The "Past" and the "Delayed-Choice" Double-Slit Experiment

John Archibald Wheeler

University of Texas
Center for Theoretical Physics
Austin, Texas 78712

Partway down the optic axis of the traditional double-slit experiment stands the central element, the doubly-slit screen. Can one choose whether the photon (or electron) *shall have* come through both of the slits, or only one of them, after it has *already* transversed this screen? That is the new question raised and analyzed here.

Known since the days of Young is the possibility to use the receptor at the end of the apparatus to record well defined interference fringes. How can they be formed unless the electromagnetic energy has come through both slits? In later times Einstein noted that in principle one can determine the lateral kick given to the receptor by each arriving quantum. How can this kick be understood unless the energy came through only a single slit?

Einstein's further reasoning as reported by Bohr (1) is familiar. Record both the kicks and the fringes. Conclude from the kicks that each quantum of energy comes through a single slit alone; from the fringes, that it nevertheless also comes through both slits. But this conclusion is self-contradictory. Therefore quantum theory destroys itself by internal inconsistency.

Bohr's reply (1) has become by now a central lesson of quantum physics. One can record the fringes or the kicks but not both. The arrangement for the recording of the one automatically rules out the recording of the other. The quantum has momentum p, de Broglie wave length $\lambda = h/p$, and reduced wave length $\lambda\!\!\!^{-} = \hbar/p$. To record for it well defined interference fringes one must fix the location of the

Copyright © 1978 by Academic Press, Inc.
All rights of reproduction in any form reserved.
ISBN 0-12-473250-X

receptor within a latitude

$$\Delta y < (\text{fringe spacing})/2\pi = (L/2S)\lambda\!\!\!{}^{-}. \qquad (1)$$

To tell from which slit the quantum of energy arrives one must register the transverse kick it gives to the receptor within a latitude small enough to distinguish clearly between a momentum $p = \hbar/\lambda\!\!\!{}^{-}$ coming from below, at the inclination S/L, and a momentum coming from above at a like inclination; thus,

$$\Delta p_y < (S/L)(\hbar/\lambda\!\!\!{}^{-}). \qquad (2)$$

However, for the receptor simultaneously to serve both functions would be incompatible with what the principle of indeterminancy has to say about receptor dynamics in the Y-direction,

$$\Delta y \Delta p_y > \hbar/2. \qquad (3)$$

Not being able to observe simultaneously the two complementary features of the radiation, it is natural to focus on the one and forego examination of the other. Either one will insert the pin through the hole shown in Fig. 1. It will couple the receptor to the rest of the device. It will give the receptor a well defined location. Then one will be able to check on the predicted pattern of interference fringes. Or one will remove the pin. Then one can measure the through-the-slot component of momentum of the receptor before and after the impact of the quantum. Then one will say that one knows through which slit the energy came.

Pin in or pin out: when may the choice be made? Must it be made before the quantum of energy passes through the doubly slit screen? Or may it be made after? That is the central question in this paper as that question first seems to impose itself. However, a closer look shows that the measurement of transverse momentum kick, in principle conceivable, is practically almost out of the question. Therefore it is appropriate to alter the idealized experiment before taking up the question of "before" *versus* "after." What is the difficulty and what is the change?

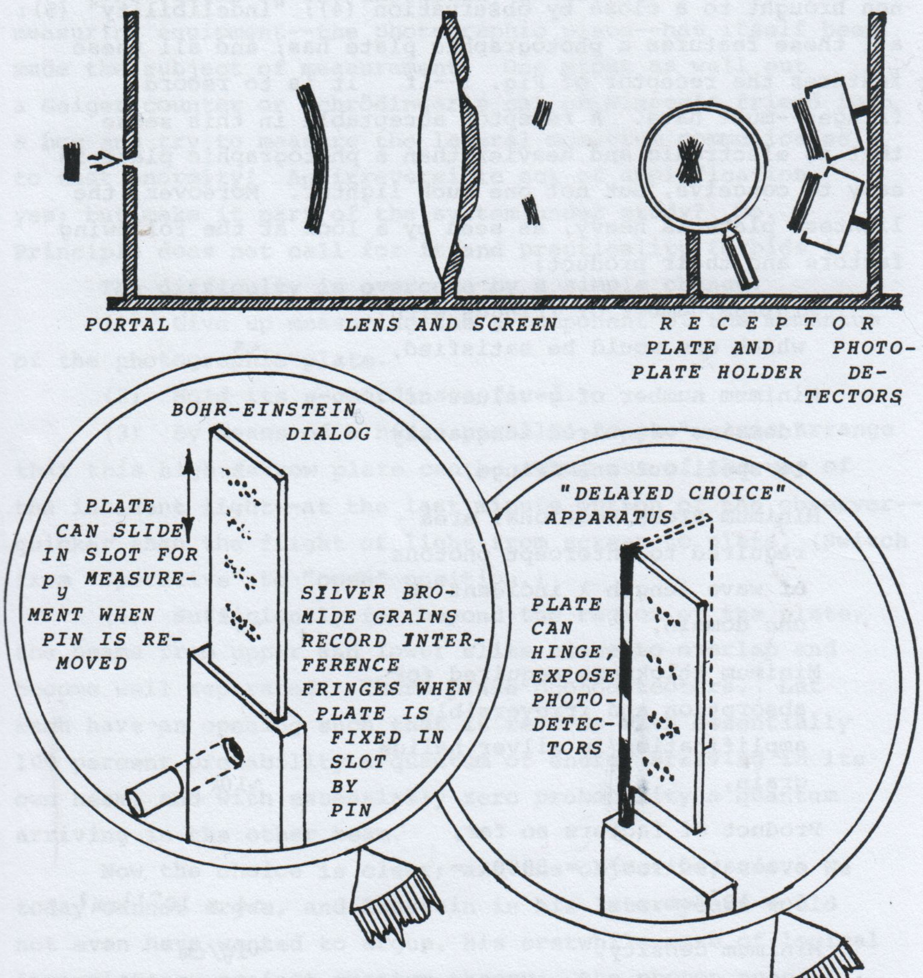

Fig. 1. Top: Idealized double-slit experiment. Distance of each slit from optic axis, S; from photographic plate, L. For simplicity, details of the plate and plate holder are omitted from the circle encompassed by the magnifying glass and are presented below, magnified and in perspective. Lower left: The version of the Bohr-Einstein dialog. The plate catches every photon. It registers precisely the y-coordinate of impact or the y-component of impulse delivered--but does not and cannot do both. Omit the photodetectors. Lower right: The present "delayed choice" version. Include the photodetectors. One or other of them is sure to catch the quantum of energy when the plate is swung aside. Whether to expose the plate or expose the photodetectors, whether thus to infer that a single quantum of energy <u>shall have</u> gone through both slits in the screen or only one, is subject to the free choice of the observer after the energy has <u>already</u> traversed the screen.

An irreversible act of amplification (2, 3); a phenomenon brought to a close by observation (4); "indelibility" (5): all these features a photographic plate has; and all these features the receptor of Fig. 1--if it is to record fringes--must have. A receptor acceptable in this sense that is electronic and heavier than a photographic plate is easy to conceive, but not one much lighter. Moreover, the lightest plate is heavy, as seen by a look at the following factors and their product:

Minimum number of fringes with which one would be satisfied,	~ 5
Minimum number of y-values or "domains" required adequately to spell out one fringe,	$\sim 2\pi$
Minimum cross sectional area required to intercept photons of wave length λ incident on one domain,	$\sim \lambda^2$
Minimum thickness required for absorption and irreversible amplification in silver halide grain,	$\sim 10\lambda$
Product of factors so far, evaluated for $\lambda = 5000A = 5 \times 10^{-5}$cm,	$\sim 4 \times 10^{-11}$cm^3
Minimum density,	~ 1g/cm^3
Minimum mass,	$\sim 4 \times 10^{-11}$g
Uncertainty, $(MkT)^{1/2}$, in momentum of so great a mass by reason of thermal agitation at $T = 300$deg$K$,	$\sim 10^{-12}$g cm/s
Lateral momentum of incident photon, say $10^{-2}\hbar\omega/c$ or less,	$\sim 10^{-24}$g cm/s
Factor by which this lateral momentum fails of being measurable,	$\sim 10^{12}$

The origin of the difficulty is not far to seek. The measuring equipment--the photographic plate--has itself been made the subject of measurement. One might as well put a Geiger counter or Schrödinger's cat or Wigner's friend into a box and try to measure the lateral momentum communicated to that enormity! An irreversible act of amplification, yes; but make it part of the system under study? No. Principle does not call for it and practicality forbids it.

The difficulty is overcome by a simple change:

(1) Give up measuring the y-component of the momentum of the photographic plate.

(2) Hold its y-coordinate fixed.

(3) By means of a hinge parallel to the y-axis arrange that this high narrow plate can be swung out of the way of the incident light--at the last minute option of the observer--quicker than the flight of light from screen to plate. (Switch from "operative" to "open" position.)

(4) Sufficiently far beyond the region of the plate, the beams from upper and lower slits cease to overlap and become well separated. There place photodetectors. Let each have an opening such that it records with essentially 100 percent probability a quantum of energy arriving in its own beam, and with essentially zero probability a quantum arriving in the other beam.

Now the choice is clear; and the objective, too. We today cannot argue, and Einstein in his later years would not even have wanted to argue, his erstwhile case of logical inconsistency against quantum theory: the photon goes through both slits, as evidenced in interference fringes, and yet simultaneously through only one, as evidenced in lateral momentum kick. Choose we know we must between the two complementary features open to study; and choose we do, by putting the plate athwart the light or turning it out of the line of fire. In the one case the quantum will transform a grain of silver bromide and contribute to the record of a two-slit interference fringe. In the other case one of the two counters will go off and signal in which beam--and therefore from which slit--the photon has arrived.

In our arrangement the photographic plate registers only the point of impact of a photon. In the earlier idealized experiment it could additionally (Einstein) or alternatively (Bohr) record the transverse momentum delivered by the impact. We have assigned the two distinct kinds of measurement to two distinct kinds of register. We have demoted the plate from a privileged status. That demotion is irrelevant to any question now at issue. Equally irrelevant is the different distance--and time of flight--from entry portal to plate, or photodetector, according as the one or other register is exposed. But the essential new point is the timing of the *choice*--between observing a two-slit effect and a one-slit one--until after the single quantum of energy in question has *already* passed through the screen.

Let the reasoning be passed in review that leads to this at first sight strange inversion of the normal order of time. Then let the general lesson of this apparent time inversion be drawn: "No phenomenon is a phenomenon until it is an observed phenomenon." In other words, it is not a paradox that we choose what *shall* have happened after "it has *already* happened." It has not really happened, it is not a phenomenon, until it is an observed phenomenon.

ANALYSIS OF THE "DELAYED CHOICE" EXPERIMENT

Whatever we now do to spell out the otherwise idealized experiment, we will leave idealized its most unusual features: the "swinging door photographic plate." That term includes the arrangement, whatever it may be,

(1) for a last minute choice, to swing the plate aside or leave it athwart the beam, after the arriving energy has already traversed the doubly slit screen, and

(2) for completion of that movement before the energy arrives at the plate. In practice it will be more reasonable to swing the beam than swing the plate. Fix the plate. Halfway from screen to it, position a Kerr cell. Apply to it a positive or a negative voltage (6) according as one wishes to record fringes on the plate, or register "which beam" on a counter. Or, still better, Manfred Fink suggests,

replace the experiment with the photon by an experiment with an electron. Then the last minute deflection of the electron beam can be accomplished by a localized magnetic field centered between screen and plate. One or another of these arrangements to swing the beam will be understood hereafter to apply in practice when in principle we speak of swinging the plate.

Now we turn to other requirements and presuppositions of the idealized experiment. First among them is the "window in time" that admits the incoming radiation to the entry portal. Well known windows today tailor the pulse of radiation that implode a pellet of *D-T* (7). The window in the present device we take to be "Gaussian in time." The incident amplitude is given by

$$E(x,t) = Re\ const\ exp\ [ik_o(x-t) - (x-t)^2/2\tau^2] \qquad (4)$$

Here the origin in x and origin in time are so chosen as to bring the center of the wave train to the doubly slit screen, $x = 0$, at $t = 0$; and the symbol *Re* stands for "real part of." In this expression and hereafter three simplifications are made. First, time is expressed in cm of "light travel time;" that is, in the same geometrical units that are employed for distance. Second, all reference is eliminated to polarization. Only that scalar function is considered which multiplies the familiar polarization factor in a full-dress treatment of a propagating electromagnetic wave. This simplification will be legitimate provided that we reduce sufficiently the angles of diffraction of the beams, and the angle of their crossing, and the consequent departures from the forward direction.

The third simplification requires more discussion. One part of the factor that multiplies the scalar amplitude (4) and subsequent expressions for diffracted waves is a quantum-field-theory creation operator, a^{+}. It will be understood so unmistakably to be a part of all such expressions that it will never even be written down. The rationale for this omission was given originally by Heisenberg and Pauli (8) and was subsequently spelled out by Breit (9). In the present context that rationale goes as follows: The wave in question is to be Fourier-analyzed into elementary waves. Each elementary wave is to bear its own creation operator. The probability for a particular silver halide grain to be transformed, or for a particular point in the photodetector to initiate an electron cascade, is evaluated by the following sequence of mathematical operations. Take the square of the amplitude of the elementary wave at the point in question. Multiply it by the number operator $N = a^{\dagger}a$ for the elementary wave in question. Sum over all elementary waves to get the desired probability for bringing about a local excitation. But this calculation of the probability is closely equivalent in its result to a calculation made along semiclassical lines, in which no mention is made of creation and annihilation operators: Take the electric field as calculated classically as a function of position and time. Square. Multiply by $2c/8\pi$ to get energy flow. Integrate over time to get energy per unit area. Divide by the average energy per photon, $\hbar c k_o$. The quotient is very nearly the probability that a photon will strike a unit once normal to the direction of flow. Close equivalence can be converted to exact equivalence by making the calculation more elaborate. For this purpose the electric field $E(t)$ at the place in question should be Fourier analyzed before squaring; and the squared Fourier amplitude, $|E(k)|^2$, should be divided by the quantum energy, $\hbar c k$, appropriate to the frequency in question before integrating--this time over frequency, rather than time--to determine the expectation value of the number of photons that arrive. However, the length τ of the wave train is envisaged here to be long in comparison with one vibration period of the radiation; or more precisely, we take the dimensionless quantity $k_0\tau$ to be large compared to unity:

$$N \equiv k_o\tau >> 1. \tag{5}$$

Therefore the spread of frequencies, $\Delta k \sim 1/\tau$, in the wave packet is small compared to the frequency k_o itself. For this reason we dispense with the complication of Fourier analysis, to make manageable the mathematics of which would force us to a more sophisticated profile for the wave packet.

In summary, our simplification gives for the probability per unit area for impact of a photon the expression

$$w = (1/4\pi\hbar k_o) \int E^2(position,\ t)dt. \qquad (6)$$

More than any other design feature, the delayed-choice double-slit experiment calls for beams directed from the two slits towards the plane of interference and--when the photographic plate normally located there is swung aside--continuing on to make clean entry into well separated photodetectors. It is appropriate to begin with the optics of a single beam. For simplicity let it be conceived to come from a slit centered on the optic axis, $y = 0$, and let it run in the x-direction. Later on the slit location can be moved and the beam direction changed to describe the beams actually desired in the double-slit experiment.

It is easy to complicate the double-slit interference pattern. For that purpose it is enough to have a complicated single-slit diffraction pattern, and let the waves from two such slits interfere. In contrast, we want fringes, when fringes are observed, to originate in, and serve as evidence for, interference. Therefore we rule out any single-slit design that gives a diffraction pattern with a zero in it. We propose to meet this requirement by operating with a Gaussian diffraction pattern, obtained from a slit with a Gaussian distribution of transmissivity,

$$(transmission) = exp(-y^2/2a^2). \qquad (7)$$

Lenses have been used in recent times with a distribution of transmissivity designed to eliminate any zero in the diffraction pattern (10), and it takes no feat of the imagination to think of applying the same technique of "apodisation" to one slit, and by extension to both slits.

The pattern of diffraction from the "Gaussian slit" defined by (7) we evaluate by decomposing the incoming wave train (4) into monochromatic constituents of the form $exp\ ik(x-t)$ with a Fourier amplitude

$$[\tau/(2\pi)^{\frac{1}{2}}]exp[-\tau^2(k-k_o)^2/2]dk, \qquad (8)$$

determining in the familiar Fresnel approximation the diffraction pattern produced by each, and recomposing these separate contributions. From the portion $d\bar{y}$ of the wave front of a monochromatic wave of wave number k and unit amplitude at location $x = 0$ the Fresnel prescription gives us at any distance $x >> \lambda\!\!\!^{-} = k^{-1}$ and at any elevation y above the optic axis the contribution

$$(2\pi i \lambda\!\!\!^{-} r)^{-\frac{1}{2}} e^{ikr} d\bar{y}; \qquad (9)$$

or, at the level of approximation at which we work,

$$(2\pi i \lambda\!\!\!^{-} x)^{-\frac{1}{2}} exp[ikx + ik(y-\bar{y})^2/2x]d\bar{y}. \qquad (10)$$

For a detailed analysis we cannot overlook the presence of a third dimension, z, perpendicular to the (x,y)-plane of Fig. 1. We tailor the transmissivity of each slit in its dependence upon z according to a Guassian law, $exp(-z^2/2b^2)$, in order to avoid any zeros in the diffraction pattern. However, the Gaussian widths, b and a, in the z- and y- directions are very different in order of magnitude ($b>>a$). Thus the angle of diffraction in the z-direction, $\sim\lambda\!\!\!^{-}/b$, may be considered to be negligible in comparison to the angle of diffraction in the y-direction $\sim\lambda\!\!\!^{-}/a$.

More important than any diffraction-induced spreading of the beam in the z-direction is the lens-induced convergence of the beam in that direction. This effect allows one to use a truly narrow protographic plate, as indicated schematically in the inset diagram at the lower right in Fig. 1. However, it is appropriate not to make the plate location, $x = L$, identical with the focal point, $x = F$, of the lens. The reason is simple. The convergence in the z-direction, coming to a head in the immediate neighborhood of the focus, produces there a sudden change in phase of $\pi/2$, analogous to the well known change in phase of π that occurs at a "two-way" focus (11, 12, 13). To operate too close to this point is to run the risk that some parts of the photographic plate will cross the place of change of phase, giving rise to anomalies in the interference pattern. Granted the elementary precaution needed to insure against this effect, $|L-F|>>\lambda F^2/b^2$, we neither want to give nor will give any more attention to the development of the beam in the z-direction. We shall idealize it as if it developed entirely in the (x,y)-plane.

The diffraction-induced spreading of the beam in the (x, y)-plane we compensate insofar as feasible with a lens-induced convergence. That means that the wave can be regarded as starting at the plane of the slit, $x=0$, with not only its amplitude dependent upon y, as described by the Gaussian of (7), but also its phase dependent upon y, as appropriate for a wave that would come to a focus at $x = F$; thus,

$$(amplitude) = \exp[-(ik/2F)y^2-(1/2a^2)y^2]. \quad (11)$$

Composing this elementary amplitude with the Fresnel propagator (10) and the Fourier coefficient (8) we have for the beam amplitude at the point (x,y) at the time t the expression

$$(beam\ amplitude) = (\tau/2\pi)\iint (k/ix)^{\frac{1}{2}}\exp\{-ikt-(\tau^2/2)(k-k_o)^2 + ikx+(ik/2x)(y-\bar{y})^2-(ik/2F)\bar{y}^2 - (1/2a^2)\bar{y}^2\}d\bar{y}dk. \quad (12)$$

In the exponent, in the coefficients of $(y-\bar{y})^2$ and $\bar{y}^2$, and also in the multiplying coefficient before the exponential, we replace k by $k_o = \lambda^{-1}$ to make the integral manageable. The resultant error is readily estimated. In the case of interest two conditions will be fulfilled:

(1) The focal length will be of the order of the distance x to the plate.

(2) The contributions to the beam width from diffraction, $x\lambda/a$, and from the original (focusing neglected) width of the beam, a, will be comparable; that is, we will operate at a distance x of the order of a^2/λ. Under these conditions we find that the error in the exponent is of the order of $\lambda/\tau \sim 1/N$. This correction can be neglected when the number of waves in the incident wave packet is sufficiently great. In this approximation the integration (12) gives the result,

$$(beam\ amplitude) = Re(1-x/F+i\lambda\!\!\!^{-}x/e^2]^{-\frac{1}{2}} exp\{ik_o(x-t)-(x-t)^2/2\tau^2$$

$$-\frac{(y^2/2)}{a^2(1-x/F)^2+(x\lambda\!\!\!^{-}/a)^2}$$

$$+ (iky^2/2)\frac{x(F\lambda\!\!\!^{-}/a^2)^2-(F-x)}{(F-x)^2+(xF\lambda\!\!\!^{-}/a^2)^2}\}. \qquad (13)$$

Every term in (13) has its simple interpretation. The first two terms in the exponent come over unaltered from the phase and the profile of the primary radiation. The next term in the exponent shows that the parameter describing the width of the beam in the y-direction is the square root of the sum of the squares of two separate width parameters. Of these one describes a diffraction-induced spreading of the beam by the amount *(distance)·(angle)* = $x\lambda\!\!\!^{-}/a$. The other describes a focus-induced convergence of the original beam width a to a figure $(1-x/F)a$. Both effects also show up in the multiplicative factor before the exponential. It describes the change of amplitude caused by the focussing and diffraction of the beam. In agreement with these considerations, the square of its absolute value, the intensity, is correctly given by dividing the original width parameter of the beam, a, at $x = 0$ by the value of the width parameter at $x = x$.

The phase of the amplitude factor evidently changes by $\pi/2$, as anticipated, when x passes through $x = F$. However, the distance over which this change of phase comes about is not of the order of the reduced wave length, $\lambda\!\!\!^{-}$, itself, as one might at first have expected. It has the much larger value $(F/a)^2\lambda\!\!\!^{-}$. What this distance means is best seen by shifting attention from the slit, $x = 0$, to the place, $x = x_{min}$, where the beam has its minimum width:

$$(width)^2 = a^2(1-x/F)^2+(x\lambda\!\!\!^{-}/a)^2,$$

$$(F^2/2)d(width)^2/dx = -a^2(F-x)+x(F\lambda\!\!\!^{-}/a)^2,$$

$$(width)^2_{min} = F^2\lambda\!\!\!^{-2}/[a^2+(F\lambda\!\!\!^{-}/a)^2],$$

at

$$x_{min} = a^2F/[a^2+(F\lambda\!\!\!^{-}/a)^2]. \qquad (14)$$

Viewed from x_{min} as origin, the beam, tracked either forward or backward, is propagating and diffracting as if it were a nearly plane wave emerging from an aperture of opening $(width)_{min}$. Therefore at a distance d from this virtual aperture there are two contributions to the width. One is this aperture opening itself. The other is diffraction, with a diffraction angle $\sim(reduced\ wave\ length)/(aperture) = ƛ/(width)_{min}$. Thus the beam width is given by the formula

$$(width)^2 = (width)^2_{min} + d^2ƛ^2/(width)^2_{min}. \qquad (15)$$

As a check, we can verify that this widening is just enough at the distance $d = x_{min}$ to send the beam back through the original slit $(width = a)$.

The transition from the "inner region" of propagation of the beam as a nearly plane wave of nearly constant width to the "outer region" of propagation as a cylindrically converging or diverging wave occurs at that distance from the point of minimum width where the two contributions of (15) to the width are of the same order of magnitude,

$$d = \frac{(F/a)^2ƛ}{1+(Fƛ/a^2)^2}. \qquad (16)$$

The "correction factor" in the denominator of (16) is of the order of unity under conditions where the diffraction-induced broadening of the beam does not substantially exceed the slit width, a, itself. Then in the present Huygens approximation, the calculated length of the inner region of propagation is of the same order as the calculated length, $\sim(F/a)^2ƛ$, of the region in which the phase of the wave rises by the amount $\pi/2$. It will be convenient for the subsequent consideration of the idealized experiment to regard the photographic plate as located in this inner region, $d \ll (F/a)^2ƛ$, and yet staying away from the focus in the z-direction by the already specified margin, $d \gg (F/b)^2ƛ$.

These two foci, for the y-convergence and the z-convergence, are not the same. This difference comes about, not because of any imperfection in the lens, but because the y-diffraction caused by the narrow slit of width a fights against the normal geometrical-optics convergence. Thus the beam reaches its minimum extension in the y-direction, not at the distance F from the lens, but at the shorter distance x_{min} of (14). Adopting for the slit width here and hereafter a figure

$$a = (F\bar{\lambda})^{\frac{1}{2}}, \tag{17}$$

we have

$$x_{min} = F/2. \tag{18}$$

The final term in the exponent of (13) tells us that the wave crests and troughs in the beam at the distance $\bar{x}$ in their progression towards the right are warped into circular arcs centered on a point, $x = \bar{x}+D$, located to the distance D to the right of $\bar{x}$, where

$$\frac{1}{D} = \frac{(F-\bar{x})-\bar{x}(F\bar{\lambda}/a^2)^2}{(F-\bar{x})^2+\bar{x}^2(F\bar{\lambda}/a^2)^2} \tag{19}$$

In the limit of geometrical optics, where diffraction is negligible ($\bar{\lambda}\to 0$), (19) gives $D = (F-\bar{x})$, so in this limit the point of convergence is located at $x = \bar{x}+D = F$, as expected. However, to operate as we propose, where the wave fronts are plane, that is to say, centered at infinity, we look for the $\bar{x}$ value that annuls the numerator of (19), thus

$$\bar{x} = F/[1+(F\bar{\lambda}/a^2)^2]. \tag{20}$$

It is not surprising that this point is identical with the point, $x = x_{min}$, where the wave front achieves its minimum width. This is one more instance of a general theorem [see for example reference (14) for a survey and citations of the literature] relating width of the wave with radius of curvature,

$$d(width)/(width) = -(dx)/D, \tag{21}$$

a theorem that one verifies explicitly by comparing (21) with (14).

There is no better epitome of these considerations on beam width, beam phase, diffraction, and curvature of the wave front than the factor

$$[(1-x/F)a+ix(\lambda\!\!\!^{-}/a)]^{-\frac{1}{2}} \tag{22}$$

which (along with the constant $a^{\frac{1}{2}}$) multiplies the exponential in expression (13) for the beam amplitude. Out of (22) are to be read out immediately the $\pi/2$ change of phase going through the region of convergence and the two separate contributions to the square of the beam width; and from this beam width the radius of convergence or divergence of the waves follows directly out of (21).

In summary, we specify the reduced wave length $\lambda\!\!\!^{-} = \lambda/2\pi = 1/k_o$ of the incident wave train and the parameter $\tau = N\lambda\!\!\!^{-}$ that measures the length of the incident wave train. We specify the focal length F of the lens. We fix the slit width a to be $(F\lambda\!\!\!^{-})^{\frac{1}{2}}$. We locate the photographic plate at the point $x = F/2$ where the beam has reached its minimum width

$$w = a/2^{\frac{1}{2}} = (F\lambda\!\!\!^{-}/2)^{\frac{1}{2}} \tag{23}$$

and where the wave fronts are planar. Well past this region of convergence the wave spreads out with an angular width parameter

$$\theta_{diffraction} = \lambda\!\!\!^{-}/w = 2^{\frac{1}{2}}\lambda\!\!\!^{-}/a = (2\lambda\!\!\!^{-}/F)^{\frac{1}{2}}. \tag{24}$$

We now replace the one slit beaming energy along the optic axis by two slits, one sending a wave from an elevation S^* above the optic axis on a downward slant, the other beaming a wave on an upward slant from a point S^* units of distance below the optic axis. Were each wave converging as predicted by geometrical optics, the two would come to a common focus at $x = F$. In actuality each comes to its narrowest width and to planarity at $x = F/2$. There the center of the one wave lies above the center of the other wave by half the original $2S^*$ separation. We want to make the wave centers coincide. Only so will we secure optimum interference between the two waves. For this purpose we saw through the lens on the plane $y = S^*/2$, also on the plane $y = -S^*/2$. We throw away the intervening slab of glass. We glue back together the upper and lower portion of the lens. Now the slit centers have the elevations $\pm S \equiv S^*/2$ with respect to the optic axis, and optimum interference is achieved.

The difference in angle of fire of the two beams,

$$\theta_{fire} = \left(\begin{matrix} difference \\ in\ elevation \end{matrix}\right) \Big/ \left(\begin{matrix} distance\ to \\ overlap\ of\ wave\ centers \end{matrix}\right)$$

$$= (2S)/(F/2) = 4S/F, \tag{25}$$

is not all pure gain in cleanly separating the beams as they pass the plane $x = F/2$ of interference (and potential location of the photographic plate) on their way to the two photodetectors. Only when θ_{fire} exceeds the diffraction angle θ_{diff} of (24) by a very substantial factor, say 20-fold, will we get from the strong Gaussian fall off of the diffraction pattern the assurance we want against a contribution from either beam to a count in the wrong photodetector. With

$$\theta_{fire} = 20\ \theta_{diffraction} \tag{26}$$

we have for the required off-axis slit distance the value

$$S = (F/4)\theta_{fire} = 5\theta_{diffraction}F = 5(2\lambda\!\!\!^{-}F)^{\frac{1}{2}} = 5\cdot 2^{\frac{1}{2}}a. \tag{27}$$

In other words, a suitable separation between the centers of the two slits is $10\cdot 2^{\frac{1}{2}}a$ or about 14 times the Gaussian width parameter of either slit individually.

Separation of the beams well past $x = F/2$ being thus assured, we turn back to the details of the interference pattern at $x = F/2$ itself. Were one of the beams directed along the optic axis, it would admit in this region of its development the simple mathematical expression

$$(electric\ field) = E_o \cos k_o(x-t)\ exp[-(y^2/2w^2)-(x-t)^2/2\tau^2], \tag{28}$$

with $w^2 = a^2/2 = F\lambda\!\!\!^{-}/2$. We get this result by substituting into our original formula (13) the expressions $x = (F/2) + \bar{x}$ and $t = (F/2) + \bar{t}$. These substitutions put the origin of the coordinate $\bar{x}$ at the plane of interference and the origin of the time variable $\bar{t}$ at the instant when the peak of the wave train arrives at this plane. For simplicity we drop the bars over the new coordinates in (28) and hereafter. We have only to rotate the x and y axes by the small angle $\phi = S/(F/2)$ or the small angle $-\phi$ to obtain the expression for the one interfering wave or the other. We add the two expressions and evaluate the sum at the plane of the photographic plate, $x = 0$. In this way we find

$$\begin{pmatrix}\textit{electric field} \\ \textit{at } x = 0\end{pmatrix} = E_0 \cos k_0(t+\phi y)\exp[-y^2/2w^2-(t+\phi y)^2/2\tau^2]$$

$$+(\textit{corresponding term with } \phi\to-\phi). \qquad (29)$$

We square this sum and integrate over time as in Eq. (6) to arrive at a measure of the probability that a photon will strike a unit area at elevation y. Dropping the factor E_0 of primary field strength and associated normalization factors, and dropping terms of the type $\exp(-k_0^2\tau^2) = \exp(-N^2)$ in comparison with unity because we take $N>>1$, and concerning ourselves solely with relative probabilities, we have

$$\begin{pmatrix}\textit{relative probability} \\ \textit{for photon to hit} \\ \textit{at the elevation } y\end{pmatrix} = e^{-y^2/w^2}\,[1+1+2\cos(2k_0\phi y)e^{-\phi^2y^2/\tau^2}]. \qquad (30)$$

Here the first "1" can be considered to measure the contribution of the lower slit; the second "1," the upper slit; and the final term, positive or negative, the effect of interference between the waves from the two slits. This interference is governed by the retardation, $2\phi y$, of the one beam relative to the other at the elevation y, one beam being tilted up at the angle $\phi = S/(\textit{distance from slit}) = 2S/F$, the other being tilted down at the same angle. The same tilt puts the peak of the one wave train behind the peak of the other wave train by the same amount. When that retardation is comparable to the length parameter of the wave train, τ, the interference drops off or practically disappears, as evidenced by the final exponential in (30). The first exponential in (30) affects all terms alike, and is a consequence of the wave's having the natural width, w.

Now measure elevation above the optic axis in units of the parameter a of slit width, which we have taken to have the value $(\lambdabar F)^{\frac{1}{2}}$; thus,

$$y = aY = (\lambdabar F)^{\frac{1}{2}} Y. \tag{31}$$

Also insert for all other parameters the values that we have assigned them for convenience and for illustration:

$$w^2 = \alpha^2 2 = \lambdabar F/2,$$
$$\phi = 2S/F = 5 \cdot 2^{\frac{1}{2}} (\lambdabar/F)^{\frac{1}{2}},$$
$$\tau = N\lambdabar. \tag{32}$$

Then the expression for the relative probability for a photon to strike at the elevation y is

$$(relative\ probability) = e^{-2Y^2}[1+1+2\cos(10\sqrt{2}Y)e^{-50Y^2/N^2}]. \tag{33}$$

When the wave train is indefinitely long ($N\to\infty$), expression (33) (divided by 4) becomes $exp(-2Y^2)\cos^2 5\sqrt{2}Y$. It is natural to write $n\pi$ for the argument of the cosine. Where the fringes would peak in the absence of modulation, $n = 0$, $n = \pm 1$, $n = \pm 2$, there the modulation makes the relative intensities 1, 0.674, 0.206. Thus there is a well developed interference pattern. Moreover, one has only to increase the separation $2S$ of the two slits or the relative inclination 2ϕ of the two plane wave fronts to have more fringes and a still more fully developed interference pattern showing up within the effective width of the beam. This modification, far from causing difficulty, only helps the photodetectors to discriminate between the two beams when the photographic plate is swung aside.

Having thus dealt with all other considerations, we conclude that the length of the wave train, $\tau = N\lambda$, is the central factor in deciding whether one can have good interference fringes and still have "delayed choice." When N is finite, the intensity no longer goes to zero at fringe minima; the fringes are more or less "washed out." Returning to expression (30) for the distribution of intensity in the interference pattern, we can use the words "purity of the interference pattern" and "background" to speak of the quantities

$$(\text{"local purity of pattern"}) \equiv e^{-\phi^2 y^2/\tau^2}, \qquad (34)$$

$$(\text{"local background"}) \equiv 1-e^{-\phi^2 y^2/\tau^2}. \qquad (35)$$

A fuller analysis would go into the information carried by the interference pattern in the technical sense in which information is defined (15, 16). However, it is enough for the present purpose to multiply (34) and (35) by the overall modulation factor e^{-y^2/w^2} and integrate each over y and compare the thus weighted integral of the "local pattern purity" (34) with the sum of the two integrals to arrive at one useful index of overall pattern purity,

$$\left(\begin{array}{l}\text{"index of overall}\\ \text{pattern purity"}\end{array}\right) \equiv \frac{[\text{weighted integral of (34)}]}{[\text{weighted integral of (34) plus (35)}]}$$

$$= \frac{1}{(1+w^2\phi^2/\tau^2)^{\frac{1}{2}}}, \qquad (36)$$

an expression which reduces in the illustrative example to

$$\left(\begin{array}{l}\text{"index of overall}\\ \text{pattern purity"}\end{array}\right) = \frac{1}{(1+25/N^2)^{\frac{1}{2}}} \qquad (37)$$

Then and only then can one speak with assurance of both slits being involved in what happens to a quantum of energy when this index is close to unity.

When the incident wave train is very long ($N\to\infty$), Eq. (37) gives a purity index of unity. In other words, fringe intensity approaches zero at the canonical locations; background in this sense is negligible; and any limitation on the number of fringes that can be seen arises solely from the finite width parameter of the beam. Thus with $N=\infty$ we found five fringes ($n=0,\pm1,\pm2$) with intensities within an order of magnitude of the intensity of the central fringe. Those five fringes still stand out with nearly that much absence of background so long as the parameter N for the length of the incident wave train is of the order of several times $N_{crit}=5$ or greater, we conclude from (37);

but when the length parameter is little by little decreased below $N_{crit}=5$, first the outer two of these five fringes are washed out and then the next two. How this effect comes about is seen from an examination of the two wave trains depicted in Fig. 1 where they cross in the zone of interference. We conclude that there is a natural limit of the order $\tau_{crit}=N_{crit}\lambda\!\!\!^{-}=5\lambda\!\!\!^{-}$ below which it is not appropriate to reduce the length of the wave train. Only by arranging for the incident wave train to be that long or longer do we obtain well defined fringes. Only so will the interference pattern convey the message, "the energy came through both slits."

The requirement

$$\tau = N\lambda\!\!\!^{-} > w\phi = N_{crit}\lambda\!\!\!^{-} \quad (N_{crit}=5 \text{ in the example}) \tag{38}$$

makes no difficulty for the new feature of the "delayed choice" in the double slit experiment. The distance $F/2$ from the doubly slit screen to the photographic plate, like $\lambda\!\!\!^{-}$, is a primary parameter in the design of the idealized apparatus. It can be made as big as one pleases. For example, it can well be 10cm or $(2\times10^5)\cdot(5\times10^{-5}\text{cm})$ or $1.256\times10^6\lambda\!\!\!^{-}>>>>N\lambda\!\!\!^{-}$. Thus the wave train travels many times its own length on the way from the doubly slit screen to the plate. In conclusion, there is ample time, after the energy of the single quantum has already passed through the screen, to decide electronically, by a random-number generator or by a plan of choice programmed in advance in any way one pleases, what kind of indelible evidence shall be produced: "which-slit" evidence, or "double-slit" evidence.

Light travels from screen to plate on a null cone. Therefore we have good reason to ask, does the passage through the screen really take place in the past light cone of the "decision," the swinging of the photographic plate into, or out of, the zone of interference? Unless the answer is clearly "Yes," we cannot speak of an "apparent inversion of the normal order of time" in the delayed-choice double-slit experiment.

As one way to put the "traversal of the screen" clearly in the past light cone of "the decision" it is enough to switch from a light beam to an electron beam travelling at a speed, v, equal for example to $c/2$. However, it avoids the introduction of the new parameter v into all the foregoing analysis to stay with light, and make a simple alteration, not shown, in the arrangement of Fig. 1. Let a mirror be introduced half way from the doubly slit screen to the zone of interference. It folds the path of travel of the light almost back on itself. It puts the zone of interference almost on top of the screen. It makes the separation in space between the event of passage of the screen and the event of arrival in the zone of interference small compared to the separation in time of those two events. In this way the decision about "what shall have happened" in the passage of the screen is placed quite clearly in the future light cone of that passage itself.

"DELAYED CHOICE" AS AN ADDITIVE OPTION IN OTHER IDEALIZED EXPERIMENTS

To be forced to choose between complementary modes of observation is familiar, but it is unfamiliar to make this choice after the relevant interaction has already come to an end. Moreover, one can assert this "voice in what shall have happened, after it appears already to have happened" in illustrations of complementarity other than the double slit, by suitable modification of the idealized apparatus.

In the gamma ray microscope as described by Bohr (17) and Heisenberg (18), a lens of angular opening ε, receiving and bringing to a focus light of the reduced wave length $\lambda\!\!\!^{-}$, tells the position of the electron that scatters the light into the lens within a latitude of the order of

$$\Delta x \sim \lambda\!\!\!^{-}/\varepsilon. \qquad (39)$$

When the lens is thus used to fix position, the quantum of energy scattered into the lens gives the electron a lateral kick, the amount of which is subject to an uncertainty of the order

$$\Delta p \sim \left(\begin{matrix}\textit{photon}\\ \textit{momentum}\end{matrix}\right)\left(\begin{matrix}\textit{angular opening}\\ \textit{of the lens}\end{matrix}\right) \sim (\hbar/\lambda\!\!\!^{-})\varepsilon. \qquad (40)$$

This magnitude is coordinated with the Δx of (39) by the usual indeterminacy relation. However, the uncertainty in the lateral kick can be reduced to a very small fraction of (40) by placing a sufficiently great collection of sufficiently small photodetectors at a little distance above the lens. Whichever one of them goes off signals the direction of the scattered photon and thus the momentum imparted to the electron. When the lens is operated in this mode it ceases to serve as a lens in the true sense of the word. Lost then is the possibility to know the position of the electron within anything like the narrowness of limits implied by the Δx of (39). All this is the standard and well known lesson of complementarity.

We now add the feature of delayed choice. Only after the quantum of electromagnetic energy has already passed through the lens do we decide which lattice work of photodetectors to swing into action. One lattice is located in the focal plane of the lens. Let it be the lattice that is swung into action. Then one of the counters in this lattice goes off. This irreversible act of amplification tells us where the electron was when it scattered the radiation, within the lattitude Δx of (39). In drawing this conclusion about resolving power we accept the fact that the radiation made use of the entire aperture of the lens.

Or let the other lattice of photodetectors be swung into action. They are located in a plane some small fraction of the way up from lens to focal plane. One of these devices thereupon registers an event. This indelible record tells us within a certain range, much smaller than the Δp of (40), what lateral kick was given to the electron in the Compton process. The reasoning is simple. The kick to the electron is deduced from the direction of the scattered photon. The direction of this photon is revealed by the coordinates of the photocounter that registered, because it responds to a quantum of energy only when that energy goes through a highly restricted portion of the aperture of the lens.

Shall we say that "the whole aperture transmitted the energy" or that "only a very small fraction of the aperture transmitted the energy?" We can freely decide one way or the other, according as we activate one set of photocounters or the other, after the lens has *already* finished transmitting the energy. That is the unfamiliar feature that "delayed choice" brings to the idealized gamma-ray microscope.

The split-beam experiment provides another example, Bohr tells us (19), "to which Einstein very early called attention and often has reverted. If a semi-reflecting mirror is placed in the way of a photon, leaving two possibilities for its direction of propagation, the photon may either be recorded on one, and only one, of two photographic plates situated at great distances in the two directions in question, or else we may, by replacing the plates by mirrors, observe effects exhibiting an interference between the two reflected wave-trains. In any attempt of a pictorial representation of the behaviour of the photon we would, thus, meet with the difficulty: to be obliged to say, on the one hand, that the photon always chooses *one* of the two ways and, on the other hand, that it behaves as if it had passed *both* ways."

Stapp (20) has proposed a more specific experimental arrangement. In it the two parts of the split beam are brought back together with a phase difference, δ, at a second half-silvered mirror. The two alternative ways for a quantum of radiant energy to leave that mirror have relative probabilities $\sin^2\delta$ and $\cos^2\delta$, as can be measured by suitably positioned counters. Alternatively the counters can be moved to intercept the arriving energy before it hits the half-silvered mirror. According as we dispose the counters in the one way or the other we observe the consequences of interference or we find out in which of the two beams the quantum arrives; but, in conformity with the principle of complementarity, we cannot do both kinds of observation on the same photon.

When we add to this experiment the feature of delayed choice, we arrive at the arrangement of Fig. 2. We swing the photodetectors into the positions C_1, C_2 or the positions C_1', C_2' at our free choice and at the very last instant. In other words, after the quantum of electromagnetic energy has *already* finished interacting with the doubly silvered mirror, *A*, we choose whether this quantum shall manifest a two-beam interference phenomenon or shall arrive via a single beam.

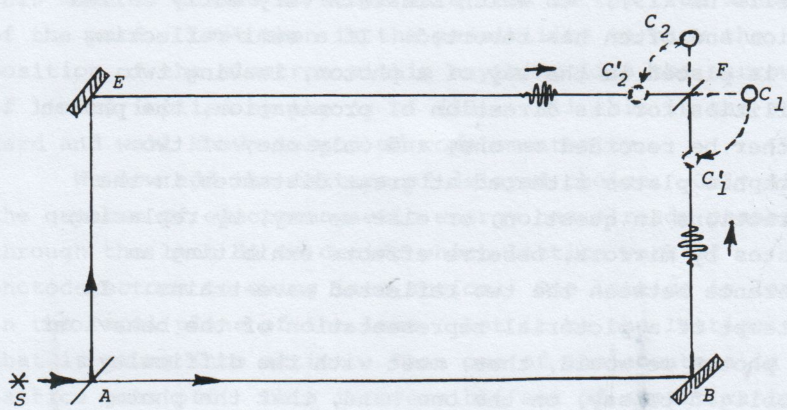

Fig. 2. The "delayed choice" split-beam experiment. S, pulsed source, giving a wave train long compared to one wave length but short compared to the total travel distance. It is operated at an intensity level so low that at most one quantum of energy is transported per pulse. A and F, semi-reflecting mirrors. B and E, fully reflecting mirrors. C_1 and C_2, photodetectors set to determine, by their relative counting rates, the phase difference between the two alternative routes to the point of observation. C_1' and C_2', the same photodetectors swung into position to determine on which route the quantum of energy arrives.

Fig. 3 shows the feature of delayed choice added to another familiar idealized experiment. Here we have to choose between measuring the time of emission of a photon and measuring its energy. When the faces of the tiltable teeth of the grating are aligned, it becomes a spherical mirror. Then the time-of-arrival counter alone is operative. The information it gives, projected back in time, tells us we had to do with a sharp pulse and tells us the time at which it left the atom. Excluded from measurement in this option, and therefore without meaning, is the energy with which the photon was emitted. In the other option we determine this energy, but forego any possibility of a precise measurement of the time of emission. For this purpose we turn the teeth of the grating--and in an improved dispensation, also adjust their shape--so that, in the language of spectroscopy, they "blaze" practically all of the energy into the first order spectrum. In this case one of the spectrum-analysis counters registers an event. From it we project back

(1) to conclude that the quantum of energy had a narrowly defined wave length and

(2) to deduce the energy it carried away from the atom.

The atom is very far away in the idealized arrangement we contemplate. There is ample time between the emission of the radiation and its arrival at the grating to choose which way to tilt the teeth. Thus after the radiation has *already* been emitted, we choose whether it is to manifest itself as pulselike or as nearly monochromatic.

In example No. 5 we make a delayed choice between measuring the direction (θ,ϕ) of an emitted photon and measuring the (l,m,π)-mode to which it belongs. For this purpose we can use a natural but rather complicated extension of the tiltable tooth arrangement to encompass the whole 4π solid angle around, and some distance from, the emitting atom. Then, according as we make the one or the other last minute choice of alignment of the tiltable teeth, we get one or other very different story--directivity or multipolarity--for the character of the radiation that had *already* left

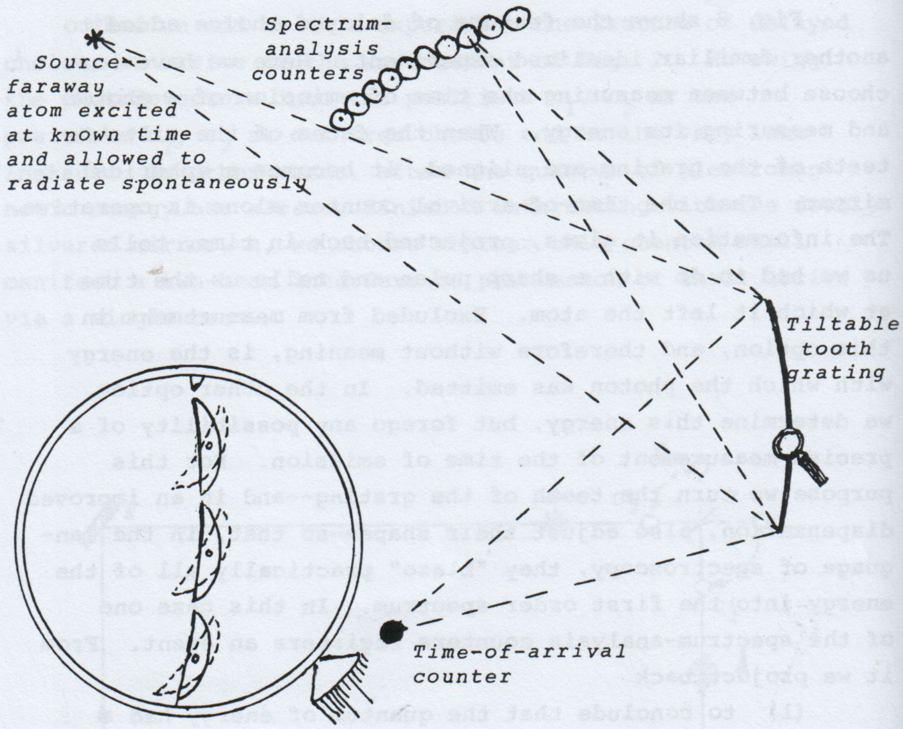

Fig. 3. Use of a tiltable tooth grating to make a delayed choice between alternative pictures (wave train versus pulse) of the radiation that the atom has already emitted. The magnified view shows a few typical teeth of the grating depicted (1) by dashed lines when the teeth are so oriented as to reflect or "blaze" the light to the spectrum-analyzing counters and (2) by full lines when the teeth are aligned. Then the grating becomes a spherical mirror, and the time-of-arrival counter alone is operative.

the atom at the time the choice was made.

Example No. 6 of delayed choice between complementary alternatives makes polarization the focus of attention. We arrange for a distant atom to be excited at time $t = 0$. We allow it to radiate spontaneously. The arriving quantum of energy we analyze with a Nichol prism. It is turned to let through radiation with an azimuth of polarization $\phi = \phi_o$. A transmitted quantum triggers a photodetector A ("accepted quantum"). A photon of the orthogonal polarization $\phi = \phi_o + (\pi/2)$ would normally run into the layer of black paint that is part of the traditional design of a Nichol prism and be absorbed. However, to provide extra assurance of a reliable analysis of polarization, we strip away that paint and arrange for such a photon to enter into, and trigger, a second photodetector *R* ("rejected quantum"). We make the choice of the acceptable azimuth, ϕ_o, and in effect[1] rotate the Nichol prism and associated *A* and *R* counters to that azimuth, after the quantum of energy has *already* left the distant atom, and before it reaches the analyzer.

As seventh and last example of adding delayed choice to a standard illustration of complementarity, consider the spontaneous annihilation of the light "atom" (e^+, e^-) in its singlet ground state. Let two analyzers be used to investigate the polarizations of the two 0.511MeV quanta given off in opposite directions in this annihilation process.[2]

Let the two analyzers first be set to "corresponding" polarizations. That means "orthogonal azimuths" in the case when both detectors are adjusted to analyze for linear polarization, as in the experiments done so far. There is a prospect that new experiments will look for circular

---

[1] *In actuality, under ordinary experimental circumstances, the equipment will be too large and fragile and the available time too short to permit any significant rotation of the Nichol prism. Therefore the equivalent of the last minute rotation is understood to be accomplished by an appropriate electronic rotator-of-polarization located just before the Nichol prism.*

[2] *Theory, (21-23); observations with two analyzers of linear polarization, (24-34); theory for the case when the analyzers can be set to accept photons of arbitrary Stokes parameters (36) observations on the polarization of the photons given out in a related process, the two-step transition of an atom from an excited state to its ground state, (37, 38); a result in such an experiment in contradiction with quantum predictions, (39); failure to confirm this experiment, (40); confirmation of quantum predictions (40-42).*

polarization. In that case "corresponding" means opposite senses of rotation of the electric vector as judged by a faraway observer of the equipment looking along the common line of travel of the two photons, one approaching him, the other receding. For the more general case of elliptic polarization, "corresponding" means that the Stokes vector (35) of the polarization accepted by the one analyzer appears to the same faraway assessor of the apparatus to be antipodal to the Stokes vector of the polarization accepted by the other analyzer. Let the information in each Stokes vector be summarized in the location of a point on the surface of the unit sphere. Then the two polarizations "correspond," Kagali shows (36), when their separation, θ, on the unit sphere is $\theta = 180°$.

When the two analyzers are set to polarizations that correspond in the foregoing sense, then the predictions of quantum mechanics are simple; and the observations, so far as they go (linear polarization) support those predictions: Every time an annihilation quantum triggers the "accept counter," *A*, of the one analyzer, its partner annihilation quantum sets off the accept register, *A*, of the other analyzer; and every time the first photon is recorded by the "reject counter," *R*, of the first analyzer, its mate triggers the "reject photodetector," *R*, of the other analyzer.

When the Stokes vectors of the two analyzers do not correspond; when their separation, θ, on the unit sphere is not 180°, but some lesser angle, then the predictions are still simple (36), and the observations for the case of linear polarization support those predictions:

When the accept counter of the first analyzer goes off,
the accept counter of the second analyzer goes off with a probability $\sin^2(\theta/2)$; and
the reject counter of the second analyzer goes off with a probability $\cos^2(\theta/2)$.

When the reject counter of the first analyzer goes off,
the accept counter of the second analyzer goes off with a probability $\cos^2(\theta/2)$; and
the reject counter of the second analyzer goes

off with a probability $\sin^2(\theta/2)$. (41)

Very different are these predictions from what one would have expected had the two annihilation photons started out from their point of origin with "already defined polarizations:" linear polarization of the one quantum at an azimuth $\phi_{emission} = \phi_e$, where ϕ_e is random; and then, of necessity, linear polarization of its mate at the perpendicular azimuth $\phi_e+(\pi/2)$. Were this a proper description of what goes on, were the polarizations thus defined independently of the act of observation, then the chance of an "accept count" in the first analyzer (set at an azimuth ϕ_1) along with an accept count in the second analyzer [set at an azimuth $\phi_1+(\theta/2)$] would have been

(chance of an "accept-accept coincidence")

$$= \cos^2(\phi_e-\phi_1)\cos^2[(\phi_e+\pi/2)-(\phi_1+\theta/2)],$$

averaged over azimuths, $\phi_{emission}$

$$= (1/8) + (1/4)\sin^2(\theta/2), \tag{42}$$

instead of the quantum-mechanically predicted--and observed--and twice as strong (*zero* to full strength) dependence of coincidence rate on relative azimuth, $\theta/2$, of the two analyzers,

$$(\text{chance of an A-A coincidence}) = (1/2)\sin^2(\theta/2) \tag{43}$$

as illustrated in Fig. 4.

Set the analyzers at one relative azimuth and measure the probability of an accept-accept coincidence; repeat for a second relative azimuth and for a third. Between these three probabilities Bell established the existence of an inequality, by now well known, and used by him and others--in conjunction with the observations--to exclude various theories of "hidden variables" and other proposed alternatives to quantum mechanics (43, 44). We have no need to examine that watch-dog inequality here. We are not interested in "alternatives." We seek, not to escape standard battle-tested quantum theory, but to learn what lesson it has to teach in "delayed choice experiments."

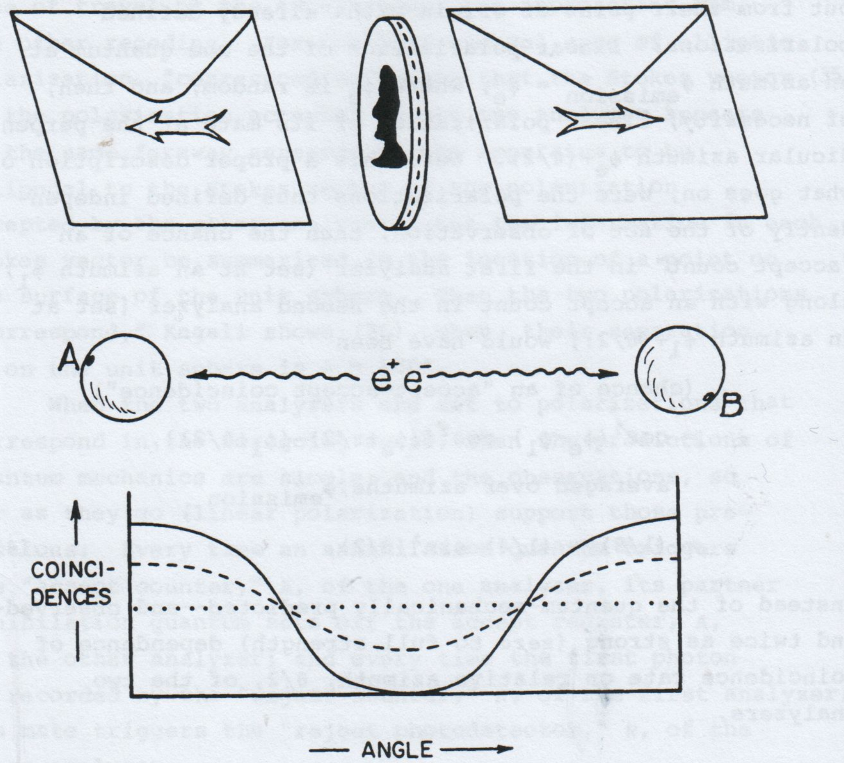

Fig. 4. In contrast to splitting a coin, putting the two slices into envelopes, shuffling them, and sending them to two remote observers (upper part of diagram) the experiment on the annihilation photons (middle part of diagram) permits a double infinity of choices for the polarization to be looked for in the right-hand (e^+e^-) annihilation photon, with corresponding consequences for the polarization that will be found for the left-hand photon. If the polarizations were determined in the act of emission ("hidden variables") the coincidences between the two photons, when both analyzers are set to accept linear polarization, would show only half the dependence on relative orientation of the two polarizations (dashed curve in lower diagram) that is predicted by quantum mechanics--and observed (full curve)

In the "delayed-choice and arbitrary-polarization" version of the experiment on annihilation quanta, we wait until the two photons are on their way from source to receptors to fix the experimental conditions. At the last instant before the one quantum arrives at the one receptor we set the one analyzer to one set of Stokes parameters (i.e., select one point on the right-hand unit sphere of Fig. 4); similarly set the other analyzer to another set of Stokes parameters (i.e., select a point on the left-hand unit sphere in Fig. 4--equivalent to a reflected point on the right-hand unit sphere). It does not matter how different the ellipticities are to which the two analyzers are set nor how late in the travel of the two photons these settings are made; the probabilities of AA, AR, RA and RR coincidences are still given by Eqs. (41). To specialize, pick at the very last minute whatever Stokes vector one pleases for the one analyzer, and the complementary polarization (antipodal Stokes vector) for the other analyzer. Then one is assured of a perfect concord between the causally disconnected records made at the two receptors. When the accept counter goes off at the one receptor, the accept counter goes off at the other receptor. When the reject register triggers at the one, the reject register goes off at the other. It is unpredictable whether the photon travelling to the left will set off the accept counter or the reject one. Whichever it is, the other photon will do the *same*. It does not matter that the two photons are out of causal contact with each other. It does not matter that only at the last instant, while they were *already* on their way, did we make that choice of complementary polarizations to which the two photons, in perfect unison, align or antialign.

LESSON OF DELAYED-CHOICE EXPERIMENTS

The double slit experiment, like the other six idealized experiments (microscope, split beam, tilt-teeth, radiation pattern, one-photon polarization, and polarization of paired photons), imposes a choice between complementary modes of observation. In each experiment we have found a way to delay that choice of type of phenomenon to be looked for up to the very final stage of development of the phenomenon, *whichever* type we then fix upon. That delay makes no difference in the experimental predictions. On this score everything we find was foreshadowed in that solitary and pregnant sentence of Bohr (45), ". . .it. . . can make no difference, as regards observable effects obtainable by a definite experimental arrangement, whether our plans for constructing or handling the instruments are fixed beforehand or whether we prefer to postpone the completion of our planning until a later moment when the particle is already on its way from one instrument to another."

Not one of the seven delayed choice experiments has yet been done. There can hardly be one that the student of physics would not like to see done. In none is any justification whatsoever evident for doubting the obvious predictions.

We search here, not for new experiments or new predictions, but for new insight. Experiments dramatize and predictions spell out the quantum's consequences; but what is its central idea? A pedant of Copernican times could have calculated planetary positions from the equations of Copernicus as well as Copernicus himself; but what would we think of him if his eyes were closed to the main point, that the "Earth goes around the Sun"?

In the absence of an equally simple statement of *its* central idea, quantum theory appears to many as strange, unwelcome, and forced on physics as it were from outside against its will. In contrast, if the essential point could be grasped in a single phrase, we can well believe that the quantum would seem so natural that we would recognize at once that the universe could not even have come into being without it.

Special relativity's findings shower out like fireworks from a single compact package, "The laws of physics are the same in every inertial reference system." No leap of the imagination to a comparably compact and explodable formulation of quantum theory being forthcoming, experience recommends step-by-step progress. Of such steps none in recent times moved our understanding forward more than the Einstein-Bohr dialog (1). Out of that dialog no concept emerged of greater fruitfulness than "phenomenon" (46): ". . .[In my discussions with Einstein] I advocated the application of the word *phenomenon* exclusively to refer to the observations obtained under specified circumstances, including an account of the whole experimental arrangement" (47). No other point does the present analysis of idealized delayed-choice experiments have but to investigate what "phenomenon" means as applied to the "past."

After the quantum of energy has *already* gone through the doubly slit screen, a last-instant free choice on our part--we have found--gives at will a double-slit-interference record or a one-slit-beam count. Does this result mean that present choice influences past dynamics, in contravention of every formulation of causality? Or does it mean, calculate pedantically and don't ask questions? Neither; the lesson presents itself rather as this, that the past has no existence except as it is recorded in the present. It has no sense to speak of what the quantum of electromagnetic energy was doing except as it is observed or calculable from what is observed. More generally, we would seem forced to say that no phenomenon is a phenomenon until--by observation, or some proper combination of theory and observation--it is an observed phenomenon. The universe does not "exist, out there," independent of all acts of observation. Instead,it is in some strange sense a participatory universe.

That present choice of mode of observation in the double-slit experiment should influence what we say about the "past" of the photon; that the "past" is undefined and undefinable without the observation, may be illustrated by a little story.

The "game of twenty questions" will be recalled. One of the party is sent out of the room. The others agree on a word. The one fated to be questioner returns and starts his questions. "Is it a living object?" "No." "Does it belong to the mineral kingdom?" "Yes." So the questions go from respondent to respondent around the room until at length the word emerges: victory if in twenty tries or less; otherwise, defeat.

Well does one participant recall the evening when he, fourth to be sent out, returned to find a smile on everyone's face, sign of a joke or a plot. He innocently started his questions. But each question he put took longer in the answering--strange, when the answer itself was only a simple "yes" or "no." At length, feeling hot on the trail, he asked, "Is the word 'cloud'?" "Yes," came the reply and everyone burst out laughing. They explained that when he had gone out, they had agreed not to agree in advance on any word at all. Everyone could respond "yes" or "no" as he pleased to whatever question was put to him. But however he answered, he had to have a word in mind compatible with his own reply--and with all the replies that went before. No wonder it took time to answer!

It is natural to compare the game in its two versions with physics in its two formulations, classical and quantum. First, the puzzled participant thought the word already existed "out there" as physics once thought that the position and momentum of the electron existed "out there," independent of any act of observation. Second, the information about the word was brought into being step by step through the questions that the interrogator raised, as the information about the electron is brought into being step by step by the experiments that the observer chooses to make. Third, if the participant had chosen to ask different questions he would have ended up with a different word, as the experimenter would have ended up with a different story for the doings of the electron if he had done different measurements or the same measurements in a different sequence. Fourth, whatever power the interrogator had in influencing the outcome for the word was partial only. A major part of the decision lay in the hands of the other participants. Similarly, the experimenter has some substantial influence on what will happen to the electron by the choice of experiments he will do on it; but he is well aware that there is much impredictability about what any given one of his measurements will disclose. Fifth, there was a "rule of the game" that required of every participator that his choice of yes or no should be compatible with *some* word. Similarly, there is a consistency about the observations made in physics. One person can tell another in plain language what he finds and the second person can verify the observation. Interesting though this comparison is between the world of physics and the world of the game, there is an important point of difference. The game has a finite number of participants and terminates after a finite number of steps. In contrast, the making of observations is a continuing process. Moreover, it is extraordinarily difficult to state sharply and clearly

(1) where the community of observer-participators begins and where it ends, and

(2) what the degree of amplification must be to define an observation: "The amplification of atomic effects, which makes it possible to base the account on measurable quantities and which gives the phenomenon a peculiar closed character, only emphasizes the irreversibility characteristic of the very concept of observation" (48).

It is not necessary to understand every point about the quantum principle in order to understand something about it. Of all the points that stand forth from comparing the world of quantum observations with the game of twenty questions, none is more central than this: As in the game no word was the word until that word had been promoted to reality by the choice of questions asked and answers given, so no phenomenon is a phenomenon until it is an observed phenomenon.

ACKNOWLEDGMENTS

The author expresses his appreciation to Ransom Marlow and other organizers of and participants in the June 2, 3, 4, 1977 New Orleans Conference on the Mathematical Foundations of Quantum Theory for discussions; and to Aage Bohr, Dieter Brill, David Deutsch, Cécile De Witt, and Manfred Fink, as well as Paul Gleichauf, Jørgen Kalckar, H. Pfister, Henry Stapp, Xerxes Tata and William Wootters for advice or help or both in the preparation of this report.

REFERENCES

1. Bohr, N., "Discussion with Einstein on epistemological problems in atomic physics,"in "Albert Einstein: Philosopher-Scientist," pp. 199-241. Library of Living Philosophers, Evanston, Illinois, 1949.
2. Bohr, N., "Atomic Physics and Human Knowledge," p. 73. Wiley, New York, 1958.
3. George, C., I. Prigogine and L. Rosenfeld, "The macroscopic level of quantum mechanics," Kgl. Danske Videnskab. Selskab, Mat.-fys. Meddelelser 38, No. 12 (1972).
4. Bohr, N., "Atomic Physics and Human Knowledge," pp. 88-89. Wiley, New York, 1958; see also extensive references to the literature on p. 174 of Petersen, A., "Quantum Physics and the Philosophical Tradition," M. I. T. Press, Cambridge, Massachusetts, 1968.
5. Belinfante, F. J., "Measurement and Time Reversal in Objective Quantum Theory," p. 39. Pergamon, Oxford and New York, 1975.
6. Duguay, M. A., "The ultrafast optical Kerr shutter," pp. 161-193 in "Progress in Optics, vol. 14 (E. Wolf, Ed.), North Holland, Amsterdam, 1976, compares the new ultrafast optical Kerr shutter to the older and slower conventional and electrically operated Kerr cell.

7. Lubkin, G. B., "Tests show CO_2 laser is suitable for fusion," Physics Today 30, 19 (September 1977).

8. Heisenberg, W., and W. Pauli, "Zur Quantendynamik der Wellenfelder," Zeits.f. Physik 56, 1-61 (1929); and "Zur Quantentheorie der Wellenfelder. II," Zeits. f. Physik 59, 168-190 (1930).

9. Breit, G. "Quantum theory of dispersion," Rev. Mod. Phys. 4, 504-576 (1932); see especially pp. 521-532 and 553-554; and "Quantum theory of dispersion (continued). Parts VI and VII," Rev. Mod. Phys. 5, 91-140 (1933); see especially pp. 104-105 and 127-128.

10. Lapedes, E. N., Ed., "McGraw-Hill Dictionary of Scientific and Technical Terms," article on apodisation, McGraw-Hill, New York, 1974; Jacquinot, P., and B. Roizen-Dossier, "Apodisation," in "Progress in Optics," Vol. 3 (E. Wolf, Ed.), North Holland, Amsterdam, 1964.

11. Gouy, L. G., "Sur une proprieté nouvelle des ondes lumineux," C. R. Acad. Sci. Paris 110, 1251 (1890).

12. Bruhat, G., "Optique," p. 59, theory; p. 95, experiment, Masson, Paris, 1942.

13. Simms, D. J. and N. M. J. Woodhouse, "Lectures on Geometric Quantization," list of references on pp. 159-163, especially citations of the work of Maslov, Springer, Berlin and New York, 1976.

14. Misner, C. W., K. S. Thorne and J. A. Wheeler, "Gravitation," section of focusing, pp. 570-583, Freeman, San Francisco, 1973.

15. Everett, H., III, "The theory of the universal wave function," Princeton University doctoral thesis, 1956; published in abbreviated form as "'Relative State' Formulation of Quantum Mechanics," Rev. Mod. Phys. 29, 454-462 (1957); and in full in "The Many-Worlds Interpretation of Quantum Mechanics," (B. De Witt and N. Graham, Eds.) Princeton University Press, Princeton, New Jersey, 1973, pp. 3-140; see there pp. 43-53 for the considerations on information theory cited in the text.

16. Wootters, W. K., unpublished October 1977 lecture on the information content of an interference pattern.

17. Bohr, N., "The quantum postulate and the recent development of atomic theory," Nature 121, 580-590 (1928); see especially pp. 582,583.
18. Heisenberg, W., translated by C. Eckart and F. C. Hoyt, "The Physical Principles of the Quantum Theory," pp. 21-22, University of Chicago Press, Chicago, Illinois, 1930.
19. Bohr, N., "Discussion with Einstein on epistemological problems in atomic physics," pp. 201-241 and specifically p. 222 in Schilpp, P. A., "Albert Einstein: Philosopher-Scientist," Library of Living Philosophers, Evanston, Illinois, 1949.
20. Stapp, H., oral report, The University of Texas at Austin, spring, 1977.
21. Wheeler, J. A., "Polyelectrons," Ann. New York Acad. Sci. 48, 219-238 (1946). On p. 235 the figures, 1.08 at 90° and 1.10 at 74°30', should be corrected to 2.6 at 90° and 2.85 at 82°, according to Snyder et al (23).
22. Pryce, M. H. L. and J. C. Ward, "Angular correlation effects with annihilation radiation," Nature 160, 435 (1947).
23. Snyder, H. S., S. Pasternack and J. Hornbostel, "Angular correlation of scattered annihilation radiation," Phys. Rev. 73, 440-448 (1948).
24. Bleuler, E., and H. L. Bradt, "Correlation between the states of polarization of the two quanta of annihilation radiation," Phys. Rev. 73, 1398 (1948).
25. Hanna, R. C., "Polarization of annihilation radiation," Nature 162, 332 (1948).
26. Vlasov, N. A. and B. S. Dzeljepov, "Poljarizatzija annigiljatzionnikh gamma-kvantov," Dokl. Akad. Nauk. SSSR 69, 777-779 (1949).
27. Wu, C. S. and I. Shakhov, "The angular correlation of scattered annihilation radiation," Phys. Rev. 77, 136 (1950).
28. Hereford, F., "The angular correlation of photoelectrons ejected by annihilation quanta," Phys. Rev. 81, 482 (see also pp. 627-628) (1951).

29. Bertolini, G., M. Bettoni and E. Lazzarini, "Angular correlation of scattered annihilation radiation," Nuovo Cim. 2, 661-662 (1955).
30. Langhoff, H., "Die Linearpolarisation der Vernichtungsstrahlung von Positronen," Zeits. f. Physik 160, 186-193 (1960).
31. Kasday, L., J. Ullman and C. S. Wu, "The Einstein-Podolsky-Rosen argument: positron annihilation experiment," Bull. Amer. Phys. Soc. 15, 586 (1970).
32. Kasday, L., "Experimental test of quantum predictions for widely separated photons," pp. 195-210 in "Foundations of Quantum Mechanics, Proceedings of the International School of Physics 'Enrico Fermi,' Course 49" (B. d'Espagnat, Ed.), Academic, New York, 1971.
33. Faraci, G., D. Gutkowski, S. Notarrigo and A. R. Pennisi, "An experimental test of the EPR paradox," Lettere al Nuovo Cimento, 9, 607-611 (1974).
34. Wilson, A. R., J. Lowe and D. K. Butt, "Measurement of the relative planes of polarization of annihilation quanta as a function of separation distance," J. Phys. G: Nucl. Phys. 2, 613-624 (1976).
35. Born, M., and E. Wolf, "Principles of Optics," 4th Ed., Pergamon, New York, 1970.
36. Kagali, Basavaraj A., "Correlation of polarizations in e^+-e^- annihilation," preprint, Center for Theoretical Physics, University of Texas (1977); submitted for publication.
37. Kocher, C. A. and E. D. Commins, "Polarization correlation of photons emitted in an atomic cascade," Phys. Rev. Lett. 18, 575-577 (1967).
38. Freedman, S. J. and J. F. Clauser, "Experimental test of local hidden-variable theories," Phys. Rev. Lett. 28, 938-941 (1972).
39. Holt, R. A., "Atomic cascade experiments," Doctoral dissertation, Harvard University, Cambridge, Massachusetts, unpublished, available from University Microfilms, Inc., Ann Arbor, Michigan 48106. (1973).
40. Clauser, J. F., "Experimental investigation of a polarization correlation anomaly," Phys. Rev. Lett.36, 1223-

1226 (1976).

41. Freedman, S. J. and R. A. Holt, "Tests of local hidden-variable theories in atomic physics," Comments on Atomic and Molecular Physics 5, 55-62 (1975).

42. Fry, E. S. and R. C. Thompson, "Experimental test of local hidden variable theories," Phys. Rev. Let. 37, 465-468 (1976).

43. Bell, J. S., "On the Einstein Podolsky Rosen paradox," Physics 1, 195-200 (1964).

44. Wigner, E. P., "On hidden variables and quantum mechanical probabilities," American J. of Physics 38, 1005-1009 (1970).

45. Bohr, N., see reference 1, p. 230.

46. Petersen, A., "Quantum Mechanics and the Philosophical Tradition," M. I. T. Press, Cambridge, Massachusetts, 1968, gives a preliminary account of the stages in the development of the concept of "phenomenon."

47. Reference (2), p. 64.

48. Reference (2), p. 89.