Scientific Explanation in Quantum Theory

Robert K. Clifton

Science and Philosophy

The media spectacle surrounding Alan Sokal's bogus article "Transgressing the Boundaries"[1] was a poignant reminder of C. P. Snow's concern about the gulf of mutual incomprehension and hostility that exists between cultural intellectuals and scientists, particularly physicists [2]. What drove Snow's concern was not academic infighting, but the consequent lack of deep reflection on the cultural and societal impact of scientific ideas. Snow's vision was that we ought to promote vigorously the cultivation of an education in which people are brought up

> *...to understand what technology, applied science, science itself is like, and what it can and cannot do.* ([2], p. 10)

This is a perfect expression of the goals of research in the philosophy of science. Disagreements about how to understand the scientific enterprise will arise, but they should never be allowed to overtake the pursuit of Snow's vision.

A more dangerous distraction from this pursuit is the distinctly twentieth century myth that there is a sharp boundary that philosophers and others must transgress in order to engage with the scientific enterprise. Descartes, Newton, Leibniz, Maxwell, Mach and Poincaré, to name only the most prominent classical physicists, all couched their ideas about the physical world in philosophical— not just mathematical—terms. And the intermingling of philosophy with physics has become even more apparent with the emergence of the kind of abstract theories that have come to dominate physics this century. For example, as a prelude to establishing in his special theory of relativity that simultaneity is not an invariant concept independent of an observer's state of motion, Einstein needed to clear the way by giving an epistemological critique of the methods observers can use to establish whether spatially separated events are simultaneous. And Einstein cleared the way for his general theory of relativity by arguing, from the way that gravity effects objects independent of their size or make up, that an object's motion under gravity is indistinguishable from the motion it would be seen to have, in the absence of gravity, from the perspective of an observer accelerating past the object. Similar epistemological critiques, such as Heisenberg's examination of the procedures by which we can determine both the position and momentum of a particle, were formative in the early development of quantum theory.

In his commentary on the Sokal affair, Steven Weinberg [3] is only willing to grant the 'jurisdictional point' that time, space and matter, things which had been thought to be proper subjects for philosophical argument, actually belong in the province of ordinary science. But, if I may appropriate Snow once more, limiting the interaction between physics and philosophy in this way surely means that

> *...at the heart of thought and creation we are letting some of our best chances go by default. The clashing point of two subjects, two disciplines, two cultures—of two galaxies, so far as that goes—ought to produce creative chances. In the history of mental activity that has been where some of the break-throughs came.* ([2], p. 23)

The clash Snow wrote about was a clash between two radically different ways of seeing the world, two rival sentiments of rationality. And I think it's fair to say that the clashing point was, and still is, scientific explanation, which has been historically vital, not just to the progress of science, but to society's view of itself. In this essay, I want to explore the novel problems for our scientific understanding of the world raised by modern physics. My chief concern will be with the seemingly intractable problems of quantum theory, a theory that was initially designed to account for atomic and sub-atomic processes but is now hailed by many to be universally valid.

Scientific Explanation

Even if we embrace the virtues of interdisciplinarity, there is no royal road towards an appreciation of the way in which modern physics understands the world. This is not just because modern physics makes extensive use of abstract mathematics, but because there is no consensus even among experts about how to characterize scientific explanation generally.

For some time the leading light on the issue was Carl Hempel [4] who held that scientific explanations come in the form of arguments to conclusions. These arguments were supposed to invoke general laws of some theory, and their conclusions were supposed to yield specific predictions of the theory. For example, suppose a cannonball falls from a tower. The relevant law is that the combination of the ball's kinetic energy of motion and its potential energy, due to its height in Earth's gravitational field, must remain constant throughout the ball's flight. Using this law, and ignoring air friction, one calculates that the ball's speed upon hitting the ground will be a function only of the height from which it fell and the constant acceleration imposed on all objects near the Earth due to gravity. So if we ask why the ball's speed at impact was such and such, the answer comes in the form of a deductive argument with the law of conservation of energy as one premise and the fact that the ball fell from such and such a height as the other premise.

Of course, Hempel recognized that not all good scientific explanations terminate, as with this example, in predictions made with absolute certainty. Nor need good explanations be based on deductive as opposed to inductive reasoning. Thus I am surely right to explain my recovery from poison ivy by citing the facts that I took the appropriate steroids and that *most* people who take them recover as I have. Still, there are many reasons to reject Hempel's identification of explanations with arguments from general laws. I shall only mention two.

First, Hempel's identification forces us to call certain arguments explanatory when intuitively they are not. For example, it appears to follow from our best cosmological theories that the initial expansion rate of the universe needed to fall within extremely narrow limits for it to have been hospitable for life. At first, some cosmologists sought to turn the explanation of the initial expansion rate around by deducing—in sound Hempelian fashion—the initial expansion rate from the fact of our existence in the universe! But of course our existence can at best provide *evidence* for the initial expansion rate, not explain it. If there is any scientific explanation forthcoming (and many have denied that there could or even should be), then one would expect it to come from detailed considerations of the mechanisms operative following the big bang. The upshot is that arguing from general laws to a specific prediction does not in itself suffice for explanation—some attention to the *causes* of events seems called for.

Second, Hempel's account of statistical explanation appears to be too weak. Consider the fact that in order to produce certain particles in the laboratory, high energy physicists need to build supercolliders that feed one beam of particles into another, and even then the events producing the particles of interest occur with very low probability. What, then, explains the occurrence of one of these rare events? An inductive statistical argument will not suffice, since most of the time the beams collide without producing the event in question. And simply deducing its low probability from the statistical laws of quantum theory doesn't capture the sense in which the rare event was *caused* by the colliding beams; for had the collision not occurred, the event could not have either.

These points suggest that one ought to try and put the 'cause' back into the 'because' in scientific explanations, even if it means entertaining fundamentally indeterministic causes that merely raise or lower the probabilities of their effects without making them certain. Probably the most ardent defender of the idea that causes are critical to scientific explanation is Wesley Salmon [5]. For Salmon, scientific explanations appeal to possibly unobservable and often indeterministic *causal processes* that underpin observed regularities in nature. But while Salmon's account of causal explanation arguably captures most classical physical explanations, can it capture the abstract explanations of modern physics?

Consider an example of the geometric explanations so distinctive of modern spacetime theories. In Einstein's special theory of relativity, if an observer moves past a metre stick at constant velocity she will find its length to be shorter (< 1m) than if it were simply at rest relative to her—the infamous effect of Lorentz contraction. In Figure 1 below the shaded strip marks the history, or 'worldtube', of the stick as it moves through spacetime, the t' axis marks the history, or 'worldline', of an observer moving towards the metre stick, and the t axis is the worldline she would follow were she at rest relative to the stick. To determine its length, each observer needs to take a snapshot of the stick at some fixed instant. Events simultaneous for an observer are picked out by lines perpendicular to her worldline. Thus events on the x axis are simultaneous for the t observer. More surprisingly, events on the x' axis

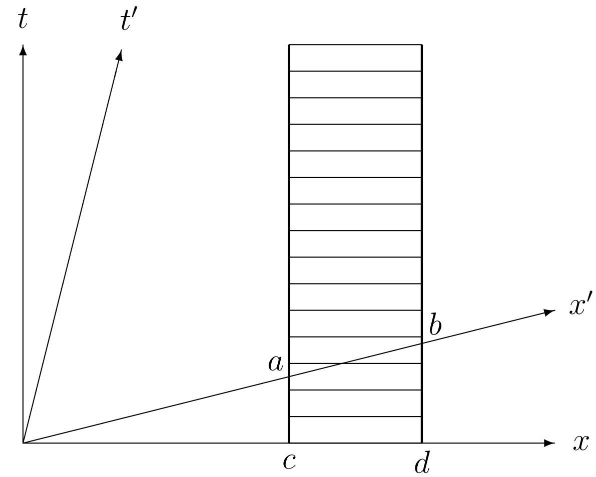

Figure 1. Sketch of the geometrical explanation of Lorentz contraction

are simultaneous for the t' observer to which the x' axis is *indeed* perpendicular, since special relativity operates within a nonstandard Minkowski geometry not reflected by the distance and angle relations we normally take to hold of points on the printed page.

Now each observer's line of simultaneity intersects the worldtube of the metre stick in a line segment. The length of that segment thus represents the observer's judgment about the stick's length. But it is a theorem of Minkowski geometry that the line segment ab is *shorter* than cd, thus the observer moving towards the metre stick finds its length to be shorter than normal. This too is not faithfully reflected by the relative lengths of ab and cd on the page, which suggest $ab > cd$. However, the claim is not that we have at hand a pictorial explanation of Lorentz contraction, but an abstract (!) geometric one. Similar geometric stories can be told for all manner of relativistic phenomena (time dilation, the Doppler effect, $E = mc^2$, etc.)

Are these stories explanatory *in themselves*, or do they just summarize in convenient geometric language the answer to any question one might ask in a relativistic world? This is a major point of dispute among philosophers of space and time. Those who hold that the ascription of geometrical structure to spacetime is genuinely explanatory have also tended to invest reality in that structure. Those for whom that makes little sense have tried to show that references to geometrical relations in spacetime explanations can in principle be reduced to, and therefore replaced by, complex assertions about causal relations between events. Although the success of this reduction is still in dispute, particularly for the spacetimes entertained by *general* relativity, we can see something of its flavour by noticing how the above geometric account of Lorentz contraction suppresses the details of how the observers take their snapshots of the metre stick's length. If we pursue those details, and the differing judgments of simultaneity on which they are based, we will be led back to considering the causal processes an observer has available to make determinations of simultaneity. And, as I noted earlier, that was precisely Einstein's starting point in setting up the theory.

As one might expect, Salmon places himself squarely within the causal reductionist approach to spacetime theories ([5], pp. 140-1). But the method of letting abstract mathematical models do nontrivial work in our understanding of the physical is not just an annoying habit of spacetime realists, but pervades all of modern physics. Generalizing from the example of Minkowski spacetime, the physicist Robert Geroch remarks

> To answer a physical question, one first translates that question into various objects (including only objects which are relevant to the question) in the mathematics, with various properties describing the physical setup. Then one manipulates these objects within the mathematics and translates the results back into physical terms. ([6], p. 86)

More recently, Salmon has written that "an adequate characterization of quantum-mechanical explanation is a premier challenge for contemporary philosophy of science", yet he confesses to the feeling that explanations in the quantum realm will ultimately boil down to mechanisms, albeit of a different sort than we're used to ([7], p. xii). I want to suggest in the rest of this essay that if we are to understand quantum theory, in which even talk of *indeterministic* causal processes breaks down, we will have to take seriously the idea that locating phenomena within a coherent and unified mathematical model is explanatory in itself.

In fact, the idea is not completely new. Michael Friedman has forcefully urged that to explain a body of phenomena *is just to* supply a theoretical unification of it [8]. For Friedman, science increases our understanding of the world precisely because it reduces the total number of independent phenomena that we have to accept as brute facts. Unfortunately, his view became bogged down almost immediately in technical difficulties (cf. [7], Sec. 3.5), at which point Philip Kitcher picked up the baton, emphasizing that a theory unifies a large body of phenomena when it exploits the same small number of argument patterns over and over again [9].

Kitcher's conception of a general argument pattern allows the geometric derivations in spacetime theories to count as such, as well as arguments employed in genetics and evolutionary biology. I think the same is true of many of the patterns of argument employed in quantum theory. But I don't want to get side-tracked spelling out what should count as an instantiation of an argument pattern, or what considerations are relevant to assessing when a collection of argument patterns unify. Rather, my primary focus will be to see what light can be thrown by ideas such as these on the notorious interpretive problems of quantum theory—particularly the quantum theory of relativistic particles and fields.

For this purpose, I'll appropriate (with minor modifications) a definition of explanation given by R. I. G. Hughes [10], which will be recognized simply as a rough-and-ready expression of the intuition behind the remarks of Geroch cited above.

> We explain some feature B of the physical world by displaying a mathematical model of part of the world and demonstrating that there is a feature A of the model that corresponds to B, and is not explicit in the definition of the model.

It is natural to call explanations based on this maxim *structural* to emphasize that they need not be underpinned by causal stories and may make essential reference to purely mathematical structures that display the similarities and connections between phenomena. The informal requirement that B not already be built explicitly into the model carries some of the burden of the Friedman-Kitcher requirement that explanations unify, since it prevents spurious unifications based on models that simply catalogue all the phenomena to be explained without in any way 'organizing' it. And the spirit of Kitcher's requirement that good explanation consists of exploiting a small number of argument patterns is reflected in the fact that the rules of discourse about phenomena will be circumscribed, more or less rigidly, by the mathematical model itself.

Structural Explanation in Quantum Theory

One often hears philosophers say that they are not ready to take lessons from quantum theory until its interpretation is sorted out. The illusion that issues of explanation can be held at bay while a quantum metaphysics is developed was to some extent in the background of my early interpretive work on the nonrelativistic version of the theory. What initially attracted me to quantum theory was that one could start with some intuitive idea about what the quantum world could be like, some idea about causation or locality or identity, then make it mathematically precise, and finally see whether the idea was compatible with the structure of the theory. But why should this method be compelling if one doesn't already take that structure seriously as delimiting the possibilities for *explaining* quantum phenomena? And if it does, why not use the structure itself as one's point of departure, rather than forever seeking ways to reconcile various views of the world to it?

Well, there is a problem. In the first instance, quantum theory is a theory about the probabilities for measurement outcomes, but it leaves the outcomes themselves unanalyzed. And when we attempt to model the operation of measurement devices within the theory, not even a structural explanation of their outcomes seems possible.

Consider the physical magnitudes one might wish to measure on a particle, its position, momentum, energy, etc. For any one of these, let's focus on position, and any possible measurement outcome, say 'the particle is in the box', there is a state $|in\ the\ box\rangle$ of the theory that the particle could occupy in which a measurement of its position will yield the answer 'the particle is in the box' with certainty. In such a situation, one attributes a definite confinement to the particle—it *is* in the box—quite independently of the act of measurement. Probabilities arise because other position states of the particle are possible, like $|in\ the\ lab\rangle$, $|on\ the\ moon\rangle$, etc., and these states are represented by vectors in the theory. Being vectors, they can be added together yielding other vectors that are also legitimate states, like

$$\frac{1}{\sqrt{2}}|in\ the\ box\rangle + \frac{1}{\sqrt{2}}|elsewhere\rangle. \tag{1}$$

In this 'superposition' of states, the probability of finding the particle in the box is no longer one, but given by the square of the coefficient that sits in front of the $|in\ the\ box\rangle$ term, which is $1/2$. The same goes for the probability of finding the particle 'elsewhere' (i.e. outside the box), it too is $1/2$. Given these probabilities, it is no longer automatic that the particle can be thought of as having a definite confinement, either to the box or elsewhere. In fact, to say that the particle *does* have a definite confinement is standardly taken to be incompatible with the particle occupying the superposition state (1). For if we take into account all measurements we could perform on the particle (not just measurements of its position), the measurement statistics dictated by state (1) turn out *not* to be identical to those we would get by saying that the particle actually occupies one of the two states $|\textit{in the box}\rangle$ or $|\textit{elsewhere}\rangle$ with equal probability of being in either state.

Now let's try and bring measurement devices themselves within the theory. We need a device sensitive to detecting the particle in the box if it's really in there, and elsewhere if it is not. So the device must have two possible recording states, $|\textit{`in the box'}\rangle$ and $|\textit{`elsewhere'}\rangle$ (the quotes distinguishing detector from particle states). And the device must interact with the particle so that if it is initially in the state $|\textit{in the box}\rangle$ the detector will go into the state $|\textit{`in the box'}\rangle$, and likewise if the particle initially occupies the state $|\textit{elsewhere}\rangle$. But if we assume that the particle starts out in superposition (1) and model the particle-detector interaction in the usual quantum-mechanical way—evolving their joint state using Schrödinger's equation—then after the detection is complete, the state of the combined particle-detector system has to be

$$\frac{1}{\sqrt{2}}|\textit{in the box}\rangle|\textit{`in the box'}\rangle + \frac{1}{\sqrt{2}}|\textit{elsewhere}\rangle|\textit{`elsewhere'}\rangle. \qquad (2)$$

We get yet another superposition, this time involving states of the detector! If, then, the usual way of thinking about superpositions is correct, we cannot say the detector has definitely detected the particle either in the box or elsewhere, and so we cannot account for its results. Worse, what could we have meant by saying that there is equal probability of detecting the particle either inside the box or elsewhere?

At this point we have reached a major fork in the road. Either we modify or replace Schrödinger's equation so that superposition (2) can be allowed to 'collapse' with equal probability onto one of the two states that figure in (2), states in which the detector *does* have a definite state of detection. Or we reject the usual way of thinking about superpositions and amend the models of quantum theory so that detectors can still be taken to register definite results, superpositions notwithstanding. But while both collapse and no-collapse theorists do successfully locate measurement outcomes within their models, on closer examination it becomes evident that their models are set up so that they contain outcomes as part of their explicit definition! The prospect of explaining measurement outcomes, even just structurally, looks bleak.

As I noted earlier, a superposition makes different statistical predictions than if we attribute a state to the system which is simply a statistical mixture of the states in the superposition. What makes collapse theories viable in the face of the overwhelming evidence for the standard theory's superpositions is that the observables whose statistics one would need to measure to tell the difference between a superposition and the corresponding statistical mixture become increasingly difficult to measure when all the interactions between the system (here, our particle) and its environment (our detector) are explicitly taken into account. (This phenomenon is known as 'decoherence'.)

Nevertheless, if one is not inclined, as I am not, to alter the empirical content of the standard theory, one should try and find a way of turning no-collapse theories into proper structural accounts of quantum measurements. The problem is that if we abandon the standard assumption that a particle definitely possesses a property only when its quantum state dictates that the property is certain to be found on measurement, we need an alternative proposal for connecting quantum states to properties that will allow us to say measurements have outcomes without having to reject any of the mathematical structures around which the standard theory is built.

An important aspect of the structure worth mentioning is that the observables of any quantum system (position, momentum, etc.) are represented by a noncommutative algebra of operators on the system's state space, the mathematical space in which vectors like $|\textit{in the box}\rangle$ and $|\textit{elsewhere}\rangle$ reside. A major part of the literature on the foundations of quantum theory has been devoted to showing that this noncommutative algebra cannot support value assignments to all observables in the standard theory (unless one wishes to revise logic, which can be seen as an attempt at structural explanation of a different sort). But no-collapse theorists do not need *every* observable to possess a definite value, only enough observables to model measurement outcomes. What no-collapse theorists need is a way of distinguishing an appropriate *subset* of all the observables in the standard theory that *can* be ascribed definite values and to which the probabilities of quantum theory *can* sensibly refer. A good deal of my post-doctoral work was devoted to identifying some of the obstacles to this program while at the same time trying to carry it through (cf. [11]).

Suppose, now, that the program is complete. Would we have an explanation of measurement outcomes? On this, not all no-collapse theorists agree. The issue is dynamics. Once we have a coherent story about which observables of the standard theory possess definite values, it seems entirely natural to extend it to a story about how those values change over time. Will their time evolution be deterministic or indeterministic? Will particles have well-defined trajectories in space? If we sever the standard connection between quantum states and possessed values, answers to these questions are no longer fixed by the standard theory, which only gives a law of evolution for quantum states. In a recent paper with Michael Dickson [12] we showed that, to account for the puzzling correlations predicted by quantum theory, even no-collapse theories with an *indeterministic* dynamics for values face problems. In particular, these theories must violate Lorentz invariance, i.e. the relativistic requirement that the dynamical stories told by observers in different states of motion should have the same form, and be related to one another by 'Lorentz transformations' (so that the observers can all agree on the speed of light). But should no-collapse theorists give in to the demand for dynamics in the first place? It seems to me that serious commitment to the idea that structural explanations are good explanations will not necessarily privilege explanations that invoke dynamical structure over other structures.

I just alluded in passing to puzzling quantum correlations. And in the previous section I mentioned that even indeterministic causal processes will not suffice to recover quantum phenomena. In Salmon's *Four Decades of Scientific Explanation* [7] he places three issues at the top of his agenda of open problems for scientific explanation in the fifth decade (which is now drawing to a close). Central to Salmon's account of causal processes is the idea of a 'conjunctive fork', in which the correlation between two or more spatially separated events is traced to causal processes propagating from some common cause in their past. It is natural, then, that Salmon's third open problem should raise the question of explaining the puzzling nature of quantum correlations between spatially separated events.

> I have no idea what an appropriate explanation would look like; we may need to know more about the microcosm before any explanation can be forthcoming. But I do have a profound sense that *something* that has not been explained needs to be explained. ([7], p. 186)

I entirely agree with this last sentiment. So I want to end this section by sketching a structural explanation of probably the most puzzling correlations thrown up by quantum theory: correlations in the vacuum state of a relativistic quantum field.

For present purposes, I need not elaborate on what a relativistic quantum field is, nor on exactly what defines its vacuum state, except to note that it is the state of lowest energy of the field. Apart from this, the vacuum is not what its name suggests—it is a sea of activity. Any detector you might design, to measure any characteristic of the field, will always have a nonzero probability (albeit, generally *low* probability) of registering a value for that characteristic. This is the phenomena of vacuum fluctuations. The interesting thing for us is that the vacuum fluctations that can occur in *any* two nonoverlapping regions of space are *maximally* correlated. This means that any measurement outcome obtained in one region could have been predicted with (arbitrarily high) certainty on the basis of a measurement result obtained in the other region.

To appreciate how surprising this is, consider the well-known correlation between the reading on a barometer and stormy weather. When the barometer's reading falls one can (usually!) predict with certainty that a storm is on its way. Thus stormy weather is maximally correlated to the drop in the barometer's reading. Of course, the explanation for this correlation is that the falling barometer and stormy weather both arise from a common cause: a fall in atmospheric pressure. But in the case of vacuum fluctuations, *every* possible occurrence in one region is maximally correlated to some occurrence in the other. It's as if each and every feature of the weather were maximally correlated to some feature of the barometer! Moreover, in the vacuum maximal correlations like these obtain between the states of *any two* regions of space.

It seems almost unimaginable that we could find a common cause or causes in the past of these maximal correlations that would explain them *all*. And even if we could, the common causes would themselves be involved in exactly the same kind of correlations between spatially separated events, because the vacuum looks the same at all times. Either we embark on an infinite regress of explanation in search of common causes, or we have to accept the correlations as brute facts. In fact, the infinite regress is stopped at the very first step by the fact that vacuum correlations violate 'Bell inequalities' that all common cause models of spatial correlations are committed to. So we're stuck with accepting the correlations as brute facts.

Or are we? A physicist might be inclined to say that the vacuum is filled with maximal correlations because they are required for the field to satisfy restrictions imposed on its energy-momentum by relativity. But while true, the formal proof of this fact is not terribly enlightening and doesn't seem to instantiate any general argument pattern! Instead, let's see if we can get a sense of how maximal correlations could possibly arise in the vacuum by looking more generally at quantum states that can give rise to correlations like that.

First, a few more words about quantum states. Recall our particle that could be either in the box or elsewhere, and our particle detector that either detected it in the box or elsewhere. After the particle-detector measurement interaction, they were left in superposition (2). The most general state they could have been left in, if they had interacted in any old way, has the form

$$\begin{aligned} c_{11}|\textit{in the box}\rangle|\textit{`in the box'}\rangle \quad &+ \quad c_{12}|\textit{in the box}\rangle|\textit{`elsewhere'}\rangle \\ +c_{21}|\textit{elsewhere}\rangle|\textit{`in the box'}\rangle \quad &+ \quad c_{22}|\textit{elsewhere}\rangle|\textit{`elsewhere'}\rangle \end{aligned} \tag{3}$$

where the coefficient c_{11} is in general a complex number whose absolute square gives the probability that the particle was in the box and the detector detected it there, with a similar interpretation applying to the other coefficients c_{12}, c_{21} and c_{22} above. Since the four terms above represent all the four possible things that could happen, the sum of the squares of their coefficients, which are the probabilities that attach to each possibility, must sum to 1. It is convenient to represent state (3) by listing its coefficients in the 2×2 array, or matrix

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}.$$

So every possible state of our particle-detector system corresponds to a matrix of this form whose squared entries sum to 1. From now on I'll take it for granted that all our matrices are to have this property.

Notice that had we assumed the particle to be in one of three nonoverlapping regions, and the detector to have one of three recording states, then their state would be given by a 3×3 matrix. More generally, if we have two systems, the first capable of occuping m different states, and the second n different states, then their joint state will be given by an $m \times n$ matrix, with m rows and n columns. Still *more* generally, if we add a third system that can occupy p different states, then the joint state of all three systems will be given by a *three*-dimensional $m \times n \times p$ matrix with mn rows, mp columns and np 'files'. (See Figure 2 overleaf.) We can also envisage states of three systems given by infinite×infinite×infinite matrices, e.g. three particles each confined to one of an infinite number of nonoverlapping boxes. And if we have more than three systems, we just have to add on more dimensions to our matrices.

The connection between maximal correlations and quantum states is established via the idea of (linear) independence. To get the basic idea, let's

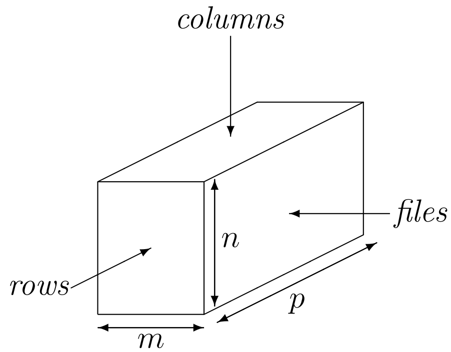

Figure 2. Three-Dimensional Matrix
(The entries would be spaced at regular intervals throughout the box.)

go back to the simple case of a 2×2 matrix state

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}.$$

Two mathematical operations can be performed on the rows and columns of this matrix. First, we could multiply all entries in a row or column by some number. Thus multiplying the row (c_{11}, c_{12}) by 17 yields $(17c_{11}, 17c_{12})$, which is also written as $17(c_{11}, c_{12})$. Second, we could add any pair of numbers to the entries in a row or column. Thus adding $(-3, 17)$ to the column (c_{11}, c_{21}) gives $(c_{11} - 3, c_{21} + 17)$, which is also written as $(c_{11}, c_{21}) + (-3, 17)$. The rows of a 2×2 matrix are called independent if for no numbers x and y is it the case that

$$x(c_{11}, c_{12}) + y(c_{21}, c_{22}) = (0, 0)$$

except when $x = y = 0$. Similarly, the columns of a 2×2 matrix are independent if the equation

$$x(c_{11}, c_{21}) + y(c_{12}, c_{22}) = (0, 0)$$

holds only when $x = y = 0$. For example, the first two of the following 2×2 matrices

$$\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \begin{pmatrix} 1/3 & 0 \\ 1/3 & 1/3 \end{pmatrix} \text{ and } \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}$$

have independent rows and columns, while the third has neither independent rows nor columns.

This concept of independence applies to matrices of any size and dimension. Thus, the three columns of a 2×3 matrix

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{pmatrix}$$

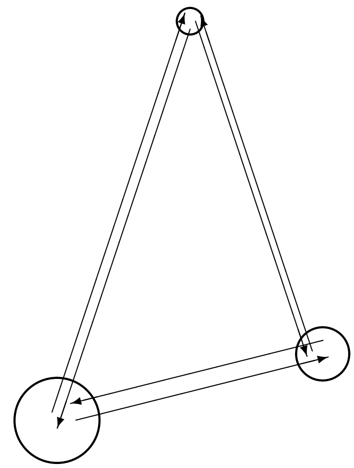

Figure 3. Maximal Correlations between Fluctuations in Three Regions

will be independent if

$$x(c_{11}, c_{21}) + y(c_{12}, c_{22}) + z(c_{13}, c_{23}) = (0, 0)$$

implies $x = y = z = 0$. In fact, it is not difficult to verify that when the number of columns exceeds the number of entries in each column, as in this case, they cannot be independent. The same would hold if the number of rows exceeded *their* length. Note, finally, that we can extend the idea of independence to matrices with infinitely many rows (columns, files, etc.) by stipulating that they are independent when any *finite* number of them are.

Now we can get some insight into maximal vacuum correlations. Focus just on the maximal correlations between three spatial regions, as depicted in Figure 3 above. (Considering more regions only complicates things without changing the fundamentals.) Treating each region as a 'system', we can assign it one of various quantum states. The three regions together will then have a state given by some three-dimensional matrix—as in Figure 2 previously. The physical requirement that vacuum-like maximal correlations obtain between the three regions turns out to be equivalent to the purely mathematical requirement that this three-dimensional matrix have independent rows, independent columns and independent files. (For the argument that establishes this for an arbitrary set of regions, and hence arbitrary-dimensional matrix, see [13].)

Now our three-dimensional matrix has dimensions $m \times n \times p$. Recall that the number of independent rows cannot exceed the length of each row, and likewise for columns and files. Thus for our three-dimensional matrix to have independent rows, columns and files we need (cf. Figure 2)

$$mn \leq p, \; pm \leq n, \; \text{and } np \leq m.$$

Multiplying both sides of the first inequality by m, using the second inequality and cancelling the n's, we get $m^2 \leq 1$ which implies $m = 1$! Using similar arguments, it follows that $n = p = 1$. But regions in the vacuum can certainly occupy more than *one* state, otherwise what sense can it make to say that *fluctuations* occur in each region? So our only option in the face of the inequalities above is take each of m, n and p to be infinite. The first thing we've learned, then, is that to sustain vacuum-like correlations, our three regions must each be able to occupy an infinite number of different states.

So think, now, of the set of all infinite×infinite×infinite matrices (whose squared entries sum to one!—so there can't be any lower bound on their size). Within this set, there *will* be matrices with independent rows, columns and files, though they are somewhat tricky to construct. Furthermore, if we start with *any* infinite three-dimensional matrix, it is always possible to find another with independent rows, columns and files that has entries as close to the entries of the first matrix as you like. (The argument for this is nontrivial; once again, see [13].) In a certain sense, then, almost every infinite three-dimensional matrix you write down is going to support vacuum-like correlations between our three regions!

I believe this provides a striking illustrating of structural explanation. We started with the puzzle of explaining maximal vacuum correlations without appeal to causal processes. To see how such correlations could be possible, we considered what the joint state of three regions of space needed to look like. We translated into the model the feature to be explained—mathematically, it is equivalent to the independence of the rows, columns and files of the state's matrix. We then noticed that only matrices with infinitely many entries can satisfy this condition, but also that once we restrict ourselves to infinite matrices the condition is satisfied by almost all of them. The upshot is that once one has committed to using the standard mathematics of quantum theory, one is *implicitly* committed to a whole host of states with vacuum-like correlations, and (I claim) a large part of the mystique surrounding those correlations evaporates.

Problems for Structural Explanation in Quantum Theory

I have yet to give any compelling argument for letting structural explanations stand alone as full-blooded explanations. I shall soon indicate some obstacles that I believe such an argument would need to face. But I first want to indicate briefly three foundational issues in the relativistic quantum theory of particles and fields that raise both opportunities and difficulties for a structural approach to explanation.

The first issue concerns problems that arise when one attempts to marry relativistic and quantum structures. If we try to realize Lorentz invariance in the state space of a free particle, it follows that the 'particle' has to be an intrinsically nonlocal object that cannot be confined to very small distances. Worse, the particle's 'position' states (relativistic counterparts to $|in\ the\ box\rangle$ and $|elsewhere\rangle$ of the previous section) allow the particle to behave in ways contrary to the spirit, if not the letter, of relativity. For one thing, these states can propagate faster than light. For another, an observer who merely changes her state of motion relative to the particle will see its state spread out over all of space, no matter what lengths are taken to confine the particle to a finite spatial region!

The standard response to these maladies is to abandon the search for a coherent relativistic quantum theory of particles and move directly on to the analysis of *fields*. But it is not clear what dictates this abandonment. Jeremy Butterfield and Gordon Fleming [14] have forcefully argued that no principle of relativity, properly understood, is violated by the strange position behaviour of relativistic quantum particles. Marrying relativistic and quantum structures gives birth to spectacular new predictions about particles which are *there in the formalism of the theory*, and so ought to be taken seriously unless we have direct evidence to the contrary. This kind of argument deserves close scrutiny, especially by those like me who profess to take the mathematical structure of physical theories seriously.

This leads directly to the next problem, which concerns how we should deal with surplus mathematical structure in our theories not instantiated in the physical world. Think again of the matrices we invoked in the previous section. In restricting ourselves to matrices whose squared entries summed to one, we were setting aside all other matrices as physically irrelevant. But this was innocuous given that for none of those other matrices would the probabilities for an exhaustive set of mutually exclusive possibilities sum to 1. After all, we wouldn't want to be put in the position of saying that half the time the particle will be found in the box and *three-quarters* of the time it will not!

Not all cases are that easy. Quantum field theory models collections of 'identical particles' by requiring that there be no quantity one can measure that will distinguish the particles from each other. This requirement puts a mathematical restriction on the states the particles can occupy. But the restriction is far weaker than the one Nature itself instantiates. We observe bosons and fermions, but never any of the other 'paraparticles' whose states would *also* prevent them from being distinguished one from another. In response, some physicists have devised clever theoretical arguments to exclude paraparticles *a priori*, but in the end they either proceed from dubious premises or beg the question. If indeed they all fail, then since there is no natural way to excise paraparticle states from the theory, a structuralist about explanation will have to accept their nonexistence in nature as a brute fact. Can a structuralist brush this off as easily as a constructive empiricist such as Bas van Fraassen [15], for whom empirical adequacy of a theory and the pragmatics of explanation are all that matters?

The third issue concerns the stance one should take towards essentially different formalisms of the same theory. Presumably for a structuralist the formalisms will *not* say the same thing and a choice needs to be made. The issue is usually discussed by philosophers in reference to choosing between a particle and field *ontology* for relativistic quantum field theory. But what I have in mind is a different sort of choice, between the standard vector space formalism and the algebraic approach to the theory. This latter is fast becoming the formulation of choice for pursuing the foundational problems of quantum field theory (on both flat *and* curved spacetime), and has its roots in classic papers written in the 30's and 40's by Jordon, von Neumann and Segal.

I do not think the suggestion is that we should all become realists about the algebraic structure of observables in quantum theories. Nor do I think (as a few commentators have worried) that the algebraic program relies on an operationalist philosophy of science in its reliance on observables as primitives. Instead, what motivates the program is the desire to capture the *intrinsic structure* of relativistic quantum field theory by associating algebras of local observables with regions of spacetime. In the 'concrete' approach to the theory, those observables are constructed out of quantum fields, but once constructed the algebraic approach counsels us to throw that ladder away. Will we then have isolated the instrinsic structure of the theory, the structure we should appeal to in our explanations? It is not clear, partly because the issue of surplus structure raises its ugly head again through the fact that the algebraic definition of a state is far more liberal than the standard one involving matrices that I employed in the previous section.

There are many more issues to be pursued than these (e.g. the challenges posed for our understanding of quantum processes by the emerging field of quantum computation). And, of course, there is much more to be done before one could take structural explanation seriously as a philosophy of physics.

Thus one could ask why locating phenomena within a unified mathematical model is anything more than a pragmatic virtue. One could also question whether structural explanations are always adequate to the task. In *How the Laws of Physics Lie* [16], Nancy Cartwright argues that idealized physical laws—and presumably this includes idealized mathematical *models*—fail to capture the richness and complexity of phenomena in the real world. Causes, even in quantum theory, must be invoked to fill the explanatory gap. On the other hand, there is no reason why mathematical models at varying 'distances' from the actual phenomena cannot share in the burden of explaining that phenomena. The problem would be to spell this out so that all contact is not lost with the model that sits at the highest level of abstraction and idealization. (A particularly stark example is the gap—indeed, chasm—that exists between the laws of atmosphere physics and models employed to forecast the weather.) I should also emphasize that I do not take structural explanation to be incompatible with explanation by appeal to causal structures. My claim is only that explanation *as* explanation does not privilege one sort of structure over any other. This attitude is, I believe, unavoidable in modern physics where all one typically has, at least at the highest level of abstraction, is a theory's mathematical structure.

There is of course another branch of philosophy in which the status of mathematical structures is controversial—the philosophy of mathematics itself. It strikes me that nothing less than a full-fledged philosophy of mathematics will be needed to make the ideas presented here compelling. In particular, there has been a recent surge of interest in explanation in *mathematics proper* which I think has great potential to shed light on explanation in modern physics. A mathematical explanation of why some theorem is true does not (cannot!) proceed by citing its causes. Nor will the explanation always be simply the deductive proof of the theorem, which might only tell us *that* the theorem is true without elucidating *why* it is true. Rather, the theorem's explanation seems to have something to do with how it fits into the larger picture of mathematical ideas. That picture I take to be best captured by category theory (about which philosophers of mathematics have largely been silent) which could be seen as playing roughly the same role in mathematics as a theoretical model does in physics. Furthermore, I suspect that a philosophy of mathematics based on category theory could well show many of the traditional problems of mathematical ontology to be misguided, just as structuralism about physics suggests a similar view about physical ontology. But this is not to suggest that structural explanation cannot be objective. In fact, my most difficult challenge will be to spell out why and *in what sense* structural explanation *is* objective.

In a wonderful little article on *Concepts of Cause in the Development of Physics* [17], Thomas Kuhn observes

> ...in physics new canons of explanation are born with new theories on which they are, to a considerable extent, parasitic. New physical theories have, like Newton's, repeatedly been rejected by men who, while admitting the ability of the new view to resolve previously intractable problems, have nevertheless insisted that it explained nothing at all. Later generations, brought up to use the new theory for its power, have generally found it explanatory as well. The pragmatic success of a scientific theory seems to guarantee the ultimate success of its associated explanatory mode. Explanatory force may, however, be a long time coming. The experience of many contempories with quantum mechanics and relativity suggests that one may believe a new theory with deep conviction and still lack the retraining and habituation to receive it as explanatory. ([17], p. 29)

I'd like to consider myself a contributor to that retraining so that modern physical explanation can be embraced as more than just force of (mathematical) habit.

References

[1] A. D. Sokal, "Transgressing the Boundaries–Toward a Transformative Hermeneutics of Quantum Gravity", *Social Text* **46/47** (1996): 217-252.

[2] C. P. Snow, *Public Affairs* (Scribner and Sons, New York, 1971).

[3] S. Weinberg, "Sokal's Hoax", *The New York Review of Books* **XLIII** No. 13 (August 8, 1996) pp. 11-15.

[4] C. G. Hempel, *Aspects of Scientific Explanation* (MacMillan, London, 1965).

[5] W. C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton University Press, Princeton, 1984).

[6] R. Geroch, *Mathematical Physics* (University of Chicago Press, Chicago, 1985).

[7] W. C. Salmon, *Four Decades of Scientific Explanation* (University of Minnesota Press, Minneapolis, 1989).

[8] M. Friedman, "Explanation and Scientific Understanding", *Journal of Philosophy* **LXXI** (1974): 5-19.

[9] P. Kitcher, "Explanatory Unification", *Philosophy of Science* **48** (1981): 507-531.

[10] R. I. G. Hughes, "Theoretical Explanation", *Midwest Studies in Philosophy* **XVIII** (1993): 132-153.

[11] R. Clifton, "Beables in Algebraic Quantum Mechanics", forthcoming in *From Physics to Philosophy: Essays in Honour of Michael Redhead*, Cambridge University Press, 1998.

[12] M. Dickson and R. Clifton, "Lorentz Invariance in the Modal Interpretation", forthcoming in Dennis Dieks and Pieter Vermaas (eds.), *The Modal Interpretation of Quantum Mechanics*, Kluwer Academic Publishers, 1998.

[13] R. Clifton, D. V. Feldman, M. L. G. Redhead, and A. Wilce, "Hyperentangled States", http://xxx.lanl.gov/archive/quant-ph (paper no. 9711020).

[14] J. Butterfield and G. Fleming, "Strange Positions", forthcoming in *From Physics to Philosophy: Essays in Honour of Michael Redhead*, Cambridge University Press, 1998.

[15] B. C. van Fraassen, *Quantum Mechanics: An Empiricist View* (Clarendon Press, Oxford, 1991) Ch. 11.

[16] N. Cartwright, *How the Laws of Physics Lie* (Clarendon Press, Oxford, 1983).

[17] T. S. Kuhn, *The Essential Tension* (University of Chicago Press, Chicago, 1977) Ch. 2.