Fundamental Physics and the Nature of Reality

Norman Gray

Borders Astronomy Club, 1 June 2000

I'm not really going to talk about Astronomy here, but instead use Astronomy and Physics as a platform for talking about the way we interpret, and indeed the way we *see*, the outside world. This may seem arcane, and even laughably removed from the certainties of the physical sciences, but by the end I hope to have persuaded you that Physics in particular, and Science in general, are less certain, and consequently in some ways more curious, than they may at first appear.

To make my points, I'm going to refer to two fundamental areas of Astronomy and Physics which, when examined closely, expose the provisionality of our answers to the question 'What is the world ultimately made up of?'

1 Fundamental physics

I am going to mention the picture that general relativity gives us of the structure of spacetime, in particular the notion of spacetime as a *geometrical* entity, from which any classical notion of 'gravity' has been erased. Then, I am going to give a *very* quick picture of the sort of picture that high-energy physics suggests to us. This talk is, in some ways, a radically compressed version of a ten-week course I give for Glasgow University's Adult-Education department: see `http://www.adultedu.gla.ac.uk/`.

1.1 General relativity and the edge of spacetime

We are all familiar, whether we know it or not, with the Newtonian picture of space and time – it is the 'common sense' picture we have of ourselves as objects within space, moving through time like boats on a river. It is the uncomplicated picture that allows us to arrange to meet a friend at a certain place and time without having to worry whether they have higher mathematical ability enough to understand the arrangement. It is an integral part of that picture, that the same rules apply at the highest speeds and the largest scales in the universe.

We now know that this picture is inadequate in fact, and that space and time behave in ways both more subtle and more interesting than that.

Special relativity is the study of the motion of objects at (high) constant velocity (this special case of zero acceleration is what gives this branch of the theory its name). It turns out that, when we do things like measure the lengths of moving objects, or examine clocks being carried on them, the answers we get are drastically at odds with the answers we would expect, based on our Newtonian intuitions about the world; specifically, we observe such extraordinary things as moving clocks both running slower than, and being measured to be shorter than, the same clock at rest at our side. There is a good

---

[Version: 1.2, 2002/05/20]

deal more to be said about special relativity, but for the present, I simply wish to remark that these extraordinary phenomena can be to some extent natural and obvious, given a suitable viewpoint.

Imagine looking at a pole sitting at some angle to you (see Fig. 1). You *could* measure the 'length', l, of the pole by measuring the distance between the two observers who can see one end of the pole directly in front of them; and they can measure the 'depth', d, of the pole by having the same two observers measure the distance to the end of the pole that they can see. Both these measurements l and d have *some* physical significance, but as long as the observers insist on using l and d as their measurement of 'the length of the pole', then they will get different answers for the pole's length depending on where they are standing when they make the measurements. This is an undesirable situation, which is only resolved when we make use of our knowledge of geometry to construct the invariant length, s, of the pole, using Pythagoras' Theorem. You do this instinctively, when you look at objects in the world around you, because you know, from age two or so, that the length of an object as it presents itself on your eyeball is not intrinsic to that object, but is instead an artefact of the angle from which you're looking at it.

In a very closely analogous way, we can say that we have no right to be surprised at the curious behaviour of objects moving at high speed, since the changes in length and the flow of time for those objects are purely an artefact of our choice to observe them while they are moving at speed. Again analogously, we can take the physical and temporal separations of two events (say two handclaps at different positions, or two explosions) and, although neither 'distance' is physically meaningful in isolation (contrary to our Newtonian instincts), we can combine them, using an expression essentially equivalent to Pythagoras' Theorem, to produce a length – called the 'invariant interval' by the cognoscenti – which genuinely is physically meaningful, and which would be agreed on by everyone who measured the same events, no matter what speed they were moving at.

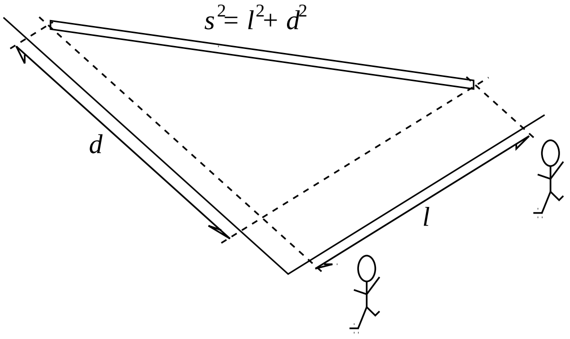

Figure 1: Observers make different measurements of the 'length' of the pole.

The connection with classical geometry is not a mere analogy. When we study the universe in relativity's domain – high speeds and cosmological distances – we are forced to regard the world as a four-dimensional *spacetime*, with three space dimensions and one time dimension not, as in the Newtonian worldview, a sort of optional extra, but instead on an equal footing with the spatial ones. This is a space with different geometrical rules from those familiar ones first written down by the Greeks more than two thousand years ago.

It is this geometrical approach which allows us to make progress when we move beyond Special Relativity, to consider accelerated motion and gravity. Hold tight: this paragraph is a little bumpy. In Einstein's *General Relativity*, objects in free fall – which means objects such as thrown balls or orbiting planets, moving under the influence of gravity alone – move in straight lines in spacetime. This requires a little explanation. Consider the surface of a globe of the Earth: this is a two-dimensional surface, because there are two independent directions you can move in, namely that of increasing longitude and that of increasing latitude, but you cannot, obviously, move in the third, 'up', direction and stay on the globe's surface. Now stretch a thread between, say, London and New York: you have traced the shortest distance between the two points, which stays entirely on the *surface* of the globe. On such a two-dimensional surface, as in the three-dimensional space of our intuitions, we term such a minimal-distance line a 'straight line'. This line really is straight, according to this definition, but we measure it as curved while we view it as sitting within our three-dimensional world. The same holds true the other way: a line which is straight within four-dimensional spacetime – that is, the path within spacetime of a freely falling body – we see as curved when viewed from three dimensions. These curves are the parabolae and ellipses which

Newtonian mechanics described so successfully, for three hundred years, as being the result of the gravitational force.

It does not matter if that last paragraph made little sense – it is exceedingly compressed. The point being made is that, where once we had an apparently successful and intellectually satisfying explanation of motion under gravity, employing Newton's gravitational force, we now have a theory of gravity in which, firstly, spacetime becomes an active participant in dynamics, rather than merely the arena which objects move through, and secondly, the 'force of gravity' which Newton introduced has completely disappeared – the curvature of the path an object takes is merely an illusion, caused by the circumstances under which we view it.

One could view this as a replacement, and say that the relativistic explanation has superceded the Newtonian one, which is demoted to a mere approximation. However, that is *not* how the relationship is actually viewed by scientists in their day-to-day work; instead Newton's gravitational theory is taught, understood, and used as an alternative explanation, more intuitive, and more mathematically tractable, than the full apparatus of General Relativity, and so to be preferred, for the insights it gives, in all those applications which do not require either exquisite accuracy or cosmological scope.

1.2 The universe on the smallest scales

The search for the elementary constituents of matter began, probably, with Democritus, who elaborated the earlier notion that the world is composed of individual and indivisible *atoms*. John Dalton (1766–1844) offered indirect evidence for atoms when he observed that atoms can combine and recombine in certain fixed proportions to make up the compounds that we see around us. Mendeleev systematised chemical knowledge by recognising the patterns within the chemical properties of matter, and laying the atoms out in a regular form – the periodic table of the elements. Mendeleev recognised that the regularity of the pattern was important, and used the presence of gaps in the pattern to successfully predict both the existence and chemical properties of hitherto unknown elements, to fill those gaps. The regularity of the pattern suggested that there was some more fundamental structure, and this was borne out in the early part of this century with the detection by Thomson, Rutherford and Chadwick of, respectively, the *electron*, the *nucleus* and the nuclear atom, and the *neutron*. At that point, all matter could be said to be composed of combinations of just three particles, the protons and neutrons within the nucleus, and the electrons orbiting around it.

In 1930, experiments on neutron decay seemed to suggest that the process violated the principle of the conservation of energy. To rescue this principle, Wolfgang Pauli proposed the existence of the *neutrino*, a ghostly particle which interacts only very weakly with other matter. It is consequently very difficult to detect, and that was only done in 1956. Dirac's relativistic mechanics introduced the idea of *antiparticles* – partners of each of the normal particles, which have opposite values for all of their quantum parameters except mass. When people were able to detect very high energy cosmic rays in cloud chambers they found the *muon* and the *pion*. The list of fundamental particles has become rather long.

Then the particle accelerator was invented.... When physicists were able to probe to higher energies, a host of new particles tumbled out, each with as much of a claim as any other to fundamental status. It started to look as some more fundamental structure had to be found.

Fundamental particles were classified, phenomenologically, into the *leptons*, particles such as the electron and the neutrino which are very light and appear to be truly fundamental, and *hadrons*, heavier particles which were further divided into heavier *baryons* and relatively lighter *mesons*.

At first merely descriptive, but later with more theoretical support, the notion of *quarks* was introduced by Murray Gell-Mann, in the 1960s. In this picture, the various groups of particles with similar properties are created by combining more fundamental objects. With three quarks, called *up*, *down* and *strange*, you can make a neutron by binding an up to two downs, or a sigma particle Σ^+ with two ups and a strange, and generally create all of the known baryons with a suitable combination of three quarks. Each of these quarks has its antiparticle, and the mesons are created by combining a quark and an antiquark – a pion π^+, for example, is an up bound to an antidown. As with Mendeleev's periodic table, the tableaux constructed by these methods had some gaps, and this allowed Gell-Mann to predict the existence and mass of the so-called Ω^-.

Quarks started off as a purely mathematical construction – purely as marks on paper – as a way of constructing things with the correct properties. No-one started off saying anything about them *existing* (and they shouldn't be observable in principle), but now physicists routinely talk and act as if they believe precisely that, in the sense that the quarks are taken to be the things within the proton and friends.

Much like the geometrical idea of spacetime, quarks almost seem to be (particularly beautiful) mathematical objects which have struggled off the page, shaken the ink from their feet, and made it out into the real world as physical objects. We seem forced towards the question 'just how real are quarks and spacetime?' Are quarks a thing which has always existed 'out there' in the world, which we have only recently invented the maths to describe, but which could not appear to us in any other form; or is the world more plastic than that? From one point of view, we can suppose that there is some reality out there with at least some fixed nature, but that the details of how it appears to us are *not* fixed, but depend largely on our mathematical history. With a different history, we might well have constructed a different, *but equally correct*, way of looking at fundamental objects. In this limited sense, we might be said to invent the world we see around us.

I am not making the more extreme point that 'all reality is socially constructed', or that 'all descriptions are equivalent'. Instead, I am drawing your attention to such pairs of theories as classical gravitation and general relativity, or the idea of electrons as particles and electrons as fields, or classical and statistical approaches to thermodynamics. In each case – and especially in the last one – we have pairs of theories with radically different fundamental objects, which are nonetheless separately successful when it comes to describing and predicting the world. In each case, it is true, the second of the pair is more successful than the first, in terms either of accuracy or fertility, but this is true only because it was precisely those weaknesses which forced the scientific community to the extreme effort required to generate such an incommensurable alternative theory. Without those weaknesses, no such effort would have been made. We have no reason to suppose that, were such an effort to be made, we could not come up with a radically different, but equally successful, alternative to, say, general relativity, or in other words, that external reality is merely a constraint on, and specifically does not control, the language and pictures we use to represent it.

John Wheeler – partly responsible, amongst other things, for the many-worlds interpretation of quantum mechanics – tells a parable of a game of Twenty Questions. Imagine you are sent out of the room while your friends agree on an object; you return, ask individuals in turn the usual yes/no questions, and successfully work out what object they are thinking of. Only then do you discover that the rules your friends had agreed on were not the usual ones: instead of agreeing on an object beforehand, each separately thought of an object, if necessary changing it to be consistent with the answers already given, and if you asked them a question, they answered with respect to the object they alone were thinking of. It is you, by choosing the questions to ask, who has created the correct answer to the game.

2 Versions of realism

The most straightforward and common-sense interpretation is also the most difficult to justify. *Scientific realism* (also known, somewhat pejoratively I think, as naïve realism) is the claim that the objects that science discusses exist, without qualification, in the world, so that things like quarks and the curved spacetime we live in actually exist, in that form, and that science has *discovered* these objects through its study of the world. This is probably the view held by most scientists, but despite this authority, it is probably the least robust position to hold, and starts to crumble as soon as it is put under pressure.

The big difficulty that realism has, is that it has to believe that the world we perceive through our senses is an accurate representation of the world as it really is. The existence of optical illusions demonstrates that the impressions we get can sometimes deceive us – such illusions are of course resolvable, or removable, with measurement, but how can we be sure that we are not in the grip of some more subtle illusion or illusions, which cannot be so easily seen through. Perhaps we are deceived, for example, in believing that events in the world are causally linked to each other.

This problem has been approached many times in the history of philosophy, but the empiricists cut through the problem by declaring that the world of sense impressions is all that we *can* know, and that to discuss the world as it really is, is metaphysical. When this approach is applied to the philosophy of science (where it manifests itself as positivism), it results in the instrumentalist proposition that scientific theories are merely structures which relate our observations to each other, and allow us to predict what will happen in future. That is, our theories are seen as *instruments*, rather than insights, and they should not be judged in terms of their greater or lesser correspondence with the truth, but purely in terms of their usefulness. For the empiricists, it is unscientific, or metaphysical, to regard objects such as electrons or fields as actually existing in the world.

A third extreme view of the problem talks of reality as *socially constructed*. A five-pound note, for example, has a real value, in the sense that I can go down to the shops and use it get real food; clearly, however, the note is real in a different way from that in which Ben Nevis, say, is real. The reality of the note's value – as opposed to the value of any other bits of paper with writing on them – is created by some process in society in the large [1]. In the sociology of science, the social constructivists extend this argument from money and marriage (where it is uncontentious) to matter; they do not deny that there is some sort of reality to the world out there, but claim that rather than being themselves independently real and discovered by science, things like electrons and the gravitational field are real only in the sense that scientists have agreed to use those terms in discussing and manipulating those particular bits of Nature. The electron is real, therefore, in the same way that the note's value is.

As an example of the middle ground between these various camps, I will briefly discuss Hilary Putnam's *internal realism*. Putnam distinguishes internal from metaphysical realism. (essentially the 'scientific realism' we described above). Internal realism is objective, in the sense that the world is experienced the same way by different observers. There is also an element of pluralism to the theory, however, in the sense that the world admits of different 'mappings' – different, and potentially incompatible, explanatory schemes which explain how the world works. Bohr's Copenhagen interpretation and Bohm's hidden-variable theories are incompatible explanations of quantum mechanics; similarly Newton's and Einstein's gravitational theories are incompatible.

This internalism is not intended to be an empiricist theory, so my understanding of Putnam's idea is that the fundamental objects of *all* these incompatible theories can usefully and reasonably be regarded as real. If this is correct, then internalism seems to be a sort of 'linguistic constructivism': there is something out there in Nature, but it is our decision to call a particular lump of that stuff an 'electron' which makes us see the world in those terms, and allows us to erect a theory to enable us to understand how these electrons move around.

In the same way, Ben Nevis only becomes a mountain, as opposed to a raised bit of Scottish landscape, when we decide to *call* it Ben Nevis. The tourist board has a

---

[1] John Searle discusses this process in detail, in John R. Searle, *The Construction of Social Reality*, Penguin 1996. Searle would emphatically not approve of the extension of the process to science.

different, incompatible, map of Scotland from the geographer: Ben Nevis also appears on their map, but as the same sort of thing as the border towns – a popular tourist attaction – and as a quite different thing from a similar mountain (to the geographer) not in a climbing area. The point here is that Ben Nevis really is a mountain, and it really is a tourist attraction, but that both of these are insightful constructions we place on something that really is just a bit of land higher than its surroundings.