

The Necessity of Mathematics

Blake Stacey

17 June 2008

Today, everything from international finance to teenage sexuality flows on a global computer network which depends upon semiconductor technology which, in turn, could not have been developed without knowledge of the quantum principles of solid-state physics. Today, we are damaging our environment in ways which require all our fortitude and ingenuity just to comprehend, let alone resolve. More and more people are becoming convinced that our civilization requires wisdom in order to survive, the sort of wisdom which can only come from scientific literacy; thus, an increasing number of observers are trying to figure out why science has been taught so poorly and how to fix that state of affairs. Charles Simonyi draws a distinction between those who merely “popularize” a science and those who promote the public understanding of it.¹ We might more generously speak of bad popularizers and good ones, but the distinction between superficiality and depth is a real one, and we would do well to consider what criteria separate the two.

Opinions on how to communicate science are as diverse as the communicators. In this Network age, anyone with a Web browser and a little free time can join the conversation and become part of the problem — or part of the solution, if you take an optimistic view of these newfangled media. Certain themes recur, and tend to drive people into one or another loose camp of like-minded fellows: what do you do when scientific discoveries clash with someone’s religious beliefs? Why do news stories sensationalize or distort scientific findings, and what can we do about it? What can we do when the truth, as best we can discern it, is simply not politic?

Rather than trying to find a new and juicy angle on these oft-repeated questions, this essay will attempt to explore another direction, one which I believe has received insufficient attention. We might grandiosely call this a foray into the philosophy of science popularization. The topic I wish to explore is the role mathematics plays in understanding and doing science, and how we disable ourselves if our “explanations” of science do not include mathematics. The fact that too many people don’t know statistics has already been mourned, but the problem runs deeper than that. To make my point clear, I’d like to focus on a specific example, one drawn from classical physics. Once we’ve explored the idea in question, extensions to other fields of inquiry will be easier to make. To make life as easy as possible, we’re going to step back a few centuries and look at a development which occurred when the modern approach to natural science was in its infancy.

¹See <http://www.sunclipse.org/?p=525>.

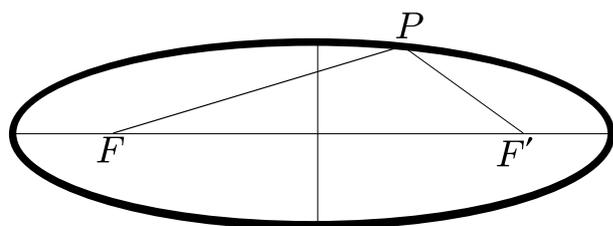
Our thesis will be the following: that if one does not understand or refuses to deal with mathematics, one has fatally impaired one's ability to follow the physics, because not only are the *ideas* of the physics expressed in mathematical form, but also the *relationships* among those ideas are established with mathematical reasoning.

This is a strong assertion, and a rather pessimistic one, so we turn to a concrete example to investigate what it means. Our example comes from the study of planetary motion and begins with Kepler's Three Laws.

1 Kepler's Three Laws

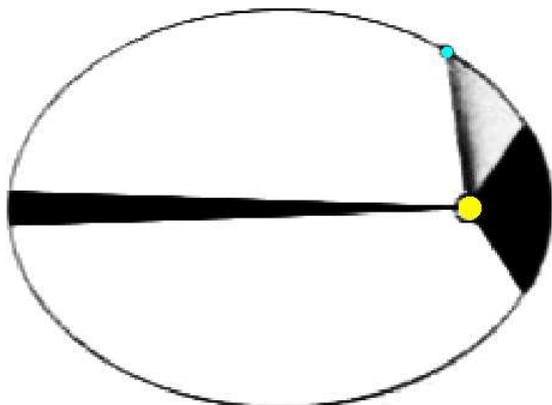
Johannes Kepler (1571–1630) discovered three rules which described the motions of the planets. He distilled them from the years' worth of data collected by his contemporary, the Danish astronomer Tycho Brahe (1546–1601). The story of their professional relationship is one of clashing personalities, set against a backdrop of aristocracy, ruin and war. From that drama, we boil away the biography and extract some items of geometry:

Kepler's first law states that a planet travels in an *ellipse* around the Sun, with the Sun placed at one focus of the ellipse. We can define an ellipse in one of several ways, all of which describe the same curve; the most familiar and convenient definition is that given two points, which we call *foci*, an ellipse is the set of all points such that the total distance from one focus to the point on the curve and then to the other focus is the same for all points on the curve. If we call one focus F and the other F' , then the distance FP plus the distance $F'P$ is the same for all points P on the ellipse. If F and F' lie on top of each other, then the ellipse becomes a *circle*: all the points on a circle are the same distance from the center. (Circles and ellipses are two kinds of curves in the *conic section* family, which consists of the curves you get when a flat plane intersects a cone. These two close on themselves, but their relatives, the parabola and hyperbola, are open-ended.) Lots of aspects of ellipses have their own names: an ellipse's deviation from being a perfect circle is its *eccentricity*, the widest part is its *major axis*, the narrowest is its *minor axis* and so forth.



Having described the *shape* of an orbit, we next describe how a planet moves along its orbital path. By sifting through Tycho's measurements, Kepler found that the planets sped up and slowed down as they orbited the Sun, and moreover, that they did so in a precise way. When a planet was near the Sun, it moved faster, and when its elliptical orbit carried it farther from the Sun, it moved more slowly. Kepler quantified this relationship by drawing an imaginary line from the Sun to the moving planet; as the planet progresses along its orbit, such a line swings along and sweeps out a region of space within the orbit. Kepler's second

law states that a line drawn from the Sun to a planet will sweep out equal areas in equal times.



Kepler's third law relates planets in different orbits. It states that the square of the planet's period (the time the planet takes to go all the way around its orbit) is proportional to the cube of the orbit's major axis. If the orbit is circular, then the square of the period is proportional to the cube of the radius.

2 Bodies in Motion

The history of classical mechanics is full of people stumbling onto bits of the truth, fighting each other for credit, making lucky guesses and turning down blind alleys.² We often obscure this history and summarize it all with a hat-tip to a few favorite names, or else we skip the history completely and present a slick, modern treatment of mechanics decorated with names of long-rotten brains who'd require a serious refresher course to follow the improvements we've made in the interim. In grandiose terms appropriate for someone seeking a "classical education," we could say that the practical, hands-on approach³ employed by some pre-Socratic Greeks joined with the geometry of Euclid and company to overturn some ideas advocated by certain *other* Greeks, in particular Aristotle's scheme of how motion worked and what the planets were doing.

Aristotle (384–322 BCE) had said that Earthly matter obeyed fundamentally different principles than celestial matter. In the Aristotelian scheme, the Sun, Moon and planets were made of a "fifth element" whose natural inclination was to move in circles, whereas terrestrial stuff fell towards the center of the Earth (which was, at that time, considered the center of everything). Aristotle was unmistakably a bright fellow, and he became the go-to guy for answering scientific questions for hundreds of years. Observation, however, trumps authority:

²A taste of this can be found at The *n*-Category Café; see http://golem.ph.utexas.edu/category/2008/01/a_tiny_taste_of_the_history_of.html.

³Exemplified by the Tunnel of Samos, <http://eands.caltech.edu/articles/LXVIII1/samos.html>.

Nicole Oresme (1323–1382) realized that Aristotle’s ideas weren’t enough to describe how motion worked, and so he added the concept of “impetus.” If you gave Earthly matter a push, you could impart an impetus, which the object exhausted as it moved, and with the impetus gone, the object would fall straight down again. Furthermore, *contra* Aristotle, Oresme deduced that gravity accelerates all bodies equally: drop a heavy ball and a light ball from the same height, and they’ll hit the ground at the same time. (He also invented a scheme for using coordinates in geometry and graphing quantities as functions of time and at least entertained the idea that the Earth turns on its axis.⁴)

Leonardo da Vinci (1452–1519) and Nicolo Tartaglia (1500–1557) noticed that cannonballs didn’t just fly in straight lines and then fall down, but instead went in curving paths. Giovanni Benedetti (1530–1590), after falling out with his old teacher Tartaglia, realized that straight-line motion could be pulled into a curve, as when swinging a rock on a string in preparation to beaming one’s little sister. When you release the rock, it starts to fly off in a straight line, but gravity takes over and soon bends its path into a different curve. Benedetti also spent time up a tower throwing balls over the side, and he found that Oresme had been right and Aristotle wrong. Simon Stevin (1548–1620), a Dutch engineer who invented decimal notation in his time off from designing better windmills, performed the same experiment from the Delft church tower and got the same result.

Three years after Stevin’s experiment, in 1590, an arrogant math professor who had made his name with a spyglass for spotting ships at sea wrote a series of essays covering his systematic investigation of motion. For doing it last — and thoroughly, and clearly, following what we now recognize as the experimental method — Galileo Galilei (1564–1642) gets all the credit. Galileo later turned his telescope to the sky, found the best evidence yet that Aristotle’s geocentric universe was wrong, got himself in trouble with the Church and became an icon, not just to scientists but to anyone who believes themselves possessed of a revolutionary idea (most of whom turn out to miss an essential part of Galileo’s fame: being right).

The fact that one man, Galileo, explored motion both in the Heavens and on the Earth provides biographical foreshadowing for the discovery that Heaven and Earth are made of the same stuff, obeying the same principles of natural law. This is one of those big ideas we owe to a great many people, but as far as gravity goes, two of the big names are Robert Hooke (1635–1703) and Isaac Newton (1643–1727). In 1674, Hooke presented a “System of the World” which, he said,⁵ was grounded on the following three basic suppositions:

First, That all Coelestial Bodies whatsoever, have an attraction or gravitating power towards their own Centers, whereby they attract not only their own parts, and keep them from flying from them, as we may observe the Earth to do, but that they do also attract all the other Coelestial Bodies that are within the sphere of their activity.

Hooke realized that the Sun, the Moon and the planets were (roughly) spherical because the gravitational attraction of each bit of a massive body pulled on every other bit, so

⁴<http://www.dam.brown.edu/people/mumford/AM18/HANDOUTS/2006PartIb.pdf>

⁵See http://www.roberthooke.com/motion_of_the_earth_015.htm.

that the whole aggregate is pulled together into a sphere. That he speaks of a limit to a celestial body’s “activity” suggests that he did not expect the gravitational force to extend to indefinitely large distances.

The second supposition is this, That all bodies whatsoever that are put into a direct and simple motion, will so continue to move forward in a straight line, till they are by some other effectual powers deflected and bent into a Motion, describing a Circle, Ellipsis, or some other more compounded Curve Line.

Nowadays, we refer to this as the property of *inertia*. Moving at an unchanging speed without wavering in direction is just as good as standing still; put another way, there exists no unique state of standing still.

The third supposition is, That these attractive powers are so much the more powerful in operating, by how much the nearer the body wrought upon is to their own Centers.

Once you start thinking in terms of forces deranging the motion of objects, I suppose it’s natural to guess that whatever the details of the forces, nearby objects will disrupt one’s motion more than distant ones. Hooke didn’t know those details, but he had high hopes:

Now what these several degrees are I have not yet experimentally verified; but it is a notion, which if fully prosecuted as it ought to be, will mightily assist the Astronomer to reduce all the Coelestial Motions to a certain rule, which I doubt will never be done true without it.

The rule which Hooke envisioned turned out to be Newton’s law of universal gravitation. This principle states that any two objects — planets, moons, stars, daffodils — attract each other with a force which grows with the product of the two objects’ masses, and falls off with the square of the distance between them. (Newton deduced this “inverse-square” dependence from Kepler’s Third Law.) We can squeeze that mouthful down into an equation,

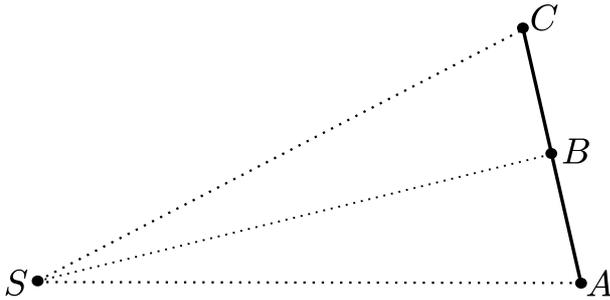
$$F = G \frac{m_1 m_2}{r^2}, \tag{1}$$

which is useful for stating the gravity law concisely, doing calculations and intimidating the reader. Objects *exert* forces following Newton’s formula, and they *react* to forces by changing their motion in the direction of the force applied. The more mass an object has, the stronger its gravitational pull, but the more it *resists* a change in motion. For some reason not known to Newton, the *inertial* mass which tells how difficult it is to change a body’s motion is always equal to its *gravitational* mass, the quantity which appears in the universal gravitation equation. This, incidentally, explains the observation of Benedetti, Stevin, Galileo and the rest that the amount of matter in an object doesn’t change how that object accelerates under gravity.⁶

⁶Briefly, one sets the F in $F = ma$ equal to the F in $F = GMm/r^2$, where M is the mass of the Earth and m that of an apple, and then one cancels the m on both sides.

3 Equivalence

Having sketched the odd and complicated history, we come to the core of our problem. We have two assertions. From Kepler, we know that a planet sweeps out equal areas in equal times. Newton tells us that the gravitational force on a planet is always directed towards the Sun. Neither of these statements are remarkably mathematical in nature: oh, we're speaking of "equal areas" and "equal times," but it's not particularly the sort of mathematics which can give the listener heebie-jeebies. Given a moderate amount of patience, we can translate both statements into "everyday language." What we *cannot* express in everyday language is the fact that these two statements *are really the same thing*.



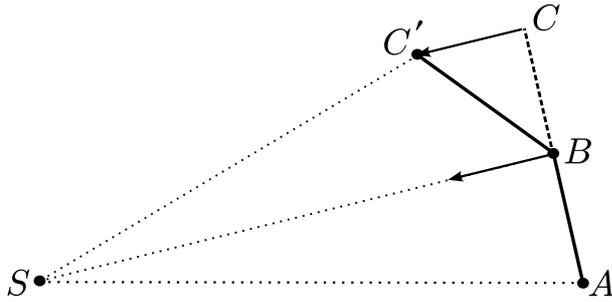
In the space of imagination, start a planet in motion around the Sun, whose position we'll call S . The planet starts at A and moves in a certain period of time, let's say one week, to a new position, B . If no force were acting upon the planet, it would continue to move in a straight line without perturbation. One week after the planet is at B , we'd find it at C , a point the same distance from B as B is from A . All three points would fall in a nice, straight line.

However, the Sun is pulling on the planet. To approximate this effect, we'll suppose that the Sun's influence applies an impulse to the planet at point B , the middle position. Once we've figured out the consequences of this impulse, we can repeat the analysis with a smaller and smaller time step, approaching the limiting case of a continuous pull.

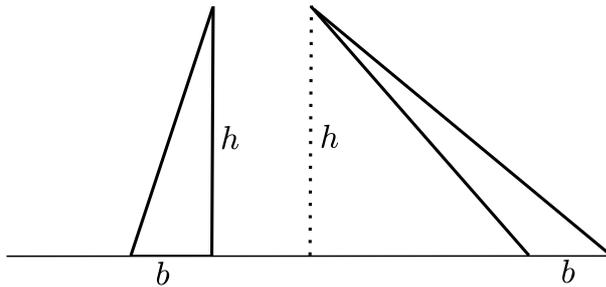
The effect of the Sun's attractive impulse is to pull the planet into a new course of motion. Instead of arriving at the point C one time step after passing through point B , the planet will go to a new point, C' . Where in space is this point? Newton's laws will tell us. C' is displaced from C by an amount which depends on the strength of the Sun's gravity, and in the direction along which the attractive force was applied. The force was applied at B , so the change from C to C' is in the same direction as the line from B to S .

Even if we don't know the *strength* of the Sun's gravitational pull, the mere fact that it is always directed inwards to the Sun tells us something interesting. First, we observe that in the case of *no* force, a certain quantity does not change over time. Consider the two triangles $\triangle SAB$ and $\triangle SBC$. We can show, readily enough, that they have the same area. Recall that the area of a triangle is half its base times its altitude,

$$a = \frac{1}{2}bh, \tag{2}$$



and that if the triangle is sheared over to the side, or “obtuse,” the altitude lies outside the triangle itself. This right triangle and this obtuse triangle have equal areas:

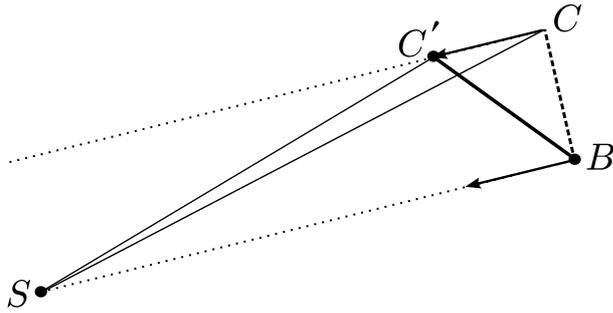


The two regions $\triangle SAB$ and $\triangle SBC$ have bases of equal length, because \overline{AB} is exactly as long as \overline{BC} . Remember, we’re talking about a planet moving at constant speed in a straight line, and the points at which we mark the planet’s path occur at equal time intervals apart (like once every week). $\triangle SAB$ and $\triangle SBC$ also have a common altitude: it’s the line from S which cuts across the line \overline{ABC} at right angles.

In equal times, then, areas of equal areas are swept out. That sounds like Kepler’s second law, but remember, we’re working in the case of no gravitational force. What happens when we add an attractive influence acting always towards the Sun? Well, consider the two triangles $\triangle SBC$ and $\triangle SBC'$. The former is the triangle swept out in the event of no force, and the latter is that shape swept out when the impulse is turned on. They have a common base, the line segment \overline{SB} , and they have the same altitude because they lie between parallel lines. Since the attraction is always “central,” always directed to the Sun, the segment $\overline{CC'}$ is parallel to \overline{SB} , and no matter where you draw an altitude cutting across two parallel lines, it always has the same length. So, $\triangle SBC$ and $\triangle SBC'$ have the same area.

This means, in turn, that $\triangle SAB$ and $\triangle SBC'$ also have the same area. In other words, the area swept out in each time step is always the same, when the force is towards the Sun. We can let the time step shrink smaller and smaller, and this relationship will always hold, down to the limiting case where the attractive impulse happens infinitely often and we have a smooth curve instead of a jumpy one with sharp corners.

We have deduced that when the attractive force is directed towards the Sun, equal areas are swept out in equal times. If we run our logic backwards, we can show that if equal areas are being swept out in equal times, then the force must be “central,” *i.e.*, directed Sun-wards. The two statements are equivalent. Both Hooke and Newton pointed this out,



and the proof sketched above appears in Newton's *Principia*.⁷ Richard Feynman picked it as an example of how statements are connected via reasoning; this essay has slavishly adopted Feynman's argument.

We've illustrated the problem using physics, and with the earliest stage of modern physics, to boot. It is perhaps most obvious in physics, but the same issue can arise in other fields, even biology. In the study of evolution, different verbal descriptions of a phenomenon can converge to the same place once they are treated with mathematical honesty. Demonstrating this requires dealing with genetics and statistics both, which means that we'd be talking about something like *covariances* instead of areas. The prerequisites are harder, but the upshot is the same. This may seem like an abstract point, but the area in question is the evolution of social behaviors like cooperation and altruism,⁸ an area of some small philosophical interest. If we lack the necessary background knowledge — which includes differential equations, game theory and, to be complete, perhaps matrix algebra as well — then the literature will be incomprehensible,⁹ and if we are not familiar with manipulating mathematical statements, then the process of exploring the issue will be largely obscured.

We could break here and expound the implications of this problem, that without mathematical reasoning one is screwed out of understanding the science, but that exposition will be a gloomy one, and it seems only fair to put it off as long as possible. We have explored a little bit of classical physics, and after understanding a little, the temptation is always to grab for more. Therefore, to illuminate the place we have staked out in science, we shall head out in a few directions through the "Mathiverse," navigating our way recklessly from our Newtonian starting point before returning to take stock of our discoveries.

4 Angular Momentum

We deduced that for a body moving under a central force, the rate of change of the area swept out by its motion does not change. This turns out to be a special case of a more general principle called the *conservation of angular momentum*. Here's the idea:

Suppose we wanted to talk about many planets all orbiting the same star. What would be the analogue of this rate-of-area-change proposition? Well, the easiest thing to do is to

⁷See <http://books.google.com/books?id=XJwx0lnKv0gC&pg=PA37>.

⁸See <http://www.sunclipse.org/?p=680>.

⁹Some relevant papers are listed and discussed at <http://www.sunclipse.org/?p=685>.

draw lines from the star to each planet, find the area swept out by each one, and add up all the rates of change. This is, almost, the angular momentum. The extra wrinkles are that, first, we ditch the factor of $\frac{1}{2}$ which comes from the triangle area formula, and more importantly, we make the heavier planets count more strongly. If a planet is twice as massive, the changing area it sweeps out contributes twice as much to the total angular momentum. So, the way to calculate the angular momentum around a point in space is this: draw lines from that point to each object, find the rate of change of the area swept out by each swinging line, multiply that by twice the mass of the object, and add all the figures up. If this total quantity does not change, then we say in physics jargon that it is *conserved*, and the system as a whole obeys the conservation of angular momentum.

5 Symmetries and Lagrangian Mechanics

To better understand this idea of quantities which remain constant during a system's development, physicists turn to a mathematical description of motion which treats the entire history of a system — say, a planet's entire orbit around a star — as a single item. Instead of the Newtonian approach, where we find the force on a body at one instant, see where it goes in the next instant, find the force at *that* position and so on, we switch to the “Lagrangian” approach, where we calculate the whole path at one go. This is quite a different attitude to take, but mathematically, it is *provably equivalent* to the Newtonian scheme. The differences lie in which approach is more convenient for a particular problem, and in the new ideas which each approach generates. Once again, disparate statements turn out to be linked, inseparably joined, through a connection we have to appreciate by starting with symbolic manipulations.

You may have heard of the rule in optics that light takes the quickest possible path from A to B . Lagrangian mechanics uses a similar idea, but instead of treating light, we apply it to solid objects. Suppose we have, again, a planet orbiting a star, and we know the planet starts at A . We also know that at some later time, the planet is found at Z . What course will it take between them? The Lagrangian method is to assign a *number* to *each possible path* the planet can take from A to Z in that time interval. This number is called the *action*, and it depends upon the influences at work in the system — in this case, the gravitational effect which star and planet have upon each other. The route the planet actually takes, the *physical path*, is the one for which the action has a certain property: that for any small perturbation of the path, the action does not significantly change.

This sounds like a completely loopy idea. How can a planet “consider” all the possible paths it could take, “smell” the one with the right action, and “decide” to go that way instead of any other? Yet, this approach really works. In a sense, it is “more right” than the Newtonian original, since Newton's laws have been superseded by Einstein's relativity and the principles of quantum mechanics, yet Lagrangian mechanics adapts quite well to relativistic regimes, and esoteric quantum field theories are often defined in terms of an action equation.

One of the most delicious fruits of Lagrangian mechanics is the relation it shows between

conservation laws and *symmetry*. In daily life, we speak of people’s faces being roughly symmetrical, since what is on the left looks like what is on the right; an orange can be rotated around its axis and look basically the same; a snowflake or a starfish can be rotated any one of several discrete amounts and look the same as before the rotation. The mathematical statement of symmetry takes this intuition and runs with it: an entity is symmetrical if it can be transformed in some way and appear the same after the transformation as it did before. We can speak of the symmetries of *objects* — crystals, Platonic solids and such — but we can also apply the changeless-after-transformation idea to more “lofty” things, such as physical laws themselves. How could *that* possibly work? Well, recall our statement about inertia: uniform motion at a constant speed in a steady line is just as good as standing still, or in other words, there is *no* unique state of stillness. If Joe and Moe are floating past each other in deep space, neither can tell who is “really” moving: each one perceives himself as staying still and the other as moving past. Neither Joe nor Moe can perform an experiment or take any measurement to tell who is absolutely, definitively stationary. There is no Great Eye of Creation toward or away from which motion can be judged, next to whom we can be truly motionless. This is, in fact, a statement of *symmetry*: we can exchange Joe’s vantage point for Moe’s, which is moving *relative* to Joe’s, without having to change our basic descriptions of physical law.

This kind of symmetry was already implicit in Newtonian mechanics. If, however, you add the postulate that the *speed of light* is one of those physical laws which Joe and Moe must agree upon, you take yourself beyond Newton, into the realm of Einstein. It was Einstein’s Special Theory of Relativity, published in 1905, which really got physicists thinking about symmetry and developing advanced mathematical tools to deal with the ways which physical laws could be symmetrical. Ten years later, Emmy Noether (1882–1935) made a remarkable discovery by looking at symmetry in the context of Lagrangian mechanics. She found that if you have a system described by some action equation, *for every symmetry of that action, the system obeys a conservation law*. This result is nowadays known as Noether’s Theorem.¹⁰

We mentioned that Lagrangian mechanics is useful because it leads to different ideas when you’re trying to generalize your knowledge and invent new theories. Here’s an example: so far, we’ve been talking about *pointlike masses*. We’ve modeled stars and planets as zero-dimensional objects with no physical size. What if we wanted to describe the motion of *one-dimensional* objects, entities with some extent to them, things which can wiggle as they move? Let’s say we want the fundamental bodies in our model to be 1D instead of 0D, and we want them to obey the principles of special relativity (after we’ve figured that out, we’d like to add quantum mechanics to the mix, too). It turns out to be much easier to use a Lagrangian approach when upping the dimension like this than to start with Newton’s laws. For various reasons, physicists started poking around with relativistic 1D wiggly things back in the 1970s — and that’s where string theory came from.

¹⁰John Baez has an introduction to Noether’s Theorem which should be accessible to physics undergraduates, at <http://math.ucr.edu/home/baez/noether.html>.

6 The Laplace-Runge-Lenz Vector

There exists another quantity of interest which is also conserved, but only for the special case that the central force varies inversely as the square of the distance. This quantity has both a size and a direction; in mathematical terminology, it is a *vector*,¹¹ a type of entity we often represent with an arrow. We can think of this particular vector as an arrow from the Sun to the point of the orbit closest to the Sun, called the *perihelion*. The Swiss mathematician Jakob Hermann (1678–1733) was the first to consider this quantity, publishing in 1710 a proof that the length and direction of this arrow did not change if the gravitational force were an inverse-square law. Johann Bernoulli (1667–1748), member of a great mathematical family, generalized Hermann’s work. Oddly enough, however, the perihelion arrow is not named for either of these two men, but for people who came to the problem later: it is now known as the *Laplace-Runge-Lenz vector*, or the *LRL vector* for short.

With a certain amount of algebra, it is possible to prove that if the LRL vector is conserved — that is, if it changes neither its size nor its direction over time — then the planet’s orbit is a conic section.¹² If the planet doesn’t have enough energy to escape the Sun’s gravity, then the orbit will be an ellipse; orbiting bodies with more energy follow parabolas or hyperbolas. Hooke, Newton and Christopher Wren (1632–1723) — yes, the fellow who designed St. Paul’s — each deduced that an inverse-square law of gravity yields elliptical orbits, though they lived too early to use the LRL vector in their proofs. In fact, the method Hooke used to deduce this result is not known.

As indicated above, if the attractive force pulling on the planet does not follow an inverse-square relationship, then the LRL vector is not conserved. The arrow from Sun to perihelion will drift over time, changing the place to which it points. The perihelion, the point of closest approach, will *precess*, so that the orbit describes a “spirograph” kind of pattern.

So far, we’ve only considered the gravitational pull of the Sun on its planets, but planets have mass too. Messrs. Hooke and Newton tell us that *all* massive objects pull on one another: for each force which the Sun applies to an orbiting planet, that planet pulls back on the Sun. Because the Sun is far more massive than any of the planets going around it, we were able to neglect its wiggling and wobbling; however, careful telescopic observation of other stars has been able to detect the wiggles in their motion¹³ caused by planets in orbit around them. Thanks largely to this technique, we now know of more than twenty times as many planets outside our solar system than in! Planets also attract one another. In the nineteenth century, it was observed that the position of the planet Uranus was not exactly what it should be, given the tugs of Jupiter and Saturn nudging it this way and that. John Couch Adams (1819–1892) and Urbain Jean Joseph Le Verrier (1811–1877) independently figured that this could be explained by another planet, orbiting out beyond Uranus. Both men calculated the unknown planet’s position, and when Le Verrier convinced an observatory to point a telescope to that location, Neptune was there, waiting to be found.

In 1859, Le Verrier published a similar calculation for the orbit of Mercury, pointing out

¹¹I wrote an introduction to vectors in a blog post, <http://www.sunclipse.org/?p=148>.

¹²See <http://www.sunclipse.org/?p=709>.

¹³See, *e.g.*, http://planetquest.jpl.nasa.gov/science/finding_planets.cfm.

that Mercury’s perihelion was precessing just slightly faster than the nudges of the other planets could account for. He worked out the discrepancy as 38 seconds of arc per century, and a later calculation refined the figure to 43 arc-seconds per century. To get an idea of how small this is, recall that a circle is divided into 360 degrees, each degree is divided into 60 minutes, and each minute of arc is equal to 60 seconds, so at 43 arc-seconds per century, you’d have to wait almost 8400 years to see a change of one full degree. But, tiny as it was, Newton’s laws and the known planets couldn’t account for it, so Le Verrier postulated that yet another unknown planet was fudging the works. (It worked the last time!) He hypothesized that the new planet, which he called Vulcan, circled the Sun within Mercury’s orbit.

Trying to spot a planet which would almost always be lost in the glare of the Sun is not easy, but try the astronomers did. Decades of searching turned up nothing. By the close of the nineteenth century, people were willing to suggest that Vulcan did not exist, and Simon Newcomb (1835–1909) proposed that a small deviation from the inverse-square gravity law could explain the anomaly. In 1915, Einstein’s General Theory of Relativity gave the world a better, stronger, harder explanation of gravity, which turned out to handle the problem nicely. (Later, it became possible to test general relativity in the laboratory, without having to squint at telescopic measurements complicated by the pulls of all the planets, but that’s a story for another day.) Einstein’s equations yielded the slight difference from the inverse-square dependence necessary to explain why the LRL vector failed, by the tiniest of smidgins, to be conserved.

The moral, we could say, is this: it’s easiest to explain a discrepancy by introducing new pieces which obey the familiar rules, but when that fails, we have to consider the chance that Nature is playing a different game. The same story played itself out once more, when TwenCen astronomers watching the galaxies found that their motions didn’t make sense. Galaxies seemed to be spinning too fast to hold themselves together, and the movement of multiple galaxies relative to each other weren’t as expected, either. One proposal had it that extra matter, “dark matter” not visible to telescopes, was exerting gravitational influence according to Einstein’s good old general relativity; others suggested that the gravity law would itself have to be modified. You have to admit, finding a better gravity law and out-Einsteining Einstein would be pretty darn cool. Nevertheless, the Universe is not constrained by our prospects for career advancement, and in August 2006, astronomers using the Chandra X-ray telescope satellite found direct evidence of dark matter in the intergalactic formation known as the Bullet Cluster.¹⁴

7 Releasing the Verbal Gerbils

Having illustrated our thesis with an example, we ask what it means. As physicist Joe Polchinski says,

¹⁴The researchers made their results openly available at <http://arxiv.org/abs/astro-ph/0608407>; a press release can be found at http://chandra.harvard.edu/press/06_releases/press_082106.html and an informative blog post at <http://cosmicvariance.com/2006/08/21/dark-matter-exists/>.

This process of translation of an idea from words to calculation will be familiar to any theoretical physicist. It is often the hardest part of a problem, and the point where the greatest creativity enters. Many word-ideas die quickly at this point, or are transmuted or sharpened.¹⁵

Polchinski is speaking about the standards one must maintain while *doing* science, but similar concerns apply to the process of *explaining* science. John Armstrong addresses the same question from the opposite direction: according to Polchinski, going from words to equations is the hard part of getting work done, while Armstrong points out that when making our explanation “kind to the lay-folk”, *that’s the very step we omit!*¹⁶ Specifically,

To avoid the mathematics in its entirety cuts the legs out from under any popularization of physics, and risks becoming *The Tao of Physics* or *The Dancing Wu Li Masters*.¹⁷

Thanks to the non-verbal reasoning which provides the connective tissue of physics, one cannot reason about physical science by thinking about the everyday meanings of words. Alan Sokal, a physicist whose name became a verb after he pulled an academic prank, tells the story of a friend, intelligent and well-read, who was quite puzzled by quantum mechanics. How could it be, the friend asked Sokal, that quantum mechanics says the world was both *discontinuous* and *interconnected*? On the face of it, these do sound like contradictory properties. We could probably construct a contorted mass of verbiage justifying how they are in perfect accord — theologians have spent centuries learning how to “resolve” worse puzzles than that! — but the *real* answer is that *as physicists use these terms*, they have precise, technical meanings. Each item of physicists’ jargon refers back to a conceptual web of equations, theorems and classic experiments, and interpreted using these definitions, those descriptions of quantum phenomena are entirely consistent.

(My best guess is that the poor friend had heard about “entanglement”, which led to “interconnectedness”, while also receiving some garbled notion of the fact that every time you peek at a system like an atom, you find it in one of a set of states, often with no intermediate possibilities allowed.)

One cannot think about physics as if it were a novel whose great themes were “chaos”, “probability” and “uncertainty”! It would be like trying to compose music, knowing only the descriptions critics gave in reviews: “bright”, “rich”, “brooding”, “colorful” and so on.

What happens when we lack this sophistication and try to manipulate the physics knowledge we’ve received from second-hand sources using the only ways we know how, *i.e.*, anthropomorphic and verbal reasoning?

When traveling through the Internet, one often encounters crackpot physics,¹⁸ and after some familiarity with the genus, one begins to organize the specimens into species. I often find

¹⁵See <http://cosmicvariance.com/2007/05/21/guest-post-joe-polchinski-on-science-or-sociology>.

¹⁶See <http://unapologetic.wordpress.com/2007/05/18/publicizing-mathematics-and-physics>.

¹⁷See <http://www.sunclipse.org/?p=99#comment-717>.

¹⁸An entertaining treatment of its common features is the Crackpot Index, <http://math.ucr.edu/home/baez/crackpot.html>.

myself remarking upon a particular characteristic of pseudo-physics: the use of verbal instead of mathematical arguments. Instead of positing a premise and doing the math to deduce consequences of that premise, we get a whole pile of jargon, pulled from different sources and decorated with equations to disguise the fact that the “arguments” are essentially plays on words, with the whole thing bent towards bolstering some pre-established conclusion. (One traveler on the Web recalls a pseudoscientist blogger who “would actively go through my rebuttals to his quirky physics posts, pick out vocabulary he thought sounded cool, and use it in the following post.”¹⁹ What do you do in order to debunk this kind of nonsense? One option is to focus on a few highlights and show that they lack validity; however, addressing *all* the points may balloon out into a full physics course of its own.

An example of this failure mode is provided by a woman whom Richard Dawkins interviews in *The Enemies of Reason*, part two (at the 16:10 point, and following).²⁰ Manjir Samanta-Laughton begins by noting, correctly,²¹ that black holes have been found at the centers of galaxies. She then asserts, “Black holes have become a creative principle,” and claims that “chakras” inside our bodies are actually small black holes. How did we get from *A* to such a ludicrous *B*? Well, some years ago, it was suggested that matter falling towards a black hole and blasting out again as a high-energy jet could impact clouds of interstellar gas and trigger star formation.²² It would be fair to say that such a black hole promotes the creation of new stars. How does the reader receive such a report? Well, if we don’t understand any details of the physics, and we aren’t capable of doing the calculations or following the equations ourselves, we key off the familiar words — and thus, the supermassive gravity sink becomes a “creative principle.” Incidentally, past a certain point, when they grow *too* supermassive, black holes disrupt their galactic environment so much that new stars *can’t* form — perhaps their jets blast away the gas from which stars form, or perhaps they heat it enough that it can’t settle down and begin condensing into stars.²³ No doubt this is an argument for keeping our chakras on a modest diet.

Once, at PZ Myers’s science blog Pharyngula, we had a troll who, insisting that only “Theism” could solve the problem of dark matter, displayed some rather confused ideas about “energy” and “gravity.” Using the actual physics, energy is interchangeable with mass — that’s the content of Einstein’s famous equation $E = mc^2$ — and we’ve known since Newton that mass is the source of gravitational attraction. Of course, we’ve updated the Newtonian force law to the Einsteinian description of curved spacetime, but still, more energy means more mass which means a stronger gravitational pull. A sealed box of hot helium gas will actually pull *very slightly harder* on nearby objects than a box of the same size containing the same number of helium atoms at a lower temperature. Yet to our persistent Pharyngula

¹⁹See <http://www.sunclipse.org/?p=436#comment-25771>. A more ominous example is L. Ron Hubbard’s book *Dianetics* (1950), for which see <http://stlouisskepticalsociety.blogspot.com/2008/01/chapter-5-summary.html>.

²⁰This video can be viewed at <http://www.sunclipse.org/?p=244>.

²¹See astronomer and writer Phil Plait’s remarks at <http://www.badastronomy.com/bablog/2008/06/02/galaxies-wrap-their-black-holes-in-loving-arms/>.

²²See http://www.space.com/scienceastronomy/mystery_monday_050207.html.

²³See <http://www.nasa.gov/vision/universe/starsgalaxies/galex-20060823.html>.

commenter, “energy” and “gravity” were diametrically opposed concepts — because, I think, when you have “lots of energy” you’re not “weighed down” but instead “bouncing around the room.”

In short, inconsistencies which appear at the verbal level turn out to be illusory when you progress to the actual science.

8 Implications for Education

Witnessing this kind of mathematics in action — the establishment of equivalences between ideas — gives us a certain perspective on what we should be putting in schoolbooks and the lessons through which we must put teachers. The tools we’ve used in this essay are not elaborate or abstruse: areas of triangles, properties of parallel lines and so forth. I was taught this kind of geometry, officially, in high school; in Alabama, it was a graduation requirement, although I’m genuinely baffled why it took until high school to get there.

But pacing is not the only problem.

All too often, discussions of what should be done with American mathematics education polarize into a debate between, essentially, pushing pencils by rote and punching calculator buttons by rote. The issue touches all parents and students, yet our attempts to figure it out get nowhere. To put the matter bluntly, we leave students completely unprepared for the type of reasoning we have seen is critical for natural science — yet we sell mathematics as part of the curriculum partly because it’s important for understanding how the world works. We can change what we teach in a great many different ways, without ever delivering on that promise.

To use mathematics in the natural sciences, we first decide how we wish to represent some aspect of the world in mathematical form. We then take the diagrams and equations we’ve written and manipulate them according to logical rules, and in so doing, we try to make predictions about Nature, to anticipate what we’ll see in places we have not yet looked. If additional observations corroborate our expectations, then we’re on the right track. (It’s rarely so clean-cut as that — the process can spread across thousands of people and multiple generations of activity — but that’s the gist of it.) Several skill sets are involved: one must know how to idealize the world, and then how to work with that idealization. Remarkably enough, our schools fail to teach either skill.

“Word problems,” which should involve the translation from the messy world to abstractions and back again, are usually invented to showcase a particular kind of formula rather than to generate insight into the world or the translation process. “Proofs,” which should be the means by which new truths are discovered from familiar ones, have degenerated into worksheet exercises where one column is headed “Statement” and the other “Reason”; students trudge towards theorems of little applicability and vanishingly small intrinsic interest, starting from postulates the choice of which is entirely impenetrable to them and, frequently, to their teachers as well. Neither utility nor conceptual beauty survive the conversion to the schoolbook and the lesson plan.

Everyone loses.

Dwelling upon this thesis also leads us to a conclusion about the *vulgarisation* of physics, and of the other natural sciences where this issue matters. The trouble is most evident with TV, newspaper and magazine reporting, although the problem also applies to books and our saintly host of websites. Judging by what we have seen, the most honest advice we can give to a reader trying to gain a working knowledge of physics from these sources is the following:

Give up. Give up now.

Consider how we might explain any recent development in physics. For example, what if we got the contract to popularize the general idea of *gauge/gravity duality* or the specific case of the *AdS/CFT correspondence*? This is the area where ideas from string theory are coming closest to experimental reach, and has even excited people working in accessible subjects like superconductivity,²⁴ so quite plainly, it *matters*.²⁵ We might start by saying, “It’s a relationship between a gravity theory and a gauge theory —”

“Wait a minute. What’s a gravity theory?”

“That’s like Einstein’s General Theory of Relativity.” Einstein is a safe reference, you see.

“Er... ”

“OK, imagine a flexible rubber sheet covered in gridlines... ”

Piece by piece, using analogies consumer-tested to produce minimum misery, we convey something of what we mean by “gravity theory.”

“Um, all right. Now, what’s a gauge theory?”

Now, we’re outside the standard playbook. For some reason, popularizers haven’t worked as hard to explain gauge theories, so we don’t have off-the-shelf vulgarizations close at hand.²⁶ Nevertheless, we persevere: depending on the audience, we might start with nuclear physics, pointing out that the nuclear force between two neutrons is the same as that between two protons, and how this is a *symmetry* (change a thing around and it looks the same as before). We could, alternatively, begin with electrical circuits, talking about how if you have

²⁴See <http://www.sunclipse.org/?p=258>.

²⁵Aaron Bergman’s comments at <http://zippy.ph.utexas.edu/~abergman/Review.pdf> are interesting in this regard.

²⁶After this essay was first posted on the Internet, John Armstrong suggested the following:

“You remember how we’ve pointed out that the laws of physics have certain symmetries? Like if I do an experiment it doesn’t matter whether I do it today or tomorrow, or whether my lab bench faces north or southeast, or whether my lab is in New York or Shanghai? Good. Let’s consider the orientation just to be specific.

“So: if I take my entire experimental setup and turn the whole thing by some angle, it’s not going to make a difference in the results I get. In fact, for a large number of experiments (*e.g.*, polarization measurements) where the apparatus comes in a number of different pieces, you can rotate each piece by the same angle and get the same results.

“Now here’s where the insight comes in: if the parts are all separated from each other by large distances, they obviously can’t tell instantly when another part of them has been turned. That would be information, and relativity tells us that information travels at a finite speed. So the old rotational symmetry has to be ‘gauged’. That is, we need to think of rotating by a different angle at each point in spacetime. Gauge theories are mathematical models physicists have developed to cope with this sort of ‘gauged’ symmetry.”

a component like a resistor, what matters is the voltage change between its ends, not an “absolute voltage” at either end. Again, this is a kind of symmetry: we can describe the same circuit using different voltage values and still be describing the same thing.

Or, suppose we judge it important to explain the “C” in “CFT”, which stands for *conformal*. We might start our tour with a magnet being heated up and about to lose its magnetism, or with water being brought to a boil and how the boiling point changes as you put on extra atmospheric pressure.

By now, you’d be forgiven if you had no idea how these different illustrations are at all related. How could they be part of the same picture, portraying the same thing? But they’re all starting points for explaining what gauge theories are and why we’d care.

We feel shaky enough getting the idea of a gauge theory across, but remember, we were trying to popularize the *duality* between gauge theories and *gravity* theories. “OK,” we say, “now let’s imagine this black-hole-like thingy in hyperspace, whose dynamics we can describe in two different ways...” And the two shaky towers of analogy we have built, leaning awkwardly towards each other, totter and fall!

The situation is just what Feynman described in the 1960s, when, incidentally, gauge theories were getting their start in life:

The layman searches for book after book in the hope that he will avoid the complexities which ultimately set in, even with the best expositor of this type. He finds as he reads a generally increasing confusion: one complicated statement after another, one difficult-to-understand thing after another, all apparently disconnected from one another. It becomes obscure, and he hopes that maybe in some other book there is some explanation... The author almost made it — maybe another fellow will make it right.

When the connective tissue of the science is cut away, we lose our ability to understand how progress is being made. What matters is no longer the *facts of the science*, but the *story it is easiest to tell*. All the madness we saw in the early days of classical mechanics — all the false starts and partial insights, all the glimpses of truth and observations which only made sense in retrospect — is still with us. (A few things have changed: we’ve learned that astrology is bunk, for example, so physicists down on their luck no longer cast horoscopes for royalty, as Kepler did; they go to work on Wall Street instead.) Suppose a novel idea comes along — or, somewhere on the Internet, an idea is presented as novel — and somebody has to decide how to write a piece about it. When a typical statement by a physicist trying to sort out the mess might read, “Any embedding of your gauge group in either noncompact real form of E_8 will always give you a nonchiral fermion spectrum,” well... the temptation will always be to plump for the “social” angle and emphasize the *personalities* of the physicists involved. This leads you to classic failure modes like the “Oppressed Underdog,” David readying his slingshot at the Goliath of the scientific establishment.²⁷ My “typical statement” is a

²⁷I made a rough attempt to catalogue these failure modes at <http://www.sunclipse.org/?p=49>.

paraphrase of Jacques Distler,²⁸ who pointed out this problem almost six years ago;²⁹ it's still alive today, and kicking more than ever. We see it frequently enough in physics, but even biology is not immune,³⁰ when the science turns on relationships among propositions. You don't *need* intimidating jargon, though it helps: all you need is for the actual relationship between David's idea and Goliath's orthodoxy to be mathematical in nature.

In defending this thesis, I recognize that I'm leaving myself open to the accusation that I am an "elitist" — indeed, even an Elitist Bastard. How dare I suggest that science is not readily transparent,³¹ that the training one receives in climbing the steps of the ivory tower is necessary to enjoy the view from the top? I cannot refute this calumny: all the water in the Tennessee River cannot wash out an MIT degree or a year spent in France. Yet, if education is to mean anything, we must confront its problems as they stand. If we do not identify the obstacles in our path and work honestly to overcome them, then the process of education will be a useless charade, and the hope of a scientifically literate populace will be more distant than ever.

9 Readings

- Feynman, Richard (1965). *The Character of Physical Law*. The original purpose of this essay was to get one of Feynman's points out onto the Net without copyright restrictions.
- Sokal, Alan (2008). *Beyond the Hoax*. The other original purpose of this essay was to get people to buy Sokal's book.
- Goldstein, Herbert, Charles Poole and John Safko (2002). *Classical Mechanics: Third Edition*. Chapter 3 covers central-force problems, and the LRL vector is in section 3.9.
- Zwiebach, Barton (2004). *A First Course in String Theory*. Chapter 4 reviews Lagrangian mechanics, chapter 6 starts the business of generalizing 0D point particles to 1D strings using Lagrangian methods, chapter 8 is a good introduction to Noether's Theorem, and chapter 16 has a bit of AdS/CFT.

Various people have taken stabs at popular, semi-popular or undergraduate treatments of gauge/gravity duality:

²⁸Specifically, Distler was investigating the mathematics of Garrett Lisi's attempt to make a "Theory of Everything." After being irritated by the press coverage and the blogospheric treatment, Distler explained how he found the mathematics wanting, at <http://golem.ph.utexas.edu/~distler/blog/archives/001505.html>.

²⁹<http://golem.ph.utexas.edu/~distler/blog/archives/000033.html>.

³⁰For example, see the reporting which Michael White criticizes at http://www.scientificblogging.com/adaptive_complexity/bad_science_journalism_and_the_myth_of_the_oppressed_underdog.

³¹Isabel Lugo observes, "[I]n some provinces of the Internet there seems to be an idea that everything should be immediately understandable to all readers at first glance." She adds, "I use Reddit for the links it gives me to interesting news, not for the comments." See <http://godplaysdice.blogspot.com/2008/05/shoup-computational-introduction-to.html>.

- Johnson, Clifford (2007). “Exploring QCD in Cambridge,” Asymptotia. <http://asymptotia.com/2007/08/22/exploring-qcd-in-cambridge/>.
- Shock, Jonathan (2007). “The gauge/gravity duality — or how dimensions may not be what they seem,” SciTalks. <http://scitalks.wordpress.com/2007/09/08/the-gaugegravity-duality-or-how-dimensions-may-not-be-what-they-seem/>.
- Guarrera, David (2008). “Physics at Strong Coupling: Why String theory and supersymmetry are ridiculously cool,” podcast video. http://web.mit.edu/guarrera/www/Daves_Podcast/Podcast/Entries/2008/6/4_Physics_at_Strong_Coupling.html.

My second figure was redrawn after episode 22 of Caltech’s TV series *The Mechanical Universe*, available at <http://www.learner.org/resources/series42.html>.

This essay first appeared on my blog, Science After Sunclipse, at <http://www.sunclipse.org/?p=663>, and it is distributed under the Creative Commons Attribution Noncommercial Share-Alike 3.0 License.