Dirac's Conception of the Magnetic Monopole, and its Modern Avatars

Sunil Mukhi

Sunil Mukhi is a Professor at Tata Institute of Fundamental Research, Mumbai. His research interests are string theory, quantum field theory, quantum gravity and supersymmetry.

Electricity, Magnetism and Duality

A curious high-school student in the early 1930's might have speculated about the existence of magnetic monopoles. There is no record of such a student, but if she had existed, she could have reasoned as follows.

She would have learned that atoms are made of electrons whirling around a central core, the nucleus. The electron has a negative electrical charge, while the nucleus has a positive charge. Opposite charges attract each other via the electrostatic force, which causes the electrons to stay in their orbits around the nucleus.

Now it is quite easy to connect a pair of charged spheres with an insulating rod. One sphere can be given a positive charge and the other, an equal amount of negative charge. Thus the whole object has no net electric charge. It is called an 'electric dipole'.

Our student's science teacher might have brought a set of such electric dipoles to class and carried out the following experiment. When the positively charged ends of two dipoles are brought close together, they repel. The same is true of the negatively charged ends. But when we bring the positive end of one dipole and the negative end of another together, they attract.

After keenly observing these experiments, the student might have noticed a similarity with magnetism. What we call the 'north' and 'south' poles of a bar magnet behave just like the positively and negatively charged ends of the electric dipole. For bar magnets, we know that like poles repel, and unlike poles attract. What is

more, the force between two bar magnets shows the same variation with distance and angle as the force between two electric dipoles.

If our student had a flair for clever-sounding expressions, she might have described this similarity as 'electric-magnetic duality'. By this, she would have meant that upon replacing all the electric objects in an experiment by their magnetic counterparts, the physical behaviour of the system remains the same.

Though pleased at her originality, her science teacher would regretfully have pointed out a flaw in the 'duality' idea. If we cut an electric dipole in two by breaking the insulating rod, we recover the charged spheres that we started with. Each of these has a net electric charge (one is positive, the other negative), and can be called an 'electric monopole'. On the other hand, when we cut a bar magnet in two, we end up not with two magnetic monopoles, but with two smaller bar magnets, each having a north pole at one end and a south pole at the other. So electric and magnetic dipoles do not behave the same way when we cut them. This seems to be bad news for 'duality'.

The student feels that her idea was too elegant to be rejected so quickly. She points out that all we need to restore electric-magnetic duality is for magnetic monopoles to exist in nature, just as electric monopoles exist. We do not find magnetic monopoles when we cut a bar magnet, but there might well be other ways of producing them.

While the above anecdote is hypothetical, one of the most creative minds of that period probably went through a similar thought process and came to a similar conclusion. But when Paul Adrien Maurice Dirac came to write about his theoretical findings in 1931, he expressed the idea rather differently.

Dirac's Monopole Paper

On May 29, 1931, Dirac submitted a paper to the Royal Society with the title 'Quantised Singularities in the Electromagnetic Field'. This was to be the first significant publication on the subject of magnetic monopoles.

Just a few years earlier, in 1928, Dirac had proposed a wave equation for the electron that satisfied the principles of special relativity. By mathematical analysis of his equation he had found, somewhat to his surprise, that it predicted the existence of a new particle, similar to the electron but of opposite electric charge. Such a particle, the 'positron', was discovered by Carl Anderson in 1932.

Preliminary experimental results had been obtained by Anderson in the summer of 1931, but it is not clear (at least to me) whether Dirac knew of these results at the time that he submitted his monopole paper. Certainly he did not refer to them. But a consistent theoretical picture of antiparticles had emerged by then, and Dirac seemed satisfied that the idea of antiparticles was correct.

The lesson that Dirac drew from the antiparticle story was that mathematical consistency could be used to predict a new elementary particle. It was becoming too difficult to develop theoretical physics in the direct way: by analysing experimental results and proposing theoretical formulae to fit them. He suggested that "the theoretical worker in the future will therefore have to proceed in a more indirect way", by trying to "perfect and generalise the mathematical formalism that forms the existing basis of theoretical physics". This could lead to new theoretical predictions, and experiments would only be invoked at a later stage to confirm them.

With this belief, Dirac decided to go deep into the foundations of quantum theory and investigate the meaning

> "The theoretical worker in the future will therefore have to proceed in a more indirect way", by trying to "perfect and generalise the mathematical formalism that forms the existing basis of theoretical physics"

of charge. He was hoping to explain the quantisation of electric charge – the fact that it always appears in whole multiples of a basic unit. He did not find such an explanation, but instead stumbled on something closely related – a law relating the quanta of electric and magnetic charge, now known as 'Dirac's quantisation condition'. In the process, he provided the first convincing argument that magnetic monopoles are consistent objects within the framework of quantum mechanics.

According to quantum mechanics, the probability amplitude for a particle is given by a 'wave function', a complex number at each point of space. The phase part of this complex number is a subtle quantity. It does not contribute to the absolute square of the wave function, which tells us the probability that a particle is in a given place. But it contributes if we first superpose two different wave functions (with different phases) and then square them. The phase is the source of all phenomena in quantum mechanics that go by the name of 'interference'.

Dirac put this phase factor under his formidable intellectual microscope. Did it have to be 'integrable'? This amounted to asking whether, if the phase of a wave function was followed while traversing a closed path, it came back to its original value or not. If not, it was said to be 'non-integrable'. Dirac concluded that non-integrable phase factors were allowed in quantum mechanics, but they could only arise if an electromagnetic field was present. An ordinary physicist might have stopped at this very reasonable conclusion, but Dirac then went one step further.

He noted that any angle, including the phase of the wave function, is defined only up to the addition of multiples of 2π. Different closed paths could in principle lead to different jumps in the wave function, as long as all the jumps were multiples of this unit. His analysis revealed

that this in turn was possible only if, at some point in space, there was a magnetic monopole. A line in space, that Dirac called a 'nodal line', would emanate from the monopole, and the phases of wave functions would jump or not depending on whether the closed path encircled this nodal line. This line was later to be dubbed the 'Dirac string'.

In order for the jumps in the phase to actually be multiples of 2π, the monopole had to have a charge g which, together with the smallest electric charge e in the Universe, obeyed the relation

$$eg = h/2,$$

where h is Planck's constant.

With considerable satisfaction, Dirac concluded that not only can magnetic monopoles exist, but if they do, then electric charge in the Universe is quantised. The minimum electric charge is determined by the value of the magnetic charge.

Having presented this extraordinary insight, Dirac concluded his paper on a modest note: the object had been to show that "quantum mechanics does not really preclude the existence of isolated magnetic poles". No change in quantum mechanical theory was required, just a deeper understanding of it. Dirac felt that this was good enough reason for magnetic monopoles to exist – they were permitted by his analysis, and he "would be surprised if Nature had made no use of it".

In his paper, Dirac also noted that the existence of magnetic monopoles would restore electric-magnetic duality. As he put it, "the theoretical reciprocity between electricity and magnetism is perfect".

Dirac's proposal of the monopole led to a number of interesting developments. One of these was the observation, due to Meghnad Saha, that a particle could in

"Quantum mechanics does not really preclude the existence of isolated magnetic poles"

"The theoretical reciprocity between electricity and magnetism is perfect"

The incorporation of special relativity inevitably required the possibility of particle creation and destruction.

principle have both electric and magnetic charge. Such a particle would be called a 'dyon'. The quantisation condition on its charges, derived by Saha, turned out to be an elegant generalisation of the one proposed by Dirac for monopoles.

Magnetic Monopoles in Field Theory

The landscape of theoretical physics changed rapidly through the middle of the twentieth century. By the 1960's it was clear that quantum mechanics alone was not enough to understand all processes involving elementary particles. The incorporation of special relativity inevitably required the possibility of particle creation and destruction. This in turn required a formalism that was not based on individual particles (as these could be created and destroyed) but rather, on fields of force. These became the fundamental building blocks, and the resulting 'quantum field theory' is the basic tool by which the physics of elementary particles is understood today.

Besides the electromagnetic field, there are other forces by which elementary particles interact, in particular the weak and strong nuclear forces. These are described by a generalisation of electromagnetism called 'gauge theory' or 'Yang–Mills theory'.

In the 1970's, Gerard 't Hooft in the Netherlands and Alexander Polyakov in Russia were studying the properties of Yang–Mills theories when they encountered a surprising result. In these theories one could find excitations of the Yang–Mills field which, at large distances, looked just like Dirac's magnetic monopoles. However, at short distances they looked very different, with all kinds of other fields besides the electromagnetic one being excited. These were novel examples of 'solitons', localised and relatively stable field excitations that behaved much like elementary particles.

A remarkable feature of these 't Hooft–Polyakov' monopoles is that they do not possess the nodal line or Dirac string. The latter had been a somewhat undesirable feature of the Dirac monopole. Although it could not be directly observed, it had to be present somewhere in space and seemed rather arbitrary. The new Yang–Mills monopoles were smooth lumps carrying magnetic charge but with no strings coming out.

The 't Hooft–Polyakov monopole later found a natural setting in the context of 'grand unified theories'. According to these rather speculative theories, the early Universe was in a highly symmetric phase where electromagnetism as such did not have a meaning. As the Universe cooled, the force of electromagnetism 'emerged' from the more complicated Yang–Mills field. When this happened, magnetic monopoles could be formed. Inside the monopoles the world would look like the symmetric phase of the unified theory, but outside the monopoles we would see ordinary electromagnetism. The monopole would be a sort of 'twist' or 'defect', in space.

In most theories, the number of such monopoles that were created when the Universe cooled would have been rather small. Since they are now distributed over the entire Universe, their density would be too small for us to ever detect them, unless we were especially lucky. Indeed, many attempts have been made to detect magnetic monopoles but to date, none have been found. Much as this is disappointing, it seems to be consistent with the calculations of unified theories. In this context we could well accept that monopoles exist in nature, but are too scarce to be observed.

Another class of unified theories that allow magnetic monopole excitations are the higher dimensional, or Kaluza–Klein, theories. It is an old speculation that the real world intrinsically has more than three spatial dimensions. The ones that we do not see are supposed

It is an old speculation that the real world intrinsically has more than three spatial dimensions.

to be 'compact', curled up on themselves in small circles. In such theories, it was found by Rafael Sorkin, David Gross and Malcolm Perry that one can find excitations of the gravitational field which behave like magnetic monopoles at large distances from a central 'core'. Inside the core, space actually appears to have a higher dimension! But outside it we are in normal space. It began to look as if monopoles were natural in any field theory, though Dirac had only postulated them in the context of the quantum mechanics of the early 1930's.

Extended Monopoles

To write a mathematical description of magnetic monopoles as field excitations, we start by assuming that the excitation is localised, like a 'lump'. Then we solve the corresponding 'field equations'. Indeed, the simplest monopoles are spherical, making the equations relatively simple to solve.

However, a different starting assumption leads to field excitations that are extended. Suppose that we look for an excitation that is narrow in two spatial directions, but extended for a large distance along one direction. It looks rather like a narrow tube. Inside this tube the fields are highly excited, just as they are in the core of a monopole. Such an object can be called a 'soliton string' because of its shape. Such things are not as exotic as we might think: for example a tornado or 'twister' is a naturally occurring object very much like a soliton string.

In suitable field theories there are soliton strings that carry a fixed magnetic charge per unit length. They are 'magnetic strings', rather like extended magnetic monopoles. Objects of this type are sometimes called 'cosmic strings', and might also have been produced in the early Universe when it cooled.

Can we find field excitations that are extended in two

spatial directions, like a sheet? Such objects too are known, and are called 'membranes' or 'domain walls' because they can extend over very large areas, like a wall dividing space into regions. Like soliton lumps and soliton strings, these too can carry magnetic charge. If we crossed the domain wall, the Universe would appear to change its physical properties. For example, on one side there could be pervasive magnetic fields which are absent on the other side, due to the magnetic charge stored in the wall. Such domain walls are popular in some cosmological theories which require different parts of the Universe to obey seemingly different laws of physics.

The most contemporary class of unified theories is the theory of strings. This theory postulates that all the fundamental particles are really tiny loops of a 'fundamental string', and that vibrations of these strings in different modes lead to the observed differences among particles. Such theories typically require space to have a large number of independent dimensions (such as nine, rather than the experimentally observed three). In these theories there is lots of room to make objects that are extended in 1,2,3... dimensions. These are called 'branes', a word derived from membranes. Some of these are magnetically charged, others have both electric and magnetic charge. They are central to the structure of string theory, which moreover has a large set of duality symmetries incorporating electric-magnetic duality. In this sense, modern avatars of Dirac's monopole play a central role in our conception of nature today.

> The most contemporary class of unified theories is the theory of strings.

Conclusions

Dirac's philosophy of mathematical elegance led to his pioneering work on magnetic monopoles and inspired a large number of subsequent developments. It remains true that magnetic monopoles have never been detected, despite numerous attempts, but after all it was Dirac who once remarked, "It is more important to have beauty

> In this sense, modern avatars of Dirac's monopole play a central role in our conception of nature today.

Address for Correspondence
Sunil Mukhi
Tata Institute of Fundamental
Research
Homi Bhabha Road
Mumbai 400 005, India.

in one's equations than to have them fit experiment". One might therefore assume that he was more than satisfied with the impact of this work.

This is not quite the case. Towards the end of his life, the complete lack of experimental evidence for magnetic monopoles began to weigh on him. In 1981, a year short of his 80th birthday, he was invited to Trieste for a conference to commemorate the 50th anniversary of his monopole paper. He declined the invitation due to the strain of travelling, but sent a letter of thanks to Abdus Salam, the Nobel Laureate and Director of the Centre at Trieste. In this letter, Dirac wrote: "I am inclined now to believe that monopoles do not exist. So many years have gone by without any encouragement from the experimental side."

Dirac passed away nearly two decades ago, but it remains to be determined who was right about monopoles: Dirac in 1931, or Dirac in 1981.

Prof. S N Bose, Dr Poornima Sinha (his first Ph.D. student) and Prof. Dirac in the Botanical Gardens, Calcutta in 1954.