Physical Theory and Mathematical Formalism

Soumitro Banerjee*

Ptolemy, Copernicus, Fourier

IN THE olden times, man's view of the universe was quite different from ours. People believed that the Earth was at the centre of the universe, and all the heavenly bodies revolved round it. That is of course what we observe: we see the sun, the moon and the stars going round the Earth. So the ancient peoples' cosmology was built on that observation.

They really did quite an impressive bit of hard observation—as much as is possible with naked eyes. For example, the ancient Greeks knew that the motion of the planets was different from that of the stars. While the stars appeared to go around the Earth, their positions with respect to one another did not change. But the planets' positions changed from night to night, and they appeared to wander in the starry background. Aristarchus and other astronomers made very detailed observation on the planets, and noticed the peculiarity of their motion.

Take, for example, the motion of the bright red planet Mars. If you keep track of its position from night to night, you would see that for some time it appears to move in one direction with respect to the background stars. But then the motion slows down, and finally stops. Then it turns back and moves in the opposite direction. After some time it stops again and continues its forward motion. The planet continues this wobbling motion, now going forward now turning back, but on an average appearing to go round the Earth (Fig. 1).

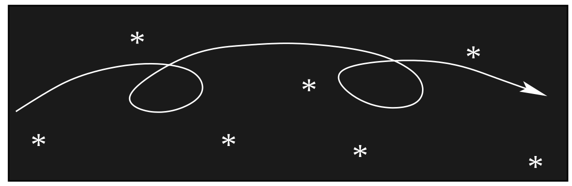


Figure 1: The observed motion of Mars in the starry background

This peculiar motion was noticed by Aristarchus and other ancient Greek astronomers. This puzzled them. While simple observation suggested a geocentric cosmological viewpoint, something more had to be said regarding the motion of the planets than simply saying that they go around the Earth.

Repeated observation that certain occurrences came in the same succession had already germinated the rudimentary ideas of causal relation. Subjective reasoning of possible causal relation drove this primitive knowledge, through extrapolation, into astrological reasoning and belief in supernatural intervention.

Another group proposed that while the sun, the moon, and the stars move in circles around the Earth, the planets do not move in circles. There are smaller circles over bigger circles—which came to be called *epicycles*—so that there are times when planets appear to move in the opposite direction (see Fig. 2). It is not known who originated this idea; but a very clear view appears in the writings of Ptolemy. That is

*Dr. Banerjee is the General Secretary of the *Breakthrough Science Society* and a Professor at the Dept. of Electrical Engineering, Indian Institute of Technology, Kharagpur.

why this picture of the universe is called the Ptolemaic cosmology.

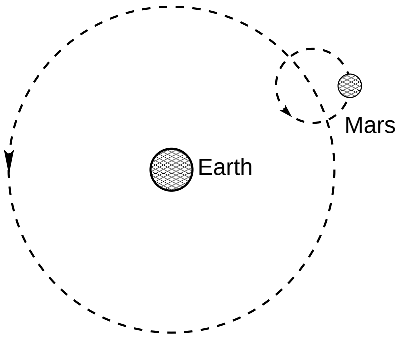


Figure 2: The Ptolemaic conception of planetary motion in epicycles.

Astronomers went on to make more and more accurate observations on the motion of the planets. Soon the picture appeared to be inadequate to account for the actual motion. Then it was proposed that there was not just one epicycle; there were epicycles over epicycles. How many? No one knew. But the picture was getting messier as the volume of observations increased.

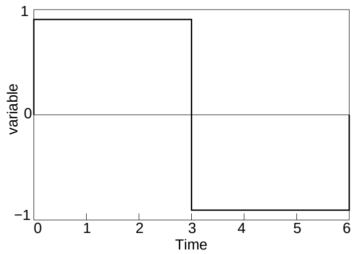
We have to jump from that point to a time a few centuries afterwards, because in the meantime nothing happened in Europe in the area of science. It was the middle age. The Ptolemaic viewpoint had become an integral part in the Christian dogma, and it was a blasphemy to question it.

The gloom of darkness began to fade at the advent of the renaissance in the fourteenth century. And Copernicus stepped into the scene. He proposed that a possible solution to the puzzling motion of the planets is to assume that the sun is at the centre of the solar system, and the Earth as well as the planets are revolving round it. He showed that the observations of a revolving planet from a revolving Earth would be similar to what is actually observed. Note, that it was just an alternative hypothesis, aimed at explaining the same observations. Only after it was supported by Bruno and Galileo that it earned some scientific respectability. Johannes Kepler improved upon Copernicus' picture by showing, using Tycho Brahe's observational data, that the planets do not move in circles. Their trajectories actually trace out elliptical paths, with the sun in one of the two foci. With this vital correction, the Copernican model could account for the apparently peculiar observed motions. The Ptolemaic picture was then jettisoned.

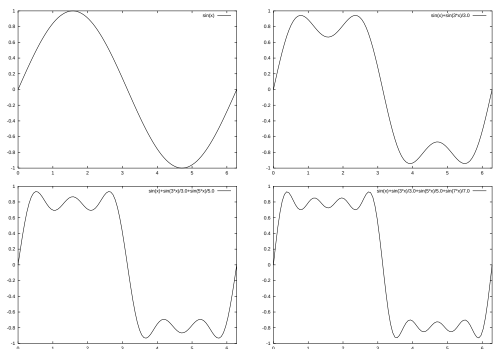
Much later, in the eighteenth century, the French mathematician and physicist Joseph Fourier (1768-1830), made an important discovery. If you imagine observations being made on any physical system as it goes through its own motion, you would obtain a waveform—that indicates how its position varies with time. In general, the observed variable need not be only position: it could also be the velocity (or momentum) of a moving body. In a waveform, the y-axis represents such a dynamical variable, and the x axis represents time.

Suppose a pendulum has a very long chord, and the bob is moved by a small angle. In that case, as the bob undergoes a to-and-fro motion, it traverses a small arc of a circle. Such an oscillation is called a simple harmonic motion. If you now plot θ (the angle of the bob from the vertical line) against time, you would get a sinusoidal waveform, that is, θ is proportional to $\sin t$.

In a physical system, the graphical representation of motion can be such a simple wave-like up and down variation. It can also represent a complicated motion. Fourier was considering the problem of representing such waveforms. He showed that any arbitrary periodic waveform can be expressed as a summation of sinusoidal functions, each with a frequency that is an inte-



(a)



(b)

Figure 3: The successive approximations of the square wave by the Fourier series.

gral multiple of a "fundamental frequency".

As an example, consider a physical system whose variable changes as shown in Fig. 3(a). For three seconds the variable has the value $+1$, and then jumps to the value -1. After six seconds it jumps back to $+1$. This is called a square-wave.

Now consider the trigonometric series

$$\sin \omega_0 t + \frac{1}{3} \sin 3\omega_0 t + \frac{1}{5} \sin 5\omega_0 t + \frac{1}{7} \sin 7\omega_0 t + \cdots$$

where $2\pi/\omega_0$ is the period. The waveforms in 3(b) show pictorial views of this series; the first one shows only the first term, the second one shows the graph of the first two terms, and so forth.

It is clear that when a larger number of terms are considered, the trigonometric series approaches the square wave. Fourier showed that *any* periodic waveform can be represented as an infinite sum of such sine and cosine terms. Fourier also showed how the coefficients of the terms in the series can be calculated for any given periodic waveform. For those who are exposed to trigonometry and complex numbers, some mathematical details of the Fourier series are given in Box-1. Other readers may skip the box and may go ahead.

The observed motion of a planet in the spherical coordinate system of the sky as seen from the Earth is also a waveform. True, a very complicated one, but a periodic waveform nevertheless. And, following Fourier, it can be said that the motion of the planets can be represented by a summation of sinusoidal waveforms.

It so happens that sinusoidal waveforms can also be represented as simple circular rotations. Imagine something rotating in circular motion about a point at constant angular speed. If the x- or y-coordinate of that body is plotted against time, you would get a sinusoidal waveform (see Fig.4), as the values of these coordinates vary as simple harmonic motion.

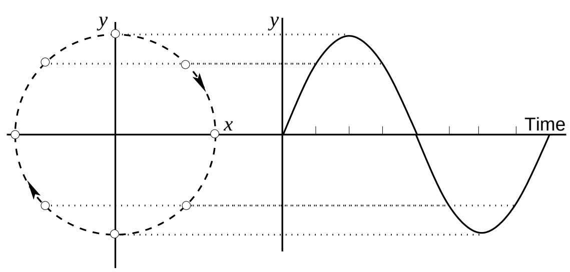


Figure 4: Simple circular motion is equivalent to a sinusoidal waveform.

Thus, the "summation of sinusoids" representation turns out to be strikingly similar to the Ptolemaic picture (see the vector-rotation picture of the Fourier series in Box-1). What turns out is that the observed motion of the planets *can* be represented

Box-1: The Fourier Series

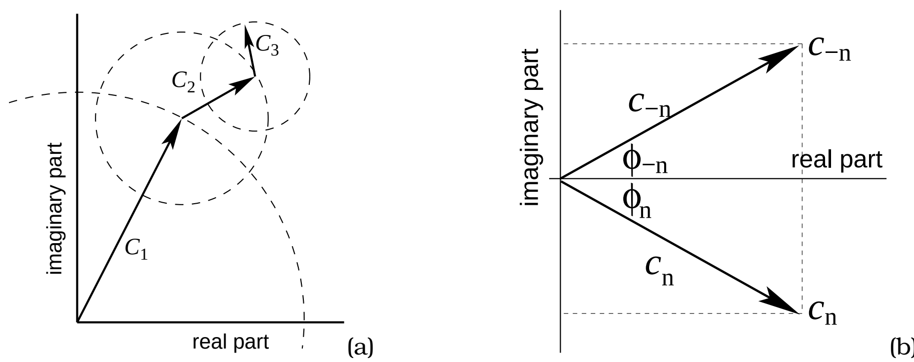
In general, the Fourier series representation of any periodic waveform is expressed as

$$\begin{aligned} x(t) &= a_0 + a_1 \cos \omega_0 t + a_2 \cos 2\omega_0 t + a_3 \cos 3\omega_0 t + \cdots + b_1 \sin \omega_0 t + b_2 \sin 2\omega_0 t + b_3 \sin 3\omega_0 t + \cdots \\ &= a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t \end{aligned}$$

where $-\infty < t < \infty$.
There is another way of representing this series—in terms of exponential terms:

$$f(t) = c_0 + \left(c_1 e^{i\omega t} + c_2 e^{2i\omega t} + c_3 e^{3i\omega t} + \cdots\right) + \left(c_{-1} e^{-i\omega t} + c_{-2} e^{-2i\omega t} + c_{-3} e^{-3i\omega t} \cdots\right) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega t}$$

This is the exponential form of the Fourier series. Here the coefficients c_n are complex numbers. To understand this series, note that a term $c_n e^{in\omega t}$ represents a rotating vector (except c_0 which is stationary). Thus the summation of the rotating vectors can be pictorially represented as shown in Fig.(a) below.



Vectors of the form $c_n e^{in\omega t}$ rotate in a counter-clockwise direction, and those of the form $c_{-n} e^{-in\omega t}$ rotate in clockwise direction. The imaginary parts cancel, giving the real value (see Fig.(b) above). Thus complex conjugate components c_n and c_{-n} must occur. The function is obtained as an infinite summation of spinning vectors rotating at speeds that are integral multiples of the fundamental frequency ω.

by considering the Earth as the centre of the solar system, and the planets moving in epicycles. You can go as close as desired to the actual observed motion by taking into account the necessary number of epicycles; and to get an exact matching you would have to consider an infinite number of them. Thus, the Copernican picture and the Ptolemaic picture (with an infinite number of epicycles) turns out to be *mathematically equivalent*. Had the discovery of Fourier come before Copernicus, what would have been our understanding of the solar system? We would have said, with all confidence, that the Earth is *really* placed at the centre, that the planets go around it in epicycles, and that there are *really* an infinite number of epicycles. We would have built our theory around this picture, enabling us to predict the position of the planets with great accuracy. Not a single observation on the motion of the planets would have contradicted this theory.

You would say that we do know that Copernicus was right. Do we really? How? Was it not a quirk of fate that out of these two mathematically equivalent formulations, Copernicus' came first?

This brings us to the central question I am trying to raise in this essay. When two formulations are mathematically equivalent, how to decide which one represents physical reality?

We will come back to this question later. First let us take a look at another such example of mathematical equivalence.

Newton, Lagrange, Hamilton

As we know, the Copernican picture of the solar system was enriched by Galileo and Kepler. When we learned how the planets move, the next obvious question was: Why? Why do the planets move the way they do?

Newton (1643-1727) solved the problem on the basis of two postulates:

1. Any two bodies attract each other with a force proportional to their masses and inversely proportional to the square of the distance between them.

2. A body's motion is governed by the equation: Force = mass times acceleration.

Since velocity v is the rate of change of position x, and acceleration a is the rate of change of velocity, acceleration of a body is the double derivative of its position. Thus the second postulate gives the equation

$$\frac{d^2x}{dt^2} = \frac{F}{m}$$

This is a differential equation, which can be solved if the initial position and velocity of the body are known. This allowed one to calculate the future positions and velocities of the body.

The causal chain became clear. Why do bodies move? Because they are acted on by forces. What creates the forces? The other bodies around them. Why does Mars move? Because it is acted on by the attraction of the sun. Why does it go from 'here' to 'there' in the sky, and not somewhere else? Because its being 'here' implies its possessing a specific position and velocity. These constitute its initial condition. With this specific initial condition, and the specific force acting on it, its motion is governed by the above differential equation. That tells how it will move in the future.

The success of this theory was spectacular. Man became able to predict the motion of heavenly bodies with great accuracy. The shroud of mystery around them was removed. Not only the motion of planets, the motion of the bodies on earth was also subject to the same laws, and could similarly be predicted.

This line of reasoning gave birth to the idea of causality as a pillar on which science built its explanation of natural phenomena. It came to be accepted that everything happening in the natural world must have a cause. It must be a natural cause (as opposed to a supernatural one), which is knowable. This confidence prompted scientists to look for the cause of every natural phenomenon—not only in the field of physics, but in all fields of human enquiry.

One important element in the idea of causality is that the cause must precede the effect in time. This was evidently built into the Newtonian formalism: The initial condition is the cause of the motion of a body in the future, and always precedes the effect.

An important improvement came in the century following Newton. Scientists were facing some practical difficulties in applying Newton's method to physical situations. The motion of most bodies had to be represented in three dimensions, and hence Newton's laws had to be written in vector form—which was cumbersome. The motions of most bodies are constrained, as a pendulum's motion is constrained to lie on the surface of a sphere. The constraints apply some force on the body which also had to be taken into account. With a large number of interacting elements in a system, the force on a body also becomes difficult to write down. Joseph Louis Lagrange (1736-1813) showed that Newton's laws can also be stated in terms of the total kinetic energy and the total potential energy (see Box-2), which makes the actual job of writing down the differential equations much simpler.

In the meantime people had noticed that in many physical situations nature follows some kind of minimization rule. For example, if a wire is bent into a loop and dipped into a soap solution, a film will form to span the loop that will minimize the area bounded by the wire. If a flexible wire is held at two ends and rotated by the middle, it would assume a shape that minimizes the surface of revolution. Some scientists wondered, is there any such minimization rule that is followed in dynamics?

A few years after Lagrange, the Irish physicist William Rowan Hamilton (1805-1865) showed that there indeed is such a minimization rule. To understand this, consider the motion of a planet like Mars in the gravitational attraction of the sun. Suppose it starts somewhere and moves to some other point in a certain amount of time (see Fig. 5). Now, you are considering why it went from point 1 to point 2 by that specific path and not by some other path. Indeed, you can imagine innumerable possible routes from point 1 to point 2. For each possible path you can calculate the kinetic energy and the potential energy at every moment of the path. Now, integrate the kinetic energy minus the potential energy over time during the whole path. It was found that the object (in this case, Mars) takes that specific path for which this integral is a minimum. In the language of Feynman, "the laws of Newton could be stated not in the form $F = ma$, but in the form: the average kinetic energy less the average potential energy is as little as possible for the path of an object going from one point to another" (*Feynman Lectures on Physics*, Vol.2, Chapter 19).

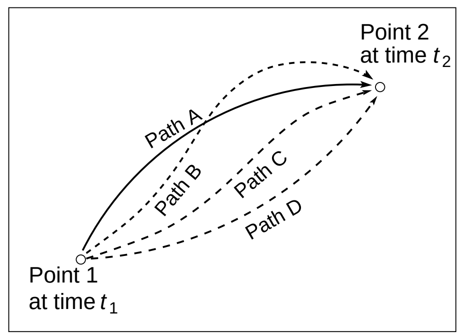


Figure 5: The path followed in moving from point 1 to point 2, and other possible paths.

The integral of the kinetic energy minus the potential energy is called the *action*, and so this minimization rule came to be known as the "least action principle". Hamilton demonstrated that it is mathematically equivalent to Newton's laws of motion, in the sense that you get the same differential equations governing a particle's motion no matter whether you start from Newton's laws or from the least action principle. In fact, the derivation is so simple that many physics teachers today prefer to derive the Lagrange equations from the principle of least action rather than from Newton's laws. For a mathematically inclined

Box-2: The equivalent approaches in mechanics

Suppose there is a body of mass m, acted on by an external force F and a constraint force F^c. Then the Newton's law would take the vectorial form

$$m\ddot{\mathbf{r}} = \mathbf{F}^c + \mathbf{F}$$

Note that the terms in boldface are vectors. If a system is composed of many such bodies (like the sun-earth-moon system), then the equation of the jth mass point will be

$$m_j\ddot{\mathbf{r}}_j = \mathbf{F}_j^c + \mathbf{F}_j$$

And the whole system of equations will have to be solved. The problem was that $\mathbf{F}_j$ and $\mathbf{F}_j^c$ were not easily quantifiable.

Lagrange showed that the same system of equations can be written in terms of the total kinetic energy T and the total potential energy V. He introduced a new function $\mathcal{L}$, now called the Lagrangian function, which is the kinetic energy minus the potential energy, that is, $\mathcal{L} = T - V$. In terms of this Lagrangian function, the Newtonian equation can be written as

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0,$$

where the q's denote the "generalized" coordinates which include the positions and momenta of the bodies involved.

The principle of least action says that if the system moved from one point at time t_1 to another at time t_2, the path in between would be the one for which the integral of the Lagrangian function

$$S = \int_{t_1}^{t_2} \mathcal{L}\, dt$$

is a minimum. Hamilton showed that one can derive the Lagrangian equation starting from this premise. Thus, the Newton's laws, Lagrange's equation, and the principle of least action are mathematically equivalent.

reader, a brief account of the Newtonian, Lagrangian, and least-action approach is given in Box-2.

Now notice the catch. For any body in motion, its initial condition *as well as the final condition* are needed to decide in what direction it should move in the next instant. The path is determined by the starting point as well as the destination. You cannot evaluate the action integral unless you already know the starting point and the ending point of the path. Since the particle moves along that path for which the action integral must be minimum, the particle must somehow "know" where its final destination is even before it starts to move.

Let us again consider the question we asked when discussing Newton's laws: Why does a particle move? A follower of the least action principle would say: Because it has to reach point 2 starting from point 1 in a specific amount of time. Thus, the destination becomes a cause of motion. The law of causality obviously becomes meaningless.

The least action principle is thus found to be of a teleological character [1]. In his textbook on mechanics (first Indian reprint, Levant books, Kolkata, 2003), the famous physicist Arnold Sommerfeld commented that in its formulation not only teleological but also theological beliefs played a role. Maupertuis and Leibniz used this principle with the assertion that it best expressed the wisdom of the Creator.

But we have seen earlier that the laws of Newton perfectly conformed to the concept of causality. Now we find that the least action principle does not. Yet, the two are mathematically equivalent to each other!

And Physical Reality

Thus we see that it is possible to have mathematically equivalent formulations that give completely different pictures of physical reality. After all, mathematical formulations have roots in human reasoning which inalienably subsume ideas and outlooks—objective or subjective. If the least action principle had been discovered before Newton, we would still be able to calculate the motion of objects. Not a single observation would have contradicted the theory. But we would have tried to build science without the idea of causality. Science would have been quite different from what we know.

Similarly, had the Ptolemaic picture been perfected on the basis of Fourier's theory before Copernicus and Galileo, we would have been able to explain all observations on the motion of the planets, but Newton's discovery would have been impossible. Science would have been completely different.

Now notice an important issue. If science is based only on observations, and the purpose of science is solely to explain the observations, the Ptolemaic picture is perfectly all right. But the moment you transcend that limited objective and set about to find what is the physical reality, the limitation of the Ptolemaic theory is revealed. If you imagine yourself placed on another planet and making observations on the motion of the other planets (including the Earth), you would still be able to explain the observations on the basis of a Ptolemy-like picture. *That* planet would then be placed at the centre of the solar system; other planets would move on epicycles that are completely different from those for Earth-based observations. That is, for every different observation from different points in the solar system, you describe a different reality. The picture of physical reality becomes different for different observers. If you want to construct a picture of physical reality that satisfies all observations but in itself is independent of the observer, you have no way but to count on the Copernican picture. That is why, in modern understanding the Copernican view of the solar system with sun at the centre is correct. This gives us the important lesson that the objective of science should be to construct a picture of physical reality that is consistent with the observations.

In contrast, there is a philosophical position called positivism, which concerns itself only with observations and observables. It argues that the objective of science is only to predict and explain the observations. According to this outlook, the source of knowledge is sense-perception, and sense-perception belongs to the objective. Many scientists in the 1920's and 1930's were influenced by this idea, and took positivism as the philosophical guideline for their science. Especially when the theory of quantum mechanics was taking shape, scien-

[1]The word "teleological" means "shaped by a purpose" or "directed toward an end."

tists like Niels Bohr and Werner Heisenberg argued in favour of it and built a mathematical formalism with which you can predict what will be the observation in a specific situation. This viewpoint came to be called the "Copenhagen interpretation". It has been very successful, as it has succeeded in explaining all observations and not a single experiment has contradicted the predictions of quantum mechanics. But that structure of quantum mechanics never talks about what is the physical reality of the microworld.

Many leading scientists of that time, including Albert Einstein, Max Planck, and Erwin Schrödinger did not agree with this viewpoint and argued that the purpose of a theory should be to grasp the nature of physical reality. If we can understand it, explanations of observations will naturally follow from that. But if the objective of science becomes only to produce a "cookbook" — a collection of mathematical techniques with which observations coming out of experiments can be predicted, the whole purpose of science is defeated.

Many people feel that the debate—led by Bohr in one side and Einstein in the other—had been settled in favour of the Copenhagen viewpoint. That is now treated as the mainstream quantum mechanics, which is taught today in the physics courses. Students learn only the technique, and are generally happy that "it works".

Yet, there is a group of scientists who feel that there must be a mathematically equivalent formalism which will reveal the physical reality of the microworld. True, that formalism has not been found yet, and the search is still on. The history of Ptolemy and Copernicus tells us that this may not be a futile exercise in the long run. We have to realize that knowledge flows *not* from observation, but from conceptual coordination and integration of the observed facts into a picture of physical reality. Knowledge has roots in reality, but is gift of human brain.

There is another aspect of mathematics that demand caution. The whole of mathematics deals with abstraction from reality. When in school we learn $2 + 3 = 5$, that itself is an abstraction. In reality there are only number of objects: two mangoes plus three mangoes equals five mangoes. When we write only the numbers without referring to whose numbers they are, we are already abstracting from reality. When we write $x + y = z$ we are generalizing the numbers, and are further abstracting from reality. There is no x in nature, yet x is related to physical reality: x represents a number, which could represent the number of something in nature. When a mathematician goes into further levels of abstraction and works with such abstract mathematical entities, playing with their own set of rules, there is always the danger of losing track of the link with physical reality. The more the abstraction, the more the danger of straying out into fiction. Many times in the history of mathematics it so happened that a mathematical concept that initially looked like pure abstraction turned out to be of great importance in understanding physical reality. Take the history of the imaginary number or tensor calculus as example. Even if it is not immediately apparent what physical reality a specific mathematical idea represents, one should have the confidence that no mathematical concept can be devoid of any link with physical reality.

Next, causality is a heuristic principle on which all of science is built. When a scientist sees a phenomenon, he looks for explanation, and asks 'Why'? Before Darwin scientists observed the multitude of life forms on Earth. The idea that species change had also beed hatched. Darwin asked: Why? Why do species change? And he proposed a causal chain that leads to speciation. That is how science is done. If a scientist does not have the confidence that a specific phenomenon has a cause, why should he or she try to look for the cause?

The instance of the least action principle teaches us that a specific theory, however successful it may be in explaining observations, cannot refute the idea of causality. There are far too many developments in science that point to the existence of a causal connection in every phenomenon. Today we see the tendency in some scientists to pronounce the priciple of causality dead, just because a mathematical formalism seemed to violate causality. When one comes across a theory that appears to violate causality, one should look for its mathematical equivalence—the way scientists continued to have confidence in the operation of causality on the basis on Newton's laws in spite of the least action principle's pointing otherwise. □