involved, and you will see that this is so.

Notice that instead of writing each eigenvector as a column matrix,

$$|V\rangle \Leftrightarrow \begin{pmatrix}1\\0\end{pmatrix},$$

we now find it useful to represent each eigenvector with a *row* matrix,

$$\langle V| \Leftrightarrow (1,0).$$

Also, notice that we have introduced Dirac's $\langle V|$ notation for a vector that is to be represented in this manner. Dirac called $|V\rangle$ the "ket" version of the vector, and $\langle V|$ the "bra" version, so that, for example, we call $|x\rangle$ a "position eigenket."[6]

So we now see that to obtain the components of an arbitrary vector (say the state vector, $|\psi\rangle$) in the position basis, we just form

$$\langle x_i|\psi\rangle = \psi(x_i),$$

where $\langle x_i|$ are the position basis vectors. That is, we "project $|\psi\rangle$ onto the position basis." We see that $\langle x_i|\psi\rangle$ is a number, and there are infinitely many such numbers, and that appropriately plotted they form the wave function (Fig. 12); that is,

$$\langle x|\psi\rangle = \psi(x). \qquad (1)$$

We will now make use of Eq. (1) in studying the effect of the translation operator T. We will only need an infinitesimal translation, so the operator can be written

$$T(\epsilon) = I + \epsilon K$$

where I is the identity (unit) operator, which of course does nothing to any vector (try it!) and ϵ is an infinitesimal real number. K is some as yet unknown operator.

Apply this infinitesimal translation operator to the state vector, and then form the new probability curve, by projecting the new state vector onto the position basis:

$$\langle x|(I+\epsilon K)|\psi\rangle = \psi'(x),$$

where $\psi'(x)$ is the dashed curve shown in Fig. 14, which has the same shape as the curve $\psi(x)$ in Fig. 12, but is shifted to the left by a certain amount, infinitesimal in the present case.

But if we know $\psi(x)$, then we *know* what $\psi'(x)$ is: Consulting Fig. 15, we see that

$$\psi'(x) = \psi(x) + \zeta\frac{d\psi}{dx} + \cdots,$$

where ζ is an infinitesimal distance that $\psi'(x)$ is shifted relative to $\psi(x)$. (Positions are represented by real numbers, so this shift in position ζ must also be a real number with "dimensions" of "length.") This is just a Taylor series. So we have

$$\langle x|(I+\epsilon K)|\psi\rangle = \psi(x) + \zeta\frac{d\psi}{dx} + \cdots$$

or

$$\langle x|I|\psi\rangle + \langle x|\epsilon K|\psi\rangle = \psi(x) + \zeta\frac{d\psi}{dx} + \cdots$$

or

$$\langle x|\psi\rangle + \epsilon\langle x|K|\psi\rangle = \psi(x) + \zeta\frac{d\psi}{dx} + \cdots$$

(since $I|\psi\rangle = |\psi\rangle$, because the identity operator I does nothing; and since ϵ is just a number) or

$$\psi(x) + \epsilon\langle x|K|\psi\rangle = \psi(x) + \zeta\frac{d\psi}{dx} + \cdots$$

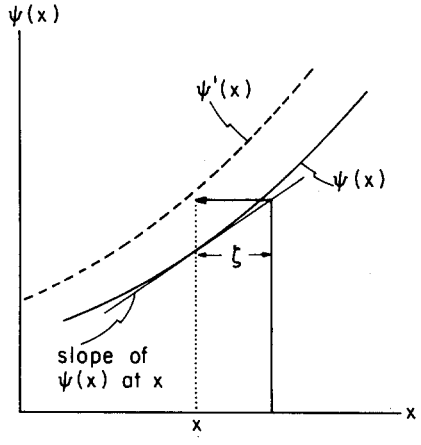

Fig. 15. The result of a shift, such as that between Figs. 12 and 14, is analyzed. The new wave function = the old wave function plus ζ times the slope of the old wave function (to adequate accuracy).

or

$$\langle x|K|\psi\rangle = \frac{\zeta}{\epsilon}\frac{d\psi}{dx}, \tag{2}$$

to first order. The number ζ/ϵ is a ratio of infinitesimals, so it is itself a finite number. The infinitesimal number ζ is real because it is an infinitesimal chunk of space, and Pythagoras tells us that space is real numbers. That is *physics*. The infinitesimal ϵ is also real, because to this point, our Hilbert space has been a real Euclidean infinite-dimensional space, as we have ourselves constructed it—no physics involved at all. In sum, then, ζ/ϵ is a real number.

Identically, if L is the operator producing translation through time, we have

$$\langle x|L|\psi\rangle = \frac{\eta}{\alpha}\frac{d}{dt}\langle x|\psi\rangle = \left\langle x\left|\frac{\eta}{\alpha}\right|\dot{\psi}\right\rangle,$$

where α is an infinitesimal quantity analogous to ϵ in Eq. (2), and η is an infinitesimal chunk of time, which will also be a real number.[7] Notice that this is a statement about the projection of certain vectors onto the position basis. Unprojected, the equation reads

$$L|\psi\rangle = (\eta/\alpha)|\dot{\psi}\rangle. \tag{3}$$

Now, what are the values of the ratios ζ/ϵ and η/α? We are free to choose these values as we please. Consider the ratio ζ/ϵ, for example. While ζ represents a certain fixed distance (the infinitesimal distance that the wave function has been shifted), ϵ is as yet undetermined. The effect of the product ϵK is to generate this specific shift, but this puts no restriction on either ϵ or K separately. We can choose ϵ to have any magnitude and units that we desire, recognizing that K will then have units reciprocal to whatever we choose for ϵ and its overall magnitude will be inversely proportional to the magnitude we choose for ϵ. In short, choosing a specific value and units for the ratio ζ/ϵ simply boils down to choosing the relative magnitudes of ϵ and K and how units are to be apportioned between them. A similar argument applies to the ratio η/α. Let us, without prejudice, use the symbol $\hbar$ to represent our chosen value for the ratio ζ/ϵ and let us also choose $\eta/\alpha = \zeta/\epsilon$.

Furthermore, I choose to divide both sides of Eq. (2) by i, and multiply both sides of Eq. (3) by i. The result of this activity should please you:

$$\langle x|(K/i)|\psi\rangle = -i\hbar\frac{d}{dx}\langle x|\psi\rangle, \tag{4}$$

$$(iL)|\psi\rangle = i\hbar|\dot{\psi}\rangle. \tag{5}$$

SIMP: Equation (5) is the time-dependent Schrödinger equation; your operator (iL) is the Hamiltonian, H. And Eq. (4) shows that your operator (K/i) is P, the momentum operator! I am indeed pleased!

SAGR: Furthermore, I noticed earlier that $\Omega|\omega\rangle = \omega|\omega\rangle$ is the time-*in*dependent Schrödinger equation, if we choose $\Omega = H$. But there is a problem! I have worked an example, $\theta = \pi/2$, using your sample operator,

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos(\pi/2) & \sin(\pi/2) \\ -\sin(\pi/2) & \cos(\pi/2) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

with the following results for $\Omega|\omega\rangle = \omega|\omega\rangle$:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix} = i\begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix}$$

and

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix} = -i\begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix}.$$

If you carry out the matrix multiplication on the left sides, you will verify that I have correctly solved the eigenvalue problem for this operator. But notice that the eigenvalues ω, which are supposed to be the values of your physical quantity, are imaginary!

SALV: That is no difficulty, Sagredo, since you are solving the wrong problem: You are starting with an operator (matrix) and finding the eigenvalues. No one has said that every matrix must represent some physical quantity! What we are doing is precisely the opposite to what you have just done: We are asserting that we *require that* the eigenvalues be real, since they are physical quantities; we are putting additional restrictions on them (conservation laws); and then we are tying to solve for (find) the *operator* involved.

But we actually do have a severe difficulty, which your example nicely reveals. Look at your eigenvectors! They include components that are imaginary! Vectors in Hilbert space suddenly have some imaginary components! How can this be? And it is a disaster for us, because our state vector's components (or their squares) are supposed to be probabilities, which must be real (and nonnegative)!

This is all very surprising, because we built Hilbert space as a *real* infinite-dimensional space. It is important that we understand *why* this has happened to us. It can only (since we have done nothing else) be because we put the physical quantities down the unit diagonal, creating the equation $\Omega|\omega\rangle = \omega|\omega\rangle$. What becomes of this equation when we change basis? Our eigenvalues are scalars, and therefore invariant (under rotations in Hilbert space). But we have no guarantee that, to maintain the constructed (and then asserted) truth of $\Omega|\omega\rangle = \omega|\omega\rangle$, the eigenvectors $|\omega\rangle$ won't develop some imaginary components. In fact, that must be what is happening.

This tells us that, in general, in a vector space where $\Omega|\omega\rangle = \omega|\omega\rangle$, the vectors have complex components. While this just about completely destroys what was left of our ability to *visualize* our infinite-dimensional space, with its state vector moving about (its projections on the eigenvectors of the various operators representing the probabilities), it certainly does not invalidate it. Our state vector still has just as many components as there are possibilities, and if we continue to represent the probabilities as functions of the projection of the state vector onto each eigenvector (summing to unity), our sole actual problem is the fact that our state vector's components are now (in general) complex, while the probabilities must be real and nonnegative.

SIMP: Can we save the situation?

SALV: Yes, we can because (remember!) we still have some freedom left in our construction of Hilbert space. We now want to use that freedom to ensure that even though the components of the state vector are complex, the probabilities are real. To accomplish this, recall that from $x + y = 1$, and $x^2 + y^2 = 1$, and $x^3 + y^2 = 1$, and so forth, we tentatively selected the "$x^2 + y^2 = 1$" version of Hilbert space. Now, remember that it is the *probabilities* (x^2 and y^2 in the case we have chosen) that must be real. We need

have no concern that the probability *amplitudes* (x and y) have turned out to be complex, because they are not observed quantities. The solution to our difficulty now becomes apparent: Instead of picking $x^2+y^2=1$, pick $x^*x+y^*y=1$ and we are back in business. We can do this, of course, only because we discovered earlier that we had a certain amount of freedom in how we constructed Hilbert space. But at this point we lose much of that freedom. It is important to notice that our new (and final) method of constructing probabilities (from complex amplitudes) is required (not an option), if we are to be able to discuss physical probabilities intelligently at all. Underlying every physical probability, we discover, lies a complex probability amplitude.

Our state vector components are now complex numbers,

$$x=a_1e^{i\theta_1} \text{ and } y=a_2e^{i\theta_2}$$

and so $x^*x=a_1e^{-i\theta_1}a_1e^{i\theta_1}=a_1^2$ which is real; so is y^*y, as required. We must have $a_1^2+a_2^2=1$, of course.

So all that remains of our freedom, in constructing Hilbert space, is the freedom to choose (or change) the phases, anywhere or everywhere, at will.

The fact that our probability amplitudes are now complex requires us to reexamine Hilbert space and see what other changes are needed. For one thing, the length of a vector (recall) was $\sqrt{x^2+y^2}$, where x and y are the probabilities (for the case $x+y=1$), or the probability amplitudes (for any other case). The special case that the vector is on the x axis ($y=0$) gave $\sqrt{x^2}=x$, $=1$ if normalized; the x axis is then the vector $(1,0)$, a basis vector of the "xy" operator. We want to keep that arrangement! Now, we've never actually had to use the length of a vector for anything other than this, so we can redefine length so as to retain the desired property. What's needed is

$$\text{length}=\sqrt{x^*x+y^*y}$$

because then, in the case of the x axis, we obtain

$$\text{length}=\sqrt{a_1e^{-i\theta_1}a_1e^{i\theta_1}+0}=a_1,$$
$$=1 \text{ normalized.}$$

It is also convenient to redefine the $\langle V|$ version of vector $|V\rangle$ to be $(v_1^*,v_2^*,\ldots)$ instead of $(v_1,v_2,\ldots)$, for if we do so,

$$\langle V|V\rangle=(v_1^*,v_2^*,\ldots)\begin{pmatrix}v_1\\v_2\\\cdot\\\cdot\\\cdot\end{pmatrix}$$
$$=v_1^*v_1+v_2^*v_2+\ldots$$
$$=(\text{length of } |V\rangle)^2$$

Something else then follows:

$$\langle V|W\rangle^*=\left\{(v_1^*,v_2^*\ldots\begin{pmatrix}w_1\\w_2\\\cdot\\\cdot\\\cdot\end{pmatrix}\right\}^*$$
$$=(v_1,v_2,\ldots)\begin{pmatrix}w_1^*\\w_2^*\\\cdot\\\cdot\\\cdot\end{pmatrix}$$
$$=v_1w_1^*+v_2w_2^*+\cdots$$
$$=(w_1^*,w_2^*,\ldots)\begin{pmatrix}v_1\\v_2\\\cdot\\\cdot\\\cdot\end{pmatrix}=\langle W|V\rangle,$$

that is,

$$\langle V|W\rangle^*=\langle W|V\rangle. \tag{6}$$

Now how about operators, which you recall are each represented, in a basis, by a matrix? Operators are of course made up of eigenvectors, that is, of vectors, and vectors are now permitted to have complex components; so we can be sure that in general the matrix elements of operators are now complex numbers.

Finally, recall that the operator $T=I+\epsilon K$ causes an infinitesimal translation through space and $I+\alpha L$ causes a similar infinitesimal translation in time. Earlier we thought of ϵ and α as being real, simply because our Hilbert space was real. How are matters changed now that our Hilbert space is complex? Just as we were free to choose the overall magnitudes and units of ϵ and α earlier, so we are free to choose the phase of ϵ and α now. The operator ϵK remains fixed in its effect, so different choices of phase for ϵ amount to different choices of overall phase for K. Nothing forces us to choose the *relative* phase of ϵ and K in any specific manner, as long as the overall phase of their product remains the same.

In particular, if we choose the phase of ϵ so that ϵ is equal to a real number times i, then the left side of Eq. (4) is effectively multiplied by i, producing

$$\langle x|K|\psi\rangle=-i\hbar\frac{d}{dx}\langle x|\psi\rangle. \tag{7}$$

Similarly, if one chooses the phase of α so that α is equal to a real number times $-i$, then the left side of Eq. (5) is effectively multiplied by $-i$, producing

$$L|\psi\rangle=i\hbar\frac{d}{dt}|\psi\rangle. \tag{8}$$

Now Simplicio will want to tell me that it is K and not (K/i) that is the momentum operator P, and it is L, not (iL), that is the Hamiltonian! This should come as no surprise: If we redefine the phase of ϵ so that it is purely imaginary instead of being purely real, the phase of K will also be redefined. We'll see why these particular choices of phase for ϵ and α are especially convenient in due course.

Now let $\Omega|V\rangle=|V'\rangle$ or, explicitly,

$$\begin{pmatrix}\omega_{11}&\omega_{12}\\\omega_{21}&\omega_{22}\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}\omega_{11}v_1+\omega_{12}v_2\\\omega_{21}v_1+\omega_{22}v_2\end{pmatrix}=\begin{pmatrix}v_1'\\v_2'\end{pmatrix}$$

and let

$$\langle V'|=\langle V|\Omega^\dagger,$$

where $\Omega^\dagger$, called the adjoint of the operator Ω, is to be determined.[8] Explicitly,

$$(v_1'^*, v_2'^*) = (v_1^* v_2^*)\begin{pmatrix}\omega_{11}^\dagger & \omega_{12}^\dagger \\ \omega_{21}^\dagger & \omega_{22}^\dagger\end{pmatrix}$$

$$= (\omega_{11}^\dagger v_1^* + \omega_{21}^\dagger v_2^*,\ \omega_{12}^\dagger v_1^* + \omega_{22}^\dagger v_{22}^*)$$

or

$$\begin{pmatrix}v_1' \\ v_2'\end{pmatrix} = \begin{pmatrix}\omega_{11}^{*\dagger} v_1 + \omega_{21}^{*\dagger} v_2 \\ \omega_{12}^{*\dagger} v_1 + \omega_{22}^{*\dagger} v_2\end{pmatrix} = \begin{pmatrix}\omega_{11} v_1 + \omega_{12} v_2 \\ \omega_{21} v_1 + \omega_{22} v_2\end{pmatrix}$$

and comparing the last, we see that

$$\omega_{11}^{\dagger *} = \omega_{11}, \quad \omega_{21}^{\dagger *} = \omega_{12},$$

$$\omega_{12}^{\dagger *} = \omega_{21}, \quad \text{and} \quad \omega_{22}^{\dagger *} = \omega_{22},$$

or, generalizing, $\omega_{ij}^\dagger = \omega_{ji}^*$. I hope this repetition of standard material from an introductory quantum mechanics course does not bore you.

SIMP: Keep going.

SALV: An operator that is self-adjoint (that is, an operator for which it happens that $\Omega\dagger = \Omega$) is called a "Hermitian" operator.

We are now in a position to tie up that loose end that we left. What the operator T does is

$$T|\psi(x)\rangle = |\psi(x+\epsilon)\rangle,$$

that is, the operator T simply shifts the wave function to the left by ϵ (you may want to study Fig. 15 again to review this).

From our definition of the adjoint of an operator, we have

$$\langle\psi(x)|T^\dagger = \langle\psi(x+\epsilon)|;$$

hence,

$$\langle\psi(x)|T^\dagger T|\psi(x)\rangle = \langle\psi(x+\epsilon)|\psi(x+g\epsilon)\rangle.$$

The right-hand side, according to our new definition of the length of a vector, is just the square of the length of $|\psi(x+\epsilon)\rangle$, which is unity for a normalized vector, so

$$\langle\psi(x)|T^\dagger T|\psi(x)\rangle = 1$$

and since $\langle\psi(x)|\psi(x)\rangle = 1$ also, we conclude that

$$T^\dagger T = I.$$

Any operator having this property we will call a "unitary" operator, and unitary operators have the important property, obviously, that they leave the lengths of all vectors they act on *unchanged*. They have a second important property: The projection of one vector on another is unchanged if both vectors are acted on by a unitary operator: The projection of $|\Omega\rangle$ on the position eigenbasis $|x\rangle$, for example, is $\langle x|\Omega\rangle$. Now act on $|\Omega\rangle$ *and* on $|x\rangle$ with a unitary operator T. The two new vectors are $T|\Omega\rangle$ and $T|x\rangle$. The projection of $T|\Omega\rangle$ on $T|x\rangle$ is $\langle x|T^\dagger T|\Omega\rangle = \langle x|\Omega\rangle$ which indeed is unchanged.

This result does finally tie up that "loose end" we left: Since T is unitary, it shifts all the eigenvectors of all possible operators in such a way that their projections on the common state vector remain unchanged, which is precisely what we wanted.[9]

Having tidied up the loose end, let's move forward.

In the case of translation through time, we are seeking the quantity (we will of course name it "energy") which is conserved as a result of time-translational invariance. Now, all we know about any projected measurement of any quantity is a curve such as that in Fig. 12 (and we *still* don't know the *shape* of the curve!) In general, Fig. 12 indicates that identical experiments can give nonidentical results, which implies nonconservation because it implies that one cannot expect time-translational invariance in the observed quantity.

This looks like a pretty fundamental setback, and if we were really trying to develop quantum physics from scratch, we might go far off the track at this point and incorrectly conclude that our dashed curve in Fig. 12 must have a sharp peak, if a conserved quantity is to exist. But starting to make assumptions about the nature of the dashed curve would be contrary to the spirit of our entire program. Fortunately, there is another way, albeit not a terribly obvious one: We will focus our interest on quantities that *on average* are conserved because of the time-translational invariance of their wave functions.

The average value of a unity we call its "expectation value." Look at Fig. 12. Suppose you had a thousand particles, each in this particular state (solidifying the dashed line for a moment!), and you measured the position of every one. What would your average result be? It would be

$$\langle X\rangle = \sum_i P(x_i)x_i = \sum_i \{(x_i|\psi\rangle^*\}\{\langle x_i|\psi\rangle\}x_i,$$

where the first bit is just the usual definition of "average" and the second bit follows from our new (and final!) choice of Hilbert space structure, namely, that probability $P(x_i) = \psi^*(x_i)\psi(x_i)$; and, finally $\langle X\rangle$ is the symbol we will use to mean "expectation (average) value of position, x."

$$\therefore \langle X\rangle = \sum_i \langle\psi|x_i\rangle\langle x_i|\psi\rangle x_i,$$

where we have used Eq. (6). Recall that $X|x\rangle = x|x\rangle$, so (since, x_i, being just a number, can be moved anywhere),

$$\langle X\rangle = \sum_i \langle\psi|x_i|x_i\rangle\langle x_i|\psi\rangle$$

$$= \sum_i \langle\psi|X|x_i\rangle\langle x_i|\psi\rangle$$

$$= \langle\psi|X\left\{\sum_i |x_i\rangle\langle x_i|\right\}|\psi\rangle.$$

But

$$\sum_i |x_i\rangle\langle x_i| = \begin{pmatrix}1\\0\end{pmatrix}(10) + \begin{pmatrix}0\\1\end{pmatrix}(01)$$

$$= \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} + \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix},$$

using the rules of matrix multiplication and addition,[10] so the sum is the identity operator I, which does nothing and may therefore be removed, so

$$\langle X\rangle = \langle\psi|X|\psi\rangle.$$

There is nothing in the foregoing that depends on any special property of the position operator, so

$$\langle\Omega\rangle = \langle\psi|\Omega|\psi\rangle,$$

where Ω is any operator.

Now let H be the operator that is associated with energy; that is,

$$H|E\rangle = E|E\rangle$$

by definition (where E is the energy eigenvalue).

"Energy" we define as that quantity that is conserved because of time-translated invariance. What we mean by this is that the time-translated state should have the same

energy (or rather, average energy) as the untranslated state. Now the untranslated state is $|\psi\rangle$, and the translated state is $(I+\alpha L)|\psi\rangle = |\psi\rangle + \alpha L|\psi\rangle$ so our requirement is that

$$\{\langle\psi| + \alpha^*\langle\psi|L^\dagger\}H\{|\psi\rangle + \alpha L|\psi\rangle\} = \langle\psi|H|\psi\rangle,$$

$$\{\langle\psi| + \alpha^*\langle\psi|L^\dagger\}\{H|\psi\rangle + \alpha HL|\psi\rangle\} = \langle\psi|H|\psi\rangle,$$

$$\langle\psi|H|\psi\rangle + \alpha^*\langle\psi|L^\dagger H|\psi\rangle + \alpha\langle\psi|HL|\psi\rangle = \langle\psi|H|\psi\rangle,$$

ignoring the second-order term $(\alpha^*\alpha)$, or

$$\alpha^* L^\dagger H + \alpha HL = 0$$

or

$$HL = -(\alpha^*/\alpha)L^\dagger H$$

and so

$$HL|E\rangle = -(\alpha^*\alpha)L^\dagger H|E\rangle$$

but

$$H|E\rangle = E|E\rangle$$

so

$$H\{L|E\rangle\} = -(\alpha^*/\alpha)E\{L^\dagger|E\rangle\}.$$

Now $I+\alpha L$ is unitary (for the same reason that $T = I+\alpha K$ is unitary). So

$$(I+\alpha^* L^\dagger)(I+\alpha L) = I$$

or

$$I + \alpha^* L^\dagger + \alpha L = I$$

or

$$L^\dagger = -(\alpha/\alpha^*)L \tag{9}$$

and our previous equation becomes

$$H\{L|E\rangle\} = E\{L|E\rangle\}.$$

The conclusion from our last equation is that the operator H, acting on $L|E\rangle$, stretches $L|E\rangle$ by a factor E. So we recognize $L|E\rangle$ as being an energy eigenket. That is, $|E\rangle$ is an energy eigenket, and when the operator L acts on $|E\rangle$ it produces an energy eigenket $L|E\rangle$. But the simplest operator that acts on $|E\rangle$ and produces an energy eigenket is H.[11]

And so we have found the operator H. It is L.

The same kind of argument applies to translation through space, meaning that P can be identified with K. So if we accept the idea that L is H and K is P, and if we choose the phases of ϵ and α as discussed above Eqs. (7) and (8), then

$$\langle x|P|\psi\rangle = -i\hbar\frac{d}{dx}\langle x|\psi\rangle \tag{10}$$

and

$$H|\psi\rangle = i\hbar\frac{d}{dt}|\psi\rangle, \tag{11}$$

which we recognize from quantum mechanics.

Equation (10) states that in the x basis, the operator P is

$$P \rightarrow -i\hbar\frac{d}{dx}.$$

Now let's project the momentum eigenvalue equation

$$P|p\rangle = p|p\rangle$$

onto the position basis, i.e., form

$$\langle x|P|p\rangle = p\langle x|p\rangle$$

or

$$-i\hbar\frac{d\psi_p(x)}{dx} = p\psi_p(x),$$

where $\psi_p(x) \equiv \langle x|p\rangle$.

The solution to this differential equation is $\psi_p(x) = \langle x|p\rangle = [1/(2\pi\hbar)^{1/2}]e^{ipx/\hbar)}$ where the $1/(2\pi\hbar)^{1/2}$ is customary normalization (you may verify this solution by substitution), and where, apparently, p is any number at all—well, no; p must be real. For if it were imaginary, that is, $p = ib$, then

$$\langle x|p\rangle = [1/(2\pi\hbar)^{1/2}]e^{-bx/\hbar},$$

which goes to *infinity* for either large positive, or large negative x (depending on the sign of b); and no $|p\rangle$ can have a projection on an $|x\rangle$ that is greater than 1, on logical grounds; so on these logical grounds, p is real. The eigenvalues of p are any real number!.

So the projections of a typical momentum eigenket $|p\rangle$, on a typical position eigenket $|x\rangle$, is

$$\langle x|p\rangle = \frac{1}{(2\pi\hbar)^{1/2}}e^{ipx/\hbar}$$

$$= \frac{1}{(2\pi\hbar)^{1/2}}\left(\cos\frac{px}{\hbar} + i\sin\frac{px}{\hbar}\right),$$

which is never zero, for any choice of real numbers x and p; that is, we have *finally* (thank heavens!) demonstrated that we are driven to "possibility 3."

SIMP: How does that show that we are in a "possibility 3" situation?

SAGR: It shows that *no* position eigenket is perpendicular to *any* momentum eigenket. (If one were, you could rotate around the momentum eigenket until your position eigenket was coincident with any one of the infinitely many momentum eigenkets that are (all) perpendicular to a given momentum eigenket, and you would see that you were in a "possibility 1" or "possibility 2" situation.)

SALV: Correct. Now let us return to the question of *why* we chose the phase of ϵ and α so that we were led to Eqs. (7) and (8), and thus to the familiar equations (10) and (11). What are the consequences of making another choice of phase? We know that the eigenvalues of P and H represent possible values of momentum and energy, respectively, so they must all be real. An operator whose eigenvalues are all real must be

SIMP: Hermitian, I know that!

SALV: Right! Now, note that Eq. (9) implies that L (which we are identifying with H) will only be Hermitian if the phase of α is $\pm i$. Similarly, K will only be Hermitian if the phase of ϵ is $\pm i$. Thus the identification of L with H, and K with P, only works if the phases are chosen this way. The choice of signs is purely conventional, though if we *do* choose the phase of α to be $-i$ and the phase of ϵ to be $+i$, the phase velocity of the wave function for a free particle is conveniently in the $+x$ direction when E and p are both positive. We are done.

SAGR: Note quite! Suppose there are *two* possibilities (for example, the familiar case, "the particle might have passed through slit A; or slit B"). How do we calculate the final probability?

SIMP: Let me answer that! The final probability is a physical probability, so underlying it is, necessarily, a complex probability amplitude. Once we have that amplitude, we have the probability! So the question is, how do we calculate the amplitude, from the probabilities and/or ampli-

tudes of the two possibilities? If we were to add the *probabilities*, we would always get a real number, so our final amplitude would not be complex as we know it must be; besides, it seems silly to add *probabilities* to get an *amplitude*. So I guess (since I must assume that the final amplitude depends *somehow* on what has gone before) that the final amplitude is the sum of the amplitudes for the two possibilities. Surely I am right?

SALV: You are right. That fact leads to many interesting consequences that do *not* appear in the case of the nonphysical "probabilities" that we are accustomed to in daily life (the probability of dealing an ace is not $\frac{1}{13}$, it is 0 or 1 depending on whether or not an ace is top card. It is only our ignorance of the facts that simulates a probabilistic situation. Again, the probability of getting heads is not $\frac{1}{2}$ when you toss a coin, it is 0 or 1 depending on exactly how you toss the coin. In this case it is lack of skill, not ignorance, that simulates a probabilistic situation).

SAGR: So, by insisting that momentum be the quantity conserved as a result of translational invariance, we find that momentum stands in a "possibility 3" relationship to position, which in turn implies that position will be uncertain if we measure momentum and vice versa. This is the core of the Heisenberg uncertainty principle!

SALV: Yes, and that is an exciting conclusion. I am sorry our result for the eigenvalues of momentum turned out to be so dull (*any* real number). But you can get a terrifically exciting result if you go on now to angular momentum—the eigenvalues are quantized! Shankar works it all out very nicely.[12] Also Shankar shows how Hamilton's equations (and hence $\mathbf{F} = m\mathbf{a}$) follow, in a certain limit, from quantum mechanics.[13]

SIMP: But Salviati, this is sad! You have shown that quantum mechanics is almost a triviality! The universe has lost its magic and its mystery!

SALV: Simplicio, you can't have it both ways. You started[3] by complaining about the unreasonableness of the universe, and now that I have shown that it is remarkably reasonable, you complain that the universe is too dull. There's no pleasing you![14]

ACKNOWLEDGMENTS

I thank W. Doyle Evans for hospitality. This work was partially supported by the United States Department of Energy, and by a NASA grant to the Johns Hopkins University.

[1] R. H. Dicke and J. P. Wittke, *Introduction to Quantum Mechanics* (Addison-Wesley, Reading, MA, 1960).

[2] R. Shankar, *Principles of Quantum Mechanics* (Plenum, New York, 1980).

[3] R. C. Henry, "Special relativity made transparent," Phys. Teach. **23**(9), 536 (1985).

[4] The sophisticated reader will recognize that we are constructing a grand Hilbert space for the particle within which all the individual eigenbases reside.

[5] J. B. Marion, *Classical Dynamics of Particles and Systems* (Academic, New York, 1970).

[6] "Braket," get it?

[7] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1970), p. 51.

[8] This causes a problem for physicists, who on entering a low-class bar, imagine that the person behind the counter is the adjoint of an operator. He is not; he is the operator of a joint.

[9] If this hadn't worked, it wouldn't have mattered. You could just reach in "by hand" and do manually what we are asking T to do. (Hilbert space is not part of nature, it is a physicist's bookkeeping device.)

[10] This two-dimensional example generalizes easily to an infinite number of dimensions.

[11] L could also be any *function* of H, of course, but that also will cause translation through time, as would any function of L.

[12] Reference 2, Eq. (12.5.15a), p. 332.

[13] Reference 2, Eqs. (6.10) and (6.11), p. 192.

[14] A brief account of the present work appears as "Teaching QM: True, trivial, inevitable," in *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafotas (Kluwer, Dordrecht, 1989), pp. 175–177.

An experiment to study localized excitations—Nonpropagating hydrodynamic solitons

Erik Winkler and Junru Wu

Department of Physics, University of Vermont, Burlington, Vermont 05405

(Received 15 July 1989; accepted for publication 10 April 1990)

A simple experiment designed to study nonpropagating hydrodynamic solitons of (0, 1) and (0,2) modes is discussed. A brief review of the properties of the soliton is provided.

I. INTRODUCTION

In 1834, the Scottish scientist John Scott Russel made the first documented observation of a solitary wave on the surface of the water in a canal. He described the wave as "...a rounded, smooth and well defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed."[1] He also coined the term "solitary wave" in his paper "Report on Waves" (1844). It was not until about 60 years later that a theory of the solitary wave phenomenon was developed. Two Dutch scientists, Korteweg and de Vries, derived the