

Misinterpreting Reality: Confusing Mathematics for Physics

by Robert H. McEachern

Abstract: In one of the great scientific tragedies of the past century, “Modern Physics” was developed long before the development of Information Theory. Consequently, the early search for the “meaning” of the equations of mathematical physics, especially quantum physics, was based on several critical assumptions about the nature of information, which have subsequently been shown to be untrue. Unfortunately, none of the physicists, at the time, (and apparently even today) recognized these assumptions, as such. Hence, the misunderstandings engendered by those assumptions have become unquestioned dogma. Equations contain very little information. This fact is what makes it possible to symbolically represent them, in a computer memory, by a very small number of bits. As a direct result of this fact, we can conclude, contrary to the fervent belief of most physicists, that equations cannot describe anything other than the most trivial physical phenomenon; those nearly devoid of all information. For complex phenomenon, it is the vast information content of the initial conditions (like the content of an observer's memory) rather than the tiny information content of equations, that really matters. Indeed, observations become “quantized” if and only if the information content of the observations is small. It is not small physical size, but small information content, that is the cause of “quantization.” More importantly, since the information content of the “meaning” of the equations is usually much greater than the information content of the equations themselves, the “meaning” cannot possibly be contained within the equations; the “meaning” has simply been made-up and slapped-on. The controversies and paradoxes surrounding non-locality, superposition, entanglement, and the uncertainty principle are all examples of this problem; the equations accurately describe the observations, it is only the slapped-on “meaning”, that causes of all the difficulties in understanding.

The Nature of the Problem - the Problem of Describing Nature

Ask a physicist why a car stops at a red traffic-light. Ask for an explanation of the physical cause, not the “agent” or driver as cause, but the underlying physics. Was it a transfer of energy, or momentum, or electric charge? Why does the light have to be red? If you press this question, you will receive no answer. The “cause” lies in the information content of the “initial conditions”, specifically, the information content of the driver's memory. It does not lie within the equations of physics. The cause is a transfer of information, not a transfer of mass, charge, energy, momentum, angular momentum or any other physical entity. It is a transfer of a symbolic entity. This symbolic entity was, of course, symbolized by some physical entity (red light), but the meaning of that symbol cannot be deduced from any attributes or properties of that physical entity, including the equations of physics used to describe its behavior.

To put it another way, physicists seek to predict how physical substances behave. But they could never have predicted that cars would stop at red traffic lights, anymore than they could predict that they would drive on one side of the road in England and on the other in the United States. Such physical behaviors do not violate the laws of physics. But they remain unpredictable, because merely knowing the mathematical equations used to describe the physics is insufficient. One also needs to know auxiliary information, such as the initial conditions. And therein lies the problem. Whenever the information content of the initial conditions dwarfs the information

content of the equations or “laws”, the laws are usually rendered superfluous; they have no predictive power whatsoever, other than predicting that they cannot be violated. When this dwarfing occurs in physics, it results in a new set of “higher laws”, namely, chemistry. When it happens in chemistry, it becomes biology, and finally, consciousness.

Consider another example of mathematical equations being used to describe observations; the now familiar JPEG image compression technique, used to reduce the memory storage requirements of a photographic image. The original photograph is symbolically represented by two parts (1) the compressed JPEG image (the initial conditions) and (2) the JPEG algorithm (the equations). Both parts are required in order to reconstruct the original image. No amount of study of the JPEG algorithm alone will ever enable one to predict or understand an image. The equations of physics are *exactly* like the JPEG equations in this regard. They cannot tell you anything *interesting* about the observations (the image) unless the image is of a scene nearly devoid of all information, like a photo of a blank wall. In a very real sense, Newton's law of gravity is perfectly analogous to the JPEG image compression algorithm. It is a “lossy” description of the original data, no more no less; the reconstructed, predicted “image” is slightly different from the original. In contrast, Einstein's theory of gravity appears to be a “lossless” compression algorithm.

All Descriptions of Reality have a Life of their Own, Independent of the Reality being Described; this includes Mathematical Descriptions

Physicists fail to distinguish between the properties of the "reality" they are attempting to describe, and the properties of their mathematical "descriptions of that reality". But they are two very different things. For example, all such descriptions, written in the English language, contain only 26 letters, a-z. Should one assume from this property of the description, that this must correspond to some fundamental property of the entities being described? Perhaps the universe is actually constructed from just 26 fundamental “letters”, in a manner similar to that in which genetic material is constructed from the 4 “letters” of the DNA code. This assumption may seem absurd, but it is no more or less absurd than the one most physicists have made. Properties like the Uncertainty Principle, Superposition and Entanglement, are all properties of the mathematical language being used to describe the world - Fourier Analysis. Consequently, it is inevitable that these properties will appear in all such descriptions of reality, regardless of whether or not they are properties of the entities being described, just as it is inevitable that the 26 letters appear in all such descriptions written in English. To subsequently *assume* that such properties must be actual properties of the entities being described, is just that, an assumption, and not a very good one. Since few physicists are familiar with the assumptions underlying Fourier Analysis, which in turn underlies quantum theory, we shall discuss these in some detail below. But first we shall briefly consider two other manifestations of the problem; Bohm's Interpretation of Quantum Mechanics and Bell's famous Inequality Theorem.

Would you like some components with that observation?

Over half a century ago, David Bohm developed a non-probabilistic interpretation of non-relativistic quantum theory. He thereby demonstrated that the very same equation being cited in probabilistic interpretations of quantum theory also had a non-probabilistic interpretation. The fact that a single equation could have such radically different interpretations should have sent up red-flags throughout the physics world. But it did not.

The problem with Bell's Theorem and the EPR Paradox is more insidious. These famous exemplars of quantum “non-locality” have been adequately described elsewhere. Here we only seek to illuminate an underlying bad assumption at the heart of the issue; namely that the number of components that exist in the mathematical description of an observation *must necessarily* equal the number of components that actually exist in the entity being observed. Our story begins with the concept of Angular Momentum, and specifically, with the components of Angular Momentum. Long before the development of quantum theory, the concept of Angular Momentum had been developed. It was described mathematically as a vector, with three components, one corresponding to each of the three spatial dimensions. Later, when “spin” was discovered, it was *assumed* to be analogous to a quantized version of Angular Momentum. Thus it was assumed to be describable via multiple components. It was soon discovered that spin did not behave like an ordinary three component vector. Instead, it was described as a multi-component “spinor”. But the key point is that no one ever asked themselves how an observation of a “single bit of information” would appear or behave. Single bits do not have multiple components; nonetheless their physical description must.

In order to make this clear, consider two idealized, microscopic entities to be observed; one is a tiny two-sided “coin”, the other a tiny six-sided “cube”. Imagine receiving one of these entities as a “message”, from which information must be recovered. That information is encoded such that a single bit of information will be recovered, in the state either “up” or “down”, from each of the observed pairs of “sides”. Since the coin possesses only one pair of sides, the maximum possible number of recoverable bits of information is equal to one. Since the cube possesses three components, three pairs of sides, the maximum number of recoverable bits is equal to three. However, any mathematical description of the “state” of the entity to be observed must take into account that the state depends on the aspect angle from which the entity is to be observed. Since this aspect angle exists in a three dimensional space, the mathematical description must contain three components, even though the coin, *by definition*, contains only one. The analogy to an observation of the spin of an electron should be apparent. It should also be apparent that the number of components in the mathematical description of any future observation of a state, is not simply related to the number of components actually existing as attributes of the entity. Rather, they are also related to things like the aspect angles from which the entity can be observed.

When one attempts to observe the spin of an electron, only a single bit of information has ever been recovered from that “message”. The question of interest is “Why?” Is it because observations of the electron behave similar to the observations of the coin, or is it because it behaves similar to a cube, in which the observation of the first component disturbs any subsequent observation of the other two components? Does it even matter? The answer to the latter question is that it does not matter at all, to either the predictions of quantum theory, or to

interpretations of quantum theory, such as the famous Copenhagen Interpretation. But it does matter, it is critically *necessary*, to Bell's purported theorem. Bell's proof of his theorem simply *assumes*, with no evidence to support the assumption, other than the existence of multiple components in the description, that multiple components must exist as actual attributes of an electron's spin. Bell's theorem cannot be proven without making this assumption, since the theorem derives a relationship between the statistics of these supposed components. Without Bell's theorem, there is no significant relationship, and consequently no issue regarding “non locality”. To the best of my knowledge, no other part of quantum theory rests upon this dubious assumption; it is unique to Bell's Theorem. And it is very probably wrong; an electron's spin is almost certainly analogous to a coin, not a cube (Ockham's razor: the simplest explanation is probably correct). Two entangled electrons are analogous to two entangled (anti-parallel) coins. The message they encode cannot be decoded (they remain in an unknown state) until the observer specifies the aspect angle that is to be used for the observation. But once that is specified, the state of both coins can be determined. There is no “spooky action at a distance”, there is only the exploitation of a priori information, that has not been properly accounted for, because of a bad assumption. Attempting to observe other “components” merely produces an observation of the same component from a different aspect angle. This may be true regardless of whether or not the first observation alters the orientation of the coin, and thereby disturbs observations of the other “components”.

To be uncertain, or not to be uncertain, that is the question

We now return to the assumptions of Fourier Analysis, and thus the hidden and long-neglected assumptions of quantum theory. Fourier Analysis was developed, two hundred years ago, for the specific purpose of injecting superposition into partial differential equations in order to aid in solving them. The principle of superposition enables one to construct all the solutions to a given problem, by combining a number of particular solutions. Fourier's idea was to find sufficiently many particular functions, that satisfy the problem, such that a linear combination of them will satisfy an initial condition, that may be any arbitrary function. Fourier discovered a method for representing this arbitrary function as a combination of simple trigonometric functions (sines and cosines).

In general, a sinusoidal function of time, $x(t)$, has three parameters, Amplitude (A), frequency (f) and phase (φ), such that $x(t)=A\cdot\sin(2\cdot\pi\cdot ft+\varphi)$. Now here's the problem; suppose you had a set of observations, $x(t)$, that appeared to be sinusoidal, and you wished to “fit” a sinusoidal function to those observations. That is, you wish to determine the values for the three parameters, A , f and φ , that would result in the best match between the equation given above and the observed function $x(t)$. The problem is that f and φ appear as arguments of the sine function. Hence, solving for them is a difficult, non-linear problem. However a trigonometric identity provides a partial solution to this difficulty:

$$A\cdot\sin(2\cdot\pi\cdot ft+\varphi) = A\cdot\cos(\varphi)\cdot\sin(2\cdot\pi\cdot ft) + A\cdot\sin(\varphi)\cdot\cos(2\cdot\pi\cdot ft)$$

Now, the right-hand-side only has one parameter, f , as an argument of a time-varying sinusoid, while the coefficients $A\cdot\cos(\varphi)$ and $A\cdot\sin(\varphi)$ represent two new constants, similar to the original constant A . Fourier's discovery was a method for solving for these two constants while *avoiding*

any attempt to solve for the unknown frequency, f . Instead, he exploited superposition to represent $x(t)$, not as a single sinusoid with its indeterminate frequency, but as a sum of many particular sinusoids, each with a pre-determined frequency, and two determinable coefficients, $A \cdot \cos(\phi)$ and $A \cdot \sin(\phi)$. Jumping forward two hundred years, and with the aid of a few mathematical tricks, these two sets of coefficients have become known as the real and imaginary parts of a Fourier Transform. The whole point of all these machinations was to *avoid any attempt to solve for the non-linear parameter, the unknown frequency, f* . This brings us to the Uncertainty Principle.

Suppose the observed function, $x(t)$, that one is attempting to represent via a Fourier Transform (transforming the set of observed values into Fourier's peculiar combination of “complex” sinusoids) was not an infinitely long function. Suppose instead that it has a duration equal to Δt . Fourier's peculiar Transform demands that in order to represent such a function, the set of “pre-determined frequencies”, noted above, must span a frequency bandwidth, Δf , equal to the bandwidth of the observed function. Furthermore, it can be shown that $\Delta t \cdot \Delta f \geq \text{constant}$, where the specific value of the constant depends on exactly how one defines the bandwidth and the duration. This is the uncertainty principle. It has *nothing* to do with physics per se, it is purely a result of Fourier's peculiar mathematical transformation, for representing arbitrary functions. It got imported into quantum mechanics, when Fourier's transforms were used to inject superposition into the wave equation. Although Heisenberg's Uncertainty Principle is usually stated in terms of distance and momentum ($\Delta x = h/\Delta p$), it is easy to see that it is really just Fourier's Uncertainty Principle, with a few changes of variables. Thus, for a photon traveling at the speed of light, c , and wavelength, λ , we have:

$$c \cdot \Delta t = \Delta x = h/\Delta p = h(\Delta \lambda/h) = \Delta \lambda = c/\Delta f \quad \text{Hence, } \Delta t \cdot \Delta f = 1$$

It should be clear from the above, that the terms in the uncertainty relationship do not represent anything like an uncertainty, or accuracy. They represent “widths of distributions”, like the duration of the observations and their bandwidth. This is fairly well-known to those familiar with Fourier Transforms. But what is not so well-known is that the entire method was based on never, ever even trying to measure one of the two “uncertainties”, under any circumstance. As described above, the whole point of introducing the trigonometric identity noted above, was to *avoid any attempt to solve for the non-linear parameter, the unknown frequency, f* . Merely attempting to measure this, violates the assumption that lead to the uncertainty principle in the first place. Consequently, the uncertainty principle has nothing whatever to say about what accuracy may be achieved when one does actually try. Instead, another principle, from Information Theory, known as Shannon's Capacity Theorem, sets the ultimate limit that can be achieved. Shannon's Capacity Theorem is similar to the uncertainty principle, but adds one more term, dependent on “signal-to-noise ratio”, S/N , to the product:

$$\Delta t \cdot \Delta f \log_2(1+S/N) = \text{constant}$$

= maximum number of bits of information that can be extracted from the observations.

Shannon's Capacity Theorem is usually thought of as pertaining to the information content of “waves” rather than “particles”. But water waves, sound waves, electromagnetic waves and the

like are all merely coordinated movements of distributions of particles. If one thinks of the “noise” in a process that counts particles, as the smallest possible non-zero error in the particle count, then $N = 1$. Similarly, the smallest possible signal is just one particle, $S=1$. Thus, when observing a single particle, the Shannon Capacity reduces to the Uncertainty Principle, since $\log_2(1+S/N) = 1$. But something else may happen when a beam of multiple particles is observed, by an observer that knows enough not to bother with inappropriate Fourier Analysis, since that assumes that you would never even try.

This may sound mysterious, but it is not. Your eyes and ears exploit this “loophole” to produce the perceptions of color and pitch, that easily exceed the limitations of the uncertainty principle, by several orders of magnitude. They accomplish this feat by cleverly avoiding Fourier Analysis, and by exploiting the fact that there is more than one particle in the signals, so $S \gg 1$. Failing to appreciate this loophole is the source of many of the so-called paradoxes of quantum mechanics, such as whenever a physicist attempts to observe a particle beam, with more than one particle in it, as in the famous double-slit experiment.

How can this be? Because the information content of equations is negligible in comparison to the information content of the initial conditions, specifically, the contents of the observer's memory. This enables an observer to exploit a priori known information, something a memoryless, elementary particle can never do. To put it simply, the equations of quantum physics were designed to model the behavior of memoryless particles. They make no attempt to model the memory content of an observer. Consequently, they can never, under any circumstances, say anything *interesting* (not devoid of information) about the behavior of such an observer. Everything of any real interest, is determined entirely by the initial conditions, not the equations. By allowing themselves to become enthralled by the “beauty of the math”, physicists have lost sight of the fact that the math is devoid of information and thus has very little to say about anything *interesting*. It is as though they spend all their time admiring the JPEG algorithm, and never bother to look at any actual image. To most other people, the image is the only thing that is *interesting*.

So, how can an observer go about exploiting this loophole? By acting like an FM radio receiver, rather than a Fourier Spectrum Analyzer. Recall that earlier, we considered a set of observations, $x(t)$, that appeared to be sinusoidal, and that we wished to “fit” a sinusoidal function to those observations. We saw that Fourier side-stepped this problem, in a clever, and very useful manner, exploiting superposition. But we can easily deduce an extremely accurate estimate for the unknown frequency, that Fourier's method avoided estimating at all, provided that we exploit a priori information; namely, that the correct model for the observations *is not a superposition*, but is indeed a single frequency wave (a single energy in the case of a particle beam). This is what an FM receiver does. It can be done in a number of different ways. One simple way involves a pair of detectors. They may be thought of as either measuring the amplitude of the sine wave at the output of a pair of bandpass filters, or as particle counters that count photons or electrons captured by the detectors. The key thing is that the detectors, like the cone cells in the retina of the eye, must behave differently towards different input wave frequencies or particle energies.

For example, let $a(f)$ and $b(f)$ be the frequency dependent amplitudes, of the wave measured at the output of the two detectors. If the detectors are designed such that they are Gaussian functions of frequency:

$$a(f) = Ae^{-(f-nD)^2/cD^2} \text{ and } b(f) = Ae^{-(f-(n+1)D)^2/cD^2}$$

where A is the input signal amplitude, c and D are constants, and n is an integer, then, taking the natural logarithm of the ratio $a(f)/b(f)$ we have: $(cD^2) \cdot \ln(a(f)/b(f)) = -(f-nD)^2 + (f-(n+1)D)^2$

solving for f yields: $f = nD + D/2 - (cD/2) \ln(a(f)/b(f))$

The accuracy with which f can be determined by this technique depends primarily on the accuracy with which the amplitudes (or particle counts) $a(f)$ and $b(f)$ can be determined, not on the uncertainty principle. In other words, it depends on the signal-to-noise ratio.

Now we can apply this to the double-slit experiment. The pair of slits act much like the pair of detectors just described. There are two main differences. First, the detectors described above are separated in frequency, whereas the slits are separated in space. Second, the slits do not have a Gaussian response; instead, the combination of their responses *looks like* an interference pattern. Nevertheless, when the observer knows, a priori, that only a single energy exists in the particle beam, he can use the resulting distribution of particle counts along the spatial axis (frequency axis in the case above) to deduce what that energy is. This is easily done via the positions of the nulls (zero particle counts). The accuracy depends on how accurate one can determine the null positions, which in turn depends on the particle counts.

So what? Well, consider:

In double-slit experiments, much is often made of the fact that the distribution of particles *looks like* an interference pattern, even when the particles are sent through the apparatus one at a time, as if that actually matters. Well, it might matter if a detector tried to measure a wave-like property like frequency or phase or a superposition. But neither of the systems just described even attempt to do that. They simply count particles and *infer* the frequency from the counts. It does not matter if the particles arrive all at once or one-at-a-time.

Why does an “interference pattern” occur at all? Because the slits do not have Gaussian responses. They have Sinc-function-like responses, whose combination just happens to *look like* an interference pattern. There is no wave behavior. There are just particles in a correlated energy/frequency state. But detectors like those described do not really distinguish between particles and waves; in effect, they just count received energy quanta and then make an inference, not a measurement; an inference based on a priori information.

Finally, note that no elementary particle could behave like these detectors, or the humans that built them; because they have no memory and thus cannot possess any information about the a priori initial conditions being exploited to enable these detectors and human observers to behave the way they do.

Lessons learned? The mathematical equations of Quantum Mechanics were developed to describe the behaviors of memoryless particles. They have nothing *interesting* to say about the behaviors of systems, like human observers, that routinely exploit vast quantities of a priori known information, like what to do when a light turns red, or how to behave towards an experimental set-up that *they know and remember*, like the double slit, or how to behave toward's Schrodinger's dying cat. Real physics, applied to such systems, needs to *explicitly* and *completely* account for *all* of the vast quantities of information in the initial conditions of the observer's memory. That has never been done. But that is the difference between doing math and doing physics. If you are going to insist on merely cogitating upon the wonderfulness of JPEG-like equations, while ignoring the JPEG-image, then you had better make sure the image being ignored is of nothing much more complicated than a memoryless, elementary particle.

It's all Relative

It is important to remember that nothing we have said alters any of the equations or predictions of the physical theories. We merely point out that the slapped-on misinterpretations of the “meaning” and “significance” of the equations is the source of all the “weirdness”. The weirdness does not exist in the equations, since they do not contain enough information to contain any significant amount of weirdness, *or anything else*. The weirdness exists entirely in an interpretation that implicitly *assumes* that the a priori information content of the memories of the people doing the interpretation is irrelevant. That has turned out to be a very bad assumption. While the space limitations of this essay do not permit anything approaching an exhaustive treatment of this issue, at least one comment concerning Relativity Theory can be considered.

In the words from my old copy of Jackson's “Classical Electrodynamics”, Einstein postulated that “The velocity of light is independent of the motion of the source.” Stated in this manner, that seems to be yet another “weird” phenomenon. But that is only because it is not a phenomenon at all; it is merely another slapped-on misinterpretation. The speed of light is not directly observable; it is not an observable phenomenon at all. Rather, it is being *inferred* from the things that actually are observable. The proper way to think about the speed of light and the Lorentz transformations is analogous to a currency exchange. In order to compare measurements (costs) obtained in different frames of references (currencies), all the measurements must be transformed to a single system. In Relativity Theory, it is often assumed that there is no preferred reference system, that all systems are equivalent. But that is not true. There is always a “privileged” observer – the source of the original transaction, the creator of the “message”. When a light wave is first created, it is created in the reference frame of this privileged observer. It is created at the frequency observed by that observer, not the Doppler shifted frequency of an observer in a different frame of reference. And the privileged observer is always at rest with respect to itself. So the privileged observer's *inferred* speed of light will always be the same. Hence, when all other observers transform their actual observables to the privileged observer's frame, they too must *infer* the same constant speed of light. But the Lorentz transformations of the observables, the observed time required to travel some observed distance, exactly cancel each other out. Consequently, inferring the speed of light from the untransformed observables yields the same constant value that would be obtained if the correctly transformed values had been used. The math is the same, but the physical interpretation is rather different, and most significantly, less weird.

Robert H. McEachern was educated as an AstroPhysicist. He then worked for several years as a Geophysicist, during which time he became interested in signal processing theory. He then spent the rest of his career developing signal processing algorithms for application to communications systems, and sensor systems. He is now retired.