International Journal of Theoretical Physics, Vol. 29, No. 12, 1990

Quantum Mechanics in a Discrete Model of Classical Physics

Geoffrey Hemion[1]

Received June 11, 1990

The role of probability theory in classical physics is examined. It is found that the probabilities for the outcomes of typical experiments depend strongly on the assumed behavior of given classical models "at infinity." A discrete classical model is introduced and it is shown that the resulting probabilities are similar to those in the usual theory of quantum mechanics.

INTRODUCTION

On the face of it, there would seem to be many differences between classical mechanics and quantum mechanics. For example, the notion of the "observer" plays a large role in the quantum theory, but it is thought to be unimportant in the classical theory. Furthermore, the classical theory is considered to be sensible and easy to understand, while the quantum theory remains mysterious. The usual ideas of probability theory seem to apply to classical mechanics, but not to quantum mechanics.

But do these common perceptions really reflect the true situation in physics? It is often said that, within classical physics, if we were to know with perfect accuracy the positions and velocities of all particles, then the future course of the universe could be calculated with absolute precision. But who is to do the "knowing"? Is it a person whose brain is also composed of particles? By posing this question, it becomes clear that many of the traditional quantum paradoxes could equally well be applied to classical physics.

Also, much is made of the fact that complex-valued amplitude functions are used in quantum mechanics. But in reality we do not experience these complex functions *directly*. Rather, they are used within the context of a certain mathematical framework to calculate real-valued probabilities for

[1]Fakultät für Mathematik, Universität Bielefeld, 4800 Bielefeld, West Germany.

0020-7748/90/1200-1335$06.00/0

the results of real experiments. In just the same way, real probabilities within the theory of classical mechanics are often calculated by means of involved mathematical reasoning. (For example, it has often been remarked that the complex-valued characteristic function of traditional probability theory can be made to resemble a quantum mechanical amplitude function.)

Contrary to the usual view, classical mechanics has never been shown to follow from simple mathematical postulates. In fact it is, like quantum mechanics, nothing more than an ad hoc collection of vague rules of thumb, combined within a certain general philosophical framework.

But what happens if we decide to apply the more stringent philosophical principles of quantum mechanics to the classical theory? This is an interesting question which seems to have received little attention. Many physicists may believe that there is nothing to be gained by thinking about the old classical theory. However, if one is prepared to consider critically the role of probability theory in classical mechanics, then a possible mechanism for quantum mechanical probabilities soon becomes apparent.

1. A HIERARCHY OF CLASSICAL MODELS

The principles of quantum mechanics are often explained in terms of the famous two-slit interference experiment. A source emits electrons which travel through an apparatus to an absorbing screen. Along the way they must pass through an obstructing screen which has two slits etched into it. Within the classical picture, the probability for the electron going through each of the slits is taken as given, and the total probability for landing at some point on the absorbing screen is the sum of the probabilities for each slit. On the other hand, in quantum mechanics one calculates the complex amplitude for each of the two possibilities, adds them together to give a further complex number, and then the probability is the square of the absolute value of this number.

The "observer" comes into the picture if we allow him to look very closely at the two slits to determine the exact path of the electron. As soon as he becomes successful, then it is found that the "classical," rather than the "quantum mechanical" calculation should be applied. But what is this classical calculation? In reality it is no calculation at all, rather it is the simple assertion that the two events of the electron going through one or the other of the slits are independent of each other. What then is the reasoning which leads to this assertion?

To begin with, classical mechanics is "deterministic." The basic object in classical physics is a mathematical model—a complete set of particle paths in space-time—which describes completely the entire universe. Given this set, then we would know the behavior of all physical systems at all

times, and thus probability theory would be meaningless. The problem is that we do not "know" all the details of this hypothetical set. On the other hand, we do know something. For example, we might know that an apparatus exists in a physics laboratory which behaves like a two-slit experiment. Within the classical picture, this knowledge can be interpreted as some kind of incomplete knowledge about the hypothetical classical model of the universe.

According to the classical philosophy, this model obeys some collection of abstract "laws of nature." One may decide that these laws are to be considered as being of primary importance. If we do this, then we no longer have one single model of the universe, but rather a different kind of "model" which is really just a general framework for many possible "configurations" of the model. The laws of nature are usually described in terms of variational principles. A value is assigned to each *imaginable* configuration, then a configuration with a "best" or "most extreme" value is identified as being an *actual* configuration of the universe.

A variational principle can also be applied in cases where our knowledge of the universe is incomplete. The information which we have might be compatible with many different configurations of the model, and each of them could plausibly be an actual configuration of the universe. Since our knowledge is incomplete, we are forced to consider all of the *possible* configurations which are compatible with our incomplete knowledge of the universe. We have no reason to say that any one of these possible configurations is more likely than any other, so this gives the rule that all of the possible configurations should have an equal statistical weighting.

Seen in this way, it becomes clear that classical physics does not really require the concept of a single deterministic model for everything in the universe. Instead, we can substitute a kind of hierarchy of increasingly more comprehensive models, representing a sequence of increasingly detailed specifications for a given experiment. There are a number of different ways of doing this. One possibility is the following.

Let us begin by specifying that "classical physics" is to consist of a collection of timelike particle paths (or "world-lines") in four-dimensional Minkowski space M^4. Denote this set of paths by Γ. A typical element $\gamma \in \Gamma$ is a mapping $\gamma: \mathbf{R} \to M^4$ of the real numbers, considered as an ordered set, into the space M^4, which also has the structure of a partially ordered set. "Timelike" then means that the mapping γ is order preserving. It may be convenient to associate further numbers—e.g., "mass" or "electrical charge"—with each particle path.

Let us imagine that we are interested in calculating the probabilities for the various possible outcomes of some particular experiment. Now it is reasonable to imagine that the experiment is located in some finite region

of space-time. The specification of the experiment represents our limited and incomplete knowledge of the situation in this finite region. To be specific, denote the region by $U_0 \subset M^4$. We will assume that U_0 is compact.

By restricting our attention to U_0, we can also restrict our attention to the intersection of the paths in Γ with U_0. In general, it is sensible to assume that only finitely many paths in Γ intersect any given compact set $C \subset M^4$. Thus, we can take Γ_0 be the finite set of paths in Γ meeting U_0. Furthermore, the paths in Γ_0 are only defined as mappings from intervals of $\mathbf{R}$ to U_0. In order to know the result of the given experiment, and in fact to know everything about this region of space-time, it is only necessary to know everything about the finite set Γ_0.

But the problem of trying to understand Γ_0 is itself a kind of "experiment," which is somewhat more ambitious than the original, given experiment. To understand this new experiment, we take a larger, compact set $U_1 \subset M^4$, with $U_0 \subset U_1$. Let the finite set of paths Γ_1 be defined to be the intersection of Γ with U_1, in analogy to the case of Γ_0.

Going a step further, we see that the problem of knowing Γ_1 is a further experiment. By continuing this process, we obtain a sequence $U_0 \subset U_1 \subset U_2 \subset U_3 \cdots$ of compact neighborhoods of the given experiment, and a sequence $\Gamma_0 \subset \Gamma_1 \subset \Gamma_2 \subset \Gamma_3 \cdots$ of finite particle paths, with Γ_i being the intersection of the paths in Γ with the set $U_i \subset M^4$. Any such sequence of neighborhoods will be called "proper" if for any given compact set $C \subset M^4$, there exists some index $i \geq 1$, with $C \subset \Gamma_i$.

But note that our definition of the set Γ_i, for each $i \geq 1$, depends on a complete knowledge of the hypothetical *actual* universe Γ. It is more realistic to admit that we do not know what Γ is. Instead, for each U_i there is a whole collection Ξ_i of possible sets of paths in the neighborhood, representing the intersection with U_i of all the possible configurations of the model which are compatible with our knowledge of the situation. Thus, $\Gamma_i \in \Xi_i$, but in general there may be many further finite sets of paths in Ξ_i as well.

Let us call the given experiment E. One might imagine that E is a property of a given set of particle paths; namely the set describes a universe containing the experiment E. Thus, each of the paths in Ξ_i, for each i, has the property E. An experiment would be meaningless if there was only one possible result. Therefore let P and Q be two different results of E. Denote by $E(P)$ or $E(Q)$ the experiment with the result P or Q, respectively. Assume that the two results are mutually exclusive, so that a set of particle paths can have either the property $E(P)$ or $E(Q)$, but not both properties.

The main question for us is to calculate the relative probabilities of the two results P and Q. Within each neighborhood U_i, it may be possible to calculate the probabilities of these two results. We obtain two numbers, $\mathcal{P}_i(P)$ and $\mathcal{P}_i(Q)$, which we call the "probabilities with respect to U_i."

There is no reason to suppose that $\mathcal{P}_i(P) = \mathcal{P}_j(P)$ for $i \neq j$. However, we might hope that a limit exists $\mathcal{P}_i(P) \to \mathcal{P}(P)$, as $i \to \infty$, for each possible result P of the experiment E. If there are limiting probabilities for all possible experimental results, and for all possible proper sequences of neighborhoods, then the experiment itself will be called "proper." From now on it will be assumed that all experiments which we consider are proper.

2. PROBLEMS WITH CLASSICAL PHYSICS

The great problem in classical physics is that we have no sensible way of dealing with infinite sets of particle paths. Imagine some commonplace event in the physical world: for example, the passage of an electron through a vacuum tube. The path might be curved, and we attribute this to the idea that it passes through an "electrical field." But where does the electrical field come from? Surely it comes from the effects of other particles in the physical world. And then these other particles are influenced by still more particles, and so on. Even such a great mathematician as Gauss struggled to find a way of dealing with these infinite-particle spaces, but without success.

A way out of this conceptual dilemma was suggested by Maxwell. His solution, which is now considered to be the standard framework for classical physics, was to think of the idea of electromagnetic fields as being of primary importance. We should no longer concern ourselves with the idea that the fields come from specific events in the past; instead, the fields are simply specified as being "there." This idea has indeed proven to be very practical, but by the same token, it sometimes leads to rather bizarre ways of thinking. Consider, for example, the standard derivation of the laws of black-body radiation. How strange it is when modern physics textbooks suddenly assert that we are living in a boxlike world whose flat walls are able to reflect electromagnetic radiation perfectly! But even beyond this, the idea of abstract fields leads to great mathematical problems. For example, Dirac has shown that the infinite "field" of a pointlike charged particle will generally lead to an exponentially increasing self-acceleration of the particle.

Few people today give much thought to these old and well-known problems in the foundations of classical physics. Instead, people occupy themselves with new problems in the modern quantum physics. These problems are also concerned with the relationships of "fields" and "particles." The modern physicist imagines that his or her methods are very sophisticated, and that the intellectual struggles of the old classical physicists are irrelevant to current work. But could it be that the new problems are really nothing more than the old problems, formulated in a strange way so that the connection has become obscured?

These questions are very important if we are to make any progress in understanding the role of probability theory in classical physics. For example, we could investigate the statistical properties of a finite, boxlike universe with mirrors for walls, containing abstractly given "electromagnetic fields." Such an investigation might lead to certain results which we would then loosely describe as the "classical statistics." But what would be the value of such work? Surely it is not enough to simply specify the electromagnetic fields in some arbitrary way. This is so because the fields come from further particles, and those particles are also involved in the statistical calculation. One might hope that the easy calculations using the "boxes" might lead to the same results as would be obtained when a more accurate method is used. But in reality this is not the case. In fact, there are important statistical effects which become lost when using the box method.

3. EQUIVALENCE CLASSES OF PARTICLE PATHS

We are concerned with calculating the relative probabilities for the different outcomes of given experiments. Let some neighborhood U_i in a hierarchy be given, together with the collection Ξ_i of possible sets of paths in U_i. Our task is to count up the number of elements of Ξ_i which have the property $E(P)$, say. At the moment we are using the traditional framework of classical physics; that is, we assume that the particles are continuous paths in a continuous space-time geometry. Of course within this framework one would generally expect to have an infinite number of different elements of Ξ_i, and there is no simple way of counting "infinity." Thus, we are confronted here with another of the traditional problems of classical physics; how should we deal with the infinite numbers of possible paths which are allowed by the classical model? The standard procedure is to define some measure on the space of possible paths. For example, Wiener measure is a popular choice. But if we do this, then we again make many "hidden" assumptions about what we believe "classical physics" is supposed to be—in a manner very similar to the use of "electromagnetic fields" in hypothetical closed boxes somewhere out in space-time. A better way of proceeding is to change from continuous paths to discrete paths, for example, as in Hemion (1988). There are many possible discretization schemes which could be reasonably considered. The main point is to proceed in such a way that we are guaranteed of always having finite sets Ξ_i. For the moment, though, the details of this discretization process are unimportant.

The important thing is that for each index i, we can define an equivalence relation between the elements of Ξ_i. Let $k > l$ be two given indices. Assume $\Theta \in \Xi_k$. Then, since $U_l \subset U_k$, we must have that $\Theta|_l \equiv \Theta \cap U_l$

is an element of Ξ_l. The set of paths $\Theta|_l \in \Xi_l$ will be called the *intersection* of Θ with with U_l. Of course, it is possible to have two elements $\Theta \neq \Omega$ in U_k, with $\Theta|_l = \Omega|_l$ in U_l. On the other hand, let Φ, $\vartheta \in \Xi_l$. We will say that they are *i-equivalent*, written $\Phi \approx_i \vartheta$, if for each possible index $k > i$, there exist two elements Φ^k, $\vartheta^k \in \Xi_k$, with $\Phi^k|_l = \Phi$ and $\vartheta^k|_l = \vartheta$, and yet $\Phi^k \cap (U_k - U_i) = \vartheta^k \cap (U_k - U_i)$. That is, within the framework of the quantum philosophy, we have no way of "distinguishing" between Φ and ϑ outside the neighborhood U_i. In particular, we can assume this means that if P is some particular result of the experiment E, then either both Φ and ϑ do, or do not, have the property $E(P)$ together.

The relation "$\approx_i$" is clearly an equivalence relation. But can we assert that this seemingly complicated definition makes much sense? For example, one could ask whether it is possible to have two distinct sets $\Phi \neq \vartheta$ which are elements of some Ξ_l, and yet $\Phi \approx_i \vartheta$, for some sufficiently large i. To answer this question in a rigorous manner, it would be necessary to make very specific assumptions concerning the details of the assumed "laws of nature" which one would like to postulate. Furthermore, assumptions would have to be made about possible restrictions on the class of configurations of the model which are to be brought into consideration. It is obvious that such an investigation would lead to a mass of uninteresting technical results, thereby diverting our attention from the main question. Surely the best way to proceed is to simply assert that the concept of "classical physics" should be defined in such a way that arbitrarily given (homogeneous) electromagnetic fields in a confined region C (in M^4) can be generated by possible particle configurations in "distant" regions of space-time (that is, in $M^4 - C$). Given this, then it must follow that the i-equivalence classes generally contain many distinct elements.

The reason for concentrating on the i-equivalence classes is that it gives us a method for counting and comparing the number of configurations of the classical model which represent different outcomes of the experiment E. It is sensible to first look at the possible equivalence classes, considering them as entities on their own. As we will see, classical mechanics is very similar to quantum mechanics in this respect. Then we estimate the number of elements in each equivalence class in order to calculate their relative probabilities.

There is also a further point, which has received too little attention in the traditional investigations of classical physics. Consider the very idea of a "variational principle." There are two different ways of looking at this idea. (I) Given a specification configuration X of the classical model, we might "vary" it to obtain a different configuration X^*. Only such "variations" of X might be admitted for comparison with X. (II) The other method is to say that we are allowed to compare *arbitrary* configurations

X and Y with one another to determine if X is most extreme. Many people might, on first thought, assert that classical physics operates according to the rule II, but of course that is quite false. The standard Lagrangian variational principles involve, among other things, the total lengths of the particle paths in the given configuration of the model. But this total length is simply infinity, and so no sensible comparisons can ever be made! In reality classical physics is based on variational principles of type I, and the varied configurations X^* can only differ from the original X within some compact region of space-time. That is to say, classical physics only makes sense within the context of *individual equivalence classes* of configurations of the model, where *equivalence* is now taken to mean the identity of two particle path spaces outside some given compact set in M^4. In classical physics it is nonsense to attempt to compare *different* equivalence classes with one another in any detailed way.

4. DEALING WITH EQUIVALENCE CLASSES

Let $\Psi, \Lambda \in \Xi_0$ be any two path spaces in the smallest neighborhood U_0 of the given experiment E. If we continue to assert that arbitrary electromagnetic fields can be generated by charged particles in distant regions of space-time, then it must be the case that for some sufficiently large i, we have $\Psi \approx_i \Lambda$. Therefore, it would make little sense to define an equivalence relation by saying that two elements Ψ and Λ are equivalent if $\Psi \approx_i \Lambda$, for some sufficiently large i. Such an equivalence relation would be trivial in the sense that all elements would be equivalent to one another. Therefore, it is necessary to argue in terms of an i-equivalence relation for various fixed values of the index i.

For fixed i, there is a simple condition which must be fulfilled if $\Psi \approx_i \Lambda$ for two given elements $\Psi, \Lambda \in \Xi_i$. Since $\Psi^k \cap (U_k - U_i) = \Lambda^k \cap (U_k - U_i)$, for all $k > i$, it must follow in particular that Ψ and Λ have the same number of path intervals in U_i, and furthermore, the ends of corresponding paths meet the boundary of U_i in identical points. Therefore, we could expect to have many different i-equivalence classes in Ξ_i.

This whole concept of dealing with the i-equivalence classes is very much removed from the usual approach in classical physics. We are used to thinking about the properties of individual configurations of the model, not equivalence classes of configurations. Is it possible to say that each i-equivalence class is "near to" some definite configuration, thus allowing us to imagine that the equivalence class is really just this configuration, together with some class of variations of it?

Let Σ be some i-equivalence class. Σ consists of a collection of particle path spaces, which are possible solutions to the given variational principle.

But of course it is impossible for us to have complete knowledge of the whole universe. Thus, as a practical matter, it would make sense to try and apply the variational principle to the individual elements of Σ, but confining the calculation to the parts of the paths *within* U_i. For example, one could assume that all elements of Σ are extended outside of U_i by a single, common set of paths, which are defined *throughout* the set $M^4 - U_i$. In this way it would be possible to compare the different elements of Σ, and we may assume that the variational principle is such that *one single* element of Σ will be found to be most extreme. Let Γ_Σ be this single most extreme element of Σ. Of course, one should remember that Γ_Σ is a set of particle paths in U_i, and the other elements of Σ are also sets of particle paths in U_i which all meet the boundary of U_i in the same way as Γ_Σ. Thus, these other elements of Σ are indeed "variations" of Γ_Σ in a very concrete sense.

In the sequel, then, we will often be dealing with i-equivalence classes of particle paths. But given such an equivalence class Σ, we will imagine that it has definite geometric properties, as given by the single most extreme element Γ_Σ.

5. CALCULATING PROBABILITIES IN A HIERARCHY

For some given outcome P of the experiment E, we must try to find a method of calculating the probabilities $\mathscr{P}_i(P)$ of P with respect to U_i. We are assuming that some discretization technique is being used to ensure that the number of different possible configurations of the model in U_i—that is, the number of elements in the set Ξ_i—is always finite. This allows us to use simple counting arguments. Once we have calculated the number $\mathscr{P}_i(P)$, it is then necessary to calculate the probabilities $\mathscr{P}_j(P)$, for all $j > i$. The true probability $\mathscr{P}(P)$ is defined to be the limit as $j \to \infty$. But the property P is always determined within the original neighborhood U_0. Thus, $\mathscr{P}(P)$ is determined by counting the numbers of configurations $\Phi \in \Xi_j$ for increasingly large j, but then concentrating on the behavior of such Φ in U_0, that is, concentrating on $\Phi|_0$.

The central idea, which cannot be overemphasized, is that within the context of classical physics it is possible to compare configurations of the model *within a single equivalence class*, but it makes no sense to attempt to compare inequivalent configurations with one another. How can this idea be applied to the calculation of probabilities for the results of experiments in classical physics?

To begin, one should recall the basic thing which distinguishes classical from quantum physics. This central difference is that each possible configuration of a classical model is taken to be independent of all other

configurations, while in quantum physics the different configurations "interfere" with one another in some mysterious way. Now we would agree that in classical physics, *inequivalent* configurations are independent of one another. But we will assert that the same cannot be said for *equivalent* configurations.

The rules for calculating probabilities within equivalence classes will be dealt with presently. For the moment, though, we should concentrate on the different equivalence classes. But we have seen that each equivalence class can be represented by its single most extreme element. Thus, we can think about the set of all such most extreme elements and assert that they are all independent of one another. This set of most extreme elements could be denoted the set of "classical paths" for a given experiment. According to this definition (which involves the use of quotation marks), each of the *other* possible configurations of the model are not "classical paths"; rather, each is a variation of some specific "classical path."

For example, we could think about the two-slit experiment. It is sensible to assert that all of the "classical paths"—that is, the most extreme configurations in each equivalence class—are independent of one another. Thus, each "classical path" goes through one or the other slit. In either case, we obtain a Gaussian distribution on the absorbing screen, and the total probability for all the "classical paths" is the sum of these two distributions. But of course this is not the total probability for the experiment as a whole.

It is not enough to simply concentrate on the set of "classical paths," which after all has been chosen from the whole set Ξ_i in a rather arbitrary way. We must count *all* of the elements of Ξ_i. That is to say, we must calculate the numbers of elements in each equivalence class. Then the statistical weight of each "classical path" should be taken to be proportional to the number of elements in its equivalence class.

6. STATISTICAL PROPERTIES OF DISCRETE PATHS

Until now we have asserted that the basic model for classical physics should involve spaces of particle paths, which are defined as *continuous*, order-preserving mappings from the real numbers **R** into Minkowski space M^4. The idea that space-time is continuous does indeed give the traditional framework of classical physics. But it is easy to argue that this continuity hypothesis is precisely the point where the traditional classical physics goes badly wrong. Which experiments could possibly disprove the contrary hypothesis, that space-time is discrete? Furthermore, the great unresolved mathematical problems of classical physics stem precisely from the hypothesis of continuity. It may be fashionable to investigate the Byzantine

complications of possible "singularities" in a continuous classical model of space-time, but in the end, what do such investigations have to do with "physics"—that is, the study of the true physical world of human experience? Finally, it is clear that quantum mechanics involves discrete, rather than continuous phenomena. The problem of trying to fit a discrete quantum theory into the traditional continuous framework of classical physics presented immense mathematical problems to the founders of quantum mechanics at the beginning of the century. After many years of intellectual effort, they finally succeeded in finding discreteness in the solution sets of certain complex differential equations. Therefore, modern physics attempts to deal with quantum (i.e., discrete) phenomena in terms of the traditional continuous framework of the old classical physics. It is certainly astonishing that such an idea almost appears to work. Of course the standard problems of singularities in continuous space-time (in the guise of "renormalization theory") continue to occupy many theorists even now. But surely anyone who is prepared to think about the overall development of modern physics would agree that a discrete mathematical description would be more appropriate.

How should we define such a discrete mathematical model? In Hemion (1989) I argued that our experience of the world involves a process of gathering new information, based on already given information. That is, our experience can be described in terms of "experiments." But each experiment has a definition, which involves the known conditions of the experiment. I argued that the set of such "known conditions" could be imagined to have an abstract geometric structure, which we could call "space-time." But this would mean that it is impossible to devise an experiment to determine the detailed properties of space-time; after all, no experiment can be devised to test the experimental setup itself! This means that the details of any discrete model for physics cannot be tested using practical real-world experiments. Perhaps, then, there may be many different possible discrete models for physics which are all equally valid, in the sense that they all describe with complete accuracy the relative probabilities for the results of physical experiments.

One possibility for such a discrete model might be obtained by retaining the continuous space-time M^4 and replacing the continuous particle lines $\gamma_c: \mathbf{R} \to M^4$ by discrete particle lines $\gamma_d: \mathbf{Z} \to M^4$, where $\mathbf{Z} \subset \mathbf{R}$ is the set of integer numbers. The discreteness condition is that the Lorentz distance in M^4 between adjacent points on γ_d is a constant (which can be related to the "physical mass" of the particle). In addition, it is necessary to replace the continuous variational principle by a new variational principle which is relevant for such discrete paths. Perhaps the simplest idea would be to simply rewrite the traditional Lagrangian expression for continuous paths

by replacing each term involving an integral with a term involving the corresponding discrete sum, for example, as in Hemion (1988).

But regardless of the details of such a process, it is easy to see that the discrete model will involve new statistical effects. For, let $\Phi \approx_i \Lambda$, where $\Phi, \Lambda \in \Xi_i$, and Ξ_i is now a set of discrete particle path spaces. The condition that $\Phi^k \cap (U_k - U_i) = \Lambda^k \cap (U_k - U_i)$, $\forall k > i$, means that the discrete set of points of Φ^k must be identical with the discrete set of points of Λ^k outside U_i. Furthermore, the condition that the Lorentz distance between adjacent points on the particles is constant provides a strong restriction on the possible paths Φ and Λ through U_i.

Consider, for example, the two-slit experiment. One might choose the first neighborhood U_0 in such a way that its boundary ∂U_0 contains both the particle source and the absorbing screen, as shown in Figure 1. Then an equivalence class Σ in Ξ_0 is such that the (Lorentz) lengths of all the particle paths in it are exact multiples of some fundamental constant.

Let P and Q be two points on the absorbing screen. P is a point of constructive interference according to the usual calculation in quantum mechanics. That is, let $L(P, l)$ be the Lorentz length of the most direct path from the source to P through the left slit, and let $L(P, r)$ be the length through the right slit. Then $L(P, l) - L(P, r) = 0 \bmod \Delta$, where Δ is some fundamental physical constant, which can be related to Planck's constant. Similarly, Q is a point of destructive interference, i.e., $L(Q, l) - L(Q, r) = \Delta/2 \bmod \Delta$.

Can we assert that, as in quantum mechanics, the probability of P with respect to U_0 is greater than the probability of Q? That is, $\mathscr{P}_0(P) > \mathscr{P}_0(Q)$? Surely this is a reasonable assertion. After all, our previous reasoning must lead to the conclusion that the number of "classical paths" (representing 0-equivalence classes) near to P is about the same as the number of "classical paths" near to Q; such paths are independent of one another, and thus we

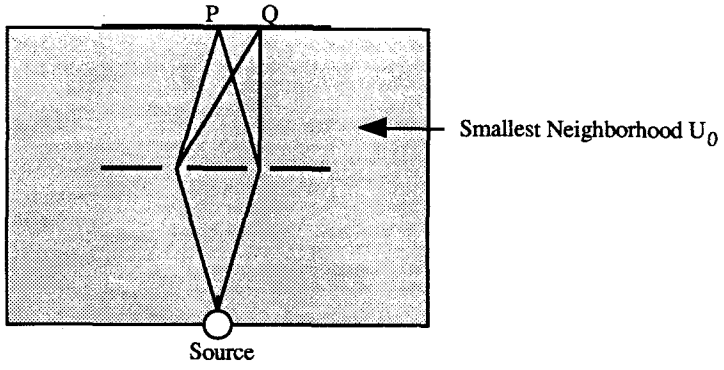

Fig. 1.

can apply the classical reasoning to them. Furthermore, the geometry of the experiment is such that the equivalence classes going through P must have more elements than the equivalence classes going through Q.

But note that this last statement needs further justification. It would follow if we can justify the following two assertions: namely (A) that the "classical path" representing an equivalence class in the two-slit experiment follows nearly straight lines from the source to the slits, and from the slits to the absorbing screen, and (B) that most of the elements in an equivalence class also follow nearly straight lines. Certainly assertion A is true for reasonable choices of the Lagrangian. But B is not so clear.

Although we have made a number of definitions and assumptions, in reality we have been avoiding the main problem of classical physics, namely the fact that it is impossible to deal with infinite particle spaces. At some point it is necessary to make definite assumptions concerning the behavior of these particle spaces "at infinity." The assumption we make is that the background electromagnetic fields—the "cosmological background," if one will—is weak and random. In the end, this assumption could be translated into a condition on the classical "model" which we are considering. But it would be pointless to try to find some complicated technical condition involving the variational principle, the class of allowed configurations of the model, etc. In reality, the assumption can best be stated in the form B above.

Given that the probability of P is greater than the probability of Q with respect to U_0, what can we say about the probabilities with respect to U_i, for $i>0$? One might at first believe that for large i, the effect which we have just described becomes small. For example, one might think that, as in the traditional view of the "classical probabilities," we might have $\mathscr{P}_i(P)\to\mathscr{P}_i(Q)$ as $i\to\infty$. But consider the following idea. Let Σ be a 0-equivalence class. Then it is a simple consequence of our definitions that for each $i>0$, there is an i-equivalence class Σ^i in Ξ_i containing Σ, in the sense that the intersection of each element of Σ^i with U_0 is an element of Σ.

The problem now is to compare the probabilities of P and Q with respect to U_i. But recall that the two different results P and Q are assumed to be mutually exclusive. The "observer" sees that either P or Q is the result, and perhaps writes this down on a piece of paper, or whatever. In any case, it is *impossible* to have an i-equivalence class containing both paths going through P and through Q. Within our framework, then, this is the true meaning of the concept of an "observer."

But now we have that the i-equivalence classes either go through P or else they go through Q. Let Σ_P and Σ_Q be typical such i-equivalence classes, going through P and Q, respectively. We can divide up the paths in Σ_P into three segments, namely the first segment from ∂U_i up to the source, then

the second segment from the source to P, and finally the third segment from P to ∂U_i. Also, the paths in ΣQ can be divided into three segments, but this time going through Q. Now the number of different segments from ∂U_i to the source is the same, both for Σ_P and Σ_Q. Similarly, we can assume that the number of different segments from P to ∂U_i and from Q to ∂U_i is the same, since the specification of the two-slit experiment gives no particular reason to favor one or the other route. Finally, we have the middle segments, but here the result P is favored, according to the ratio of $\mathscr{P}_0(P)$ to $\mathscr{P}_0(Q)$. In summary, then, the ratio of $\mathscr{P}_0(P)$ to $\mathscr{P}_0(Q)$ should be expected to be similar to the ratio of $\mathscr{P}_i(P)$ to $\mathscr{P}_i(Q)$ for all $i>0$.

7. FURTHER ASSUMPTIONS CONCERNING THE STATISTICS OF DISCRETE SPACES

The considerations of the last paragraph show that a discrete model for classical physics can have statistical properties very much like those in the theory of quantum mechanics. But there is more to quantum mechanics than this. It may be true that P is favored over Q in the context of an experiment where P and Q are mutually exclusive results. The existence of the "observer" ensures, by definition, that paths through P cannot be in the same equivalence class as paths through Q.

Going beyond this, quantum mechanics deals with larger experiments, where possible paths leading to larger results can either go through P or through Q. That is, one imagines that the "observer" decides not to look, and so the paths through P and Q can "interfere" with one another. Now it is true that all of this quantum reasoning can easily dissolve into the traditional, spongy quantum "philosophy." But the one *precise* condition of quantum mechanics is that the *same* amplitude function should describe the situation, both in the case that the "observer" decides to look and in the alternative case that he decides to look away.

Can the same condition be shown to hold in discrete classical physics? For this it is necessary to make a number of further general assumptions concerning the experimental process.

(A) Let any finite collection of spaces of finite particle paths be given in some compact region of space-time. Then the first assumption is that there is an "experiment" which admits only spaces in this collection. Put another way, the condition is that the concept of an "experiment" can be completely defined by simply specifying in a compact region the set of possible configurations of the model which are to contain the experiment, i.e., an "experiment" *is* nothing more than a finite list of possible configurations of the model.

(B) Let us say that in a given experiment, a single particle can follow a number of possible well-defined paths to different endpoints; for the sake of argument let P and Q be two possible endpoints which are unrelated to one another (i.e., neither $P < Q$, nor $Q < P$). Assume that the particle has a periodic structure which repeats itself at equal Lorentz intervals of length Δ. Assume further that there are an equal number of possible paths leading to P and to Q in such a way that, according to the "classical picture," there is nothing to favor one result over the other. We can then assert that the probability of P is greater than the probability of Q if the paths leading to P have lengths differing from one another by a multiple of Δ, while the paths leading to Q do not have this property.

(C) Let χ and ζ be two possible paths through the experiment which can be compared with one another. That is, χ and ζ are in the same equivalence class. Then the important thing when making the comparison is that we should look at the difference in the path lengths, modulo Δ. This is a number which we call the *phase difference* between χ and ζ. These phase differences only involve particle pairs. Clearly, the number of path spaces in an equivalence class is influenced by the phase differences between pairs of possible particle paths through the experiment. Thus, the probabilities in the experiment are to be calculated by examining the set of phase differences between all possible *pairs* of particle paths.

(D) Take χ and ζ as above, and assume that they lead to one outcome P. The "influence" of the pair (χ, ζ) is such as to give a greatest possible increase in the probability of P if the difference in the path lengths is an exact multiple of Δ. On the other hand, the influence of (χ, ζ) gives a smallest increment to the probability of P if the difference in the path lengths is $\Delta/2$ modulo Δ.

(E) Let us now use the the letter Ξ to denote the set of all possible paths in the experiment. We define an "influence function" $F: \Xi \times \Xi \to \mathbf{R}$ to document the effect of the set of pairs on the probabilities of different experimental outcomes. Let P be one possible outcome, and let $\Xi_P \subset \Xi$ be the set of paths leading to P. Then it will be assumed that the "quantum mechanical" part of the probability of P, i.e., the part which is exclusively due to the discreteness of the underlying model, is given by

$$\mathcal{P}(P) = \sum_{\gamma, \xi \in \Xi_P} F(\gamma, \xi)$$

(F) Note that it is possible to return to the traditional "classical" calculation of probabilities if one simply chooses $F(\gamma, \xi) = 0$, for $\gamma \neq \xi$. But these classical probabilities do not describe correctly the quantum world. At the moment we are only interested in the "quantum" component of the influence function; therefore we will assume that $F(\gamma, \xi)$ depends only on the phase difference, modulo Δ, between the path lengths of γ and ξ.

(G) Thus, we can equally well consider F as being a mapping $F: [0, \Delta) \to \mathbf{R}$. Of course the number Δ, representing the constant Lorentz distance in the discrete structure of a given particle, provides us with a basic unit of measurement. We can choose our units of measurement in such a way that the number Δ appears to be simple. For example, choose the basic units of space-time so that $\Delta = 2\pi$. Then to simplify matters further, consider F as being a function $F: [-\pi, \pi) \to \mathbf{R}$.

(H) When thinking about our previous arguments, it is clear that F must have a maximum at 0 and a minimum at π. Furthermore, we will assume that F should be continuous and it should be monotone between 0 and π. Finally, an obvious symmetry is $F(\vartheta) = F(-\vartheta)$, for all $\vartheta \in [-\pi, \pi)$. This reflects a symmetry in the definition of all discrete particle path models.

(I) One experiment can be considered as being embedded within another. Eventually it is possible to imagine arbitrarily large experiments, containing any given experiment which we might practically consider. Thus, the function F will be assumed to have a universal property: it is the same for all possible experiments.

Note that the assumption of continuity rules out a special weighting for pairs of the form $F(\gamma, \gamma)$. This means that we are ignoring "classical" effects, which would be concerned with properties of individual paths. Instead, F is only concerned with the "quantum" statistics which are due to the effects of pairs of paths. Perhaps a most general framework would allow "singularities" in the definition of the function F—reflecting "classical" effects—and then a continuous and monotone F, as in property G above, to reflect the "quantum" effects. But since we are only concerned with explaining the quantum part, it is sufficient for us to ignore the "classical" effects, and this means we may assume that F is continuous.

8. CAUSALITY

The idea of cause and effect plays a large role in physics. Any physical process is the cause of various disturbances in space and time which influence further processes in the future world. One might say that any model for physics which is based on the idea of a partially ordered set must satisfy the condition of "causality." The ordering itself can be thought of as being a "causal relationship," and so philosophical speculation on the idea of "time" is reduced to an elementary mathematical definition. But if we were simply to *define* time in this way, then we would ignore important aspects of the experimental process itself.

Given that physics is concerned with the prediction of probabilities for the outcomes of experiments, then the idea of cause and effect in physics should also be concerned with a condition on these probabilities. However,

in saying this, the condition becomes obvious: it is that the probability calculated for a given experimental outcome (meaning a point in the partially ordered set) should only depend upon the specifications of the experiment in the past with respect to that point. We will call this the condition of *strong causality.*

This definition may seem to be quite simple and natural, but in fact it is extremely restrictive. Our analysis of discrete statistics has shown that the quantum mechanical probabilities are very strongly influenced by the conditions in the future world, i.e., the phase differences between possible paths at various points in the future. Thus, the condition of strong causality is the requirement that this influence must be "masked" in such a way that we are unable to see it. What does this mean? Essentially it means that the discrete effects in the statistics must be completely "smoothed out." Technically speaking, it means that the statistics must be such that they can be calculated using the complex "amplitude functions" in the traditional description of quantum mechanics. The next few sections will be devoted to establishing this interesting new result.

The condition that *probabilities* obey a causality condition might be thought of as being stronger than the basic requirement that all models for physics should be partially ordered sets. It is easy to see that a model which is not a partially ordered set leads to paradoxes which are impossible to resolve, and thus it should be dismissed. But what about the more general case of *probabilities* being influenced by conditions in the future? It is no longer possible to cite, for example, the absurdity of an experimenter who murders his parents. Perhaps the parent who learns that there is an increased *probability* of the child being a murderer will thereby become influenced to violence! Thus, mere probabilities do not lead to definite paradoxes. Nevertheless, the condition of strong causality is just as necessary as the simpler condition of elementary causality. The fact is that we are dealing here with the concept of probabilities. Probabilities must be calculated using *all* the information available to the experimenter. Thus, it is nonsense to say at first that the probability has been calculated to have some value, but then after the experiment has been performed, to declare that one was really interested in some other experiment, and one had meant to calculate the probabilities differently!

9. STRONG CAUSALITY IN PARTICLE PATH SPACES

An appropriate technical condition for the concept of strong causality can be formulated as follows. What is an "experiment"? It is a finite collection of paths, forming a partially ordered set. But we may also consider

another experiment. This is another finite collection of paths. The collection may contain some points and paths of the original experiment, thus allowing various points of correspondence between the different experiments. How should the probabilities for different outcomes of an experiment be calculated? Keeping in mind the conclusions of Section 7, I will assume that if an experiment involves determining whether or not a particle travels from a point p to a point q, then the probability of this event occurring is to be calculated by considering the set Ξ of all possible paths in the experiment which contain both p and q. (Note that we are assuming that $p < q$ here. The points p and q are very much different from the points P and Q—representing alternative results of a given experiment—which were used in the previous arguments.) Specifically, there exists some function $F: \Xi \times \Xi \to \mathbf{R}$ such that the probability $\mathcal{P}(p, q)$ of a particle going from p to q is given by

$$\mathcal{P}(p, q) = \sum_{\gamma, \xi \in \Xi_{p,q}} F(\gamma, \xi)$$

where $\Xi_{p,q}$ is the set of all possible paths in the experiment from p to q.

When considering particle interactions in classical physics, the total action on a particle is taken to result from the sum of all pairwise interactions with the other particles. But we should not make the mistake of interpreting $F(\gamma, \xi)$ as a measure of such an interaction. Remember that γ and ξ represent two different possible paths through a given experiment, that is, two different configurations of the underlying model of the experiment. Of course, these different possibilities cannot interact directly with one another. The argument of the present section is formulated in terms of the particle paths of noninteracting particles. But this is not to say that it is irrelevant for describing interacting particles. In the interacting case, a "particle path" could be taken to be a Feynman diagram. The finite set of space-time points where interaction occurs in such a diagram can be thought of as building the partially ordered set.

The condition of strong causality can now be stated as follows. We have the given experiment; if another experiment is also considered which contains the original experiment, then within the larger experiment the probability of a particle going from p to q can be calculated by adding up the statistical weightings of the paths containing the points p and q. The model will be said to satisfy the condition of strong causality if—under the assumption that we have no particular knowledge about the larger experiment which would tend to favor one or another of the possible results of the given, original experiment—then the probability which is calculated in the larger experiment must always be the same as that which is calculated using the expression above.

10. STRONG CAUSALITY PLACES A CONDITION ON THE INFLUENCE FUNCTION

Let three points p, q, $r \in M^4$ be given with both $p < q$ and $p < r$, and such that q and r are not related to one another in the causal ordering. We shall imagine that corresponding to *any possible* finite set Ξ of paths from p to q or r there exists an experiment containing just those paths, which can be denoted $E(\Xi)$. Within the framework of particle path spaces, this is surely the most general possible way to describe the idea of an "experiment" which has two possible different outcomes. More general experiments, with many different possible outcomes, can be constructed from these elementary experiments.

We will now set ourselves the task of describing a very special class of "experiments" which are designed to illustrate the use of the principle of strong causality. For these experiments we will take three points p, q, and r as above and imagine them to be fixed. The simplest such experiment consists of just two straight paths connecting p to q and r, respectively. Let Ξ be the set consisting of these two paths.

In addition to Ξ there are many more experiments in the special class. A typical experiment will be denoted by $E(\Xi_{\neq})$. Let two further points $x < p$ and $y > \{q\} \cup \{r\}$ be given. Then $\Xi_{\neq}$ is a set of paths from x to y such that each path in Ξ is contained in a path in $\Xi_{\neq}$, and conversely any path in $\Xi_{\neq}$ which passes through q or r corresponds with a path in Ξ. In addition, $\Xi_{\neq}$ contains a large, but finite, number of paths which go directly from x to y, without going through $\{q\} \cup \{r\}$. A typical such path will be denoted by ζ. Regardless of whether the result of the experiment $E(\Xi)$ is that the particle goes through q or r, still each path such as ζ is also a possibility for the larger experiment $E(\Xi_{\neq})$ (Figure 2).

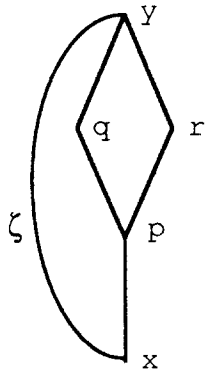

Fig. 2.

It may be useful to imagine these particle paths as being embedded within Minkowski space M^4. But for the purposes of the present section it is not necessary to assume very much about the geometry of the experiment. The important point is that each path, such as γ, can be assigned some number $L(\gamma)$, the "length" of γ, and, as in Section 7, condition F, the function F will be assumed to depend only on the lengths of the various paths. That is to say, one can equate $F(\gamma, \xi)$ with, say, $f(L(\gamma), L(\xi))$, for some suitable function f: $\mathbf{R} \times \mathbf{R} \to \mathbf{R}$. (This idea also makes sense in the case when a "path" is taken to be a Feynman diagram. Here one is interested in the space-time positions of possible measurement points, and thus length simply measures the relative positions of such points; it is not to be thought of as being the total length of all free particle paths in the diagram.)

We have specified a formula for calculating $\mathscr{P}(p, q)$ and $\mathscr{P}(p, r)$ within the experiment Ξ. Assume that the geometry of the experiment is such that the path length from p to q is the same as the path length from p to r. We assume that the larger experiment $E(\Xi_{\neq})$ is such that the paths in $\Xi_{\neq}$ consist of the paths such as ζ, which go directly from x to y, plus the paths in Ξ, but extended by a single path segment from x to p and from either q or r to y. The extension from q to y has a length which differs from the length of the extension from r to y by some definite value, call it δ. Using our formula, one calculates that in the experiment $E(\Xi)$ the probability of a particle traveling from p to q is

$$\mathscr{P}(p, q) = \sum_{\text{all paths in } \Xi_{\neq} \text{ through } q} f(L(\gamma), L(\xi)) = f(L(\gamma_q), L(\gamma_q))$$

where γ_q is the unique path in Ξ through q.

Furthermore, we have

$$\mathscr{P}(p, r) = \sum_{\text{all paths in } \Xi_{\neq} \text{ through } r} f(L(\gamma'), L(\xi')) = f(L(\gamma_r), L(\gamma_r))$$

The assumption that the path lengths are equal for each result means that $\mathscr{P}(p, q) = \mathscr{P}(p, r)$.

On the other hand, we can consider the experiment $\Xi_{\neq}$. The new terms in the formula for the probability that a particle travels from x to y involve the "influence" terms between the paths such as ζ, which go directly from x to y, and the two special paths through q and r. These terms are

$$\mathscr{P}_{\neq}(q) = \sum_{\gamma, \zeta \in \Xi_{\neq} \text{ with } \gamma \text{ or } \zeta = \gamma_q} f(L(\gamma), L(\zeta)) = 2 \times \sum_{\zeta \in \Xi_{\neq}} f(L(\gamma_q), L(\zeta))$$

and

$$\mathscr{P}_{\neq}(r) = \sum_{\gamma, \zeta \in \Xi_{\neq} \text{ with } \gamma \text{ or } \zeta = \gamma_r} f(L(\gamma), L(\zeta)) = 2 \times \sum_{\zeta \in \Xi_{\neq}} f(L(\gamma_r), L(\zeta))$$

We would like to assert that $\mathscr{P}_{\neq}(q) = \mathscr{P}_{\neq}(r)$, which would allow us to conclude that the probability of q is equal to the probability of r also within the larger experiment $E(\Xi_{\neq})$. Surely this is the condition of strong causality, as stated in Section 9. But of course one should keep in mind the fact that the condition $\mathscr{P}_{\neq}(q) = \mathscr{P}_{\neq}(r)$ is really a condition on the set of paths in $\Xi_{\neq}$. What does it mean? It is that within the larger experiment $E(\Xi_{\neq})$, the total influence of all paths in $\Xi_{\neq}$ on the path γ_q is assumed to be equal to the total influence of all paths in $\Xi_{\neq}$ on the other path γ_r. This is just the condition that we have no particular knowledge about the larger experiment which would tend to favor either the result q or r.

But the paths in $\Xi_{\neq}$ were chosen in such a way that the formula

$$\sum_{\zeta \in \Xi_{\neq}} f(L(\gamma_q), L(\zeta)) = \sum_{\zeta \in \Xi_{\neq}} f(L(\gamma_q) + \delta, L(\zeta))$$

must hold, for arbitrary "phase angles" δ, and arbitrary experiments $\Xi_{\neq}$ in the special class which we are considering in this section.

At this point it is appropriate to recall a result from Hemion (1988).

Theorem. Let $u: [-\pi, \pi] \to [-1, +1]$ be a continuous function. Assume that $u(\vartheta) = u(-\vartheta)$ for all $\vartheta \in [-\pi, \pi)$, and that $u(0) = 1$, $u(-\pi) = -1$, and u is monotone in the domain $[0, \pi)$. Assume also that if $J = \{j_1, \ldots, j_{2n}\}$ is some finite set of numbers with $j_i \in [-\pi, \pi)$ for all $i = 1, \ldots, 2n$ such that (1) $j_i = -j_{i+n}$ for each $i = 1, \ldots, n$ and (2) if $\sum u(j_i) = 0$, then $\sum u(j_i + \varphi) = 0$ for any "phase angle" $\varphi \in [-\pi, \pi)$. If these conditions are fulfilled, then $u(\vartheta) = \cos \vartheta$.

This theorem can be applied to the influence function. Condition 1 corresponds to the symmetry requirement of Section 7, condition G. Condition 2 is the condition of strong causality. We have assumed in Section 7, condition F, that F is a function from the interval $[-\pi, \pi)$ to the real numbers $\mathbf{R}$. In the theorem, we replace the range by the interval $[-1, +1]$. But this is mere convenience. In general, the range of F could be any compact interval $[a, b]$, and then the theorem would imply that F would be the cosine function with period 2π, normalized to vary between a and b, rather than -1 and $+1$.

11. AMPLITUDE FUNCTIONS

The result of the last section implies, in turn, that the probabilities for the experiment can be described using a (suitably normalized) amplitude function of pathlength. One need only take the amplitude function $g: \Xi \to \mathbf{C}$ given by

$$g(\gamma) = e^{iL(\gamma)} \qquad \text{for any} \quad \gamma \in \Xi$$

Then [ignoring special weightings for the terms of the form $f(\gamma, \gamma)$; they cannot be accounted for in an amplitude theory], summing over the paths through q, we have

$$\begin{aligned}\mathcal{P}(p, q) &= \sum_{\gamma,\xi\in\Xi} f(\gamma, \xi) \\ &= (\text{constant}) \times \sum_{\gamma,\xi\in\Xi} \cos\{L(\gamma) - L(\xi)\} \\ &= \frac{\text{constant}}{2} \times \sum_{\gamma,\xi\in\Xi} e^{i\{L(\gamma)-L(\xi)\}} + e^{i\{L(\xi)-L(\gamma)\}} \\ &= (\text{constant}) \times \left(\sum_{\gamma\in\Xi} e^{iL(\gamma)}\right)\left(\sum_{\gamma\in\Xi} e^{-iL(\gamma)}\right) \\ &= (\text{constant}) \times \left|\sum_{\gamma\in\Xi} e^{iL(\gamma)}\right|^2\end{aligned}$$

Thus, probabilities can be calculated by summing the complex amplitudes over all possible alternative paths and then taking the square of the absolute value. This is the usual prescription which is to be found in any textbook on quantum mechanics.

It is interesting to consider whether the converse result—namely that a particle-path model which admits an amplitude function is strongly causal—also holds. But in a sense, this follows directly from the definition of an amplitude space. Probabilities are defined in terms of the amplitudes which are assigned to the paths. Thus, the definition leaves no scope for questions of alternative settings in a given experimental situation. By default, then, one might conclude that the property of strong causality must hold.

Such questions can only be sensibly posed if the probability is derived from some concrete underlying model. It is possible to *define* an amplitude function on the set of paths in a particle-path space in such a way that intermediate segments along the paths are also given definite values. If this is done in a consistent way (for example, the usual Feynman path-integral prescription), then one might say that the definition itself satisfies the condition of strong causality.

But returning to the main theme, the natural question is, are (discretized versions of) the models which are usually considered in classical physics strongly causal, and thus is it a reasonable idea to assume that amplitude functions describe well their probabilities? Unfortunately, this question appears to be very difficult to resolve. In general, it is easy to *define* a number which we choose to call a "probability." But the question of *deriving* true values for the probabilities from given models is a very different

proposition. However, even if the models are not strongly causal, still we might expect to have a reasonably smooth function f, which is a (more or less good) approximation to the smooth cosine function, and thus to the probabilities calculated using the standard methods of quantum mechanics.

12. SHOULD PROBABILITIES BE DEFINED—OR DERIVED—IN A MODEL?

In the traditional, classical physics, one begins by imagining some model for the physical world. The next step is to define a collection of mathematical rules, usually formulated in terms of differential equations, which the objects in the model are to obey. A "theory of physics" is then the idea that some aspect of the physical world can be best understood by thinking about the model.

For example, a model for a gas can be obtained by imagining that some region of space-time is enclosed in a perfect box. The interior of the box is then filled with a finite number of perfectly round and elastic billiard balls which are given some initial configuration. The mathematical properties of this model can often be successfully used to explain aspects of the physical world. Another example, which is very often discussed, is the pendulum. The object of the model is now a simple mechanism, which is described using a differential equation.

Both of these models illustrate two important ideas about classical model building.

1. Each model describes only a small part of the physical world. But more than that, each model is truly just a *model*, it is obviously not reality itself! After all, who could possibly believe that the world is composed of unobservably tiny billiard balls? The pendulum model has even fewer pretenses to being complete in any sense—the pendulum is itself composed of matter which could be taken apart (thus destroying the clock, which was the subject of the model). Furthermore, there is clearly more to the universe than a single pendulum!

2. The second idea is that the models do not make any explicit statistical predictions on the results of practical experiments. Statistics come into the gas model through the assumption of some range of possible values for the initial configuration. Thus, the model itself does not produce statistical uncertainty; rather a given initial statistical distribution is transformed by the mechanism of the model into a definite final distribution. The pendulum model has even less to do with predicting the probabilities of experiments. The clock may go irregularly, but this will be attributed to "outside influences" which disturb the perfect workings of the mechanism. Thus, in both of these examples, the idea of actively and *nontrivially* observing things in

the physical world plays no essential role. In order to account for observational deviations from the models, one imagines that the models are incomplete. A more perfect model will involve more details of the physical world—for example, friction in the bearings of the pendulum, or electromagnetic and gravitational fields which may penetrate the walls surrounding the box of gas.

But according to our point of view, it is possible to consider a sort of hierarchy of models of the physical world. Each model in the hierarchy is an extension and refinement of the previous model. While each single model may be complete and fully determinate, nevertheless at each step an uncertainty is present which is then attributed to the idea that one should really be considering things in the next higher model in the hierarchy. Another way of doing things is to redefine the hierarchy of models in such a way that the uncertainty is specified at each stage. Surely this later method reflects more nearly our usual practical experience. One might object that it is less definite than the traditional type of model building, but is this true? Whether the uncertainty is attributed to an indefiniteness in the embedding within a larger model or is simply defined in the original model, the practical result in the prediction of the results of experiments is the same.

There is, however, one problem with this idea. If we agree to include uncertainty in the *definition* of a model, then we return to the usual picture of physics in terms of a mysterious conception of the amplitude function. But the real problem is that we want more of a model than just the calculation of the probabilities for the results of experiments; namely a model should also help us to *understand* the physical world.

Of course this can be a rather nebulous idea. For example, it is surely the case that the pendulum model of an old-fashioned clock does not help us to understand its workings. The model is nothing more than an idealized *description* of what we imagine the motion should be. Many people may say that a mere description is all that can be expected of a physical theory, but surely an inquiring mind demands more. One would like to understand what reason there is for defining the probabilities in the quantum mechanical way. The advantage of having a hierarchy of definite models is that the embedding of each model within a larger one should bring with it an understandable mechanism for producing the uncertainty at each stage. That is precisely what is missing in the standard formulation of quantum mechanics.

One way to proceed is to refer to the argument of the previous section. It was shown there that, within the framework of a discrete particle-path model, the quantum statistics are necessary in order to ensure that the predictions of a theory of physics obey the principle of causality. Thus, we can say that (1) if a discrete classical model which obeys the rules in Section

7 is chosen, then "interference" effects should be considered when calculating the probabilities for the results of experiments. Given this, then (2) it is reasonable to assume that causality holds, and thus the quantum statistics. But if we simply proceed to *define* a given hierarchic model, together with a set of rules for calculating probabilities at different stages, then, as far as an understanding of the reasons for the quantum statistics is concerned, we become involved in a questionable argument. It is namely the case that the condition 2 follows from the condition 1, not vice versa.

Therefore, the question becomes, what model, and therefore what classical variational principle, should we choose? Unfortunately these questions are, in practice, too difficult to deal with: the mathematical complexity simply becomes too great. Given this, then, the best method to proceed is perhaps the following. First, a definite variational principle for a given model may be considered. Then the next task is to justify the assertion that the conditions of Section 7 are valid, and thus that the range of possible extensions of the model to larger models is such as to produce the quantum statistics.

13. JUSTIFYING THE USE OF FEYNMAN DIAGRAMS

What are Feynman diagrams? The conventional answer is that they are pictorial representations of a formal infinite series—the "perturbation expansion" of a hypothetical "amplitude function" for a given quantum mechanical process. But then in the same breath it is asserted that the series is meaningless, in some profound—and thus meaningless—sense! On the other hand, all practical physicists (and almost all theoretical physicists) think about Feynman diagrams in a completely different way. In fact they imagine that each Feynman diagram represents a possible configuration of a classical model, in the sense which we have been discussing in this paper. Thus, for example, they speak about the "probability for a photon to be exchanged between two electrons" and so forth.

It is true that physicists have been taught to respect the philosophy of Nils Bohr, as it is formulated within the "Copenhagen interpretation" of quantum mechanics. Perhaps they do this by qualifying each definite statement with "so to speak," as for example: A photon «so to speak» travels «so to speak» from point A to point B. But 60 years of deep philosophical debate have only served to confirm the view that such qualifiers are nonsense! In reality Feynman's achievement was to provide a practical and simple mathematical model for the discrete world of quantum physics.

What is this model? It consists of a set of possible configurations: namely the set of finite, connected, directed graphs, together with various further properties. Let us concentrate on the theory of QED, where the

graphs can have two different kinds of edges—photon and electron lines. Now the graphs are finite in the sense that they have only finitely many vertices. The vertices (one might call them "events") determine the structure of the physical world, at least as far as it is being modeled by quantum mechanics. In QED the "internal" vertices meet precisely three edges: namely, one photon line and two electron lines. A graph may also have external edges, whose "outer" endpoints are vertices meeting no other edges. But then a hierarchy of graphs can be considered, and it is found that in larger graphs in the hierarchy, these external vertices become internal ones. The external photon lines represent the "external electromagnetic fields" according to Maxwell's philosophy. But following the arguments in the present paper, it might be better to construct hierarchies in such a way that at each stage, only external electron lines appear.

The electron lines represent freely moving electrons: i.e., electrons which do not interact with themselves or with other particles. We may assume that such free electrons have some further discrete structure which occurs at equal Lorentz spacings along the electron line. [This could be modeled according to the ideas in Section 6, or perhaps a better model would be obtained using a spiral structure for free electron paths, as in Hestenes (1985).] We have shown that given a realistic set of statistical assumptions, then the free particle statistics are given by the usual complex amplitude functions of quantum mechanics. But remember of course that these statistical assumptions cannot be deduced from the model itself! After all, we can assert that the *actual* world is fixed and definite, and thus not subject to statistical considerations. Statistics only come into things when we impose ourselves—fallible human beings with only a partial knowledge of the situation—onto the model.

As most people understand the Feynman diagram model, the free particle edges represent electrons traveling in straight lines through the usual continuous model of space-time given by Minkowski space M^4. However, I have argued in Hemion (1989) that the real-world geometry is determined by the fact that there must be an isomorphism between the group of linear transformations of the real world of known events which define a given experiment, and the group of linear transformations of the complex world of amplitude functions representing the possible outcomes of experiments. This means that it is not appropriate to impose the geometry of M^4 on the set of graphs from the outset. It would be better to allow a more general discrete geometry, for example, as described in Hemion (1988). Given some arbitrary set of "known events," then the most general idea of an "experiment"—defined in terms of its possible experimental results—would be obtained by specifying which known events should or should not precede different possible results. (This new definition can be contrasted

with the definition given in Section 7 of the present paper.) We are led to the idea of a geometry of "positions," as was explained in the previous paper. Perhaps then the geometry of the *actual* world is not that of M^4, but at least—following the arguments in Hemion (1989)—we can assert that it is probably *near* the usual geometry of four-dimensional Minkowski space.

Some readers may object that it is not natural to remove the framework of Feynman graphs from its usual continuous geometry. But they should remember that many of the traditional paradoxes of quantum mechanics are related to the fact that our usual geometric reasoning is wrong in the quantum world. For example, let an electron line be given which connects two points P, $Q \in M^4$. According to the classical picture, the momentum of the electron is thus fixed. But in quantum mechanics, all momenta are possible; indeed, all momenta are equally likely! This example shows quite clearly that, in reality, it is unnatural to impose the geometry of M^4 on each individual Feynman graph. This geometry appears more naturally as a property of the ensemble of *all* possible Feynman graphs!

14. DISCRETE PARTICLE INTERACTIONS

The statistics of free particle motion is given by the free particle Dirac equation. One can *define* the idea of "free particles" to mean that the amplitude function is a plane wave in space-time. The free particle Dirac equation can then be *deduced* in the standard way dealt with in most textbooks [for example, see Chapter 2 of Ryder (1985)]. Of course this equation is nothing more than a particular way of formulating the classical isomorphism between $SL(2, C)$ and the restricted Lorentz group. [See van der Waerden (1948) for a most general treatment of isomorphisms of linear groups.] The free particle Dirac equation is $i\not{\nabla}\psi = m\psi$. It provides a condition on possible complex-valued amplitude functions ψ.

How does the situation change if we no longer have free particles? Quite obviously the free particle Dirac equation will then no longer hold. Thus, we can write

$$i\not{\nabla}\psi = \not{B}\psi$$

This is the Dirac equation for interacting particles—formulated in an unusual way. In contrast to the free particle Dirac equation, which placed a definite condition on ψ, this new equation is nothing but a definition. It is namely the case that a certain 4-vector, which we call B, is defined in terms of the derivative of ψ.

Our equation does not as yet look like the version which is to be found in most books. We have defined $\not{B}$, and also we have an idea of the free

particle mass m (as expressed in the free particle Dirac equation). Therefore, it is possible to write $m + e\not{A} = \not{B}$. The number e is taken to be fixed, and is called the "electrical charge" of the particle. Seen in this way, we have defined a new quantity, $\not{A}$, which provides us with a new 4-vector, namely A, the electromagnetic vector potential. But then having *defined* $\not{A}$, we can also *define* the electrical current using the standard rule $j^{\mu} = \partial_{\upsilon} F^{\upsilon\mu}$, where $F^{\upsilon\mu} = \partial^{\upsilon} A^{\mu} - \partial^{\mu} A^{\upsilon}$.

The usual step now is to go ahead and identify the electrical current with the Dirac current, representing the "flow" of the electrons. But what is the result of this identification? The matter current is usually interpreted in terms of the probability of finding particles at the given point of space-time. If the current becomes very large, then this means that it is probable that there are many particles near the point. But many particles should then produce a more intensive electrical current. This way of thinking leads to the idea that the two coupled currents must be involved in an endless process of intensifying one another. Once again, we encounter the mysterious idea of an infinitely intense "singularity" in a field theory. Therefore, this simple idea of identifying the electrical and the Dirac currents seems to lead us into great problems.

Fortunately, though, there is a much simpler and more elegant way of proceeding; namely that which is *actually* followed in the theory of QED! Recall that the matter current does not really describe a flow of matter through smooth Minkowski space. In fact, a true flow of matter, such as the movement of water through a pipeline, cannot be described in terms of a simple amplitude function, as here. On the contrary, the amplitude function only has meaning within some specific experiment, and the particle experiments which involve simple Feynman diagrams are very different from the commonsense idea of an experiment with smoothly flowing, continuous matter.

What is the true meaning of the Dirac current? It is defined in terms of the amplitude function on Minkowski space. But Minkowski space itself is only the geometry of the ensemble of all possible configurations of the underlying, discrete model. In QED, this underlying model is the set of all possible Feynman diagrams. Within QED, the amplitude function is defined in such a way as to enable us to calculate the relative probabilities to be assigned to the various diagrams; that is, the probabilities that particles travel freely from vertex to vertex, according to the details of some given diagram. It may be that this underlying model is governed by some definite rules which determine the form of the particle interactions. If so, these rules must determine the structure of each *single* diagram; only in an indirect way can they determine the statistical rules which govern the ensemble of all possible diagrams. Thus, it is nonsense to say that the statistical "flow

of matter" generates electrical currents which influence each individual diagram in the ensemble!

Of course, QED does recognize the idea of an electron self-energy. But this "self-energy" is something which is much simpler: namely the recognition that many different diagrams can describe a single experiment. Thus, the actual probabilities are not those which would be calculated if we decided to choose only the simplest imaginable diagram and ignored all others. Indeed, by choosing only a single diagram, one would be denying the importance of statistics in physics. Thus, leaving out such a "self-energy" would be as false as the more usual mistake which people make of confusing an ensemble of possible configurations of a model with the configurations themselves.

What is the rule in QED for describing the electrical interaction generated by a particle? The rule is that each particle proceeds freely from vertex to vertex; that is, the interactions can only occur at the vertices. Along the line segments between vertices, there can be no interactions. However, if we restrict the electrical "current" to the discrete set of vertices in a given graph, then the possibilities for the electromagnetic vector potential A become strongly reduced. In fact it is clear that electromagnetism must be gauge invariant. (This will be dealt with further in the next section.) There are only few possible solutions for gauge-invariant fields propagating in a vacuum from a point of singularity. The choice made in QED for describing these fields is justified in a standard way in most textbooks.

For some strange reason, many contemporary theoretical physicists profess themselves to be dissatisfied with this simple rule. Having a discrete electrical current gives us discrete photons. The alternative to this rule would seem to be some sort of new theory involving the concept of "continuous" photons. Is this truly the goal of contemporary theoretical physics? Surely even such theorists would be prepared to accept the empirical fact, as demonstrated long ago by Einstein, that light is quantized. They may argue that such concepts as "photons" or "electrons" play no role in the pure Copenhagen philosophy, and thus the quantization of light should follow from more subtle and indirect principles. But by the same token, such arguments hardly provide a compelling reason for rejecting the existing discrete quantum field theory, based on Feynman diagrams.

15. GAUGE INVARIANCE

What is the reason for gauge invariance? We have defined the vector potential, and thus the electromagnetic field, to be a measure of the departure of the amplitude function from the constant wavelength in the free particle case. The amplitude function is only of importance in that it determines

the probabilities for the different results of given experiments. In particular, if different amplitude functions lead to identical probabilities, then it must be that the electromagnetic field corresponding to the change from one amplitude function to the other cannot affect the probabilities.

There is one class of variations of the amplitude functions which do not affect the probabilities. Let Λ: $M^4 \to \mathbf{R}$ be some smooth function. Now change the amplitude function by the substitution $\psi \to \psi \cdot \exp(i\Lambda)$. This means that each component of ψ is to be multiplied by the same factor at each point of space-time. When calculating the probabilities in practical experiments, the rule is that the amplitude function is first to be calculated, then the probability that a particle appears at a given point is obtained by summing over the squares of the absolute values of the components of the amplitude function. Clearly, this process is invariant with respect to the gauge transformation Λ.

Note in particular that ψ is generally obtained by summing over a number of discrete paths, say from a starting point X to an endpoint Y. Let us consider just paths from X to Y through two possible intermediate points P_1 and P_2. By possibly adding in a constant factor, we may assume that the gauge function Λ vanishes at X. The amplitude before the gauge transformation at P_i might be denoted ψ_i. Afterward it is $\psi_i \cdot \exp(i\Lambda_i)$, where $\Lambda_i = \Lambda(P_i)$. Similarly, the untransformed amplitude from P_i to Q is, say, φ_i. After the gauge transformation it is $\varphi_i \cdot \exp[i(\Lambda_Q - \Lambda_i)]$, where $\Lambda_Q = \Lambda(Q)$. Thus, the total amplitude from P to Q is first

$$\psi = \psi_1 \cdot \varphi_1 + \psi_2 \cdot \varphi_2$$

and after the gauge transformation, it is

$$\psi_1 e^{i\Lambda_1} \cdot \varphi_1 e^{i(\Lambda_Q - \Lambda_1)} + \psi_2 e^{i\Lambda_2} \cdot \varphi_2 e^{i(\Lambda_Q - \Lambda_2)} = \psi e^{i\Lambda_Q}$$

One sees from this example that such gauge transformations are consistent with the process of forming the amplitude function by summing over different discrete paths.

The existence of these gauge transformations shows that when we use amplitude functions in order to calculate probabilities, then we must accept the fact that this brings with it a certain redundancy; a whole class of different amplitudes all describe one single probability distribution. On the other hand, given a particular amplitude function ψ, we can look at the Dirac equation and use it to assign some electromagnetic vector potential A to ψ. But then we can take an arbitrary gauge transformation and apply it to ψ. The result is that we must expect to obtain a different electromagnetic vector potential

$$A_\upsilon \to A_\upsilon + \frac{1}{e} \partial_\upsilon \Lambda$$

The principle of gauge invariance is the assertion that the "physics" of the situation remain unaffected by such transformations. It is clear that this principle is a basic consequence of the quantum mechanical description of nature. Put another way, any attempt to use quantum mechanics to describe non-gauge-invariant fields would simply bring us into a state of logical contradiction!

Of course it is important to note that there is more redundancy than this in the amplitude functions. The fact that the amplitude functions which we are now dealing with have essentially two components means that gauge "rotations" in these components through space-time can also be introduced, without affecting the calculated probabilities. The situation becomes still more interesting if the redundancies of amplitude functions with three or more components are considered. This leads in the usual way to the theories of the weak and strong forces.

Finally, the role of renormalization theory in gauge theories should be discussed. In a discrete geometry, such as that dealt with in Hemion (1988), there is a smallest possible distance for the separation of two events—that is, two vertices in a given diagram. This can be formulated in terms of a momentum cutoff, precisely according to Feynman's original formulation! Therefore, in contrast to the usual ideas of theoretical physics, we find that renormalization is an essential and natural component of the theory. But there is more to renormalization than this. A theory is said to be "renormalizable" if the details of the cutoff can never be investigated by means of practical experiments. This leads us back to the discussion in Section 6. Therefore, it is hardly surprising that it is precisely the gauge theories of physical relevance which can be shown to be renormalizable. After all, a theory of interaction which allows some method of determining the discrete structure of space-time would obviously fail to be gauge invariant!

16. DOES QUANTUM MECHANICS ACCOUNT FOR ALL PARTICLE INTERACTIONS?

In Section 6 we have described a certain statistical effect which should influence the probabilities in discrete particle systems. Can there be other "classical" effects which could also affect the probabilities? According to the theory of quantum mechanics, all phenomena in the physical world can be described purely in terms of quantum statistical effects. Thus, it is appropriate to investigate the question of whether or not it is possible to have "nonquantum" statistical effects in our discrete classical physics.

But what would it mean to have nonquantum statistical effects? To answer this question, it is important to think carefully about the role of statistics in physics. To begin with, it is clear that statistics is used to describe

many different kinds of situations where our knowledge is incomplete. For example, statistics is often used in biology, or in economics. But obviously it would be absurd to apply the theory of quantum mechanics, say, to the analysis of the sales of some commercial product! In reality, quantum statistics is only relevant when it comes to describing the "basic" phenomena of physics.

But what are these basic phenomena? Certainly we can assert that our knowledge of a given physical process is determined by the model which we use when we think about the process. At the present time, people think in terms of a model involving pointlike particles moving about through space-time. Thus, the "basic" experiments in physics must be concerned with testing our uncertainty about the details of these particle paths.

In our previous arguments we have often referred to the two-slit interference experiment. Is this a "basic phenomenon" in physics? Surely not, since it involves the use of such complicated apparatus as absorbing screens, slits cut into thin sheets of physical material, and so forth. To fit these ideas into our usual model of physics, it would be necessary to reduce everything to a description in terms of particles alone. But it would be pointless to attempt to redefine the two-slit experiment in such a manner. Such an attempt would miss the whole point of our discussion of the two-slit experiment! Its purpose was to express the idea that an "experiment' involving particle paths can be *defined* in terms of the set of possible paths which the particles can follow. (This is assumption A of Section 7.) Beyond this, we did not make further use of the two-slit experiment.

In fact, the main assumption in Section 7 was I, namely that there is a *universal* influence function which only depends on the phase differences in the path lengths at the point of measurement. We now assert that this assumption should hold in all "basic" experiments in physics, i.e., those experiments whose formulation involves particles alone. But this means that we automatically include the case of particle interactions; after all, what would be the use of restricting ourselves to "free particles" alone? Surely we can only be aware of a particle if it interacts with other particles! Thus, the assumption I only makes sense if it is allowed to hold for all situations where particles interact with one another. Assumption A then becomes the assertion that there are many possible particle interactions, which give a large range of possible phase differences in the path lengths for the experimental particle.

But now we return to the previous arguments, which show that the statistics for such basic experiments can be described using an amplitude theory, and thus that they must be gauge interactions. The point at which nontrivial assumptions are made is in the argument in Section 14. According to the usual formulation of QED (and also QCD, for that matter!), the

particle interactions occur at *discrete* points along the particle paths and they are modeled by gauge particle lines in the appropriate Feynman diagrams.

In Section 7 it was asserted that special statistical weightings for the individual paths due to "classical" effects should be ignored. This was achieved by the device of assuming that the influence $F(\gamma, \xi)$ should be a *continuous* function of the difference of path lengths modulo Δ, where Δ is a constant. But then we can decide to consider such hypothetical effects by allowing the possibility that the "diagonal elements" of the form $F(\gamma, \gamma)$ have special weightings.

What does it mean to assign a special statistical weighting to some particular path, such as γ? It means that *before* the experiment is performed, we must have had some particular prior knowledge about the path γ, which affects our estimate of its probability during a run of the experiment. Put another way, we can say that the special value of $F(\gamma, \gamma)$ is a part of the *definition* of the experiment. But in general we are free to define things as we please—perhaps changing the units of measurement from one place to another throughout the experimental apparatus. This freedom is only concerned with the way we think about a given experiment; thus, such details should not be put into the category of "basic phenomena" in physics. Truly basic phenomena should be independent of these arbitrary experimental definitions. Given this, then it makes sense to exclude special weightings for the diagonal elements, and so we are left with the amplitude functions alone.

17. CONCLUSION

At the present time, theoretical physics is usually taught to students in the following way. First the old classical physics is dealt with in order to make the student familiar with the ideas of continuous space-time, variational principles, and the derivation of "laws of nature"—which are differential equations. As a next step, quantum mechanics is introduced in the form of a collection of such laws of nature. Most students (as well as their teachers) find these laws to be difficult to understand. They are told that the laws are supposed to be "beyond understanding" in some profound and mystical sense. Next, variational principles are deduced, from which the laws of quantum mechanics can be derived in an analogous way to that in the method of classical physics. It is then shown that these variational principles lead to "divergent"—that is, ill-defined—expressions. Then finally the Feynman diagrams are introduced as a way of trying to make mathematical sense out of the whole theory.

How dismaying such a course of study must be. Yet the students who do persevere are rewarded with the realization that the Feynman diagrams are really very easy to understand. Furthermore, these diagrams model, in a simple and practical way, the events which any practicing physicist experiences in the physics laboratory. True, the theoretical physicist may continue to condemn the use of Feynman diagrams, asserting that they are "impure" in some vague philosophical sense. However, the practical physicist soon learns to ignore such discordant sounds, and, if only secretly, to think of the world purely in terms of these diagrams.

The ideas in the present paper suggest an alternative pedagogical plan. After a study of the old classical physics, the student could think more generally about the role of mathematical models in physics. Statistics and probability theory could then be studied. Probability theory would be shown to be important in physics precisely because it describes our true experience of the world, as reflected through the experimental process. But this must mean that the probabilistic conditions to be imposed on a model must be formulated in terms of practical human experience with real-world experiments. Quantum physics could be introduced as evidence that nature is discrete, rather than continuous. Various methods of describing a discrete physics could be discussed, and it would be shown that the Feynman diagrams provide us with the most sensible model. The statistics of Feynman diagrams would be shown to obey the "laws of quantum mechanics." Finally, as a matter of historical interest, it would be shown that there exist variational principles—similar to those in the old classical physics—which, when applied in a formal (but formally invalid) sense, give the quantum mechanical laws of nature.

REFERENCES

Hemion, G. J. (1988). *International Journal of Theoretical Physics*, **27**, 1145.
Hemion, G. J. (1989). *International Journal of Theoretical Physics*, **28**, 1371.
Hestenes, D. (1985). *Foundations of Physics*, **15**, 63.
Ryder, L. H. (1985). *Quantum Field Theory*, Cambridge University Press, Cambridge, U.K.
Van der Waerden, B. L. (1948). *Gruppen von linearen Transformationen*, Chelsea, New York.