The Cathedral of Physics

Sverker Johansson **ur: Larsson, H-A (red) *Vetenskapens ansikten* Jönköping University Press, 1997**

For many people, physics is something abstract and incomprehensible, taught in school by distant professors filling the blackboard with weird equations, that no ordinary mortal could ever aspire to understand. This is a caricature of traditional physics teaching, but a caricature with an embarrassingly large amount of truth in it. A fair number of physicists of the old school did indeed promulgate, consciously or not, an image of physics as something that only a select few men (don't even mention women!) were worthy of, a subject best studied with great reverence as a work of abstract art, a great cathedral of Science with a capital S.

For me, as a physicist, physics is neither abstract nor incomprehensible. On the contrary, physics comes alive only when it leaves the realm of abstraction, when it is realized that physics is actually about something, that the equations are not just abstract art, but actually depictions of the very world we live in. I can also feel reverence, and perceive a cathedral — but not the same one. My reverence is for the magnificent universe around us, as depicted by physics, and for the genius that went into the construction of our remarkably comprehensible[1] physics image of this universe.

Physics touches upon all aspects of the material world around, from the smallest atom to the largest galaxy, from the most mundane practical problems to the deepest philosophical musings about the meaning of it all. And it addresses all these diverse issues within a unified framework of ideas and concepts and basic principles that are within reach of anybody who takes the time and trouble to penetrate them. The study of physics should always start in the everyday world around us. But it needn't end there — often it will take us on a journey into quite different realms, where our mundane phenomena are revealed as surface manifestations of new, often subtle and counterintuitive, principles that nevertheless deepen our understanding of the world we live in.

In this essay, I would like to take you on a two such journeys through the world of physics, give a guided tour of a small part of *my* cathedral of physics, and hopefully let you experience some of the combined sense of wonder and comprehension that is the true goal of physics for me.

Falling until the end of time

”Things fall down.” Obvious and self-evident to everybody, and may well be a baby's first physics discovery. There's even fairly strong evidence that this knowledge is inscribed in our genes[2]. But how and, more important, why do things fall, and what is really going on inside this simple process that we call ”falling”? That question will be the point of departure for our first journey into the world of physics.

---

[1] Albert Einstein once said: ”The most incomprehensible thing about the universe is that it is comprehensible.”

[2] I'm referring here to the innate ”falling reflex” of a newborn baby, in which a baby which feels itself falling will reach out and grab anything handy, to arrest the fall. This reflex exists long before the baby has any apparent conscious awareness or experience of falling, and can reasonably be explained as of genetic origin, a product of our evolutionary history, where such a reflex gave a strong selective advantage to our arboreal ancestors.

From common sense to science

As adults, our concept of falling is somewhat more sophisticated than ”Things fall down”. We have experienced a wide variety of objects, including ourselves, falling under various circumstances, and have deduced some general principles from our observations. In making these deductions, we are doing physics research, even though we’re not calling it that. Nevertheless, that is where it starts, with observations, and conclusions from what is observed. Physics it is, but not quite science yet. We’ll start calling it science when the observations are more systematic, and, more importantly, when we, after having drawn our conclusions, submit those conclusions to further tests. To some extent, this is done informally even in our everyday observations of the world, but it is surprising how strongly cemented a conclusion can become, even in the face of considerable evidence to the contrary, unless we make a specific effort to unmask any and every flaw in our deductions, through tests and further tests. This refusal to settle down and accept our findings as the final truth, this willingness to search for new evidence, is what distinguishes science from other human activities. And the result of this refusal to accept any knowledge as final is, paradoxically, a body of knowledge much more solid and reliable than any other.

Back to falling: we ”know” from our own experience a fair bit about falling. We ”know” that some things, such as rocks, fall faster than others, such as feathers. We ”know” that some things don’t fall at all, but float upwards instead, like balloons, or like air bubbles in water, or like the moon in the sky. We ”know” that things hit the ground harder the longer they fall. We ”know” a whole lot of things like that, for which I have consistently put quotes around ”know”, because this knowledge of ours has not been subjected to the systematic scrutiny and testing needed to weed out subtly erroneous conclusions. For example, a common generalization from the observation that stones fall faster than feathers (correct), is the principle that heavier objects always fall faster than lighter objects (incorrect). This principle will survive a fair bit of normal life without leading to major mishaps, unless we stop to think about it, and consider the logical consequences of the principle.

Let us consider how we could examine this principle, that heavier objects fall faster than lighter objects. There are two different approaches open to us, either theoretical or experimental. In the theoretical approach, we try to deduce various logical consequences of the principle, and see if they lead to contradictions, or to unreasonable or otherwise unexpected behaviour. In the experimental approach, we try to devise direct tests of the principle, where we try it out on the purest possible examples, in this case comparing the falling speed of two object differing only in weight, but otherwise identical. This purification will make the results clearer, and less open to conflicting interpretations.

Theoretical considerations can lead us to thoughts like the following: suppose I drop some suitable rock. It will fall with some speed determined by its size. If I then cut the rock in half, and drop the two halves, what will happen? Obviously, since we now have two pieces of rock, each of them half the weight of the original rock, they will both fall at half the speed. OK, that’s possible... but what if I then connect the two rock pieces with a string? Will this count as one big rock, falling fast, or two smaller rocks, falling slowly? Or what if I simply hold the two rock pieces together, so that they touch along the plane where I cut them apart a minute ago, but are not in any way attached to each other. This will look exactly like the original rock, but with a crack through the middle. Will this combination fall as one large or two small rocks? Similar lines of reasoning can be pursued to even weirder consequences, clearly leading us to the intuitive feeling that our principle, that heavier objects fall faster, is not entirely satisfactory.

But the final arbiter is the experimental test. It is quite possible that the world we live in really is such a weird place, and that our intuitive feelings about falling rocks are unreliable. Unless we actually find solid logical contradictions, the theoretical deductions cannot on their own show that a principle is either right or wrong. What the theoretical analysis can do, however, is to point out weak spots in our ideas, and indicate which types of experiments may be useful as tests of the principle. So since bizarre things turn up in the theory when we consider half-rocks, it seems like a good idea to get a bunch of rocks, and start cutting them in half to see what happens when they fall.

OK, we did the experiment (though not with actual rocks, it's too much work to cut them; Lego blocks are easier to handle, and hurt less when they fall on your toes). Result: it doesn't make any significant difference if two blocks are dropped assembled or separate; our intuition was right, and the principle that heavier things fall faster appears to be wrong. If this is tried with a wide variety of objects, it will be found that most of the time the weight of the object will not make any difference. The exception is feathers and such, which do indeed fall slower than rocks or Legos, but any reasonably compact object will fall at the same speed. So our conclusion will have to be that whatever the rules for falling objects may be, they do not include any simple dependence on the weight of the object.

So the principle that heavier objects always fall faster than lighter objects can be quickly and easily demonstrated to be false; it can be *falsified*. Falsification is a very important concept in the logic of scientific discovery[3], much more important than *verification*, demonstrating that something is *true*. At first, this appears to be backwards — isn't it more interesting to find out what is true, than to find out what is false — but it turns out that the concept of verification is so full of logical traps and pitfalls that it is practically useless. Falsification is on much more solid ground, as our simple example shows.

From earthbound falls to universal forces

This art of systematically testing ideas, and eliminating weak ones through falsification, is basically what modern science is all about. It is somewhat similar to evolution through natural selection; the fittest ideas survive, and become part of what we call scientific knowledge. The method has been used by scientists at least since the 16th century, though its logic and philosophy wasn't properly analysed until the 20th. But the Greeks, working 2000 years earlier, did *not* in general submit their ideas to this kind of testing, and this led to many dubious notions surviving, even though they could have been falsified quite easily. The one discussed in the previous section, about heavier things falling faster, is a case in point; it was promulgated by none less than Aristotle, and remained unquestioned and untested until 2000 years after his death.

The actual falsification of Aristotle's "law of gravity" was done by Galileo in the late 16th century, pretty much as I've described it (though without Lego blocks). But Galileo wasn't satisfied just with eliminating an incorrect idea — he wanted to find a better one. So he kept on working, trying out different ideas, testing them and (usually) rapidly falsifying them. In this way he could figure out two important things about falling:

- The weight of an object doesn't matter at all as far as gravity alone is concerned. The apparent weight effect that Aristotle and others had noticed was due entirely to air resistance, which disproportionately affects light fluffy objects.
- Things do not fall at a constant speed (again in the absence of air), but at a constant rate of speed *change*, a constant *acceleration*

Today this main idea of Galileo, that air resistance camouflages much of what really goes on in a fall, can be tested without too much trouble simply by removing the air from a suitable chamber, and try dropping things in there. Lo and behold! The feather does indeed fall just as fast as the rock, in vacuum. This is a standard experiment, performed in just about every 7th-grade classroom, showing

---

[3] "*The Logic of Scientific Discovery*" is also the title of sir Karl Popper's book, where he analyses in considerable detail the philosophy behind the description of science that I have given.

quite clearly that Galileo was more nearly right than Aristotle. (Though it's far from always that the significance of the experiment is both clearly explained and clearly understood....)

Galileo's ideas about gravity are highly accurate descriptions of what goes on here on Earth, and are perfectly adequate even today for everyday applications near the ground. But what happens elsewhere, or is falling a purely earthbound concept? That question is not as silly as it seems, for until the days of Galileo the very concept of natural laws was limited to the Earth; the heavens were the domain of the gods, to whom the earthbound laws need not apply. Aristotle invoked an entirely new element as heavenly matter, apart from his usual four (earth, water, fire, air), called the fifth element (the ”quintessence”, a word which is used in modern English (or Swedish) in an entirely different sense), because surely the heavenly bodies couldn't be made of anything as mundane as earth, water, fire or air.

But the astronomical discoveries of the sixteenth and seventeenth centuries cast some doubt on the old notion of the heavens as an entirely separate domain. The Earth was shown to move around the sun just like the other planets, and those other planets, when seen through a telescope, turned out to be spherical bodies pretty much like the Earth. So why shouldn't the same natural laws apply up there? Shouldn't things fall down on the Moon, or Mars, or the Sun, just like they do here on Earth? Besides, what made the planets stay in their orbits. Why didn't the Moon simply fall down to the ground, to Earth?

Musings like these lay behind Isaac Newton's development of a new theory of falling. His started with the assumption that the same laws apply everywhere, and that things ”fall” (or are drawn towards) all heavenly bodies, in pretty much the same way as things fall down here on Earth. This means also that the heavenly bodies are drawn towards each other, which can be used to explain their orbital motion, as an eternal falling without ever getting down. Like this: if you just drop a rock, it will fall straight down and hit your toes. But if you throw it with some speed horizontally, it will move along the ground while falling, and hit the ground some distance away. Now apply this to the Moon; it moves around the Earth at some modest speed, carrying it all the way around in one month. With the sideways speed it has, it won't fall straight down from its present position, but move a bit sideways while falling, enough so that it will miss the Earth entirely, and end up ”beside” the Earth. From there, it can repeat the process, falling down but with enough sideways speed so that it will always miss the Earth. After a few ”falls” like that, the Moon is back where it started, having ”fallen” all the way around the Earth, falling all the time, but with just the right amount of sideways motion to cancel out the falling, so that it doesn't actually get closer to the ground.

Newton calculated just how strongly the Earth would have to pull at the Moon to achieve the desired effect. It turned out to be much less than the pull of gravity here at ground level, so apparently the strength of the Earth's pull becomes weaker further away from the ground. If it is assumed that the pull decreases as the square of the distance, everything comes out just right, not just the Moon, but also all the planetary orbits around the sun, and just about everything else in the heavens. Newton's simple formula gave a highly accurate description of all heavenly motions known at the time, and works perfectly fine for our satellites and space shuttles today, which ”fall” around the Earth exactly like the Moon does.

But Newton's law of gravity wasn't supposed to apply only to planets and stars, but to *everything*. All bodies, all forms of matter, attract all other bodies, according to the same formula. This isn't observable in everyday life, since the Earth completely dominates all the gravitational pulling done around here. In principle, the book (or screen) that you're reading this text from attracts you by gravitational pull, but the pull is way too small to be noticeable. Newton *predicted* that all these tiny gravitational forces between objects existed, but they couldn't be measured until 100 years after his death. When they finally did, it was a nice confirmation of his theory of universal gravitation.

Curving space and closing time

The final fate of Newton's theory of gravity is a nice illustration of the process of progress through falsification discussed above. It is both a lesson in humility, and an example of the strength of the scientific process. During two centuries after Newton's death, his theory was the unchallenged Truth about gravity and falling, gaining strength as new phenomena were predicted, discovered, and explained. No theory had ever been more strongly verified, and no one seriously doubted that this was actually the correct explanation of falling. No one, that is, until the end of the 19th century, when a collection of miscellaneous bothersome facts started to make physics interesting again... Not all of the trouble was over gravitation, much of it seemed completely unrelated (and some of it was indeed unrelated to gravitation; we'll come back to that in the next chapter), but all of it challenged the Newtonian hegemony, potentially falsifying the very foundations of the physical sciences. By the year 1900, it was becoming pretty obvious that something was seriously wrong with the theories of that time[4].

The only problem directly associated with gravitation, was some minor anomalies in the orbit of the planet Mercury. As it moved around the Sun, it didn't return to exactly the same place after one circuit, as Newton predicted[5], but missed its destination by a kilometer or so, after a 150-million-km journey. Not much, one may think, but in comparison with the precision of astronomical measurements even then, and with the accuracy of theoretical calculations, it was a glaring error.

The key to the Mercurial problem turned out to be the behaviour of light, up until then considered to have no connection whatsoever with gravity. Light was another troublesome area in physics, to which we'll return below, involved in a variety of anomalous observations. The anomaly that concerns us here is that the speed of light in vacuum was constant.... *too* constant. There's nothing strange about something keeping a constant speed, normally. But we do expect the relative speeds of different things in motion to follow certain rules. For example, if a car is going at 100 km/h down the highway, we expect to see it pass us very slowly if we're doing 95 km/h ourselves in the same direction — and conversely, it'll zip by at twice the speed relative to us if we're going in the opposite direction. Constant speed with respect to the road, and consistent relative speeds with respect to other cars, that's what all our experience leads us to expect. But light rays don't behave like that! Their speed is constant, always. No matter how the light source moves, no matter how *you* move with respect to the lamp and light ray, you will always perceive the light moving past you at the same speed relative to *you*. The two cars meeting on the highway will both see the light leave their own headlights with the same speed as the light arriving from the other car's headlights, and never mind that the cars and headlights are moving at 200 km/h relative to each other. The two drivers may measure the speed of the same lightray at the same moment, and measure the *same* speed, each one relative to himself. Using the normal rules for figuring speeds, they can both calculate how the light moves relative to the road — and get completely different results!

There's no way to make sense of this crazy behaviour of light within the framework of Newtonian physics. But it took a great genius, Albert Einstein, to break through our deeply rooted intuitive feelings about the basic nature of the world we live in, and realize which key concepts had to be modified. The first victim was *time*. For most of us, time is something that simply passes, one second per second, always and everywhere at the same speed. The very notion of time running differently for different observers goes against all our experience. But an absolute "speed of time" could not be reconciled with

---

[4] For those of you who studied physics in high school, it may be of interest that 90% of what you were taught then, what many high school teachers would present as gospel truth, was based on these theories, which were in the process of being falsified 100 years ago.

[5] This is a considerable simplification, neglecting the influence the different planets have on each other, but that doesn't change the thrust of the argument: there was a deviation, not only from an exact circuit, but also from the full-blown Newtonian prediction.

the observed absolute speed of light, and when sufficiently accurate measurements could be made, it did indeed turn out that time passed at different rates under different circumstances[6]. Under reasonable circumstances, the difference is too small to be noticed, but it is real.

Space, and distances, are affected in a similar manner. The ”same” distance, between the same two points, is perceived by different observers as being of different length. Again, the difference is too small to be noticed in everyday life, and again the notion is intuitively perceived as utter lunacy, but it is indeed real when accurately measured.

Furthermore, Einstein united his new, modified concepts of space and time, into the single concept of *space-time*. What we perceive as space and time, separately, are simply different reflections of the underlying space-time. Space-time is universal, the same for all observers, but different observers, moving at different speeds, will perceive different portions of space-time as space, or time, explaining the differing views of space and time.

So far so good... but so far nothing relevant to gravitation. Einstein’s *special relativity*, as the theory described so far is called, concerns the behaviour of space, time, and light, for observers moving at different speeds[7]. But what happens if an observer *changes* her speed — accelerates? The inclusion of accelerated motion into the theory required some further modifications of space-time, entailing even worse perversions of our everyday notions of time and distance, and even basic geometry.

Einstein started his work on accelerated motion by noting that the effects of acceleration are very similar to those of gravitation. Our perception of being pushed backwards in our seat when a car accelerates, is practically indistinguishable from our perception of being pushed down in our seat by gravitation. This led him to build a new theory around the assumption that gravitation and acceleration are *truly* indistinguishable inside the car — postulating that you can’t tell by any means whatsoever, without looking outside, whether you’re pushed backwards in your seat because the car accelerates, or because a second Earth has sneaked up behind you, adding its gravitational pull to the regular planet’s.

From this assumption of equivalence between gravitation and acceleration, a number of interesting consequences can be derived. By contemplating the effect of the car’s acceleration upon a lightray passing through the car, it can be deduced that light falls down when exposed to gravity, just like other objects, which is not so remarkable. But exploiting the falling of light, one can also show that time is affected, not just in an interchange between time and space as in special relativity above, but in a more absolute sense. Time actually moves slower for an observer who is either accelerating, or feeling the pull of gravity.

But it should be emphasized here that it is not *falling* under gravity that it is equivalent with acceleration; it is the *perception* of being pulled down, and pushed into the seat or the ground or whatever happens to be beneath you, that feels like acceleration. While actually falling you do not feel accelerated; in fact you feel just like you would floating freely in empty space, far away from any source of gravitation. Something funny is going on here... while falling freely downwards here on Earth, you accelerate with 9.81 m/s^2, at least that is what you were taught in school. But that’s acceleration, right, and acceleration is supposed to feel like being pushed down by gravity but *not*

---

[6] An elegant experiment demonstrating this difference was performed in the 1970s: to put it simply, two very accurate clocks were synchronised, after which one of the clocks was sent on a trip around the world in an airplane, the other one staying behind at the airport. According to Einstein’s theory, the clocks ought to differ by about a ten-millionth of a second when they were rejoined. And indeed they did!

[7] It should be noted that the speed of normal physical objects is never absolute; it has to be measured with respect to some frame of reference, and the choice of reference is completely arbitrary — there is no such thing as a universal reference, unmoving in any absolute sense. This is true of Newtonian physics as well as Einsteinian physics (even though Newton remained philosophically wedded to the concept of an absolute space).

falling. So, falling should be the equivalent of not falling, according to this logic. But if you jump out the window you'll quickly find out the difference. "Falling doesn't hurt, it's reaching the ground that hurts!", as the saying goes.

Of course Einstein had a solution to this as well. But it required a redefinition of acceleration, as well as a complete reconstruction of our image of gravitation. In Newton's view, gravitation was a force that every body in the universe exerted upon every other body. Nothing was said about how the force was formed, or how it got from the source-body to the body that was being pulled. Einstein turned this view completely upside down, abolishing the very concept of a *force* of gravity. Instead, Einstein's concept of space-time, introduced to clear up the mess with the speed of light, was modified to handle gravitation as well.

Until this time (and still today, for most of us), space and time, as well as their union in space-time, didn't really participate in what went on in the universe, they were just a background. Space-time didn't *do* anything, it just *was*. A theatre might be a suitable analogy, where the actors were all the physical objects of the universe, but space-time was simply the stage on which the play was performed. In Einstein's new version, however, space-time is no mere stage any longer, but plays a major part in the action. Physical objects affect space-time, and space-time affects objects, and this interaction between space-time and things is what we perceive as gravitation.

In order to understand the way in which space-time participates, an excursion into geometry is required. Geometry, the branch of mathematics dealing with lines and circles and triangles and stuff, is like all the rest of mathematics based on a small number of *axioms*, supposed to be self-evidently true. Everything else is then derived through solid deductive reasoning from these axioms, meaning that *if* the axioms are true, *then* all other geometry is true as well. But what if they're not? All the axioms do appear self-evident to us, but our notion of an absolute time appeared self-evident as well, until we found out that it's false... It is found that the axioms of geometry are true *only* in the old non-participating space-time — "real" geometry (as opposed to pure abstract mathematics) is a fundamental property of space-time, and if space-time interacts with matter, then the axioms of geometry have to be replaced, and all the geometry we learnt in school has to be modified accordingly. The consequences are minor in practice — the sum of the angles of a triangle isn't *exactly* 180^o anymore, but the difference is too small to measure unless the triangle is of astronomical size — but philosophically they are at least as revolutionary as the variable speed of time.

The interaction between matter and space-time is roughly as follows:

- Any physical object affects the geometry of the surrounding space-time.
- The heavier the object, the larger the departure from "classical" geometry in its neighbourhood.
- Objects moving through space follow the geometry of the local space-time, modified as it may be by other objects.
- Non-accelerated motion means following the equivalent of a straight line in this new funny geometry (note that such a line is *not* what we would intuitively perceive as straight).
- Accelerated motion means departing from that "straight line", because of an external force.
- The geometrical modifications effected by a heavy object is such that "straight lines" generally go downwards, towards the heavy object.
- Following such a line downwards is what we call falling, which is thus *non*-accelerated motion, whereas what we perceive as standing still on the surface of the Earth is a departure from our line downwards, and thus *acceleration*. Hence the equivalence between acceleration and gravitation without falling.

These deformations of the geometry of space-time are frequently referred to as "curved space", which is not a bad analogy, as the motion of an object in modified space-time is in many respects similar to the motion of a ball rolling on a curved surface. But the analogy is not perfect, and we must take into account that it is space-*time* that is "curved", both the usual three spatial dimensions, and time.

Furthermore, we must not forgot Einstein's old special relativity; space and time are still exchangeable for different observers in the manner it describes. All in all, this makes full-blown curved space-time nearly impossible to visualize. The basic principles are simple enough, if counterintuitive, but the applications are *not*. As long as the geometry of space-time isn't modified too much, Newtonian gravity is still a good approximation, even though it has been falsified. But the geometry *is* modified "too much" in quite a few places in the universe, where there are nearby heavy objects (such as our sun — remember Mercury), and in those places the errors in Newton are obvious. It has been amply confirmed, both in laboratory experiments and in astronomical observations, that Einstein's curved space is a much better "theory of falling".

Einstein's theory is, in fact, perfectly adequate as an explanation of everything we know about falling and gravitation, and in the 80 years since it was published we haven't found any contrary evidence. But when the theory was applied in other areas, unrelated to falling, some interesting consequences turned up. One area, which we shall not pursue here, is the question of what happens when the curvature of space becomes very large. Black holes belong in this context, but there's even weirder stuff apparently permitted by Einstein's theory; space-time may be curved in such a fashion that it is effectively a time machine, or curved so that two parts of space-time, very distant in space and/or time become connected — in effect, a shortcut through the universe.

Another interesting area where we can apply Einstein's theory of space-time is the universe as a whole. The universe — all the space and time there is — can be regarded as one single huge, possibly infinite, piece of space-time, subject to Einstein's equations. What can we say about the *general* geometry of the universe, apart from all the local deviations caused by planets and stars and galaxies?

One thing we can say about the universe, which Einstein himself at first refused to accept, is that as a piece of space-time it cannot be stable. If it is not "doing" anything to begin with, the matter in the universe will cause space-time to curve, making all matter move closer together, making space-time curve even more, making matter move even closer, making ..., and so on until it collapses completely, curled up into an infinitesimally tiny little ball. The only way out of this collapse is to have space-time expand instead, stretching and growing in all directions, so that new space is added faster than the matter can curl it up. So we can have either expansion or contraction, but not stability.

Until that time, it had been implicitly assumed in astronomy that space *was* stable and pretty much eternal. Interestingly enough, however, in the same decade, the 1920s, as the discovery of the instability of Einsteinian space-time, evidence was found in outer space that the universe was indeed expanding; distant galaxies were moving away from us in just the pattern expected in an expanding universe. Further astronomical observations have lent strong support to the notion of an expanding universe, and there is a very nearly total consensus among astronomers today that we live in a universe that is currently expanding. Furthermore, if it is expanding, that means it used to be smaller than it is, and the evidence indicates that once upon a time, some 10-20 billions of years ago, it was very small indeed, probably pointlike and certainly less than pocket-size —instead of all the uncountable trillions of miles of empty space we have today, our piece of space-time was curled up in a package that would fit in your pocket with room to spare.

The universe today is very large; it is an unresolved question whether it is infinitely large or not. With our traditional notion of space, we would expect the universe to be infinite, since it would otherwise require an edge. But with Einstein's curved space-time the universe can have a finite size without an edge, since it is quite possible for space-time to curve around and meet itself — not just possible, but natural and expected if there is enough matter in it to make it curve that much. The situation is much the same as it was for old-time sailors, who were afraid to venture too far out in the ocean, for fear of going over the edge of their flat world[8] — but a round Earth, like the one we actually live on, with a curved surface, is safe with no edges, even though it has a finite size. If the universe is similarly finite, one could, in principle, circumnavigate it with a spaceship, going one way and coming back from the opposite direction. In practice, however, one could not, and not just because we can't build adequate spaceships; we must not forget the time part of space-time. If the universe is curved around and closed in space, then it is also similarly closed in time, at least in Einstein's basic theory. A universe with a finite size also has a finite lifetime. It starts out small at the beginning of time, grows larger for a while (billions of years in our case), and then reverts, starts shrinking again, until it is back where it started, curled up in a tiny ball, the end of space, and the end of time.

Whether time and space are actually closed and finite, depends on if space-time is curved enough to meet itself, which in its turn depends on how much matter is around to make it curve. The weight (or more properly density) of the universe determines it size — the heavier it is, the smaller it is in space and time. And if it is not heavy enough to curl up, then it is infinite in space and time, going on forever in all directions. We have a fairly good estimate of the density of our universe today, and it turns out to be remarkably close to the critical limit, where it is just barely light enough to be infinite — or possibly just barely heavy enough to be finite. It is so close that we can't tell which way it'll be. Further investigations of the weight of the universe, and the way things fall out there, will likely bring us the answer, but we're not there yet.

And that the brings us to the end of our journey through the world of falling. It has brought us a long way — from contemplations of what happens when you drop a rock on your toes, we reach conclusions about the very nature of the universe, and touch upon the end of the world. Starting with the most humble and commonplace phenomena around us, physics can bring us to a deep understanding of the subtler nature of this magnificent world we live in.

The light outside Plato's cave

Our next journey starts with light, ordinary light from a lamp or from a sun. What is the nature of this thing we call light? As with gravity, the Greeks had considered the question, but didn't really get very far beyond clothing our everyday experience in philosophical language. And, again not coincidentally as with gravity, serious scientific investigation of this issue began in the 16th and 17th century. Newton is a prominent name here as well, spending at least as much time on optics (or Opticks, as he called it) as on gravity, with detailed investigations of a number of optical issues. All the stuff we learn in school about the behaviour of light, how it behaves in lenses and mirrors, how to split up a sunray in different colors, and so on, was pretty much established by the year 1700. But this still doesn't tell us very much about the nature of light. We learn what light *does*, not what it *is*. Of course, there was plenty of speculation, but precious little solid evidence one way or the other in those days. From the 17th-century speculations, two main hypotheses concerning the nature of light emerged:

- Particles: Light is composed of small particles, like tiny little dustgrains, made of some special ”light-substance” but nevertheless basically material objects. A lightray is simply a stream of such particles. Different particles have different colors, giving the light its color.

[8] A popular, but false, myth is that Columbus was the first to realize that the world was round, and that India could be reached across the Atlantic, and that others of that time expected him to go over the edge of the world. That the world is round was known by the Greeks 2000 years earlier; at the time of Columbus no eductated person thought the world flat. The size of the world was also well-known, about 40,000 km around the equator, meaning that India was far out of range for a sailing ship going westwards, so no sensible person tried. But Columbus *incorrectly* believed that the world was much smaller, and that India was within reach. If the Americas hadn't been there, he and his men would have died of hunger and thirst long before they got anywhere near Asia.

- Waves: Light is composed of waves, moving through some universal but not really material substance called the *ether*[9], pretty much like ocean waves move on water, or sound waves move through the air. Different colors correspond to waves of different wavelength[10].

Newton championed the particle hypothesis, calling the light particles ”corpuscles”, and his authority was such that his support lent the corpuscle hypothesis a lot more credence than it deserved on empirical grounds. Physicists actually working with optics, such as Christian Huygens, generally found the wave model more fruitful. But the controversy remained alive in its original form for more than a century, before any clear evidence became available.

Is light a particle?

Some of the observed behaviour of light is easier to explain with a particle model, for example the fact that light rays generally travel in a straight line, unless disturbed, much like ordinary particles do, and do not, like sound waves, travel around corners. Others, like reflection in a mirror and refraction in a lens, can be explained just about equally well with either waves or particles, though the explanations might look quite different. For example, refraction of light in a lens can be explained if light waves slow down as they enter the glass of the lens — or if light particles speed up instead! This might have been a way to distinguish between the wave and particle hypotheses, if only somebody could have thought of a way to measure the speed of light in glass with 17th-century equipment[11].

More difficult, however, is to explain with particles the reason why some of the light hitting a window is reflected, and some goes straight through — every particle following the same path ought to do the same thing. This is not a problem with waves, though, as can be easily verified with water waves in bathtub — just let a towel or something hang vertically down into your bath, like a curtain, and send a wave towards it with your hand.

But the real clincher didn't come until 1802, when a physicist named Young made some experiments with divided lightrays. Light from the same source was split up into two rays, and then allowed to rejoin, after following different paths for a while. If you do this with particles, nothing remarkable ought to come out of it; half the particles go one way, and half the other, so that when they rejoin you once more have the full number. But adding waves is not quite so straightforward — waves have both crests, that are higher than the surrounding water (or whatever), and troughs, that are deeper down than the surroundings. If two waves join so that the crests from one meet the crests from another, then the result is one wave twice as tall, reasonably enough — but if the waves join so that the crests of one wave meet the troughs of the other, then the sum is *nothing*. The lower parts of one wave cancel out the higher parts of the other wave, and vice versa. If the waves match exactly, the result is a flat surface for water waves — or darkness for light waves!

And light+light=darkness is exactly what Young found. When he added two lightrays from the same source, the result was a mixture of light spots where crest met crest, and dark spots where crest

---

[9] Not to be confused with the organic chemical ether, which many of us remember from a variety of medical applications, for example as an early anesthetic.

[10] The wavelength of a wave is the distance from one crest (wave-top) to the next.

[11] A direct measurement of the speed of light requires a distance which it takes the light a measurable time to travel. In the 17th century, that meant at least a few seconds, which meant astronomical distances, making measurements in glass obviously impractical — the first successful measurement (in vacuum) was done in 1675 by Ole Rømer, a Danish astronomer studying the moons of Jupiter, yielding a result of around 230,000 km/s. This is substantially less than today's accepted value of 299,792.5 km/s. This discrepancy was caused by the imperfect astronomical knowledge of the time, and has subsequently been corrected to a value not significantly different from today's (*not* to a value higher than today's, the claims of certain crackpots notwithstanding.)

met trough. Nobody could have asked for clearer and more solid evidence that light is some kind of wave. Light cannot possibly be a particle, if it displays such obviously wavelike behaviour.

Is light a wave?

Well, we've just *proven* that light is a wave, haven't we, so why am I heading this chapter with such a silly question?

For a hundred years after Young's announcement it was indeed a silly question. Everybody knew that light was a wave, and a multitude of new discoveries and experiments gave the wave hypothesis ample corroboration. Furthermore, work on electricity and magnetism led, through Maxwell's Equations, to the theoretical prediction that electromagnetic fields could move through space in a wave-like motion. Interestingly enough, Maxwell could even calculate the speed of these hypothetical electro-magnetic waves, which turned out to be around 300,000 km/s, suspiciously close to the measured speed of light... The conclusion that we had finally found the true nature of light, as electromagnetic waves, wasn't all that difficult to reach. End of story — or so everybody thought, just like everybody thought that Newton's work was the last word about gravitation...

But Newton's theory of gravitation held for more than two centuries; for light we had the "final" answer for just a few decades, before problems started cropping up. Minor things at first, just like the perturbations of Mercury's orbit that finally brought down Newton...

One problem was the behaviour of hot objects. If a thing gets hot enough, it starts to glow, like the business end of a cigarette, or a plate left on too long on an electric stove. This glow-light is emitted from the glowing body in a process that is easy to explain through the electromagnetic theory of light, and it is easy to calculate just how much light ought to be emitted, and what colour. There's just one problem — the result of the calculations was all wrong. We all know from everyday experience that things start glowing weakly in red at modest temperatures, and that the glow intensifies and becomes first yellow and then whiter, as the temperature rises. What could be calculated from the theory was that all objects, regardless of temperature, ought to glow with infinite amounts of violet light. It seems there is indeed a minor problem here.

One of the physicists working on the glow problem towards the end of the 19^{th} century was Max Planck, a senior German conservative professor of physics. He found a solution, kind of, but wasn't at all pleased with it. His "solution" went beyond Maxwell's well-established theory, and imposed new rules for light waves, quite arbitrary rules apart from the fact that they made the theory agree with the observed glow-light. According to Maxwell, as well as to our normal understanding of waves, light-waves could be emitted in any size and any number. But Planck's new rules set down limits for the emission of lightwaves; only certain sizes were permitted. And the permitted size was different for different colors. Red light could be emitted in small "wave packets", but violet light, even though its waves were just half the length of red lightwaves, had to be emitted in packets twice as big as the red ones. With the introduction of this large minimum size for blue and violet light packets, the problem with the supposed infinite violet glow went away. But Planck still wasn't happy — the packets were just his arbitrary invention, but if they nevertheless were to have any basis in physical reality, they smelled too much of light particles, which were proven not to exist.

Another problem was found in the interaction of light with atoms. The concept of atoms had been around for a long time, and was pretty well established. But it wasn't until near the end of the 19^{th} century that physicists acquired the equipment necessary to probe the interior of atoms, and discover things like electrons floating around in there. While trying to figure out what electrons were doing inside atoms, it was found that light could kick them out. Shine a lamp on a piece of metal, and electrons will come out. So far, no problem; light was known to be composed of electromagnetic fields, and these could easily interact with the electric charge that was the electron's main characteristic.

The problem came when the experiment was repeated with light of different colors and intensities. To kick out an electron required a certain amount of energy, so the expected result was that electrons would start coming out as soon as the light was intense enough to deliver the energy needed, and that the electrons would come out with higher speed (more ”spare” energy) with higher light intensity. No such luck — in the actual experiment the electrons didn’t come out until light of a certain *color* was used. Blue light, no matter how faint, kicked out electrons, whereas red light, no matter how intense, didn’t dislodge a single one. And the speed of the emerging electrons also depended on the color rather than the intensity — the shorter the wavelength of the light, the faster the electrons moved.

In hindsight, the solution is simple enough, but at the time (1905)[12] it took an Einstein to figure it out. Our old friend Albert realized that the energy transferred to an electron when it’s kicked out, corresponded exactly to the energy contained in one of Planck’s wave packets of blue light. But red light came in smaller packets, and a single packet of red wasn’t enough to dislodge an electron. So nothing happened with red light. Here was strong evidence that wave packets actually were real objects, and that light in this context behaved like a particle, not like a wave.

And further experiments with light and atoms just made the particle picture clearer. It turned out to be possible to *bounce* light particles off electrons, with the light and the electron behaving just like two billiard balls hitting each other — definitely not normal behaviour for a light *wave*! So, in the previous section we proved that light couldn’t be a particle but had to be a wave — and in this section we’ve proven that light couldn’t be a wave but had to be a particle. Hmm.... something funny is going here.

So what *is* it then — and what are you?

It is quite evident by now that whatever light really is, it does *not* correspond to our everyday concept of either a wave or a particle. Both hypotheses lead to contradictions with evidence. Nevertheless, light does have some wave-like properties, but also some particle-like. In certain contexts it behaves like a wave, and in other contexts it behaves like a particle. Is there any reasonable way to combine the two, and resolve the apparent paradox?

There are not really any experimental difficulties or subtleties to blame — both the wave and particle aspects of light can be demonstrated clearly and unambiguously in any high school physics classroom. And the interpretation is obvious enough as well, the wave aspects through comparison with water waves, and the particle aspects through some variant of Einstein’s experiment, or even through the counting of individual particles with a sensitive detector. If you study physics, you’ll get to verify all this for yourself, with your own eyes.

No, the paradox will instead have to be resolved through the abandonment of one or more of our underlying premises. Just like we resolved the paradoxes surrounding the speed of light by abandoning our normal everyday concepts of absolute space and time, we will have to find a similarly drastic way to cut the Gordian knot of light.

---

[12] For a physicist 1905 is a ”magical” year, when the up to then utterly unknown young Albert Einstein during a single year published the solutions to three completely independent problems that had puzzled and defeated the best and most famous minds of the time. Any single one of those solutions would have earned him a Nobel prize and instant fame — and here he came with three at the same time. The first was, of course, his theory of special relativity, the second the electron-kicking light (more formally known as the photoelectric effect) mentioned here, and the third was an explanation of Brownian motion (the random motions of tiny dust grains in a liquid) that gave solid and tangible evidence that atoms actually existed (which was not universally accepted at the time).

One important breakthrough came in an experiment that at first just seemed to make things go from bad to worse. The French prince[13] and amateur physicist Louis de Broglie, fed up with trying to figure out how light waves could act like particles, instead suggested that maybe the objects that we all thought were ”normal” particles, normal everyday material objects, actually were some kind of wave as well. Now, testing this hypothesis with full-sized objects, like students, is a bit difficult — but with electrons it could be done. So Young’s experiment, with two light waves from the same source rejoining after following different paths, was redone with electron ”waves”. And to everybody’s surprise electrons exhibited exactly the same wave properties as light, giving the same kind of dark/light pattern characteristic of wave addition. The electron, just like light, sometimes behaved like a particle, and sometimes like a wave[14].

As far as we can tell, all objects exhibit these wave-like properties in the right circumstances. The wave-particle paradox is not limited to light, but appears to affect everything around us. Just as with Einstein’s oddities, the wave effects of normal everyday material objects are much too small to notice, as are the particle aspects of sound waves and the like — but they are still there.

The apparent universality of the wave-particle duality led several physicists of the time (1920s) to consider ways and means to rebuild the foundations of physics, with the intention of finding a replacement that would resolve the perceived paradox. In 1925 two different solutions were invented independently by two young German[15] physicists, Werner Heisenberg and Erwin Schrödinger. On the surface, there were no similarities between their solutions. Both of them threw away all the old physics, but that was the only point they appeared to have in common. After a few years, however, they realized that their two solutions actually were equivalent, with the same physics expressed in different mathematical languages. For mathematical they were, particularly Heisenberg’s version was far more abstract than even Einstein’s curved space-time.

As usual in these physical journeys, we have now reached the point where it turns out that the universe is both weirder and subtler than we had imagined., a point where another one of our most cherished common-sense concepts will have to be abandoned as too crude and primitive to represent the real world. Absolute time and space and geometry and the force of gravity are already gone, through Einstein’s ministrations, and replaced with more powerful substitutes, giving us a universe more elegant and awesome than our predecessors could imagine.

The concept to be left by the wayside at this point, is that of a physical object as something that actually exists here in our world. One might even say that the very concept of *existence* becomes problematic, and has to be modified. In Quantum Mechanics, as the Heisenberg/Schrödinger-theory became known, what we perceive as a physical object is no longer a fundamental entity, existing in its own right. Instead, the fundamental reality belongs to a certain mathematical abstraction, in Schrö-

---

[13] A distant relative of the king who lost his head in 1792. The royal dynasties of France are a mess; there are currently *three* unemployed monarchs claiming the throne, two different descendants of Louis XVI and one descendant of Napoleon.

[14] One amusing detail is that G.P. Thomson, who did the electron-wave experiment and won the Nobel Prize for it, was the son of a certain J.J. Thomson who, one generation earlier, had identified the electron as a particle, and won the Nobel Prize for *that*. And to add to the confusion, there was a third, even more famous, physicist named Thomson, not related to the other two, alive at the same time: William Thomson, better known as Lord Kelvin.

[15] The world of physics was dominated by Germany for the entire period between the two world wars, much like the United States has dominated science ever since. Apart from occasional foreigners like the Thomsons, just about all the foundations of modern physics were laid by Germans (Einstein as well was born in Germany). And when the focus shifted to the U.S. during WWII, exiled German physicists played a major role, not least in the discovery of nuclear fission and the Manhattan Project. An interesting (but horrifying) speculation is what the world would look like today if Hitler had had sense enough not to exile the leading physicists of the day — the first nuclear weapon may well have been his! But since Einstein and several others were Jews, their work was just ”dirty Jewish pseudoscience”, of no possible value to the Reich. So Hitler lost the war...

dinger's version called a *wave function*. Wave functions cannot be perceived directly, but manifest themselves indirectly instead whenever we try to probe their world. And these indirect manifestations, these secondary shadows of wave functions, are what we perceive as objects in our universe. It is quite possible and normal for a wave function to manifest itself in different ways depending on how we choose to probe it, thus explaining the troublesome wave-particle duality. What we sometimes perceive as a light particle and sometimes as a light wave, are actually different manifestations of the same underlying wave function. The wave function itself is well-defined and free from contradictions, but its different manifestations do not all correspond to the same intuitive category for our senses to perceive, thus giving us the illusory impression of paradox.

The name ”wave function” does seem to imply that the wave nature is somehow more fundamental, but the waves that the wave function refers to are most definitely *not* the kind of physical electromagnetic waves that light was supposed to consist of. The ”wave” part of the function is just from the mathematical imagery used by Schrödinger — it could equally well have been called a ”matrix function”, from Heisenberg's.

But what *is* this wave function then? Good question... *Very* good question. Let's start with the points that everybody agrees on, few as they are:

- It is *not* anything tangible in our universe; we have no direct access to the wave function itself, only to its surface manifestations.
- It is an excellent mathematical tool for describing what happens when we do things to electrons and atoms and lightrays and other tiny objects. Calculations using the mathematical machinery of quantum mechanics consistently reach the right answer. Experimental results are predicted with unprecedented precision. In fact, quantum mechanics has been verified to a much higher accuracy that any other branch of physics — the deviations, if any, between theory and reality are less than one part in a billion. It is on much more solid empirical ground than Newton ever was, or Einstein is.

The latter aspect leads many physicists and, particularly, engineers to the pragmatic attitude often called ”Shut up and calculate!”. Quantum mechanics *works*, so who cares what a wave function *is*? And it does work, works very well indeed — just about every 20^{th}-century technology that I can think of is utterly dependent on quantum mechanics, from computers to lasers to television to solar cells to nuclear power, none of them would have been possible if we hadn't had quantum mechanics.

Nevertheless, I find this attitude highly unsatisfactory — I want to *understand* the world, not just use it, no matter how successfully. ”Shut up and calculate” means philosophical and intellectual capitulation. The founders of quantum mechanics weren't entirely happy either; the development of quantum mechanics from 1925 onwards is a story of consistent success from a practical point of view — and a story of consistent failure to reach agreement on the underlying nature of reality. Even today, the issue of the philosophical interpretation of quantum mechanics is far from settled.

One central problem is that of *uncertainty*, which appears to be at the heart of quantum mechanics. Classical Newtonian (or for that matter Einsteinian) physics is basically deterministic; it is possible, at least in principle, to know exactly where a certain particle is and where it is going, and to predict exactly what will happen to it. Not so in quantum mechanics; the wave function in itself is by and large deterministic, but since we have no access to it that doesn't help very much. And when we look at the various surface manifestations of the wave function that we *can* access, the stuff that we interpret as objects in our normal world, we get random results. The wave function contains in some way a *probability* that it'll give a certain result when probed in a certain way — if we try to measure the position of what we perceive as an electron, the result will sometimes be that it's here, and sometimes that it's over there, without either the electron or the wave function actually having moved. The system behaves as if somebody inside the wave function were flipping a coin to decide which manifestation to show you. We can calculate and predict exactly the probability of getting a certain result from a measurement, , or the average value of a large number of measurements, but there appears to be *no* way of predicting in advance the result of an individual wave-function ”coin toss”. There is something truly random at the bottom of the wave function. And worse, the wave function has some perverse mathematical properties that makes it impossible to measure several things at the same time and get reasonable results. If we do an experiment in which a certain electron behaves like a particle, we cannot at the same time test the same electron’s wave aspects, because one measurement sabotages the other. Likewise, we cannot simultaneously measure the position and the speed of a particle, because one measurement causes error in the other one, and the more precisely we measure one quantity, the worse the error that it causes in the other.

For most practical purposes, though, this uncertainty is so small that we don’t notice it. It is significant on an atomic scale, but if there is an atom-sized uncertainty in *your* position not even your most intimate friend will notice. There may be a finite probability that your wave function will suddenly flip its coin and have it land on an edge, changing its manifestation of you so that you are found sleeping in the wrong bed, but the probability is so minuscule that you can sleep soundly for the lifetime of the universe. Nevertheless, there is a fundamental difference between knowing that you *are* here, and knowing only that your wave function is such that it is very likely that someone probing it will find you here. In one case, you are *you*, in the other case you are just a surface manifestation of some spooky mathematical abstraction, a fleeting shadow of the real world, with the real world forever out of reach. In many ways, this is reminiscent of the Greek philosopher Plato, who wasn’t into quantum mechanics, but did have a world view in which the world that we perceive as real is a mere shadow of the *really* real world — he is famous for his cave metaphor, where we are likened to prisoners in a cave, chained out of sight of the real world outside the cave. The only things we can see are shadows on the cave wall, shadows cast by whatever is going on outside the cave. From these shadows we try to deduce what’s out there, but a shadow never shows all aspects of its source — we never get the whole truth, but only one slice of it at a time.

Is this really the world we live in, this dark cave, one step removed from reality? Well, we don’t really know; as I said, the philosophical aspects of quantum mechanics are not resolved, and are not likely to be anytime soon. There is an ongoing debate between physicist, both in formal contexts and over the fourth and fifth beer late at night, concerning different interpretations of quantum reality. One thing is clear, though: what it means to exist, and to be real, is not as obvious as we usually assume.

This journey ends here, on a rather less happy note than the preceding one. Through gravitation, we could follow a thread of reasoning until we literally reached the end of the universe — but following the thread starting with light, we nearly unravelled the fabric of our own existence. Neither result was possible to anticipate at the departure, but each thread could be followed through careful research, through experimentation, hypothesizing, deduction, testing, and more experimentation and exploration, with many blind alleys and pitfalls along the way. And we are certainly not at the end of the journey of science — what I have presented here are just two small appetizers, just scratching the surface, just a quick walk down the nave of the cathedral of physics. I hope this quick tour has given you some idea of the magnificence of the cathedral in all its splendour, and maybe even a yearning to explore a bit more of it, and learn more of this marvellous universe that we live in.