Henry Margenau

CAUSALITY IN QUANTUM ELECTRODYNAMICS

Quantum mechanics, even in its early and simple phases, has often been regarded as a non-causal discipline. The argument supporting this view cites the uncertainty principle as prohibiting the ascertainment of complete knowledge concerning physical states upon which causal prediction could be based. Recent developments in atomic physics have added new and puzzling features to the problem of causality insofar as they operate, not only with intrinsically unmeasurable states, but also with time reversals which have been interpreted to mean that the effect can be prior to the cause. Feynman's theory of quantum electrodynamics is particularly rich in unorthodox suggestions which tantalise philosophers. The purpose of the present paper is to exhibit them, appraise their methodological function and see in what manner they violate the rules of causal description. This purpose, it seems, is best achieved by a sequential discussion of three questions: What does causality mean in physics? What is the new method of quantum electrodynamics? Is this new method compatible with the causal doctrine in some satisfactory form?

https://doi.org/10.1177/039219215400200605 Published online by Cambridge University Press

I. *The Meaning of Causality*

When modern science speaks of causal connexions, it has reference to one or the other of two quite different relations between events or observations. The first is illustrated by such a chain of events as this: appearance of a cloud in the sky—darkening of the sun—lowering of the temperature on the earth—people putting on coats, etc.; or another, perhaps more scientific: emission of light from a star—propagation of an electromagnetic wave through space—absorption of light by a metal—ejection of an electron, etc. The events composing these chains form a sequence of *continuous action*; we know precisely, in terms of visible or postulated agencies, how the appearance of a cloud leads to the obscuring of the sun, how this in turn brings about a lower temperature, how this makes people uncomfortable and induces them to put on their coats. The other sequence can be traced in a similar manner; the events it connects, while occurring in widely different places and in totally different objects, are linked by some continuous action, some pervasive influence the details of which are understood. If there were gaps in this understanding, missing links in the chain of continuous action, the term causal would not be applied to it.

I shall speak of that meaning of causality which these examples illustrate as continuous action in time and space; it adverts to little more than relatedness by scientific agents and therefore makes causality tautologically equivalent to scientific understanding. It is a variant of the Humean doctrine of invariable sequence, refined by the inclusion of connective agencies between the members of that sequence. This interpretation of causality is large and generous, enjoys favour chiefly in the non-physical sciences and, of course, in everyday language; it is the stock-in-trade of lawyers and biologists. However, it is difficult to formulate with precision, and the difficulty resides in the circumstance that the view at issue places no restrictions upon the location and the kinds of events which are connected into a single chain. The emission of light can be on Sirius, the absorption can take place in some photocell on earth, and so forth. The only supposition is that the effect is later than the cause.

Physics, while at times espousing the continuous action view (often without being aware of the difference which I am exposing), is partial to another meaning of causality, a meaning first clearly formulated by Kant and Laplace. To wit: A physical system is described in terms of *states* which change in time. For example, the state of a body undergoing thermal changes may involve its temperature, its volume and perhaps its phase, and these variables are said to be variables of state, or variables defining a state. Another physical system, called an elastic body undergoing deformations, has states which are defined through stresses and strains; an electromagnetic field (which is also a physical system in our sense of the word, for physical systems need not be material!) assumes states specifiable by an electric and a magnetic vector. Common to all these instances is the supposition that the variables of state, however defined, change in time in a manner conforming to certain equations which are ordinarily called laws of nature. Future states are therefore predictable if a complete present state is known. A prior state of a given system is called the cause of a later state, the later state the effect of a former.

The principle of causality, in this sense, asserts the existence of a determinate temporal continuum of evolving states, all referable to the same physical system. Between a given cause and a given effect there is an infinity of other causes of the same effect, though only one cause at one time. The advantage of this view, which I shall designate by the label of *unfolding states*, resides in its greater logical precision and in the uniqueness it confers upon the causal relation. When a cloud appears in the sky, that cause (in the former sense) has *many* effects at a given later time, one of which is a lowering of the temperature; when an elastic body has a given distribution of stresses and strains, a *single* definite distribution at a specified time is its effect. The continuous action view permits many causal chains, the unfolding state view only a single train of evolution.[1]

The simplest physical system, indeed the one for which a causal theory of the latter type was first developed, is the moving mass point. Its states are pairs of variables, positions and momenta, and the law of nature governing their evolution is Newton's second law. The latter is a differential *equation of the second order* requiring two constants of integration in its complete solution. Position and momentum of the particle at a fixed time can serve as constants of integration and therefore determine the solution of all times. States and laws of a causal theory must always have this internal

[1] Our survey of the meanings of causality is not quite complete. One deficiency lies in its failure to analyse further the laws which connect the states. They must in some sense be invariable, or time-free. This point has been discussed in my book, *The Nature of Physical Reality* (McGraw-Hill, 1950), where further reflections concerning the suitableness of probabilities to function as state variables will be found.

Also omitted has been a version of causality which, though extremely limited, has found its way into the technical literature under the label 'causality conditions' (see, e.g., Van Kampen, Phys. Rev., 89, 1072, 1953). It is nothing more than the requirement of relativity limiting the speed of a wave packet to the speed of light and says, in effect, that a cause at one point at time t cannot produce an effect at another point, a distance r from the first, at a time earlier than $t+r/c$.

affinity. The states must be so chosen that they provide the information demanded by the initial conditions that make the solution of the law complete. It follows from this circumstance that the definition of states in a causal theory cannot be arbitrarily altered without corresponding changes in the law of nature, and a change in the law will generally necessitate a redefinition of the state of a system.

Newton's theory of the motion of particles is the prototype of all causal description, and the laws and states it demands, rather than its formalism, have come to be regarded as essential elements of causality. This misunderstanding, or, at any rate, this inflexible identification of states, has led to the belief that quantum mechanics is no longer a causal discipline. Let us recall the important details: Newton's law was found to fail for atomic particles, and Schrödinger succeeded in replacing it by a new equation. But that equation did not have solutions specifiable by the old positions and momenta. Heisenberg discovered through his uncertainty principle that these variables had furthermore lost their soundness as universally observable quantities, for they cannot both be known with precision. This seemed to spell the doom of causality because the states it involved are neither theoretically significant nor observationally available.

It is not idle, however, to ask the question whether Schrödinger's law selects, or is compatible with, states in terms of other variables, and whether these variables permit a causal description in the second, more formal sense of our principle. That is, in fact, the case; only it is the misfortune of these variables, or rather of the states which they define, to be somewhat strange and elusive when judged from Newton's familiar standpoint. They turn out to be probabilities.

We shall need a little of the mathematical context of the quantum theory in the next section and, therefore, do well to be explicit at once.

Schrödinger's equation is $$i\frac{\delta \Psi}{\delta t}=H\psi \qquad (1)$$

H, the Hamiltonian operator,[2] contains co-ordinates and the time, and is in general of the form $T + V$, symbolising kinetic plus potential energy. The arguments of the function ψ are therefore likewise the space co-ordinates and the time. If the functional form of ψ were known at any time t_1, it could be computed from equation (1) for all other times. In other words, if ψ can be regarded as a state of the particle, quantum mechanical description is causal.

[2] Here and everywhere else in this paper, energies are understood to be frequencies, i.e., every energy is divided by Planck's constant.

https://doi.org/10.1177/039219215400200605 Published online by Cambridge University Press

The identification of ψ with observable matters was made by Born and Jordan: $|\psi(x,y,z,t)|^2$ is the probability that, when the particle is looked for in suitable ways, it will be found as x,y,z,t. If probabilities are not decent physical variables—and many physicists do regard them as loathsome—then this interpretation of states and its causality must be rejected. Further probing into why probabilities are objectionable reveals that they cannot be determined by a single measurement but require many observations. The position of a particle, or its momentum, can each be measured in a single act. But if the condition of the particle is unspecifiable by statements saying where it is and how fast it is going; if it is found sometimes here and sometimes there, then an aggregate of measurements must be performed, and the interesting information is the relative frequency, i.e., the probability, with which it is found here or there. It is difficult to see why an observable should lack the fitness to serve as a variable of state if its determination requires more than one measurement. These measurements can be performed, it should be added, in a way that will yield the probability at a definite time, so that the reference to an instant, which is crucial to the causal sequence, is not lost.

Subject to the acceptance of probabilities as physically meaningful states, quantum mechanics remains a causal theory. Henceforth we shall take this stand and proceed to show that the latest advances in quantum electrodynamics leave this status unchanged.

II. *Summary of Some Recent Innovations*

I shall attempt to sketch here the theory proposed by Feynman, partly because of its successes and partly because of its richness in stimuli for philosophic reflection. Among the many approximation methods for solving equation (1) Feynman selects one which represents ψ as an integral over the initial state with the use of a Green function or kernel, K:

$$\psi\ (\lambda_2 t_2)=\int K(\lambda_2 t_2,\lambda_1 t_1)\ \psi\ (\lambda_1 t_1)d\lambda_1 \qquad (2)$$

For simplicity, we have abbreviated the co-ordinates x,y,z into the single symbol, λ. The kernel K can be found if the form of $H=T+V$ is known.

For a free electron, $V=O$, and K therefore has a definite representation which we call $K^{(0)}$. As we shall make no explicit use of it we need not write it down.

If V is small but finite (it is understood to be a function of x and t) $K^{(0)}$ is its dominant part, and it is possible to write a series

$$K=K^{(0)}+K^{(1)}+K^{(2)}+\ .\ .\ . \qquad (3)$$

in which successive members decrease in magnitude because they involve

https://doi.org/10.1177/039219215400200605 Published online by Cambridge University Press

V in increasingly higher powers. Each constituent of the kernel is, of course, a function of two sets of co-ordinates, $K^{(i)}=K^{(i)}(x_2t_2,\ x_1t_1)$ and we shall now abbreviate it by writing $K^{(i)}$ (2, 1). Feynman shows that the terms in series (3) can be computed as follows:

$$K^{(1)}(2,1)=-i\int K^{(0)}(2,3)\,V(3)\,K^{(0)}(3,1)\,dx_3\,dt_3 \qquad (3a)$$

$$K^{(2)}(2,1)=(-i)^2\iint K^{(0)}(2,4)\,V(4)\,K^{(0)}(4,3)\,V(3)\,K^{(0)}(3,1)\,dx_3\,dt_3\,dx_4\,dt_4 \text{ etc.} \qquad (3b)$$

These formulas have an interesting and suggestive interpretation.

If $\psi(x, t)$ were the probability that an electron be situated at the world point x, t, and K $(x_2t_2,\ x_1t_1)$ the probability that an electron makes a passage from x, t, to x_2t_2, equation (2) would be the relation connecting these quantities, as a little reflection will show. To be sure, ψ is not a probability but a 'probability amplitude' (since $|\psi|^2$ plays the role of a probability) and the same must be said about K. To carry through the interpretation, however, we will ignore this distinction.

We have recognised K(2, 1) as the probability of passage of an electron from point 1 to point 2. Such a passage need not be direct. Indeed it is reasonable to classify all passages according to the number of times the electron changes its direction of motion. Thus we define $L^{(i)}$ (2, 1) to be a path leading from 1 to 2 and having i corners. $L^{(0)}$ (2, 1) is the characteristic path of a free particle, i.e., of an electron in the absence of a

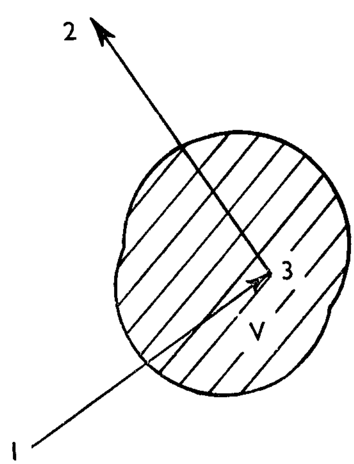

FIG. 1. *Pictorial illustration of* $K^{(1)}$. *An electron going from point* 1 *to point* 2 *is scattered once on the way by the potential* V *extending over the shaded region.*

https://doi.org/10.1177/039219215400200605 Published online by Cambridge University Press

potential energy V. A possible $L^{(1)}$ (2, 1) is drawn in fig. 1; the corner at the point (3) lies within the space-time region in which V is finite, since otherwise a deflection could not occur. What is the probability of $L^{(1)}$ (2,1)? It is proportional to three quantities: (i) the probability of passage from 1 to 3; (ii) the probability of a deflection at 3, and this might be supposed to be proportional to $V(3)$; (iii) the probability of passage from 3 to 2. When these three quantities are multiplied together and integrated over all intermediate points 3, the result is the probability that the electron will go from 1 to 2 via *any* path $L^{(1)}$ (2, 1). But the indicated operations are exactly those defining $K^{(1)}$ (2, 1), equation 3a (except for the factor $-i$, which need not concern us here). We conclude, therefore, that $K^{(1)}$ represents the probability of any passage in which the electron suffered *one* deflection.

An extension of this reasoning serves to show that $K^{(2)}$ (2, 1), as defined in 3b and as illustrated in fig. 2, represents the probability of *any* passage from 1 to 2 in which the electron suffered two deflections.

Equation 3a makes no specifications as to where the points 1, 2, and 3 shall be; it asks us to compute values of $K^{(1)}$ (2, 1) for all possible points and includes paths via all intermediate points 3. One *mathematically* possible path is depicted in fig. 3, another in fig. 4. But what is the physical meaning of such diagrams?

In fig. 3 an electron starts out at 1 and goes to 3, moving to the right. At 3 it continues to the right but *goes backward in time.*

In fig. 4 it starts going backward in time and at the point 3 it takes on a reasonable behaviour. A path on which time is reversed, i.e., a leg of a diagram that is directed downward, seems to be obvious nonsense.

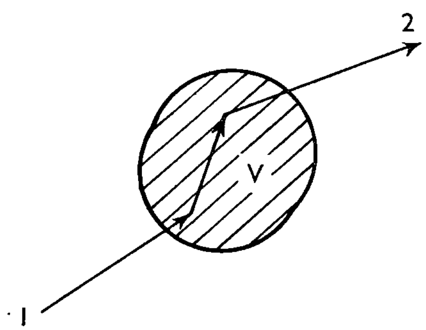

FIG. 2. *Pictorial illustration of $K^{(2)}$. An electron going from point 1 to point 2 is scattered twice on the way by the potential V extending over the shaded region.*

https://doi.org/10.1177/039219215400200605 Published online by Cambridge University Press

At this juncture, however, the theory of relativity has something important to say: *The world line of an electron moving backwards in time represents a positron moving forward.* Thus, if we reverse the arrow on the nonsense leg of fig. 3 and direct it from 2 to 3, that leg represents a positron moving from 2 to 3. The whole diagram, then, depicts an electron and a positron converging toward 3 where, since there are no lines at times later than t_3, they cease to exist. The diagram corresponds to the annihilation of a pair. The reader will have no difficulty in seeing that fig. 4 represents pair production.

The contents of this section, when briefly summarised, might be put as follows. Quantum electrodynamics portrays the motion of electrons as a series of broken passages, zigzag transits from one point to another, with calculable probabilities assigned to each possible leg of a journey. It interprets world lines with times reversed as belonging to positrons. Unless this interpretation is made, many legs whose consideration is required by the mathematical formalism remain devoid of meaning and encumber an otherwise satisfactory theory by their unwanted presence. When time reversals are allowed, the theory becomes correct and powerful in its predictions.

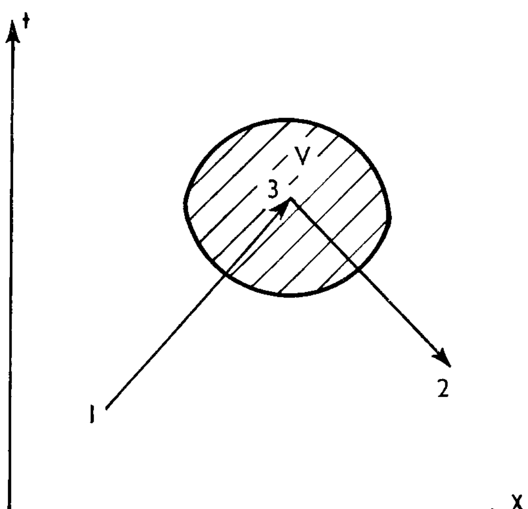

FIG. 3. *Pair annihilation.*

III. *The Causal Status of the Theory*

Diagrams such as those in figures 1 to 4 present no problem to causal analysis. But there are others, encountered in the study of $L^{(2)}$ (2, 1), which are hopelessly acausal when viewed from the standpoint of classical, or Newtonian physics. Consider, for example, fig. 5. A single electron starts from point *A* at time t_A and moves to *B*, where it meets a positron and is annihilated. This positron was created, together with a second electron, at *C*; the time of its birth was later than t_A. The second electron moves on toward *D*. It is interesting to note that the single world line *ABCD* (the fact that it has sharp corners is not significant, for they can be rounded without detriment to our interpretation) represents the fate of three different particles. Between t_A and t_C there exists but one electron, between t_C and t_B there are three particles; after t_B again only one. Observation may not disclose all these events. It will in general tell that an electron starts at *A* and emerges at *B*, for it cannot distinguish one electron from another. The occurrence of pair production and annihilation remains hidden, and the whole process is called 'virtual pair production'.

The Newtonian state of the electron at point *A*, its position and momentum, is completely specified by the component and the slope of its world line at *A*. But the fact that at a time t_C, later than t_A, a pair was produced

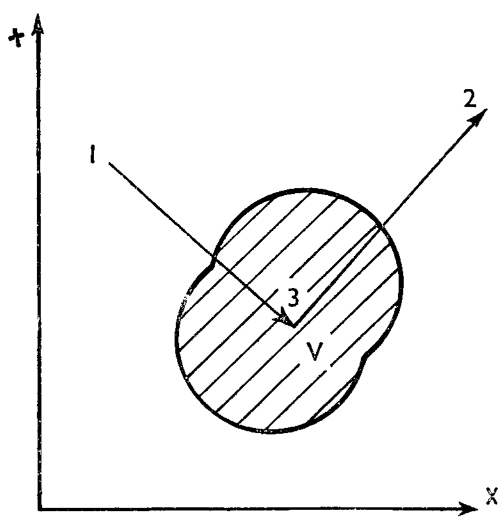

FIG. 4. *Pair production.*

https://doi.org/10.1177/039219215400200605 Published online by Cambridge University Press

with a positron moving to annihilate our electron at B, is not implied by the state at *A*, nor could it have been predicted from any property possessed by the electron at *A*. The birth of a pair at t_C and the motion of an electron at t_A are unrelated events. Hence the diagram does not describe the deterministic behaviour of a single electron starting out at *A*. Causality in the Newtonian sense thus clearly fails.

One might try to restore it by enlarging the physical system, saying that fig. 5 describes the motion not of one particle but of three. Thus if the state of the electron at *A* *and* that of the pair at *C* were known the rest of the diagram could be filled in. This is true. But the principle of causality asks that we specify the state of a system at *one* time only, whereas the procedure in question makes reference to two instants, t_A and t_C. As it stands, then, the diagram is still not causal; to make it so we must cut off its lower portion along a horizontal line through *C*—a part which the theory is not willing or able to sacrifice.

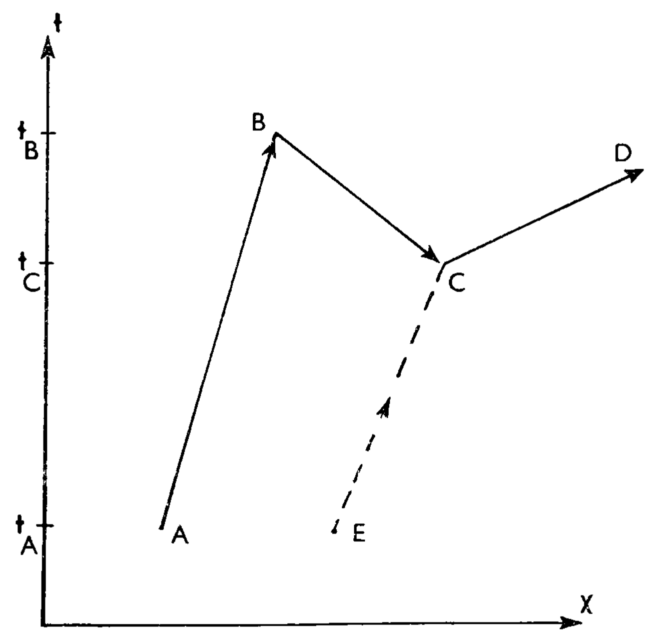

FIG. 5. *The events represented by this world line, when regarded from the point of view of classical mechanics, do not form a causal sequence.*

Finally, it might be supposed that the diagram is incomplete in failing to state the cause from which the pair results. Perhaps it should include another feature, the presence of a photon at E moving toward C. For it may have been this photon that produced the pair at C. While this is a possible conjecture, neither theory nor observation can tell us, at time t_A, that the photon actually will produce a pair at C. The photon may indeed create a pair anywhere along its path or none at all. It is clear that the device of introducing a photon into the diagram, while satisfying our desire for a more complete description of the state at time t_A, does not yield a state in Newton's sense which is connected with a determinate future state through Newton's law or, indeed, through any law we know of. Hence the principle of causality still fails.

No such disaster occurs if the probabilistic interpretation of states is adopted. In that case a single world line says nothing about a positive, actual occurrence; it merely presents a sample of what might occur, a hypothetical instance to which probabilities can be attached. World lines do not lose their meaning, any more than points of space lose their meaning in the more orthodox form of quantum theory. For while this latter theory denies that under certain conditions an electron can be said to be at (x, y, z), it nevertheless needs that point as a peg for its probabilities. In the same way, the Feynman theory denies that a given diagram is a positive portrayal of reality, but it needs that world line as a carrier for its probabilities. Nothing more is implied in the use it makes of the K integrals; it integrates over all Ks to get a state in accordance with equation (2); it regards them as possible samples of what might occur without committing itself in Newtonian fashion as to actual paths. Equation 1 remains its basic law.

That equation induced a new definition of physical states, as we have seen. Quantum electrodynamics leaves this definition unchanged, and those who adopt it must regard it as a causal discipline.[3]

[3] The leisure for reflections that have led to this and other publications was afforded by the tenure of the Hill Foundation Visiting Professorship at Carleton College, for the award of which I am grateful.