PROBABILISTIC CAUSALITY IN SPACE AND TIME

It is my intention in this paper to survey results and problems in the theory of probabilistic causality with an emphasis on the theory in the context of space and time. In view of this emphasis, the first section is devoted to problems of space and the second to problems of time. Here, space and time are construed in the sense of classical physics, although in some of the examples considered no real physics will enter. The third section is devoted to space-time, but the problems considered are restricted to those that arise in the framework of special relativity. Questions about probability and causality are difficult enough in this framework without considering the still more difficult case of general relativity.

As in inevitable in such surveys, the work has too much emphasis on problems that I have been concerned with myself. I therefore lay no claim to offering a balanced objective overview, but only one of some topics that I think are of general interest.

1. SPACE

The fundamental problem of probabilistic causality in the framework of classical space can be formulated this way: explain simultaneous but distant correlated events. The general methodology goes back at least to Reichenbach and can be summarized under the slogan: find a common cause.

I begin with a theorem that shows how intimately the existence of a common cause is related to the joint distribution of random variables. There are two things to note about the theorem. First, it is restricted to random variables that are two-valued – the kind of random variable that would ordinarily be used to represent a qualitative property rather than a quantitative one. Second, the criterion for the existence of a common cause that is generally accepted is the conditional independence of the random variables given the common cause. Note that in the case of two-valued random variables this can be expressed just in terms of the conditional expectation of the random variables. For

random variables with more values, considering just their expectations will not be sufficient to express their conditional independence.

**THEOREM 1** (Suppes and Zanotti 1981). *Given phenomenological random variables* \(X_1, \ldots, X_N\), *then there exists a common cause* \(\lambda\) *such that*

\[
E(X_1, \ldots, X_N | \lambda) = E(X_1 | \lambda) \cdots E(X_N | \lambda)
\]

*if and only if there exists a joint probability distribution of* \(X_1, \ldots, X_N\). *Moreover, \(\lambda\) can be constructed as a deterministic cause, i.e., for* \(1 \leq i \leq N\), *the variance of* \(X_i\) *given* \(\lambda\) *is zero.*

What is important about this theorem is that the only condition on the existence of a common cause is the phenomenological condition of having a joint probability distribution.

We also then have the following corollary.

**COROLLARY.** *Any joint distribution of* \(X_1, \ldots, X_N\) *can be represented as a mixture of conditionally independent distributions:*

\[
F(X_1, \ldots, X_N) = \int_{-\infty}^{\infty} F_1(x_1 | \lambda) \cdots F_N(x_N | \lambda) \, dG(\lambda).
\]

There are various consequences of the theorem that are worth examining but that I will not formulate in formal corollaries. Some applications are examined.

*Quantum mechanics.* By virtue of Theorem 1 the existence of hidden variables in the sense of classical quantum mechanics is equivalent to the existence of a joint probability distribution for two-valued random variables. This fact has not been sufficiently emphasized in foundational discussions of hidden variables.

Bell (1964) derived his well-known equations as phenomenological conditions on the existence of a hidden variable. The equations were put in better form by Clauser et al. (1969). Most surprising of all, Fine (1982) showed that Bell’s inequalities were a sufficient condition for the existence of a joint probability distribution. It is easy to show that Bell’s inequalities are a necessary consequence of the existence of a joint probability distribution.

To be technically more explicit, these matters may be summarized in the following way. Bell’s inequalities are specifically formulated for
measurements of spin of pairs of particles originally in the singlet state. Let $A$ and $A'$ be two possible orientations of apparatus I, and let $B$ and $B'$ be two possible orientations of apparatus II. Let the measurement of spin by either apparatus be 1 or -1, corresponding to spin $1/2$ or $-1/2$, respectively. By $E(AB)$, for example, I mean the expectation of the product of the two measurements of spin, with apparatus I having orientation $A$ and II having orientation $B$. By axial symmetry, we have $E(A) = E(A') = E(B) = E(B') = 0$, i.e., the expected spin for either apparatus is 0. It is, on the other hand, a well-known result of quantum mechanics that the covariance term $E(AB)$ is:

$$E(AB) = -\cos \theta_{AB},$$

where $\theta$ is the difference in angles of orientation $A$ and $B$. Again, by axial symmetry only the difference in the two orientations matters, not the actual values $A$ and $B$. (To follow the literature, we begin with the notation $A, B, A'$ and $B'$ for phenomenological random variables, rather than $X_1, \ldots, X_N$, which we go back to later.)

On the assumption that there is a hidden variable that renders the spin results conditionally independent, i.e., that there is a causal hidden variable $\lambda$ in the sense of Theorem 1, Bell (1964) derives inequalities equivalent to the following.

$$\begin{cases} -2 \leq -E(AB) + E(AB') + E(A'B) + E(A'B') \leq 2, \\ -2 \leq E(AB) - E(AB') + E(A'B) + E(A'B') \leq 2, \\ -2 \leq E(AB) + E(AB') - E(A'B) + E(A'B') \leq 2, \\ -2 \leq E(AB) + E(AB') + E(A'B) - E(A'B') \leq 2. \end{cases}$$

This form of the inequalities is due to Clauser et al. (1969).

**THEOREM 2** (Fine 1982). *Bell’s inequalities are necessary and sufficient for the existence of a joint probability distribution compatible with the six given covariances of the phenomenological random variables $A, A’, B,$ and $B’$ occurring in the inequalities (1).*

On the other hand, pursuing these results about hidden variables, it is natural to ask what is the causal situation when more random variables are considered. We know from much experience with conditions on joint distributions that the conditions must expand with the number of variables, i.e., it would be the most improbable event imaginable to have simply a fixed condition on subsets of four random
variable be sufficient for a joint distribution of all the random variables. We know that we must consider subsets of increasingly large size and nonreducible conditions on these larger and larger subsets. This general experience is well borne out here. Garg and Mermin (1982) have given a counterexample to Bell's inequalities being sufficient for eight random variables when what we term Bell covariances are given. By Bell covariances we mean covariances $E(X_iX_j)$ for $1 \leq i \leq i_0 < j \leq N$, for some integer $i_0$. Garg and Mermin's counterexample is for $N = 8$ and $i_0 = 4$. Let $E(X_1X_5) = E(X_2X_6) = E(X_3X_7) = 1$ and otherwise $E(X_iX_j) = -1/3$ for $1 \leq i < j \leq N$. Then it is easy to show that for the quintuple $(X_1, X_3, X_4, X_6, X_8)$ all covariances must be $-1/3$. But it follows at once from Theorem 3 stated below that with $N = 5$, $a_i = a_j = 1$, for existence of a compatible joint distribution $\Sigma E(X_iX_j) \geq -2$, and so there can be no joint distribution compatible with the given covariances all equal to $-1/3$.

It is natural to ask more generally under what conditions given covariances of pairs of phenomenological random variables are such that they are compatible with the existence of a joint distribution of the random variables. Fortunately, a rather simple condition can be given, although the proof of its correctness is rather intricate.

**THEOREM 3** (Suppes and Zanotti, to appear). *A necessary and sufficient condition that there exist a joint probability distribution compatible with the given covariances of all pairs of $N$ phenomenological random variables is that*  

$$\sum_{{i<j}} a_i a_j E(X_iX_j) \geq (1 - n)/2$$  

*for all subsets of odd cardinality $n \leq N$ and with $a_i, a_j = \pm 1$.*

The various hidden-variable theorems relevant to quantum mechanics have been formulated with the weakest possible conditions on the causal variables. The reason is certainly because everyone has expected that quantum mechanics and relevant experimental data would probably support quantum mechanics. Therefore, the strongest possible test of quantum mechanics versus causal theories with hidden variables is to impose the weakest possible conditions on the causal theories, and still show their lack of empirical support.

On the other hand, it is also of interest to try to ask what are reasonable conditions on causal hidden variables in order to make
them physically acceptable. Without restrictions, highly artificial mathematical constructions can be given, as is the case for the deterministic hidden variables of Theorem 1.

An example of such a theorem is one that imposes some natural symmetry conditions both at a phenomenological and at a theoretical level. The main principle of symmetry I shall use is that of exchangeability. Two random variables $X$ and $Y$ with possible values $\pm 1$ are said to be exchangeable if the following probabilistic equality is satisfied:

$$P(X = 1, Y = -1) = P(X = -1, Y = 1).$$

The theorem I state shows that if two random variables are exchangeable at the phenomenological level then there exists a hidden causal variable satisfying the additional restriction that they have the same conditional expectation if and only if that correlation is non-negative.

THEOREM 4 (Suppes and Zanotti 1980). If $X$ and $Y$ are exchangeable, then three exists a hidden variable $\lambda$ such that

(i) \[ E(XY|\lambda) = E(X|\lambda)E(Y|\lambda) \]

(ii) \[ E(X|\lambda) = E(Y|\lambda) \]

if and only if \[ \rho(X, Y) \geq 0. \]

Certainly it is possible to criticize this theorem as being too strong in its symmetry requirements from a physical standpoint. This has been done by Shimony (1981). More can be said about these matters but in this survey it seems more useful to turn to entirely different fields of investigation where similar ideas still apply.

Social sciences. Among the most widely used models in the social sciences are latent trait models. The applications range from the theory and analysis of mental tests to the analysis of sociological surveys on a variety of attitudes toward topics of political or social importance. Let us take ability in the case of mental tests and attitude in the case of surveys. In such applications, the latent trait represents an ability or attitude measured by the test or survey instrument. The data consist of individuals’ responses to the test or survey. It is understood that the
responses are of a yes–no variety for the purposes of the present discussion. A latent trait model is in principle just like a hidden-variable model in quantum mechanics. The objective is to identify theoretical variables that render the probabilities of responses to the various items of the test or survey conditionally independent, i.e., conditionally independent given the theoretical trait variable. It is the job of the latent trait model to predict the observed probabilities, and to explain the phenomenological correlation between items by the standard common cause approach, i.e., by rendering the items conditionally independent.

Theorem 1 shows that if no restrictions are placed on latent trait models then as long there is a joint probability distribution such latent traits can always be found. This fact has not at all been generally recognized in the extensive literature on latent trait models. The theory has not been oriented toward general causal questions as in the case of quantum mechanics, but to the testing of specific models. Unlike the case of quantum mechanics, in most ordinary applications of latent trait models one expects the test or survey, simply on the principles of experimental design, to have a proper joint distribution of item responses. So in principle no problem arises of finding hidden variables, and the very weak theory, interesting because of the expected negative results in the case of quantum mechanics, does not admit of the same interest in this social science context. Holland and his associates (Holland 1981, Cressie and Holland 1983, and Rosenbaum 1984) have studied when various latent trait models are consistent with observed data. In this case, additional restrictions are placed on latent traits. For example, one standard restriction is that the conditional distribution of an item must always be monotone in the latent trait variable $\lambda$.

Some technical details are needed to formulate the ideas explicitly. I follow the excellent general analysis of these matters to be found in Holland and Rosenbaum (1986). First, we say that the distribution of a random vector $X$ is associated if $\text{Cov}(f(X), g(X)) \geq 0$ for all non-decreasing, bounded functions $f$ and $g$ of $X$. Obviously this is a concept of positive association for the components of $X$. (For notational purposes it is simpler to refer to the random vector $X$ rather than the unidimensional variables $X_1, \ldots, X_N$.) Second, the distribution of a random vector $X$ is conditionally associated if, for any partition $(Y, Z)$ of $X$ and any function $h(Z)$, the conditional distribution of $Y$ given $h(Z)$ is associated. Third, latent conditional independence is just the
kind of independence expressed in Theorem 1 but generalized to more than two values. In particular, it is the condition that

\[ F(x_1, \ldots, x_N | \lambda) = \prod_{j=1}^{N} F_j(x_j | \lambda) \]

for all \( x_1, \ldots, x_N \) and \( \lambda \), where \( F \) is the conditional distribution of \( x_1, \ldots, x_N \) given \( \lambda \). Fourth, a latent variable model \( (X, \lambda) \) satisfies latent monotonicity if the functions \( 1 - F_j(x | \lambda), j = 1, \ldots, N \) are all non-decreasing functions of \( \lambda \) for all \( x \). In case \( \lambda \) is a vector, \( 1 - F_j(x | \lambda) \) must be nondecreasing in each of the coordinates of \( \lambda \). Fifth, a model \( (X, \lambda) \) is unidimensional if \( \lambda \) is a scalar, i.e., is unidimensional.

THEOREM 5 (Holland and Rosenbaum 1986). If a latent variable model \( (X, \lambda) \) satisfies the conditions of latent conditional independence, latent monotonicity and unidimensionality then the distribution of \( X \) is conditionally associated.

In contrast to Theorem 1 the three conditions on the common cause \( \lambda \) given in Theorem 5 impose strong phenomenological conditions of positive association on the observable variables.

It is not always the case that joint distributions of the phenomenological variables are guaranteed by experimental design in test theory. For a discussion of the problems that are due to practice effects, see Holland and Thayer (1985). Let \( X \) be a random vector representing a test score and \( X_p \) the test score after taking a practice test. Holland and Thayer point out that we cannot collect data in any direct way on the joint distribution of \( X \) and \( X_p \).

There are a variety of other examples in which various general restrictions are imposed. These general restrictions usually make a good deal of intuitive sense and thus render the search for a common cause empirically testable and interesting.

Distance as a cause. One kind of standard model of spread of epidemics makes the fundamental assumption that the probability that an individual becomes infected is monotonically decreasing in distance from the source of infection. This of course is an assumption that will hold in the large. If one is looking at the spread of an infection over a large geographical area it is quite a reasonable one that we would expect to be well substantiated by the data. Formally, distance is for this model the common cause rendering the probabilities of individuals’ becoming infected conditionally independent.
From the standpoint of the philosophical discussion of causality most of us are unhappy with simply thinking of distance as a cause. On the other hand, matters are rather subtle here. We certainly think of the force of gravity as being inversely proportional to the square of the distance. In fact, a whole class of models of classical mechanics are concerned with the study of forces that are a function of distance only. So it is clear that this epidemic model that uses distance as a causal variable is in the great mechanical tradition as well. On second thought, probably most of us can easily absorb the usage of saying something like the cause is weaker because of the distance from the object. We still think of some object as being the cause, and not the distance, but the power of the cause is affected directly by the distance.

At bottom here is the traditional problem of action at a distance. Substantively, distance like time is never believed to be in and of itself a cause. Models in which distance becomes a cause are recognized by all and sundry as purely formal in character. Moreover, most of us have faith that such models can be expanded to show that distance as such is a spurious cause.

The deeper mystery is what to make of action at a distance, but it is beyond the scope of this survey to enter into this tangled problem. (I have done so previously in Suppes 1954; 1970, pp. 84–88.) The one point to remark on in the present context is that we certainly cannot give an adequate causal account of gravitational forces in the sense of classical physics if we have as an apparatus only the probability of discrete events. Fortunately, even ordinary talk is saturated with the qualitative idea of a continuously varying quantity. For example, As he came nearer, his expectation of success increased. An introductory probabilistic causal account of quantitative variables is given in my 1970 monograph, but a more thorough development in relation to substantive physical theories is needed.

Medical example. A standard approach to diagnosis and treatment in medicine is to assume that there exists a set of symptoms that serve as “first level” common causes for the sickness. In this case, the symptoms are highly restricted in character, for example, a high temperature, diarrhea, etc. At the next level we look for “real” common causes. For example, we may find the presence of some bacterium or some virus. But this identification of “real” common causes — being a bacterium, for example — is not one that is the least bit satisfactory from the standpoint of fundamental molecular biology. One wants to
understand the molecular action of the bacteria: exactly how do the bacteria cause collections of cells to show phenomenological symptoms. Notice in this case there is no real question of spatial distance. We very much think of the action of the bacteria in the cell as being by ordinary scale a case of contiguous action, not action at a distance.

Classical statistical example. A classical example of common causes thought of in terms of a normal distribution at a fixed time and therefore, in the general sense we are using here, a spatial distribution, is this. Let us restrict ourselves for purposes of notation to two phenomenological random variables $X$ and $Y$, which have a bivariate normal distribution. We search for a common cause $Z$ under which the phenomenological correlation of $X$ and $Y$ become zero, i.e., the partial correlation of $X$ and $Y$ with $Z$ held constant is zero. To construct such a common cause $Z$ we proceed as follows:

Set

1. $E(Z) = 0, E(Z^2) = 1$.
2. If $\rho(X, Y) \geq 0$, set
   \[ \rho(X, Z) = \rho(Y, Z) = \sqrt{\rho(X, Y)}. \]
3. If $\rho(X, Y) < 0$, set
   \[ \rho(X, Z) = -\rho(Y, Z) = \sqrt{|\rho(X, Y)|}. \]

Then $X, Y,$ and $Z$ have a multivariate normal distribution with

\[ \rho(XY \cdot Z) = 0, \]

where the period in the middle of the line is used to denote partial correlation with $Z$ held constant.

2. TIME

At a fundamental philosophical and scientific level, probably none of us believes in temporal action at a distance. In principle, the structure of an object or of a phenomenon at a given time completely determines what will happen next, including the laws governing inherent randomness. At a fundamental level, we might well claim that direct causes never recede backward in time. History, from this causal perspective, is a trivial subject because of the absence of knowledge of structure. Put another way, this philosophical claim is an expression of
the deeply entrenched view that phenomena, at the fundamental level, are entirely Markovian in character. This same assumption certainly dominates the general theoretical ideas of classical physics.

On the other hand, we all recognize the scientific hopelessness of such a position. No one would begin to suggest that we could in fact eliminate the historical view as the proper way to study a wide variety of phenomena. I am not talking simply about history proper here but about a wide range of disciplines from meteorology to economics. For example, if I want to predict what the annual rainfall will be for the coming year in a given spot, it is out of the question to examine only the current state of the atmosphere to make the prediction. The only chance of coming reasonably close is to look at data from past years, possibly modified with consideration of a few current local parameters. But certainly the dominant term in the prediction will be the data from the past.

Grand example. The marvelous success story of the view that current structure determines current properties and is accessible at least indirectly to determination is the theory of atomic structure. To a very large extent, the atomic structure of many objects determines in a way open to investigation their main physical properties independent of the history of the objects. One might say that what I have just said is not true of very many natural objects. It is certainly true of a variety of "pure" objects of great importance in the modern development of physical theory and also in modern engineering. The example of the theory of atomic structure is a great scientific success story. Philosophically, we might say it is not of great general significance because we all believe anyhow that current structure determines present and future behavior, but there are in fact few real cases that are computationally feasible. Yet those few cases have turned out to be of such importance scientifically and technologically that they constitute collectively the most surprising results of the last three centuries. Notice without being finicky how we think about causes in this instance. We think directly of the structure of the atoms as determining the present and future behavior of objects – as being the common cause of observable properties. The many subtle and intricate results of current solid state physics show very well that the theory of atomic structure is a subject that is very much alive and well and will be as important to the future as it has been in the past.

Chains of infinite order. Let us return to those many phenomena
that cannot be given a working scientific analysis in terms of their current structure. A general framework for treating such examples is as a chain of infinite order. Put in probabilistic terms, suppose we have an infinite sequence of random variables $X_1, \ldots, X_n, \ldots$, with in general the outcome of random variable $X_n$ being dependent on the outcomes of the preceding $n - 1$ random variables. We have in mind here the sequence unfolding in time, and we could just as well use a continuous-time index as the discrete one. (The notation will be somewhat simpler if time is indexed by discrete trials rather than continuously; I also omit the generalization to an infinite past.) What we often aim at scientifically is to obtain a Markov truncation of a chain of infinite order by introducing a concept of current state with the standard common cause property: once we conditionalize on this current state no further knowledge about the past will be of any use.

Unfortunately, however, unless we put some restrictions as in the earlier case we can trivialize the problem of finding a Markov truncation. Let us write the probability of response $j$ on trial $n + 1$ conditional on the preceding $n$ outcomes in the following fashion:

$$P(X_{n+1} = j | x_n, \ldots, x_1) = P(X_{n+1} = j | [x_n]),$$

where $[x_n]$ is a standard notation for the sequence of actual responses up to trial $n$. Obviously we get a Markov truncation by simply introducing at each stage states that are isomorphic to the sequence of first $n$ responses, i.e., the Markov truncating state at trial $n = [x_n]$. Notice how uninteresting this result is, even more so than Theorem 1 about common causes, for here states are unbounded in number and have no natural intrinsic structure. We just keep introducing new states at each trial as we need them. The scientific objective, of course, is to do something very different (for more detailed discussion, see Suppes 1986).

In certain situations a change in random variables can make a chain of infinite order a Markov chain. For example, a model of learning or adaptive control that is linear in response probabilities can take as new random variables the probabilities of responses, rather than the responses themselves. One of the early examples of the effective use of this strategy is Karlin's (1953) derivation of detailed limit theorems for certain processes. More recently this approach has also been extensively used by Norman (1972). On the other hand, as already pointed out by Lamperti and Suppes (1959), this strategy will not work when
probabilities of correction or adjustment depend on responses on previous trials. In such instances there is no apparent effective way to reduce the chains of infinite order to a Markov chain by introducing new random variables.

Stochastic processes. The widespread use of stochastic processes of all sorts, Markovian or not, is testimony to the great importance in current science of the intensive study of probabilistic theories of causality with explicit emphasis on temporal development of the causal processes under study. There is one important point of a general philosophical nature I want to make about such processes. A great variety of processes that are considered of central importance scientifically are extraordinarily complicated in detail. A familiar classic example is that of Brownian motion. For many kinds of processes of a simpler nature we are interested in studying the detailed sample paths of the particles or other objects in the process. What is important, however, about complicated processes like those of Brownian motion is the hopeless character of a detailed study of the actual sample paths. Imagine for a moment the totally unmanageable complexity of analyzing the actual paths of a few billion particles in a small sample of a gas — even if we could overcome the practical problems of making the observations we would not really know what to do with all the details of the actual paths once we had recorded them. This means that the theory of probabilistic causality often comes especially into its own when we are not looking at detailed sample paths but looking at various probabilistic features of some complicated phenomena. Brownian motion is an important example from classical physics with applications in many other domains, including quantum physics. An example of a very different sort is modeling something complicated like the activity of a telephone exchange. Here also the interest is not really in the details of each call but rather a good understanding of the dynamic behavior of the distribution of calls.

As perhaps is evident, the sermon I want to preach on this point is that it is a conceptual mistake to think we are going to ever understand the details of many kinds of complex phenomena. One of the important kinds of applications of probabilistic causality in actual scientific work, both theoretically and experimentally, is to the analysis of such complex phenomena. The important point is the identification of those random variables which represent aggregate behavior in some form and which can be studied in a meaningful fashion.
As remarked earlier, I restrict myself here to special relativity only, because I know of scarcely any detailed discussions of probabilistic causality in the context of general relativity — there are few enough combining both probability and special relativity.

First, it can be observed that the general theory of probabilistic causality goes through in special relativity with the restriction that temporally earlier causes must lie in the back light-cone. In Figure 1 we are considering event $A_t$. The back light-cone of $A_t$ is what is shown.

![Diagram](image)

Fig. 1 Restriction of earlier causes to back light-cone.
as the open triangular region immediately below $A_t$ bounded by light lines on each side in this one-dimensional spatial case, with the horizontal axis representing space and the vertical axis representing time. Note that the event $B_{t'}$ can be a cause of $A_t$ because it lies in the back light-cone, but on the other hand, $C_{t''}$ cannot because it lies outside the back light-cone. This restriction of special relativity on causal action is one that we could easily impose on probabilistic causality and represents in principle only a slight modification in the general theory.

As long as we are dealing with the probability of point events and just looking at the conditional probability of one event on another, for example, no new fundamental problems are created, but it is quite another matter when we turn to what are meant to correspond to continuous distributions of properties in space. Note that in a process of classical physics, at time $t$ a quantity may well be distributed over all of space and what we are studying is the dynamic change of that distribution with time. The immediate question for probabilistic causality in the framework of special relativity is, what is the proper domain of such a continuous distribution. (I have discussed these matters earlier (Suppes 1981), but I want to expand here on what I had to say.)

Let us take a look first at the one-dimensional case. So suppose we have a univariate normal distribution. In the classical case that distribution has as its domain simply the one dimension of space. In special relativity we would naturally pick one of two types of lines, either a separation line or an optical line (by separation line I mean a line that connects two points that cannot causally influence each other). Such lines are also called spatial lines. Already even in the one-dimensional case, if we pick optical lines we have immediately the question of which optical line through a given point, as shown in Figure 2.

This concept of an optical line generalizes to an optical plane or an optical threefold. Let $A$ and $B$ be two optical lines that are parallel and neutral, i.e., no element of $A$ is before or after any element of $B$. Then the plane containing $A$ and $B$ is an optical plane. We generalize in the obvious way from an optical plane to an optical threefold.

But there is, I think, a better solution — closer to the space of Euclidean geometry and classical mechanics. First, it will be desirable to sketch the geometry of classical, i.e., Galilean, space-time. We need to assume, or prove, among other things, that we have a four-dimensional affine space, based, let us say, on the ternary relation $B$ of betweenness. We also need a concept of simultaneity $\sim$, which is a binary relation between point events, and as well a quaternary relation
of equidistance or congruence. Let $A$ be the set of all (point) events. Then for any event $a$ we define the set of events simultaneous with $a$: $[a] = \{b : b \in A \& b \sim a\}$. Then, an axiom or a theorem is that the structure $([a], B, \sim)$ for each event $a$ is a three-dimensional Euclidean geometry. In the classical case, probability distributions are defined on these spaces, and the stochastic problem is to understand their dynamics in time.

In the relativistic case we have, of course, no direct concept of simultaneity of distant events, but we do have something to replace it. Some preliminary definitions are needed. An optical parallelogram is a parallelogram whose sides are segments of optical lines. An inertial
line and a spatial line are orthogonal if and only if they form the diagonals of an optical parallelogram. Now, in the relativistic case we select not only a point event $a$ but also an inertial line $L$ through $a$. We then define:

$$[a]_L = \{b : b \in A \& ab \text{ is a spatial line orthogonal to } L\}.$$ 

Note that if $b \in [a]_L$, the line $ab$ must intersect $L$ at $a$. The axiom or theorem of the geometry of Minkowski space we now need is this. For every event $a$ and inertial line $L$ through $a$, the structure $([a]_L, B, \approx)$ is a three-dimensional Euclidean space. Gaussian or normal probability distributions, so familiar in physics, have then such a standard spatial domain, even in the relativistic case.

Of course, we could pick as the domain the forward or back light-cone of a point, but then the standard affine properties of the normal distribution are lost. The spatial choice indicated seems to be the right one for physical purposes, at least for mean distributions.

The reservation just indicated is important. There are difficulties in extending in any direct way the full theory of Brownian motion to special relativity (for a review of the problems, see Nelson, 1978). But the causal theory of classical quantum mechanics is not so detailed as that of classical Brownian motion. Particles do not have well-defined trajectories. From a probabilistic standpoint what we mainly get are mean distributions, and the Schroedinger equation provides the causal theory of how the mean distribution changes under the action of a given Hamiltonian or potential. Relativistic quantum mechanics shares these same features. The Dirac theory of the electron is a standard specific example in which a probability density defined over the whole of space as characterized above arises in a natural way along with the probability current.

It is essential to conserve as long as possible in any theoretical developments the affine spaces that support the multivariate normal distribution. Detailed dynamic, i.e., causal, theories for any other family of continuous multivariate distributions scarcely exist. Holding on to ordinary Euclidean space within the framework of special relativity remains a critical matter for probabilistic causality as expressed in relativistic quantum mechanics, even if such distributions do not have all the invariant features we would like.

Stanford University
REFERENCES

Suppes, P. and Zanotti, M. (to appear). 'Existence of joint distribution with given covariances and Bell’s inequalities in quantum mechanics.'