18

Inner Vision, Outer Truth

REUBEN HERSH

There is an old conundrum, many times resurrected: why do mathematics and physics fit together so surprisingly well? There is a famous article by Eugene Wigner, or at least an article with a famous title: "The Unreasonable Effectiveness of Mathematics in Natural Sciences." After all, pure mathematics, as we all know, is created by fanatics sitting at their desks or scribbling on their blackboards. These wild men go where they please, led only by some notion of 'beauty', 'elegance', or 'depth', which nobody can really explain. Wigner wrote, 'It is difficult to avoid the impression that a miracle confronts us here, quite comparable in its striking nature to the miracle that the human mind can string a thousand arguments together without getting itself into contradictions, or to the two miracles of the existence of laws of nature and of the human mind's capacity to divine them.'

In Lobachevsky's non-Euclidean geometry, or Cayley's matrix theory, and Galois' and Jordan's group theory, and the algebraic topology of the mid-twentieth century, pure mathematics seemed to have left behind any physical interpretation or utility. And yet, physicists later found these 'useless' mathematical abstractions to be just the tools they needed.

Freeman Dyson writes, in his Foreword to Monastyrsky's *Riemann, Topology, and Physics*, of 'one of the central themes of science, the mysterious power of mathematical concepts to prepare the ground for physical discoveries which could not have been foreseen or even imagined by the mathematicians who gave the concepts birth.'

On page 135 of that book, there is a quote from C. Yang, co-author of the Yang-Mills equation of nuclear physics, speaking in 1979 at a symposium dedicated to the famous geometer, S.-S. Chern.

"Around 1968 I realised that gauge fields, non-Abelian as well as Abelian ones, can be formulated in terms of nonintegrable phase factors, i.e., path-dependent group elements. I asked my colleague Jim Simons about the mathematical meaning of these nonintegrable phase factors, and he told me they are related to connections with fibre bundles. But I did not then appreciate that the fibre bundle was a deep mathematical concept. In 1975 I invited Jim Simons to give to the theoretical physicists at Stony Brook a series of lectures on differential forms and fibre bundles. I am grateful to him that he accepted

the invitation and I was among the beneficiaries. Through these lectures T. T. Wu and I finally understood the concept of nontrivial bundles and the Chern-Weil theorem, and realized how beautiful and general the theorem is. We were thrilled to appreciate that the nontrivial bundle was exactly the concept with which to remove, in monopole theory, the string difficulty which had been bothersome for over forty years [that is, singular threads emanating from a Dirac monopole].

"When I met Chern, I told him that I finally understood the beauty of the theory of fibre bundles and the elegant Chern-Weil theorem. I was struck that gauge fields, in particular, connections on fibre bundles, were studied by mathematicians without any appeal to physical realities. I added that it is mysterious and incomprehensible how you mathematicians would think this up out of nothing. To this Chern immediately objected. 'No, no, this concept is not invented-it is natural and real."

Why does this happen?

Is there some arcane psychological principle by which the most original and creative mathematicians find interesting or attractive just those directions in which Nature herself wants to go? Such an answer might be merely explaining one mystery by means of a deeper mystery.

Or perhaps the "miracle" is an illusion. Perhaps for every bit of abstract purity that finds physical application, there are a dozen others that find no such application, but instead eventually die, disappear and are forgotten. This second explanation could even be checked out, by a doctoral candidate in the history of mathematics. I have not checked it myself. My gut feeling is that it is false. It seems somehow that most of the mainstream research in pure mathematics does eventually connect up with physical applications.

Here is a third explanation, a more philosophical one that relies on the very nature of mathematics and physics. Mathematics evolved from two sources, the study of numbers and the study of shape, or more briefly, from arithmetic and visual geometry. These two sources arose by abstraction or observation from the physical world. Since its origin is physical reality, mathematics can never escape from its inner identity with physical reality. Every so often, this inner identity pops out spectacularly when, for example, the geometry of fiber bundles is identified as the mathematics of the gauge field theory of elementary particle physics. This third explanation has a satisfying feeling of philosophical depth. It recalls Leibnitz's "windowless monads", the body and soul, which at the dawn of time God set forever in tune with each other. But this explanation, too, is not quite convincing. For it implies that all mathematical growth is predetermined, inevitable. Alas, we know that is not so. Not all mathematics enters the world with that stamp of inevitability. There is also "bad" mathematics, that is, pointless, ugly, or trivial. This sad fact forces us to admit that in the evolution of mathematics there is an element of human choice, or taste if you prefer. Thereby we return to the mystery we started with. What enables certain humans to choose better than they have any way of knowing?

A good rule in mathematical heuristics is to look at the extreme cases – when a small parameter becomes zero, or a large parameter becomes infinite. Here, we are studying the way that discoveries in "pure"mathematics sometimes turn out to have important, unexpected uses in science (especially physics). I would like to use the same heuristic – "look at the extreme cases". But in our present discussion, what does that mean, "extreme case"? Of course, we could give this expression many different meanings. I propose to mean "extremely simple". To start with, let's take counting, that is to say, the natural numbers.

These numbers were a discovery in mathematics. It was a discovery that much later became important in physics and other sciences. For instance, one counts the clicks of a Geiger counter. One counts the number of white cells under a microscope. Yet the original discovery or invention of counting was not intended for use in science; indeed, there was no "science" at that early date of human culture.

So let us take this possibly childish example, and ask the same question we might ask about a fancier, more modern example. What explains this luck or accident, that a discovery in "pure mathematics" turns out to be good for physics?

Whether we count and find the planets seven, or whether we study the n-body problem, where n is some positive integer, we certainly do need and use counting – the natural numbers – in physics and every other science.

This remark seems trivial. Such is to be expected in the extreme cases. We do not usually think of arithmetic as a special method or theory, like tensors, or groups, or calculus. Arithmetic is the all-pervasive rock bottom essence of mathematics. Of course it is essential in science; it is essential in everything. There is no way to deny the obvious fact that arithmetic was invented without any special regard for science, including physics; and that it turned out (unexpectedly) to be needed by every physicist.

We are therefore led again to our central question, "How could this happen?" How could a mathematical invention turn out, unintentionally, after the fact, to be part of physics? In this instance, however, of the counting numbers, our question seems rather lame. It is not really surprising or unexpected that the natural numbers are essential in physics or in any other science or non-science. Indeed, it seems self-evident that they are essential everywhere. Even though in their development or invention, one could not have foreseen all their important uses.

So to speak, when one can count sheep or cattle or clam shells, one can also count (eventually) clicks of a Geiger counter or white cells under a microscope. Counting is counting. So in our first simple example, there really is no question, 'How could this happen?' Its very simplicity makes it seem obvious how 'counting in general' would become, automatically and effortlessly, 'counting in science'.

Now let's take the next step. The next simplest thing after counting is circles. Certainly it will be agreed that the circle is sometimes useful. The Greeks praised it as 'the heavenly curve'. According to Otto Neugebauer, "Philosophical minds considered the departure from strictly uniform circular motion the most serious objection against the Ptolemaic system and invented extremely complicated combinations of circular motions in order to rescue the axiom of the primeval simplicity of a spherical universe" *(The Exact Sciences in Antiquity)*. I. B. Cohen wrote, "The natural motion of a body composed of aether is circular, so that the observed circular motion of the heavenly bodies is their natural motion, according to their nature, just as motion upward or downward in a straight line is the natural motion for a terrestrial object" *(The Birth of a New Physics)*.

And here is a more detailed account of the circle in Greek astronomy: "Aristotle's system, which was based upon earlier works by Eudoxus of Cnidos and Callippus, consisted of 55 concentric celestial spheres which rotated around the earth's axis running through the center of the universe. In the mathematical system of Callippus, on which Aristotle directly founded his cosmology of concentric spheres, the planet Saturn, for example, was assigned a total of four spheres, to account for its motion 'one for the daily motion, one for the proper motion along the zodiac or ecliptic, and two for its observed retrograde motions along the zodiac" (E. Grant, *Physical Science in the Middle Ages)*.

In recent centuries, other plane curves have become familiar. But the circle still holds a special place. It is the 'simplest', the starting point in the study of more general curves. Circular motion has special interest in dynamics. The usual way to specify a neighborhood of a given point is by a circle with that point as center.

So we see that the knowledge of circles which we inherited from the Greeks (with a few additions) is useful in many activities today, including physics and the sciences. I suppose this is one reason why 10th grade students are required to study Euclidean geometry.

Again, we return to the same question. How can we explain this 'miracle'?

Few people today would claim that circles exist in nature. Any seemingly circular motion turns out on closer inspection to be only approximately circular.

Not only that. The notion of a circle is not absolute. If we define distance otherwise, we get other curves. To the Euclidean circle we must add non-Euclidean "circles". If the Euclidean circle retains a central position, it does so because we choose – for reasons of simplicity, economy, convenience, tradition – to give it that position.

We see, then, two different ways in which a mathematical notion can enter into science. We can put it there, as Ptolemy put circles into the planetary motion. Or we can find it there, as we find discreteness in some aspect or other of every natural phenomenon.

Let's take one last example, a step up the ladder from the circle. I mean the conic sections, especially the ellipse. These curves were studied by Apollonius of Perga (262-200 B.C.) as the "sections" (or "cross sections" as we would say) of a right circular cone. If you cut the cone with a cutting plane parallel to an element of the cone, you get a parabola. If you tilt the cutting plane toward the direction of the axis, you get a hyperbola. If you tilt it the other way, against the direction of the axis, you get an ellipse.

This is "pure mathematics", in the sense that it has no contact with science or technology. Today we might find it somewhat impure, since it is based on a visual model, not on a set of axioms.

The interesting thing is that nearly 2,000 years later, Kepler announced that the planetary orbits are ellipses. (There also may be hyperbolic orbits, if you look at the comets.)

Is this a miracle? How did it happen that the very curves Kepler needed to describe the solar system were the ones invented by Apollonius some 1800 or 1900 years earlier?

Again, we have to make the same remarks we did about circles. Ellipses are only approximations to the real orbits. Engineers using earth satellites nowadays need a much more accurate description of the orbit than an ellipse. True, Newton proved that 'the orbit' is exactly an ellipse. And today we reprove it in our calculus classes. In order to do that, we assume that the earth is a point mass (or equivalently, a homogeneous sphere). But you know and I know (and Newton knew) that it is not.

Kepler brought in Apollonius's ellipse because it was a good approximation to his astronomical data. Newton brought in Apollonius's ellipse because it was the orbit predicted by his gravitational theory (assuming the planets are point masses, and that the interactive attraction of the planets is 'negligible'). Newton used Kepler's (and Apollonius's) ellipses in order to justify his gravitational theory. But what if Apollonius had never lived? Or what if his eight books had been burned by some fanatic a thousand years before? Would Newton have been able to complete his work?

We can imagine three different scenarios: (1) Kepler and Newton might have been defeated, unable to progress; (2) they might have gone ahead by creating conic sections anew, on their own; (3) they might have found some different way to study the dynamics of the planets, doing it without ellipses.

Scenario three is almost inconceivable. Anyone who has looked at the Newtonian theory will see that the elliptic trajectory is unavoidable. Without Apollonius, one might not know that this curve could be obtained by cutting a cone. But that fact is quite unnecessary for the planetary theory. And surely somebody would have noticed the connection with cones (probably Newton himself).

Scenario one, that Newton would have been stuck if not for Apollonius, is *quite* inconceivable. He, like other mathematical physicists since his time, would have used what was available and created what he needed to create. While Apollonius' forestalling Kepler and Newton is remarkable and impressive, from the viewpoint of Newton's mechanics it is inessential. In the sequence of events that led to the Newtonian theory, what mattered were the accumulation of observations by Brahe, the analysis of data by Kepler, and the development by Barrow and others of the "infinitesimal calculus". The theory of the conic sections, to the extent that he needed it, could have been created by Newton himself. In other words, scenario two is the only believable one.

If a mathematical notion finds repeated use, in many branches of science, then such repeated use may testify to the universality, the ubiquitousness, of a certain physical property – as discreteness, in our first example. On the other hand, the use of such a mathematics may only be witness to our preference for a certain picture or model of the world, or to a mental tradition which we find comfortable and familiar. And also, perhaps, to the amiability or generosity of nature, which allows us to describe her in the manner we choose, without being "too far" from the truth.

What then of the real examples – matrices, groups, tensors, fiber bundles, connections? Maybe they mirror or describe physical reality "by lucky accident", so to speak, since the physical application could not have been foreseen by the inventors.

On the other hand, maybe they are used as a matter of mere convenience – we understand them because we invented them, and they work "well enough".

Maybe we are not even able to choose between these two alternatives. To do so would require knowledge of how nature "really" is, but all we can ever have are data and measurements and hypotheses in which we put more or lees credence.

In fact, it may be deceptive to pose the two alternatives – true to nature, like the integers, or an imposed model, like the circle. Any useful theory must be both. Understandable – i.e., part of our known mathematics, either initially or ultimately – and also "reasonably" true to the facts, the data. Both aspects – man-made and also faithful to reality –must be present.

These self-critical remarks do not make any simplification in our problem.

The problem is, to state it for the last time, how is it that mathematical inventions made with no regard for scientific application turn out so often to be useful in science?

We have two alternative explanations, suggested by our two primitive examples, counting and circles. Example one, counting, leads to explanation one: That certain fundamental features of nature are found in many different parts of physics or science; that a mathematical structure which faithfully captures such a fundamental feature of nature will necessarily turn out to be applicable in science.

According to explanation two, (of which the circle was our simple example), there are several different ways to describe or "model" mathematically any particular physical phenomenon. The choice of a mathematical model may be based more on tradition, taste, habit, or convenience, than on any necessity imposed by the physical world. The continuing use of such a model (circles, for example) is not compelled by the prevalence of circles in nature but only by a preference for circles on the part of human beings—, scientists, in particular.

What conclusion can we make from all this? I offer one. It seems to me that there is not likely to be any universal explanation of all the surprising fits between mathematics and physics. It seems clear that there are at least two possible explanations; in each instance, we must decide which explanation is most convincing. Such an answer, I am afraid, will not satisfy our insistent hankering for a single simple explanation. Perhaps we will have to do without one.

Bibliography

1. I. Bernard Cohen, *The Birth of a New Physics*. (Doubleday & Company, 1960).
2. Edward Grant, *Physical Science in the Middle Ages*. (Cambridge University Press, 1977).
3. Michael Monastyrsky, *Riemann, Topology, and Physics*. (Birkhauser Boston, 1987).
4. O. Neugebauer, *The Exact Sciences in Antiquity*. (Dover Publications, Inc., 1969).
5. Mark Steiner, *The Applicability of Mathematics as a Philosophical Problem*. Harvard University Press, 1988.
6. Eugene Paul Wigner, *Symmetries and Reflections*. (Bloomington, Indiana University Press, 1967).