The Miraculous Consilience of Quantum Mechanics

Malcolm R. Forster
Department of Philosophy,
University of Wisconsin-Madison
March 21, 2005

ABSTRACT: Two events are said to be (positively) correlated when the occurrence of one increases the probability of the other. If neither event causes the other, then a causal model "ties correlated events together" by postulating the existence of a common cause, or a hidden variable. But, Bell-type examples present multiple correlations that common cause models cannot explain because they fail to tie the *correlations* together *in the right way*. On the other hand, quantum models succeed in this task. The successful unification of disparate phenomena is a feature of good scientific theories that William Whewell referred to as the consilience of inductions. This essay describes how quantum mechanics achieves this consilience, and then presents this an independent argument for abandoning the interpretation of quantum probabilities as measures of our ignorance of hidden variable values.

1. Baby Epistemology

Think back to a time when mirrors were a new experience to us as children. How did we learn that what we saw in mirrors was not a different world, but a reflection of the world we already knew? How did we acquire the parsimonious view that reflections were independent views of the same objects?[1] Presumably, it has something to do with the way that mirror images are correlated with our more direct perception of the objects.

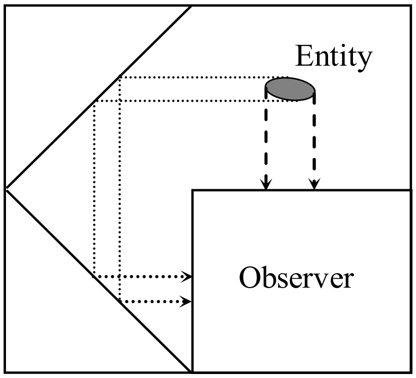

Figure 1: Reichenbach's cubical world.

Reichenbach (1938) elucidated a similar idea by imagining an observer enclosed in the corner of a cubical world, where objects in the external world cast shadows on the walls of the enclosure. It seems plausible that much more information about external world would be available if external objects cast two independent shadows on two walls of the enclosure, rather than a single shadow, or copies of a single shadow. It seems that the epistemological engine inside our heads tends to favor the judgment that they are the shadows of a single object, rather than shadows of different objects.

There is no good formal theory of how such inferences work, although they seem to conform to informal principles such as the principle of parsimony, or Occam's razor, which is usually taken to state that "Entities are not to be multiplied beyond necessity."[2] The rule is vague in crucial ways. For one, what counts as an entity? And under what conditions does it become necessary to multiply an entity?

Consider the first question more carefully. Do probabilities count as entities? Is there, in other words, a principle that says that probabilities are not to be multiplied beyond necessity? In section 3, I argue that such a principle is useful in causal modeling for it provides an account of the time asymmetry of cause and effect that does not rely on the concept of manipulation or intervention. On the other hand, there are well known circumstances in which causal modeling fails. In section 4, I show that the application of Occam's razor to probabilities leads to false predictions in the infamous double-slit experiment. The question then arises: Does Occam's

---

[1] This mirror example is used in Hung (1997).

[2] The principle is attributed to William of Ockham (~1280 - 1347 AD), although it has since taken on a life of its own.

razor apply to quantum mechanical probabilities? Or is quantum mechanics (QM) parsimonious with respect to other kinds of entities? This is a question that I attempt to answer affirmatively in sections 5, 6, and 7, at least in the context of a concrete example.

The concrete example is interesting because it is similar to the example used by Bell (1964, 1971) to challenge local hidden variable interpretations of QM. I provide a more recent version of Bell's derivation of a false prediction in section 6, while in section 7 I show that it is actually very close to the way that QM makes the correct prediction (just replace variables with operators).

Bell's theorem can be viewed as an argument that at least some quantum mechanical phenomena have no causal explanation (van Fraassen 1983), which agrees with the two theses argued in sections 3 and 4: (1) The constancy or invariance of probabilities is an essential component of causal modeling, and (2) the assumed invariance of probabilities leads to false predictions in the double-slit experiment. The most obvious conclusion is that there is no causal explanation of the double slit phenomena, and the advantage of Bell's argument is that it is harder to wriggle out the conclusion (though not impossible). But does QM therefore fail to conform to Occam's razor? Does it fail to describe the world behind the shadows in a unified way? The main goal of this essay is to argue, in at least one concrete example, that it is an exemplary kind cubical world inference, which does conform to Occam's razor

I have not provided a formal account cubical world inference because there is no such account in my view. Causal modeling is one kind of cubical world inference, and QM modeling is another, and two are formally quite different. But this is not to say that there is no overarching epistemological theory that covers both cases. In fact, a good informal description of what might be viewed as cubical world inference was published by William Whewell in 1858!

2. Whewell's Consilience of Inductions

Long before the discovery of quantum mechanics, William Whewell (1858) claimed that what he called the colligation of facts proceeds in three steps: (1) the Selection of the Idea, (2) the Construction of the Conception, and (3) the Determination of the Magnitudes (see Butts 1989, 223-237). In curve-fitting, for example, the selection of the idea is the selection of the independent variable (x) from which we hope to predict the dependent variable (y). The construction of the conception is determined by the choice of the formula connecting the two variables (family of curves). And the determination of magnitudes is what statisticians now refer to as the estimation of the adjustable parameters. Whewell then makes an insightful claim about curve-fitting:

> If we thus take the *whole mass of the facts*, and remove the errours of actual observation, by making the curve which expresses the supposed observation regular and smooth, we have the separate facts corrected by their general tendency. We are put in possession, as we have said, of something more true than any fact by itself is. (Whewell, quoted from Butts 1989, 227.)

In other words, the purpose of the curve, or the formula, is to describe the regularity behind the data, the signal behind the noise, which is "more true" than the sum of the observed facts. Unification is therefore the key to scientific innovation:

> The particular facts are not merely brought together, but there is a New Element added to the combination by the very act of thought by which they are combined. There is a Conception of mind introduced in the general proposition, which did not exist in any of the observed facts. ... The pearls are there, but they will not hang together until some one provides the string. (Whewell, quoted from Butts 1989, pp. 140, 141.)

In the case of curve-fitting, the conceptual string is the curve, but Whewell also insists that the same metaphor applies to all instances of scientific induction.

The best examples of science consist in much more than the mere colligation of facts (which is Whewell's name for 'induction'). It also strives towards what he calls the *consilience* of inductions (note the plural). The mark of a good theory lies not in the relationship between the theory and its data *in a single narrow application*, but in the way it succeeds in 'tying together' *separate* inductions. A good theory is like a tree that puts out runners that grow into new trees, until there is a huge forest of mutually sustaining trees. The 'tying together' can be achieved in either of two ways: (a) The theory accommodates one set of data, and then predicts data of a different kind. (b) The theory accommodates two kinds of data separately, and then finds that the magnitudes in the separate inductions agree, or that the laws that hold in each case 'jump together'.[3] In either case, there is a *consilience of inductions*, which enables the theory to provide a more unified and parsimonious description of the world.[4]

The main purpose of this essay is to illustrate how QM achieves a consilience of many inductions, which would be miraculous if there were no reality behind the observed phenomena. But whatever the nature of this reality, it does not appear to conform to our common sense causal picture of the world.

The right kind of consilience provides a positive criterion for the success of causal modeling; but it cuts both ways. When the right kind of consilience fails to occur, there is good reason to look for something different.

3. The Consilience of Causal Models

Consider a very simple example—the jackpot machine. The machine has two input states: Either one euro coin is placed in the machine, or two euro coins are placed in the machine, then a handle is pulled. The output state is either 'win' or 'lose'. For the sake of concreteness, suppose 'win' refers to the event that machine delivers 10 euros as a 'jackpot', and a loss leads to a zero payout. In this example, the input states do not *determine* unique output states. The 'mechanism' built into the machine is best described in terms of the constancy or invariance of *probabilities*.

To make this point, introduce two variables, X and Y. Upper case letters are used in statistics to denote variables that have probabilities associated with their possible values. Such variables are called *random variables*. X represents the input state of the machine in some particular trial of the experiment, while Y represents the output state in the same trial. The possible states of the machine can be represented by assigning arbitrary numerical values to these variables. Let $X = 1$ denote the event that only one euro is placed in the machine before the handle is pulled, while

---

[3] When the magnitudes agree, then we have what is more commonly referred to as an agreement of independent measurements. It is no coincidence that Newton scholars, such as Harper (2002), emphasize the importance of the agreement of independent measurements in Newton's argument for universal gravitation. For Whewell was also primarily concerned with the explication of Newton's methodology. See also Myrvold and Harper (2002) for an argument that this kind of evidence is not properly taken into account in standard statistical methods of model selection.

[4] My purpose is not to argue for some particular historical or exegetical thesis about Whewell's notion of consilience, but to use (and adapt) Whewell's idea for the purpose of explaining how probabilistic theories work in general, and how quantum mechanics works in particular.

$X = 2$ denotes the event that 2 euros are placed in the machine. $Y = 0$ denotes the event that there is no payout, while $Y = 0$ denotes the event that the jackpot (of 10 euros) is paid out.

(Alice)	loss	win
1 euro	90	10
2 euros	80	20

The standard ‘forward’ causal model says:

$$P(\text{win}\,|\,1\,\text{euro}) = \alpha\text{, and } P(\text{win}\,|\,2\,\text{euros}) = \beta\text{, for all trials } i.$$

That is, different trials have something in common; namely, that the values of the forward conditional probabilities are the same in each case. Note that the model postulates constant values for *all* forward probabilities because

$$P(\text{loss}\,|\,1\,\text{euro}) = 1-\alpha\text{, and } P(\text{loss}\,|\,2\,\text{euros}) = 1-\beta\text{, for all trials } i.$$

These four probabilities are *postulated* by the model. They count as *theoretically* ‘entities’. Their measurement, or estimation, is determined from the available data in the same way that any theoretical quantity is measured. Consider a typical set of data. Suppose that Alice plays the machine 200 times, and we record the input and output state on each trial. The data consist in a sequence of data ‘points’, which come in four flavors: $(1,0)$, $(1,1)$, $(2,0)$, or $(2,1)$. Because the model says that the temporal order of the trials does not matter, the data are adequately recorded in the table of the observed frequencies.

Notice that Alice plays the machine with 1 euro half the time and with 2 euros half the time (100 trials each). Out of all the times she plays the 1 euro version of the game, she wins the jackpot 10 times, thus earning 100 euros, which is the same amount that she paid to play. Out of all the times she pays 2 euros to play, she wins twice as often, and earns a total of 200 euros (twice as much). But she paid twice as much to play, so she still earned the same as what she paid to play. On the basis of the data, the machine appears to be fair.

Whewell’s three steps in the colligation of facts apply to this example in the following way. Step 1 consists in the selection of X and Y as the relevant quantities to be considered. Step 2 introduces the formula, which in this case is probabilistic in nature. It introduces a family of probability distributions parameterized by the adjustable parameters α and β. These determine the Conception. In this example, the conception is probabilistic in nature. Step 3 is the determination of the magnitudes α and β from the data. In contemporary statistical theory, this is achieved by the method of maximum likelihood estimation (MLE).

To understand how MLE works, first note that each pair of values assigned to α and β picks out a particular probabilistic hypothesis in the model. The fit of each hypothesis with the data is defined by its likelihood, which, by definition, is the probability of the data given the hypothesis (this should not be confused with the probability of the hypothesis given the data, which is a distinctly Bayesian concept). The greater the likelihood of a hypothesis (the more probable it makes the data) the better the hypothesis fits the data. The hypothesis that fits best is, by definition, the maximum likelihood hypothesis. For arbitrary values of α and β, the likelihood is:

$$\text{Likelihood}(\alpha, \beta) = (1-\alpha)^{90}\alpha^{10}(1-\beta)^{80}\beta^{20}.$$

The probabilities are multiplied together because each trial is probabilistically independent of all the others (according to the model) in the same way that coin tosses are independent. Mathematically speaking, maximizing the likelihood is the same as maximizing the log-likelihood.

$$\text{log-Likelihood}(\alpha, \beta) = [90\log(1-\alpha)+10\log(\alpha)]+[80\log(1-\beta)+20\log(\beta)].$$

Again, the terms in the square brackets can be maximized separately by differentiating with respect to α and β and putting the resulting expressions equal to zero. (You need to know that $d\log x/dx = 1/x$.) After multiplying by the factors $\alpha(1-\alpha)$ and $\beta(1-\beta)$, respectively, the equations simplify to:

$$-90\alpha+10(1-\alpha)=0 \text{ and } -80\beta+20(1-\beta)=0.$$

These equations yield the estimates: $\hat{\alpha}=0.1$ and $\hat{\beta}=0.2$. Accordingly, the theoretically postulated probabilities are estimated by the natural relative frequencies in the data, just as one would naïvely expect.

The question is whether the ‘forward’ model, which is the standard model, has evidence in its favor that the ‘backward’ model does not. The backward model seeks to predict values of X from values of Y (in the same probabilistic sense of ‘prediction’). In other words, it postulates backward probabilities as follows:

$$P(2\,\text{euros}\,|\,\text{loss})=\gamma, \text{ and } P(2\,\text{euros}\,|\,\text{win})=\delta, \text{ for all trials } i.$$

(Bob)	loss	win
1 euro	90	10
2 euros	160	40

Using the same methods as before, the parameters of the backward model are estimated by the corresponding relative frequencies in the data. The estimated values are: $\hat{\gamma} = 0.47$ and $\hat{\delta} = 0.67$. There is nothing that points to any empirical difference between the models. In fact, they cannot be compared with regard to their fit to the data because they tackle very different prediction tasks. The fit of the forward model is measured in terms of *Y*-values, while the fit of the backward model is measured in terms of *X*-values. In a sense, the two models are incommensurable.

So, why don't the two models happily co-exist? There is a sense in which they do co-exist, for the forward model is capable of making backward predictions if it is provided with information about the relative frequency of 1 euro versus 2 euro trials. With this information, together with the estimated forward probabilities, the forward model can calculate the backward probabilities, and gets the same answer as backward model. This is because both models accommodate the same table of data. But this only serves to deepen the puzzle. If both models can be seen as predictively equivalent, why interpret one as causal and not the other?

This is the familiar puzzle about cause and correlation: The evidence for causation cannot be exhausted by the correlations in the data, for correlations are symmetric, while causation is not. Either there is no additional empirical evidence, in which case causal inference is based on non-empirical criteria (or psychological habit!), or else there is other evidence that breaks the symmetry.

One solution is to look at how the models predict data of a different kind. Suppose that Bob plays the jackpot machine, and he happens to play with 2 euros twice as often as with 1 euro. Given that the forward model is true, the frequency that Bob wins will conform to the same forward probabilities, modulo a fluctuation in the data due to sampling errors. The sampling fluctuations are not relevant to this discussion, so imagine that there are no sampling errors. Then data from Bob's trials are described by the natural frequencies in the table.

Given the way that the example was set up, it is hardly surprising that the independent measurements of α and β agree. The key question is whether the same is true for the backward model. If it is not, then the symmetry between the models is broken.

A simple calculation shows that the Bob's estimates of the parameters of the backward model are $\hat{\gamma} = 0.64$ and $\hat{\delta} = 0.80$, which are quite different from the previous values, which were

$\hat{\gamma} = 0.47$ and $\hat{\delta} = 0.67$. Therefore, the backward model fails the test of consilience, and the symmetry between the forward and backward model is broken on empirical grounds (Forster 1984, Sober 1994, Arntzenius 1997).

The symmetry between the models is restored if Alice's and Bob's data are pooled—in fact, the models are always symmetric with respect to any single data set, as was already shown by considering Alice's data. The evidence is relational; and the consilience of inductions is a relation between different colligations of facts.

4. A Failed Consilience in the Double-Slit Experiment

The postulation of hidden variables is a way of interpreting the probabilities of quantum mechanics as 'measures of ignorance'. Einstein, for example, believed that a future physics would reveal the existence of such hidden variables, and the decay of radioactive particles, for example, could be predicted exactly. Bohr, on the hand, thought that quantum physics was complete in the sense that quantum probabilities are here to stay.

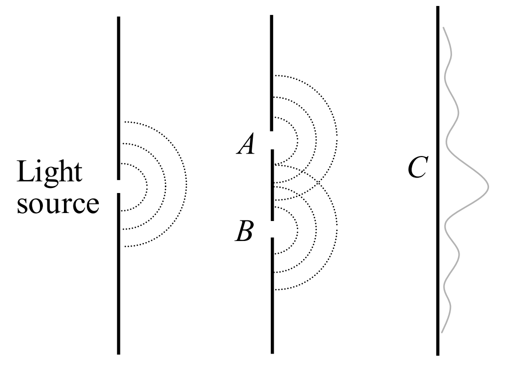

Figure 2: In the double-slit experiment, the assumption that individual photon travel through either slit *A* or slit *B* leads to the false prediction that double-slit pattern is the sum of two single-slit pattern.

At the time of the debate, a common example was the double-slit experiment. Consider a particle of light (photon) that leaves the light source and travels through either slit *A* or slit *B* (and not both). We may represent this event in terms of a random variable *X*, which can have one of two values, x_A and x_B.[5] For our purposes, it doesn't matter what numerical values we use. Now introduce a second random variable *Y* such that $Y = 1$ if the particle is detected in some region C, and 0 otherwise. In a more natural shorthand notation, let *A* stand for the event $X = 1$, *B* of for the event $X = -1$, and *C* is the event $Y = 1$. The probability of *C* given *A* is written $P(C|A)$. Similarly, $P(C|B)$ is the probability that it arrives at *C*

---

[5] Recall that a random variable is, by definition, any variable that has a probability distribution associated with it.

given that it is passes through slit *B*. $P(A)$ is the probability that the particle passes through slit *A* given that it passes through any slit. Then the probability of *C* (given that it arrives somewhere on the screen) is, according the axioms of probability:

$$P(C) = P(A)P(C\,|A) + P(B)P(C\,|B)\,.$$

	C	Not-*C*
A	40	10
not-*A*	20	30

It is worthwhile working through the proof of this theorem, because it introduces some elementary concepts of probability theory that are controversial in QM. Let's suppose that the particle passes through slit *A* or *B*, but not both. Then *B* is logically equivalent to the event not-*A*. Suppose that 50 particles pass through slit *A* and 50 particles pass through slit *B*. Of the 50 particles that pass through slit *A*, 40 arrive at *C*. Of the 50 particles that pass through slit *B*, 20 arrive at *C*. Therefore, a total of 60 out of 100 particles arrive at *C*. In symbols, $P(A) = 50/100$, $P(C\,|A) = 40/50$, $P(B) = 50/100$. $P(C\,|B) = 20/50$. Therefore,

$$P(A)P(C\,|A) + P(B)P(C\,|B) = \frac{50}{100}\frac{40}{50} + \frac{50}{100}\frac{20}{50} = \frac{60}{100} = P(C)$$

$P(A\,\&\,C)$ is the probability that the particle passes through slit *A* *and* arrives at *C*. According to the table, this joint probability is $40/100$. The argument assumes that joint probabilities, such as $P(A\,\&\,C)$, exist.

So far, there is no problem. But now postulate that the probabilities $P(C\,|A)$ and $P(C\,|B)$ do not depend on whether the other slit is open or closed. This is often called a *locality assumption* because it assumes that what happens to a particle passing through one slit is unaffected by what is happening non-locally (at the other slit in this case). It is also an invariance assumption—it is an attempt to unify the phenomena of the single-slit and double-slit experiments. It is might be viewed as an application of the principle that "Probabilities are not to be multiplied beyond necessity."

Unification is important because it leads to predictions. But in this case it leads to the false prediction that the double-slit pattern is the average of the two single slit patterns. If the slits are a certain distance apart, then is double-slit interference pattern has its brightest spot at *C* (see Fig. 2), whereas any average of the single slit patterns has the brightest spots directly in front of the two slits. The invariance of the probabilities provides a *potential* consilience of inductions, but when the consilience fails, the assumptions on which it is based are called into question.

There are ways of resisting this conclusion. The first obvious response is that different photons are interacting with each other after they pass through the slits. This possibility is highly implausible in light of the fact that the interference pattern is exactly the same if the intensity of the light is so low that only one photon passes through the slit at a time.

Another possibility is that there is some kind of continuously emitted pilot wave that guides the particles to their appropriate destinations. The pilot wave is affected by whether the second slit is open or closed. This hypothesis is ad hoc if there is no way of independently detecting the existence of the pilot wave. But it doesn't lead to any false predictions. To that extent it solves the problem. But it does not produce a new consilience.

The argument that probabilities should be invariant across the single-slit and double-slit experiments is based on the assumption that the particles pass through one slit or the other in every instance. It is an important prediction of quantum mechanics that if the precise trajectory is experimentally determined, then the classical prediction *is* correct—that is, in this case the pattern on the screen will be the sum of the two single-slit patterns. For example, electrons can be detected going through a particular slit by the electric current they induce when they pass through a wire loop. So if we place wire loops behind the two slits, then the sum of the two single slit patterns will be observed.

The problem with the causal explanation is that it is *sometimes* wrong. So, how does the quantum mechanical formalism succeed where the causal account fails. The causal model unified the phenomena by assuming that the *probability distributions* in the two single slit experiments *add* together to produce the probability distribution in double slit experiment. Quantum mechanics replaces the additivity of probabilities with the additivity of wave functions. Let's introduce a random variable Y, where y denotes an arbitrary value of Y, to represent the possible points on the screen at which the particle may be detected. Now consider the single slit experiment in which slit B is closed. Suppose a quantum mechanical model entails a wave function that does not depend on time, and has a complex number $\psi_A(y)$ associated with each point on the screen. Then the model implies that the probability that a particle is detected near the point y is proportional to $|\psi_A(y)|^2$. Note that the square of the magnitude of complex number is a non-negative real number. Similarly, the probability of a particle landing near y is $|\psi_B(y)|^2$ in the other single slit experiment. Then the probability of a particle landing near the same point in the double slit experiment is:

$$|\tfrac{1}{\sqrt{2}}\psi_A(y)+\tfrac{1}{\sqrt{2}}\psi_B(y)|^2=\tfrac{1}{2}|\psi_A(y)|^2+\tfrac{1}{2}|\psi_B(y)|^2+\text{ interference terms.}$$

If the interference terms are zero, then the prediction is that same as the hidden variable prediction. QM allows for the additivity of probabilities, but the additivity, or superposition, of wave functions is more fundamental. The QM model succeeds in unifying a wide range of disparate phenomena. It leads to a consilience within a vast network of inductions.

The QM prediction that there will be interference of the statistical patterns can be traced back to the replacement of variables with operators, which I call the miraculous conception of QM. This point will be explicit in the Bell-type of example involving electron spins examined in following sections.

In summary: QM modeling is very different from causal modeling because it uses operators in place of variables, and uses wave functions (state vectors) rather than probabilities to represent the constancies of nature. Nevertheless the conceptual novelties of QM produce an impressive consilience of a variety of inductions, which is the strongest kind of evidence that any scientific theory can have.

5. Sequential Spin Measurements on a Single Electron

In the quantum mechanical model of electron spin, there is a QM spin observable corresponding to spin in every direction in 3-dimensional space. Spin in the classical world has just one axis of rotation, so we should not try to read too much into the word ‘spin’ in quantum mechanics. For us, it is just a name of a new kind of QM entity.

There are really only two facts that you need to understand about quantum mechanical spin. The first is that if an electron passes through a non-linear magnetic field produced by a Stern-Gerlach magnet, then the exiting electron will be detected in one of two possible paths, the ‘up’ path or the ‘down’ path, provided that detectors are placed there. We shall always assume that the electron is always traveling in the z direction and all Stern-Gerlach magnets are aligned in a direction perpendicular to z (for example, x or y).

In a *sequential measurement* the election is first ‘prepared’ by passing it through one magnetic field, after which it is directed through a second Stern-Gerlach device. We need three devices in all, and there 4 possible exit paths, and the particle is detected in exactly one path by the detectors placed there (see fig. 3). When a particle is detected, it is destroyed, so that it cannot be subjected to further measurement.

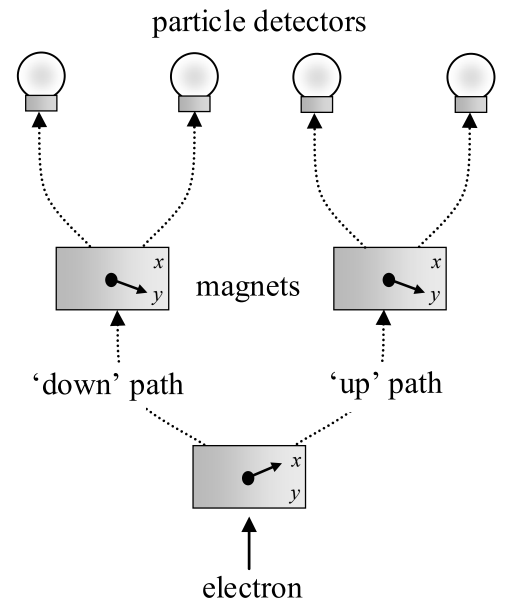

Figure 3: Sequential spin measurements on a single election that lead to four possible outcomes.

If the particle is detected by the Geiger counter placed in the 'down path of the magnet placed in the 'up' path of the first magnet, then we say that the outcome of the experiment is 'up-down'. In a classical framework, we infer from this outcome that if detectors had been placed in the exits paths of the first magnet, it would have been detected in the 'up' path. We may represent this conclusion by assigning a value to a variable. Assume that the first magnet is oriented in the x-direction, and introduce to mean that if the detectors had been placed directly in the exit paths, without any intervening device, then it would have been detected in the 'up' path. While we may think of the experimentally determined value of X_0 as a measurement outcome, it is important understand that it is based on a theoretical assumption; namely that the variables represent real properties of the election. This seemingly innocuous assumption is not used in the QM analysis.

Now consider the fact that the outcome was 'up-down' rather than 'up-up'. By convention, the statement $Y_1 = -1$ means that the electron would be detected in the 'down' path after passing through a device oriented in the y-direction. Further suppose that the second device was actually oriented in the y-direction. Then we infer that $Y_1 = -1$. Therefore, from the experimental outcome, we infer that $X_0 = 1$ and $Y_1 = -1$. There are four possible conclusions drawn from four possible outcomes.

From repeated trials of this experiment we discover that each of the four outcomes occurs equally often. This statistical fact is represented in terms of probabilities:

$$P(X_0 = 1) = \tfrac{1}{2} \text{ and } P(Y_1 = 1 \,|\, X_0 = 1) = \tfrac{1}{2} = P(Y_1 = 1 \,|\, X_0 = -1)\,.$$

This implies that the two variables are probabilistically uncorrelated. Various other facts emerge as well, such as the fact that the statistics are insensitive to the distance separating the second magnet from the first or how far the detectors are placed from the second magnet. By varying the orientations of the magnets, we also discover that the correlation between the variables depends only on the angle between them. When the magnets are oriented in the same direction we say that the electron is subjected to a repeated measurement. In that case, the variables are perfectly correlated. That is, $X_1 = 1$ if and only if $X_0 = 1$, or in probabilistic terms,

$$P(X_1 = 1 \mid X_0 = 1) = 1 \text{ and } P(X_1 = 1 \mid X_0 = -1) = 0.$$

This hidden variable representation of the observational facts can be extended to an arbitrary number of sequential measurements. For example, if we 'measure' the values of X_0, Y_1 and X_2, then we discover that the statistical relationship between X_0 and Y_1 is the same as before. It is not changed by the third measurement. Moreover, the statistical relationship between the last two variables is independent of the inferred value of the first variable. In symbols,

$$P(X_2 = 1 \mid Y_1 = 1, X_0 = 1) = P(X_2 = 1 \mid Y_1 = 1, X_0 = -1) = \tfrac{1}{2},$$

and so forth. But the relationship between the first and the third variables is changed by the second measurement. In particular, we find that $P(X_2 = 1 \mid X_0 = 1) = \tfrac{1}{2}$ and $P(X_2 = 1 \mid X_0 = -1) = \tfrac{1}{2}$, which is different from the relationship between X_0 and X_1. X_0 and X_2 are uncorrelated, whereas X_0 and X_1 were correlated. In fact, the relationship between X_0 and X_2 depends on the existence of intermediate device and its orientation. For example, if the intermediate measurement were in the x-direction, then X_0 and X_2 would be correlated. All this implies that the probability distributions depend on certain features of the total measurement setup. Once this caveat is in place, there are no false predictions that can be derived from the hidden variable representation of sequential spin measurements, as far as I'm aware.

The experimental facts might be summarized in the following way. Suppose that a measurement in the *u*-direction is followed immediately by a measurement in the *v*-direction, with associated variables U and V. Let θ be the angle between *u* and *v*. Then

$$P(V = 1 \mid U = 1) = \cos^2(\theta/2) \text{ and } P(V = 1 \mid U = -1) = \sin^2(\theta/2).$$

Notice that these forward probabilities are invariant, as required of the very best causal models. In fact, this quantum mechanism would be ideal for building a jackpot machine because the probabilities are precisely controlled by the orientations of the magnets. However, is the underlying mechanism really causal? Any model must be subjected to ever broadening standards of consilience. If some new-fangled QM model preserves the consilience of sequential spin phenomena and extends it predictions to a broader domain, then it trumps the causal model.

This is exactly what happens. Hidden variable models make false predictions about ‘entangled’ pairs or triplets of electrons, as Bell (1964) first proved. But QM models not only make the correct predictions about entangled states, but they do so *on the basis of what is learned* from sequential measurements. As Whewell’s might put it, the QM induction of sequential data explains and predicts data of a different kind, and it does so by superimposing a new conception on the facts. The remainder of section is designed to deepen our understanding of the colligation facts in QM.

Consider the ‘empirical fact’ $P(X_1 = -1 \mid X_0 = 1) = 0$. QM introduces vectors x^+ and x^- to represent the states $X_0 = 1$ and $X_1 = -1$, respectively, and then derives the probability $P(x^- \mid x^+) = 0$ from mathematical properties of vectors. The probability is 0 because the projection of the vector x^+ onto the vector x^- is 0, which is to say that x^+ and x^- are orthogonal vectors. The same argument applies to the two states associated with an arbitrary direction of measurement v^+ and v^-, where the physical angle between the two directions is θ. The question is whether there exists QM model in which all four vectors are in the same 2-dimensional space? If there exists such a unified model, then the four vectors would appear as drawn in Fig. 4, for some adjustable angle ϕ. In that case, any vector in the space can be written as a linear combination of x^+ and x^-. That is, all states would be superpositions of the states x^+ and x^-. Superposition is therefore a unifying conception in QM.

The crucial question is whether it is possible to adjust ϕ so that the ‘cross’ probabilities, $P(v^+ \mid x^+)$ and $P(v^+ \mid x^-)$, are related in the right way to the physical angle θ. To answer this question, we need to understand more about how probabilities are calculated in QM

Clearly, want to calculate probabilities in such a way that $P(v^+ \mid x^+) + P(v^- \mid x^+) = 1$, since any electron prepared in the x^+ state will go ‘up’ or ‘down’ after exiting a Stern-Gerlach magnet oriented in any direction. A trivial instance of this is $P(x^+ \mid x^+) + P(x^- \mid x^+) = 1$, which implies that $P(x^+ \mid x^+) = 1$. The vector space formalism attributes this probability to the geometric fact projecting x^+ onto x^+ leaves x^+ unchanged. But how do we extract the number 1 from this geometric fact? QM says in general that $P(v^+ \mid x^+)$ is the dot product of the vector x^+ and the projection of x^+ onto v^+. In the special case in which $v^+ = x^+$, this is the dot product of x^+ times x^+, which is just the square magnitude of the vector x^+. Clearly, this 1 if and only if x^+ is a unit vector, so all state vectors in QM are unit vectors.

Now reconsider $P(v^+ \mid x^+)$. Since all vectors are unit vectors, the dot product of x^+ and the projection of x^+ onto v^+ is equal to $\cos\phi\, v^+ \cdot x^+ = \cos^2\phi (x^+ \cdot x^+) = \cos^2\phi$. Clearly, we will match the empirical probabilities if and only if $\phi = \theta/2$. Incidentally, this is why it's called spin ½ in QM.

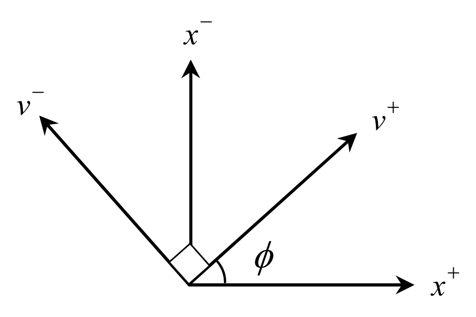

Figure 4: In QM, the transition probabilities between states can be generated from the geometrical relationship between vectors.

In place of arithmetic variables, QM makes use of geometric objects. What's remarkable is that these geometric objects have properties that not only colligates one kind of phenomena, but extend also to disparate classes of facts.

Vectors may be represented as column matrices or as row matrices; we need a notation that reflects the difference. Instead of x^+, we shall write $\left|x^+\right\rangle$ for the column vector, while $\left\langle x^+\right|$ stands for the corresponding row vector. The dot product of $\begin{bmatrix} u_1 & u_2 \end{bmatrix}$ and $\begin{bmatrix} v_1 & v_2 \end{bmatrix}$ is equal to $u_1 v_1 + u_2 v_2$, which is the result of multiplying a row vector with a column vector. That is why the dot product of u^+ and v^+ is written $\left\langle u^+ \middle| v^+ \right\rangle$. The dot product of a vector with itself is just the squared magnitude of the vector; in symbols, $\left\langle x^+ \middle| x^+ \right\rangle = \left\| x^+ \right\rangle\right|^2$. In Dirac's notation, $\left\langle x^+ \middle| x^+ \right\rangle$ is a bracket (bra-ket) and hence $\left\langle x^+\right|$ is called a bra-vector and $\left|x^+\right\rangle$ a ket-vector.

Since the vector space is two-dimensional, $\left|x^+\right\rangle$ and $\left|x^-\right\rangle$ form a complete set of mutually orthogonal unit vectors, called an orthonormal basis. That is, every column vector can be expressed as a linear combination (superposition) of the vectors $\left|x^+\right\rangle$ and $\left|x^-\right\rangle$. Similarly, the bra- vectors $\langle x^+|$ and $\langle x^-|$ form an orthonormal basis of the dual space of row vectors. In this basis,

$$|x^+\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } |x^-\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and an arbitrary vector can be expressed as the superposition $|v\rangle = \cos\phi |x^+\rangle + \sin\phi |x^-\rangle$, which is a unit vector because $\cos^2\phi + \sin^2\phi = 1$. But what is the operator, denoted by P_v that projects other vectors onto $|v\rangle$? . Any projection operator must have the following properties:

$$\mathrm{P}_v|v\rangle = |v\rangle,\ \langle v|\mathrm{P}_v = \langle v|, \text{ and } \mathrm{P}_v|v^\perp\rangle = |0\rangle,\ \langle v^\perp|P_v = \langle 0|,$$

if $|v^\perp\rangle$ is orthogonal to $|v\rangle$. Also, It is easy to see that

$$P_v = |v\rangle\langle v|$$

has the required properties, because $|v\rangle\langle v|v\rangle = |v\rangle$, and so on. So, if $v = [\cos\phi \quad \sin\phi]$, then associated the matrix that projects any vector onto v is

$$P_v = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} [\cos\phi \quad \sin\phi] = \begin{bmatrix} \cos^2\phi & \cos\phi\sin\phi \\ \cos\phi\sin\phi & \sin^2\phi \end{bmatrix}.$$

In particular,

$$\mathrm{P}_{x^+} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ \mathrm{P}_{x^-} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \mathrm{P}_{y^+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},\ \mathrm{P}_{y^-} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

where we have used the fact that $\phi = \theta/2 = \pi/4$ and $\cos\pi/4 = \sin\pi/4 = 1/\sqrt{2}$.

All QM observables are definable in terms of the projection operators. To find the observable corresponding to 'spin' in the *x*-direction, we construct an operator that has the right mean value in every state. If spin 'up' is associated with the number +1 and spin 'down' associated with −1, then the expected value of the spin observable σ_x in state *v* is

$$(+1)P(x^+ \mid v) + (-1)P(x^- \mid v) = \langle v|\mathrm{P}_{x^+}|v\rangle - \langle v|\mathrm{P}_{x^-}|v\rangle = \langle v|\mathrm{P}_{x^+} - \mathrm{P}_{x^-}|v\rangle.$$

This proves that the operator that has the correct mean value is

$$\sigma_x = P_{x^+} - P_{x^-} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

This is one of the three Pauli spin matrices (modulo a constant factor). The second Pauli matrix is

$$\sigma_y = P_{y^+} - P_{y^-} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

It is now trivial to prove that anti-commutation of σ_x and σ_y;

$$\sigma_x \sigma_y = -\sigma_y \sigma_x .$$

This is property of spin observables that leads to the correct predictions in the three-particle experiment described in the next section.

The important lesson of this section is that anti-commutation relation follows necessarily from a simple QM model that introduces just one adjustable feature, the angle ϕ between vectors, which fits a huge variety of facts when we set $\phi = \theta/2$.

6. Hidden Variables Again: Another Failed Consilience

In the original EPR thought experiment (EPR 1935), two particles fly apart to opposite ends of the universe.

1. QM predicts the total momentum is zero, but says nothing about the separate momenta, except that they are equal and opposite.
2. After momentum of one particle is measured, the outcome of a momentum measurement on the other particle can be predicted with probability equal to unity.
3. The EPR criterion of reality: There is an element of reality corresponding to any physical quantity that can be predicted with probability equal to unity.
4. Therefore, there is an element of reality corresponding to the momentum of the second particle after the first measurement.
5. Locality: The element of reality belonging to the second particle is not created by the act of measurement on the first particle (because they are space-like separated).
6. Therefore the element of reality existed prior to the first measurement.
7. Therefore there exist elements of reality that are not represented in QM; that is, QM is incomplete.

This is a simplified version of the 1935 EPR argument. The shocking fact is that the argument is unsound, as was originally proved by Bell in 1964. This section presents a more perspicuous version of the proof (GHZ 1989; Mermin 1990). Whereas the next section explains how the QM formalism succeeds where hidden variables fail.

The argument takes the form of a reductio ad absurdum proof: It assumes that the hidden variable theory is true, then allows it to accommodate the first set of experimental facts, and then shows that it makes a false prediction. Therefore, the hidden variable theory is false.

Here is a dry run. Suppose that couples are interviewed in a psychology experiment. Each partner is taken from the waiting room to two sound proof interview rooms, and each is asked one of two questions according to separate coin toss. Nobody knows in advance which question will be asked, although everyone knows that it will be question *X* or question *Y*. Suppose that the experiment shows that when both partners are asked different questions, they answer either yes-yes or no-no. That leaves it open whether they would give the same answers if they were asked the same question. Why is that? It could be that the two questions are opposed, so that someone answering truthfully would answer yes to one if and only if they answer no to the other. That could explain the facts so far if the questions ask something that couples always disagree about. Under this scenario, the prediction is that the answers will always be yes-no or no-yes when the same question is asked both partners. On the other hand, another hypothesis is that questions are essentially the same, in which case the agreement in the answers already seen will imply that the answers are yes-yes or no-no when they are asked the same question, whether it is question *X* or question *Y*. There is no scenario in which we could get a perfect agreement when question *X* is asked, and perfect disagreement when question *Y* is asked.

But suppose that is what experimenters find. It is possible that the couples are playing a prank? How could this be possible if the couples don't know what question they are going to be asked? To implement such a conspiracy, they would have to leave the waiting room with some set of instructions. Let $X_1 = +1$ denote the instruction that partner 1 will answer yes to question *X*, and $X_1 = -1$ the instruction that partner 1 will answer no to question *X*. The variables Y_1, X_2, and Y_2 are defined in the same way. To account for the observed correlations when different questions are asked, we have to suppose that $X_1 = Y_2$, and $Y_1 = X_2$. Now add the experiment fact that they give the same answers to question *X*. This implies that $X_1 = X_2$. But it now follows deductively from these three equations that $Y_1 = Y_2$. This is clearly inconsistent with the fact that $Y_1 = -Y_2$. To explain such facts we seemed forced to postulate the existence of some kind of telepathic communication that would allow the couples to change their instruction after they know the question being asked. Or else they have foreknowledge of the outcomes of the coin tosses. Either way, the standard causal story stands refuted.

Here is the QM version of the same argument. Three electrons fly apart towards three widely separated Stern-Gerlach magnets, labeled 1, 2, and 3. Each is aligned in one of two directions (x or y) orthogonal to the path of the incoming electron. Each device contains two particle detectors, one placed in the 'up' path and one in the 'down' path exiting the magnet, such that if the electron is detected in the 'up' path, the light bulb attached to the device flashes red and if it is detected in the 'down' path, the same light bulb flashes green. Every particle is detected by one or other of the particle detectors. So, each bulb flashes either red or green. We may say that the outcome is 'spin-up' if the light is red and 'spin-down' if the light is green.

Consider the first set of experimental facts: When any 2 of the measurement devices are set to y and the third is set to x (that is, for settings y-y-x, y-x-y, or x-y-y) there is always an odd number of red lights flashing in every trial of the experiment—that is, either all 3 lights flash red or one light flashes red and the other two flash green.

A local hidden variable model accommodates these experimental facts in the following way: Suppose that each particle, after separation from the others, carries with it a set of properties that determine which light bulb will flash for every possible settings of the device it enters. Let's represent the property that particle 1 would cause the red bulb to flash if the device 1 were set to position x by $X_1 = +1$, and the property that the green bulb would flash were device 1 set to position x by $X_1 = -1$. According to the hidden variable story, the particle has the property $X_1 = +1$ or the property $X_1 = -1$, but not both.

Note that while the value of the variable *determines* the experimental outcome, the variable is not being used to represent the observable outcome. The existence of these variables is *postulated by the theory* to 'explain' the observed outcome. The theory assumes that the hidden variables have values even when they are not measured.

Similarly, let $Y_1 = +1$ and $Y_1 = -1$ represent the two properties that determine the outcome when the

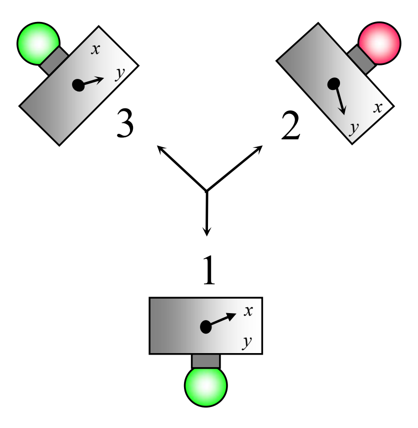

Figure 5: The GHZ version of Bell's experiment.

measuring device 1 is set at *y*. Then, a particle heading towards device 1 will have exactly one of 4 possible sets of properties, which Mermin (1990) refers to as "instruction sets": Either $\{X_1 = +1, Y_1 = +1\}$, $\{X_1 = +1, Y_1 = -1\}$, $\{X_1 = -1, Y_1 = +1\}$, or $\{X_1 = -1, Y_1 = -1\}$. Similar hidden variable states are assigned to particles 2 and 3. There are therefore 6 hidden variables that collectively play the role of a common cause, and these may adjusted in whatever way is needed to accommodate the facts.

Note that in a single run of the experiment, a measurement device cannot be oriented in two directions simultaneously, so we cannot determine all the hidden variable values by direct measurement. For instance, if we see the bulb flash red when device 1 is set to *y*, then we would only know that the instruction set was either $\{X_1 = +1, Y_1 = +1\}$ or $\{X_1 = -1, Y_1 = +1\}$. Some of the variables are always *hidden*. But their existence still has empirical consequences.

What needs to be accommodated is that fact that the outcome for the third particle must be *R* if and only if first two outcomes are either *R*-*R* or *G*-*G*. The constraints that are necessary and sufficient to accommodate all three regularities are:

$$X_1 Y_2 Y_3 = 1\,,\; Y_1 X_2 Y_3 = 1\,, \text{ and } Y_1 Y_2 X_3 = 1\,. \tag{1}$$

For example, if the setting is *y*-*x*-*y*, then the second equation tells us that if the outcome for particles 1 and 3 is *R*-*R* , then $Y_1 = 1 = X_2$, and therefore $Y_3 = 1$, and so the outcome for particle 3 will be *R*.

If the laws in (1) are assumed to apply in all experimental situations, then the following deductive consequence of the laws holds in all situations. Multiply the 3 equations in (1) together, and simplify using $Y_1^2 = Y_2^2 = Y_3^2 = 1$, to obtain

$$1 = (X_1 Y_2 Y_3)(Y_1 X_2 Y_3)(Y_1 Y_2 X_3) = X_1 X_2 X_3\,.$$

The equation, $X_1 X_2 X_3 = 1$, implies that there will also be an odd number of red flashes when all 3 devices are set to *x*-*x*-*x*. This prediction is dramatically different from the prediction made by QM, which predicts that there must be an *even* number of red flashes in this experiment! QM is right and the hidden variable prediction is wrong![6]

---

[6] As far as I'm aware, this exact experiment has not been performed. My confidence in making this claim is based on the proven consilience of QM (next section).

7. The Consilience of Quantum Mechanics

I will now *derive* the correct prediction from quantum mechanics. Instead of using *variables*, quantum mechanics assigns spin *observables* to each particle for every possible orientation of the Stern-Gerlach magnet. So, for example, the observable σ_x^1 replaces the hidden variable X_1, and the observable σ_y^3 replaces the hidden variable Y_3, and so on. The empirical probabilities are then determined in the way described in section 5.

There is another technical point that will prove useful here. Let us define an observable to be *dispersion-free* in state ψ if and only if the probability of all outcomes is zero except one, which has probability one. In classical physics, a variable is dispersion-free if and only if its variance is zero. Similarly, a QM observable $\hat{A}$ is dispersion-free in state ψ if and only if its variance is zero, where the variance is defined as the expected value of $\left(\hat{A}-a\right)^2$, where a is the mean value of the observable in state ψ.[7] It is relatively easy to prove that $\hat{A}$ is dispersion-free if and only if $\hat{A}\left|\psi\right\rangle = a\left|\psi\right\rangle$ for some real number a. In words, $\hat{A}$ is dispersion-free in state ψ if and only if ψ is an eigenstate of $\hat{A}$. The proof is found in Khinchin 1960, 54-55.

For example, if the system were in a +1 eigenstate of the observable σ_x^1, then we would predict with probability 1 that device 1 will flash red. If the system is in the −1 eigenstate of the same observable, then the green light will flash, and so on. If the system is not in an eigenstate of σ_x^1, then the probability that device 1 flashes red is still inferred from the mean value of σ_x^1.

There are 6 spin observables involved in our story: σ_x^1, σ_y^1, σ_x^2, σ_y^2, σ_x^3, σ_y^3. The new feature of this example (called the GHZ example) is that we can also construct new observables by considering products and sums of the 6 observables. The product of two observables is itself an observable if and only if the two observables commute.[8] It is therefore important to note that operators pertaining to different particles always commute. For example, σ_x^1 commutes with σ_y^2, and the product observable $\sigma_y^1\sigma_x^2$ commutes with σ_y^3, and so on. By using these facts alone, it follows that every product such as $\sigma_y^1\sigma_x^2\sigma_y^3$ is a QM observable. The product observables that correspond the three random variable products that appear in (1) are $(\sigma_x^1\sigma_y^2\sigma_y^3)$, $(\sigma_y^1\sigma_x^2\sigma_y^3)$, and

---

[7] In the general case in which observables involve complex numbers, only Hermitian operators count as observables, because only they have expected values that are real numbers in every state.

[8] This is because the product of two Hermitian operators is Hermitian if and only they commute.

($\sigma_y^1\sigma_y^2\sigma_x^3$). Each observable has well defined mean values and variances in every quantum state.

It is interesting to note that these observables are mutually incompatible in the sense that one and only one can be measured. For example, ($\sigma_x^1\sigma_y^2\sigma_y^3$) is measured if and only if the devices are in the *x-y-y* setting. Yet, a simple calculation shows that any two of these products commute because we end up applying the anti-commutation relation twice. Therefore, there exists a quantum state that is an eigenstate of all three observables simultaneously. If the system is in this state, then the outcome of each product observable can be predicted with certainty. But none of the individual observables can be predicted with certainty. In fact, individual spin outcomes have probability ½ in this state (see Appendix). So there are examples in QM in which correlations are predicted with certainty, but the correlata are completely random.

The quantum mechanical story begins with the assumption that all the electron triples are prepared in the same quantum state, $|\psi\rangle$. The fact that there is always an odd number of red flashes in the settings *y-y-x*, *y-x-y*, and *x-y-y* tells us that the product observables are dispersion-free in the state $|\psi\rangle$, which implies that $|\psi\rangle$ is an eigenstate of the three product observables. In other words, the first set of experimental facts is accommodated by the supposition that $|\psi\rangle$ is in a +1 eigenstate of all of the observables ($\sigma_x^1\sigma_y^2\sigma_y^3$), ($\sigma_y^1\sigma_x^2\sigma_y^3$), and ($\sigma_y^1\sigma_y^2\sigma_x^3$). Therefore (1) is replaced by the laws:

$$\sigma_x^1\sigma_y^2\sigma_y^3|\psi\rangle=|\psi\rangle,\ \sigma_y^1\sigma_x^2\sigma_y^3|\psi\rangle=|\psi\rangle,\ \text{and}\ \sigma_y^1\sigma_y^2\sigma_x^3|\psi\rangle=|\psi\rangle. \tag{2}$$

What predictions can be made from these quantum mechanical constraints? The fact that any spin operator times itself is equal to the identity operator (as can be verified directly by squaring the matrices derived in section 5) and the fact that operators pertaining to different particles commute, proves that

$$(\sigma_x^1\sigma_y^2\sigma_y^3)(\sigma_y^1\sigma_x^2\sigma_y^3)(\sigma_y^1\sigma_y^2\sigma_x^3) = (\sigma_x^1\sigma_y^2\sigma_y^3)(\sigma_y^3\sigma_x^2\sigma_y^1)(\sigma_y^1\sigma_y^2\sigma_x^3) = \sigma_x^1\sigma_y^2\sigma_x^2\sigma_y^2\sigma_x^3.$$

Furthermore, the anti-commutation property of the spin operators proves that

$$\sigma_x^1\sigma_y^2\sigma_x^2\sigma_y^2\sigma_x^3 = \sigma_x^1(-\sigma_x^2\sigma_y^2)\sigma_y^2\sigma_x^3 = -\sigma_x^1\sigma_x^2(\sigma_y^2\sigma_y^2)\sigma_x^3 = -\sigma_x^1\sigma_x^2\sigma_x^3,$$

where the minus sign arises from the anti-commutation of $\sigma_x{}^2$ and $\sigma_y{}^2$. Recall that the anti-commutation law is required in order to accommodate the facts about sequential measurements (section 5).

Finally, from (2), it is also true that

$$(\sigma_x{}^1\sigma_y{}^2\sigma_y{}^3)(\sigma_y{}^1\sigma_x{}^2\sigma_y{}^3)(\sigma_y{}^1\sigma_y{}^2\sigma_x{}^3)\left|\psi\right\rangle = \left|\psi\right\rangle, \tag{3}$$

and therefore,

$$\sigma_x{}^1\sigma_x{}^2\sigma_x{}^3\left|\psi\right\rangle = -\left|\psi\right\rangle. \tag{4}$$

This proves what I promised: The *only* way for the quantum model to accommodate the first set of experimental facts is to assume that (2) is true, which implies (4), which then implies that there will always be an *even* number of red flashes when the magnets are set to *x*-*x*-*x*. This is the prediction that the hidden variable theory got wrong.

It is interesting to compare the hidden variable prediction with the QM prediction in this example. The hidden variable deduction proceeds in a completely analogous way up to the step where the anti-commutation relation is used. That step is replaced by the commutation relation $Y_2X_2 = X_2Y_2$, which then leads to a plus sign instead of the minus sign. All variables commute, while operators may or may not. In this respect, QM is more versatile. But don't mistake flexibility for weakness; for the flexibility allows us to embed all the spin ½ states in a single 2-dimensional vector space, and the tightness of the colligation leads to precise predictions about a variety of phenomena.

It is equally interesting to note the *similarities*. The predictive power of both theories relies on the idea that the variables or observables have values even when they are not measured. In the hidden variable prediction, this assumption comes into play when we assume that $Y_1^2 = Y_2^2 = Y_3^2 = 1$. For this to make sense, we assume that each variable has a value, and then argue that the square is 1 no matter which value it has. In the QM derivation, we assume that $\sigma_y^1\sigma_y^1 = \sigma_y^2\sigma_y^2 = \sigma_y^3\sigma_y^3 = I$, where *I* is the unit operator. Every state is an eigenstate of the unit operator, so this observable has the value 1 in every state. We are not assuming that σ_y^1, σ_y^2, and σ_y^3 have values in the sense of having dispersion-free probability distributions. But we are assuming that σ_y^1, σ_y^2, and σ_y^3 apply to situations in which they are not measured.

In fact, spin observables are extraordinarily projectible. For the state $|\psi\rangle$ is *uniquely* determined by the equations (2) *alone* (modulo a complex phase factor, which makes no difference to any prediction). This unique state $|\psi\rangle$ then determines the full probability distributions for *any* measurement settings whatsoever, including cases in which the three magnets are aligned in directions different from x and y. This is worth proving in detail because it illustrates the predictive power of the quantum mechanical colligation (see the Appendix). In particular, it is surprising to see that the unified QM model *requires* that each individual event is random (has probability ½). QM isn't forced to explain why two people meet at the market by explaining why each will be there at that time. In comparison, the local hidden variable theory has very little predictive power; and to insult to injury, the one prediction that it does make turns out to be false!

It is not my purpose to disparage hidden variable theories in general. After all, the unknown microstate of a thermodynamic system plays the role of a hidden variable in statistical mechanics. Rather, the point is that the hidden variable model does not compete with the unity of the QM model. The existence of these variables does not lead to very much predictive power, even though many philosophers consider them to be highly *explanatory*. And at the same time, QM is considered to be mysterious and un-explanatory despite the fact that it is predictively very powerful. So much the worse for explanation. Unification is a far better criterion of success.

After working through an example like this, after seeing how tightly quantum mechanics ties the phenomena together, wouldn't it be miraculous if this huge body of data were to fit the predictions of quantum mechanics without there being some way of explaining this fact in terms of a reality behind the observed phenomena? Indeed, the argument for a realist view of the QM properties of spin looks similar to the argument for the existence of Newtonian mass. In both cases, there is an empirical overdetermination of the values and properties of the postulated entities. There are many voices, and the theory predicts that they will sing in harmony. Hidden variable theory doesn't hear the same voices, and what it does hear is not very harmonious.

Yet the EPR argument is still puzzling, for it seems to provide a conclusive reason for believing in hidden variables. Where does the argument go wrong? In my view, the mistake occurs around steps 3 and 4:

3. There is an element of reality corresponding to any physical quantity that is predicted with certainty.

4. Therefore, there is an element of reality corresponding to the momentum of the second particle after the first measurement.

It will take several paragraphs to explain why 4 does not follow from 3.

Prior to the first measurement, we can predict the correlation with certainty, so there is an element of reality associated with that quantity. This is some kind of relational property X that does not supervene on its relata (Teller 1986). That is to say, the pair of particles has a total momentum value p that is *not* the sum of the momenta of the two particles. The X-element inferred from the EPR criterion of reality is ontologically very strange. It seems to defy the logic of the word 'total'. Yet it is all that is required by the conservation of momentum applied to the situation in which one particle explodes into two particles. The same point applies to the conservation of spin.

EPR actually use a subtly different criterion of reality that does imply that the separate momenta exist *prior to the first measurement*. For them it is sufficient that a quantity can be predicted with certainty in the counterfactual sense that it could be predicted with certainty if the first measurement had already been made. It is this stronger criterion of reality that leads to the existence of hidden variables, which then leads to a contradiction a la Bell, which then leads us to reject the stronger criterion of reality. The conclusion of the previous paragraph is that the weaker and more acceptable criterion of reality implies that the total momentum exists *without implying the separate momenta exist*. It follows that the whole is ontologically more the sum of its parts.

This is the bitter pill to swallow, but once swallowed, it is not logically necessary to insist that the momentum of the second particle exists *until after it is actually measured*. Why is this? After the first measurement we can predict an *outcome* of a momentum measurement on the second particle with certainty, so the weak criterion of reality does force us to introduce a new element of reality corresponding to this quantity. Call it element S. But it does not force us to conclude that this element is *located at the second particle*. It is consistent with the EPR criterion of reality to suppose that element S is located at the first particle, where the measurement took place. In other words, the X-element is the total momentum p, and the S-element is the momentum of the first particle p_1. This implies that the outcome of the second measurement is $p - p_1$ without having to assume the existence of p_2. Why posit three elements when two will do? This leads to what I call the *local interpretation* of wave function collapse.

Why are holistic properties so hard to swallow? Aren't they commonplace? After, a causal relation between two events doesn't seem to supervene on its relata. There are no intrinsically non-relational properties of a cause and its effect that determine the causal *relation*. Consider a simple example. Suppose that a light is on if and only if the switch is up. The causal relation does not determine whether the switch is up or down, or whether the light is on or off; nor does the state of switch and the light determine that the relation is causal.

When the switch is flipped up, then we can predict with certainty that the light will come on. Therefore, there is an element of reality corresponding to this quantity. But the new element of reality need be nothing more than the state of the switch. For when this is combined with the causal relation, we have everything that is required to ontologically determine that the light will come on. Nothing more need be said about the mechanistic details. Even when we do analyze the causal connection in terms of some chain of causes, the causal *relation* between consecutive events in the chain is still non-supervening.[9] I'm not endorsing this view of cause. I'm just pointing out that it is an example of a holistic relation that is already familiar to metaphysicians.

My recommendation is that we interpret QM as introducing something ontologically similar, at least with respect to their failure to supervene of properties of the parts. Once holistic relations are allowed into our ontology, there is no need to postulate additional elements of reality that serve no other purpose than to violate the principle of locality and challenge Einstein's theory of relativity. And good reasons for not doing so.

8. Why Causal Explanations Fail

There is a wrong way and a right way of thinking about Reichenbach's cubical world. The wrong way is to view it in terms of Reichenbach's principle of common cause, which may be made more precise in the following way. For any two physical quantities X and Y that exhibit a statistical correlation, such that the correlation does not arise from X directly causing Y or Y directly causing X. Then, the only alternative appears to be that the correlation is explained by a common cause variable Z, such that Z causes X and Z causes Y.[10] This very powerful idea is severely challenged by QM examples.

---

[9] All this is familiar to us from Hume's analysis of causation, except that his modus tollens is our modus ponens.

[10] There are a number of other scenarios that need to be ruled out as well. Arntzenius (1993) provides a concise list of these. Also see Sober (1984), Cartwright (1989), Eells (1991), Hausman (1998), and

A slight modification of the GHZ example brings the challenge into focus. Suppose that one measurement device is placed in Alice's laboratory, while the other two are placed in Bob's laboratory. The magnets are all aligned in the *x*-direction, so there is always an even number of red flashes. Bob hooks up a master light bulb that flashes red if and only one of his devices flashes red and the other flashes green. This means that Alice's light flashes red if and only if Bob's master light flashes red. There is a perfect positive correlation between what Bob sees and what Alice sees. Moreover, if we imagine that Alice and Bob measurements are performed simultaneously (in their frame of reference), then the theory of relativity appears to rule out the existence of any direct causal interaction between the events. The only causal option left is ruled out by the Bell argument (section 6).

The solution argued for in section 7 was the view that are holistic properties of entangled quantum systems that cannot be reduced to the property of the parts. But maybe this view can be recast in more familiar causal terms by supposing that there is some direct causal connection between measurement events? That must depend, in part, on how far one is willing to stretch the everyday notion of cause. My reason against this move is that such a generalized notion of cause would be so dissimilar to the everyday notion that it would (and does) produce more smoke than fire. For, there is the idea that controlling a cause is a way of controlling its effects. This is, if one wiggles the cause, then the effect will wiggle. This does not apply in the quantum case because there is no way of manipulating the events that events that are correlated. The only thing that is subject to control are the measurement settings. Alice's outcome has the same probability no matter what measurement settings Bob chooses (see Appendix). Bob cannot exploit the phenomenon to send superluminal signals to Alice, or vice versa. So, if it is causal, then it is unlike other causal connections.

The fact that each individual measurement is uncontrollably random also proves that there is no reason to believe that the probabilities are invariant in the same way that the jackpot probabilities were invariant. For that required that we compare two situations in which the probability distributions of the exogenous variable (the cause variable) were different. The each spin outcome is uncontrollably random (see Appendix) means that there is no way of testing the invariance claim.

Woodward (2003) for extensive discussions of causal inference and explanation, and related philosophical issues.

There is an even deeper reason why causal models fail to apply. They represent physical quantities in terms of variables, whereas QM models do not. The use of operators in place of variables should be taken seriously in any interpretation of QM because the analysis of the GHZ example shows that it is the properties of spin operators that it is responsible for the miraculous consilience of QM.

It's worth mentioning here that we lose nothing by taking QM seriously. There is no argument that there is no such thing as causation if the world is quantum mechanical. Bell argument applies only to composite systems in entangled states. Other QM systems produce correlations that can be explained in causal terms. QM enjoys the best of both worlds.[11]

The right way of thinking about Reichenbach's cubical world is in terms of the consilience of inductions. There is no categorical imperative that we must represent the external world in terms of variables. If we go by the consilience of inductions, then local hidden variable theories do not succeed and non-local hidden variable theories fail to compete with QM. The miraculous consilience of QM not only explains the growing consensus in favor of QM, but it also explains why cubical world inference is so hard to characterize in general terms. For the inference that is ultimately convincing depends on the invention of new theoretical concepts or new mathematical formalisms, which are not easy to anticipate in advance. Such is the case in QM.

The good news is that there do appear to be epistemological principles that apply equally well to the old and the new physics. As in science itself, successful predictions are impressive. In 1858, Whewell's concluded that the consilience of inductions is the best indicator of good science. The theories have changed, but the standards are the same.

Appendix

The purpose of this appendix is to prove that the quantum mechanical accommodation of the first three experimental facts in the GHZ experiment leads to far stronger predictions than those provided by the hidden variable theory.

Because there are 8 possible outcomes, the state vector $|\psi\rangle$ is a vector in an 8-dimensional Hilbert space. This space is constructed out of the three 2-dimensional vector spaces used to represent the states of electrons separately. Let us choose the basis vectors for the first Hilbert

---

[11] On the other hand, I don't want to claim that there are no problems about how to interpret QM. There is the infamous measurement problem, which I have not address explicitly.

space to be $\left|x^{+}\right\rangle_{1}$ and $\left|x^{-}\right\rangle_{1}$, etc., where the subscripts keep track of the vector space in question. It is now possible to prove that the eight vector products, $\left|x^{\pm}\right\rangle_{1}\left|x^{\pm}\right\rangle_{2}\left|x^{\pm}\right\rangle_{3}$, form a natural basis for the 8-dimensional vector space. It is convenient to denote these eight basis vectors by $\left|\pm\pm\pm\right\rangle$.

From section 5, $\sigma_x\left|x^{+}\right\rangle=\left|x^{+}\right\rangle$ and $\sigma_x\left|x^{-}\right\rangle=-\left|x^{+}\right\rangle$. In addition, σ_y has the property

$$\sigma_y\left|x^{+}\right\rangle=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}=\begin{bmatrix}0\\ 1\end{bmatrix}=\left|x^{-}\right\rangle.$$

Similarly, $\sigma_y\left|x^{-}\right\rangle=\left|x^{+}\right\rangle$. In the context of the 3-electron system, a spin operator like $\sigma_y{}^3$ operates on the third part of the vector product, so for example,

$$\sigma_y{}^3\left|+++\right\rangle=\left|++-\right\rangle.$$

We now have everything in place that we need to apply the equations in (2) to an *arbitrary* state vector

$$\left|\psi\right\rangle=\sum_{j,k,l}c_{jkl}\left|jkl\right\rangle,$$

where the indices *j*, *k*, and *l*, range over the values + and –. Calculating the constraints on the coefficients c_{jkl} implied by (2) is tedious, but it involves nothing more than algebra. Since, the equation $\sigma_x^1\sigma_x^2\sigma_x^3\left|\psi\right\rangle=-\left|\psi\right\rangle$ follows from (2), we get $jkl\,c_{jkl}=c_{jkl}$. In all the cases in which the product *jkl* is –; that is, when there is an odd number of +'s, we have $c_{jkl}=0$. This reduces $\left|\psi\right\rangle$ to a superposition of four basis vectors:

$$\left|\psi\right\rangle=c_{++-}\left|++-\right\rangle+c_{+-+}\left|+-+\right\rangle+c_{-++}\left|-++\right\rangle+c_{---}\left|---\right\rangle.$$

The result is easy to remember because there is an even number of +'s in each term. Now apply the requirement that $\sigma_x^1\sigma_y^2\sigma_y^3\left|\psi\right\rangle=-\left|\psi\right\rangle$. This implies $jc_{jkl}=c_{jk'l'}$, where $k'=-$ if and only if $k=+$, which yields eight equations, four of which are distinct: $c_{+++}=c_{+--}$, $c_{++-}=c_{+-+}$, $c_{-++}=-c_{---}$, $c_{-+-}=-c_{--+}$. Two of these are trivially satisfied because the coefficients are 0. The remaining two are $c_{++-}=c_{+-+}$, $c_{-++}=-c_{---}$. Therefore,

$$\left|\psi\right\rangle=a\left|++-\right\rangle+a\left|+-+\right\rangle+b\left|-++\right\rangle-b\left|---\right\rangle.$$

Now apply $\sigma_y^1\sigma_x^2\sigma_y^3|\psi\rangle = -|\psi\rangle$, which implies that $a = b$. Finally, the requirement that the vector has unit length shows that, modulo an arbitrary phase factor $e^{i\theta}$,

$$|\psi\rangle = \tfrac{1}{\sqrt{4}}|+\,+\,-\rangle + \tfrac{1}{\sqrt{4}}|+\,-\,+\rangle + \tfrac{1}{\sqrt{4}}|-\,+\,+\rangle - \tfrac{1}{\sqrt{4}}|-\,-\,-\rangle\,.$$

Notice that we have not used the third equation in (3), which shows that any of three of the equations is sufficient to determine a unique state vector.

Recall from section 5 that

$$P_v = \begin{bmatrix} \cos^2\phi & \cos\phi\sin\phi \\ \cos\phi\sin\phi & \sin^2\phi \end{bmatrix},$$

so that

$$\langle + | P_v | + \rangle = \langle - | P_v | - \rangle = \cos^2\phi \text{ and } \langle + | P_v | - \rangle = \langle - | P_v | + \rangle = \sin^2\phi\,.$$

Therefore,

$$\langle \psi | P_v^1 | \psi \rangle = \tfrac{1}{4}(\cos^2\phi + \sin^2\phi) + \tfrac{1}{4}(\cos^2\phi + \sin^2\phi) = \tfrac{1}{2}\,.$$

In words, the probability of ‘up’ for any electron in any measurement direction orthogonal to z is equal to ½.

References

Arntzenius, Frank (1993): “The Common Cause Principle.” *PSA 1992 Volume* **2***:* 227-237. East Lansing, Michigan: Philosophy of Science Association.

Arntzenius, Frank (1997): “Transition Chances and Causation.” *Pacific Philosophical Quarterly* **78**: 149-168.

Bell, John S. (1964). “On the Einstein-Podolsky-Rosen Paradox”, *Physics* 1: 195-200.

Bell, John S. (1971), “Introduction to the Hidden Variable Question”, in B. d'Espagnat (ed.) *Foundations of Quantum Mechanics*, N. Y.: Academic Press.

Butts, Robert E. (ed.) (1989). *William Whewell: Theory of Scientific Method.* Hackett Publishing Company, Indianapolis/Cambridge.

Cartwright, Nancy (1989): *Nature’s Capacities and their Measurement.* Oxford: Oxford

University Press.

Eells, Ellery (1991): *Probabilistic Causality.* Cambridge: Cambridge University Press.

Einstein, Albert, B. Podolsky & N. Rosen. (1935). "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?". *Physical Review* **47:** 777-780.

Forster, Malcolm R. (1984): *Probabilistic Causality and the Foundations of Modern Science.* Ph.D. Thesis, University of Western Ontario.

Goodman, Nelson (1965): *Fact, Fiction and Forecast,* Second Edition. Harvard University Press, Cambridge, Mass.

Greenberger, Daniel M., Michael A. Horne and Anton Zeilinger (1989): "Going Beyond Bell's Theorem" in M. Kafatos (ed.), *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, Kluwer Academic Publishers, pp. 69-72.

Harper, William L. (2002), "Howard Stein on Isaac Newton: Beyond Hypotheses." In David B. Malament (ed.) *Reading Natural Philosophy*: *Essays in the History and Philosophy of Science and Mathematics*. Chicago and La Salle, Illinois: Open Court. 71-112.

Hausman, Daniel M. (1998): *Causal Asymmetries*. Cambridge University Press.

Hung, Edwin (1997): *The Nature of Science: Problems and Perspectives*, Wadsworth Publishing Co.

Khinchin, A. I. (1960): *Mathematical Foundations of Quantum Statistics*, Dover Publications, Inc., Mineola, New York.

Mermin, David N. (1990) "Quantum Mysteries Revisited", *American Journal of Physics*, August 1990, pp.731-4.

Myrvold, Wayne and William L. Harper (2002), "Model Selection, Simplicity, and Scientific Inference", *Philosophy of Science* **69**: S135-S149.

Reichenbach, Hans (1938): *Experience and Prediction*. Chicago: University of Chicago Press.

Reichenbach, Hans (1956): *The Direction of Time*. Berkeley: University of California Press.

Sober, Elliott (1984): *The Nature of Selection: Evolutionary Theory in Philosophical Focus.* MIT Press, Cambridge, Mass.

Sober, Elliott (1994): "Temporally Oriented Laws," in Sober (1994) *From A Biological Point of View - Essays in evolutionary philosophy*, Cambridge University Press, pp. 233 - 251.

Teller, Paul (1986). "Relational Holism and Quantum Mechanics." *British Journal for the Philosophy of Science* **37**: 71-81.

van Fraassen, Bas (1980), *The Scientific Image*, Oxford: Oxford University Press.

van Fraassen, B.C. (1983), "The Charybdis of Realism: Epistemological Foundations of Bell's Inequality Argument", *Synthese*.

Whewell, William (1858): *Novum Organon Renovatum*, Part II of the 3rd the third edition of *The Philosophy of the Inductive Sciences*, London, Cass, 1967.

Woodward, James. (2003): *Making Things Happen*: *A Theory of Causal Explanation* . Oxford and New York: Oxford University Press.