PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Solar System in the next millennium

Andrew J. Coates

Phil. Trans. R. Soc. Lond. A 1999 **357**, 3299-3317
doi: 10.1098/rsta.1999.0495

THE ROYAL SOCIETY

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Solar System in the next millennium

By Andrew J. Coates

Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking RH5 6NT, UK

When this journal was founded the telescope had only just been invented. Humankind knew of six planets including our own. The next three centuries added Uranus, Neptune and Pluto to the known list as well as the many moons, asteroids and comets that we know today. Discoveries such as that Earth was not the centre of the universe and that planets orbit the Sun were key steps in increasing the understanding of Man's place in space. But it is only in the latter part of the 20th century that we have been privileged to carry out *in situ* exploration of the planets, comets and the solar wind's realm and to begin to understand the special conditions on Earth that enabled life to start here.

In this article, we briefly review our current knowledge of the Solar System that we inhabit. The Sun itself, the planets, comets and asteroids are all discussed in general terms, together with the important discoveries from space missions which have led to our current views. For each of the bodies and for the interplanetary medium we present the current understanding of the physical properties and interrelationships and present questions for further study.

We describe the solar wind and the way that it interacts with the planets and comets that it encounters. The importance of the obstacle in the magnetized plasma flow is explored in particular detail. We identify the gaps in our knowledge in each case.

What is in store for planetary exploration and discoveries in the next millennium? Already, a sequence of Mars exploration missions including sample return, a landing on a comet, further exploration of Saturn and the Jovian system and the first fly-by of Pluto are planned. We examine the major scientific questions to be answered and speculate on possible space exploration in the future.

Keywords: Solar System; Sun; planets; comets; solar wind; magnetospheres

1. Introduction

Astronomy is one of the oldest observational sciences, existing for about four millennia. This time is estimated on the basis that names were given to those northern constellations of stars that were visible to early civilizations. Two millennia ago, a difference between the more mobile planets and comets on the one hand and the fixed stars on the other was realized by Ptolemy but not understood. At the dawn of the present millennium, six planets (Mercury, Venus, Earth, Mars, Jupiter and Saturn) were known. Aurorae were seen on Earth and recorded but understanding still eluded us. In the first decades of this millennium, comets were seen as portents of disaster (see figure 1). Little further progress was made during the Dark Ages.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

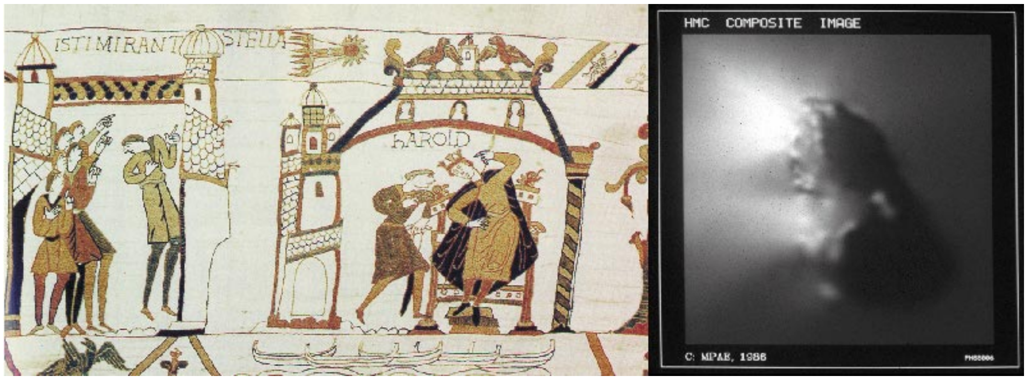

Figure 1. Our view of comets at the beginning and end of this millennium is illustrated by a scene from the Bayeaux tapestry (left) and an image (right) from the camera on ESA's Giotto spacecraft in which the Sun is towards the left (Giotto image courtesy of MPAe/ESA).

Science began to take giant leaps forward in the 16th century. Copernicus realized that Earth was not at the centre of the Solar System. Tycho Brahe's planetary observations enabled Kepler to formulate laws of planetary motion in the 17th century. In the same century, Galileo invented the telescope and Newton developed his theory of gravitation. Seven years after this journal was founded, Newton demonstrated his interest in technology by inventing his reflecting telescope. The 18th and 19th centuries saw increasing use of this new technology, which deepened and increased understanding of the objects in the sky. Among many other discoveries were the periodicity of Halley's comet and the existence of the planets Uranus, Neptune and Pluto; one planet per century including the present one. The observations were all made in visible light, to which our eyes are sensitive and which is transmitted through our atmosphere. The 19th century also gave us the basics of electromagnetism.

In the 20th century we observed many important scientific and technological advances. In terms of astronomy we have gained the second major tool for the exploration of the universe in addition to the telescope: namely, the spacecraft. Space probes have not only opened up the narrow Earth-bound electromagnetic window, which only allows us to detect visible light and some radio waves from the ground, but it has also allowed *in situ* exploration and sampling of our neighbouring bodies in the Solar System (see figure 1 (right)). Using the techniques of remote sensing to look back at the Earth has added a new perspective. For the first time, we can now begin to understand our place in the universe and the detailed processes of the formation of the universe, our Solar System and ourselves. We are truly privileged to be able to use these techniques to further this scientific understanding.

We now know that, far from being empty, interplanetary space is filled with a hot fast-flowing plasma, the solar wind. A plasma, sometimes called the fourth state of matter, is a collection of electrically charged particles that behaves, on the large scale, as a magnetized fluid, while on the small scale the motion of individual charged particles is important. The Sun and other stars are so hot that their material is in the plasma state; in fact, over 99% of the volume of the universe is plasma. Studying the solar wind and how it interacts with Solar System obstacles *in situ* allows us to study this important state of matter without the walls that confine and dominate Earth-based plasmas. As a result, we can begin to understand how aurorae and

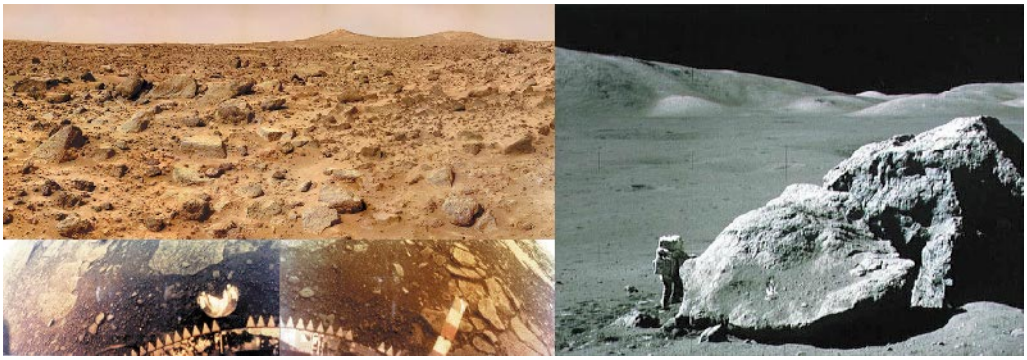

Figure 2. Images from the three bodies (Mars, Venus, the Moon) visited by humankind or his robot landers so far. Upper left is from Mars Pathfinder (courtesy of NASA JPL), lower left is from Venera 13 at Venus (courtesy of Russian Academy of Sciences), on the right is Apollo 17 astronaut Harrison Schmitt (courtesy of NASA).

comet tails form and we can study the potentially dangerous effects of solar storms on humankind.

As we approach the beginning of another millennium, the sense of wonder in looking at the night sky has not changed. As we use better ground- and space-based telescopic techniques and more detailed *in situ* exploration one can only feel a sense of excitement at the discoveries that the next millennium may bring. In this paper, we will review our Solar System in general terms and, in particular, consider the interaction of the solar wind with the various planets and comets that inhabit the Solar System. We consider scientific questions for the future and speculate a little on how they will be answered.

2. Objects

The objects in the Solar System now reflect the history of its formation over 4.5 billion years ago. Because the planets are confined to a plane, the ecliptic, it is thought that the Sun and planets condensed from a spinning primordial nebula. The heavier elements in the nebula are thought to be present due to earlier nearby supernova explosions. As the nebula collapsed and heated, the abundant hydrogen fuel ignited in the early Sun. Gas further from the centre of the nebula became progressively cooler and condensation occurred onto dust grains. This caused differences in composition due to the progressively cooler temperatures away from the Sun. Gravitational instabilities then caused the formation of small solid planetesimals, the planetary building blocks. Accretion of these bodies due to collisions then formed the objects familiar to us today.

This model predicts different compositions at different distances from the Sun, and this is seen in the different classes of Solar System objects today. The outer planets are associated with their much colder planetesimals, some of which remain as comets. The inner, rocky planets are associated with their own planetesimals, of which the asteroids are the partly processed survivors in the present Solar System.

For each of the present-day Solar System objects we now briefly review the information which has been found from space missions and consider the major outstanding questions for future exploration (see also Beatty *et al*. 1999; Lewis 1997).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

3. The Sun

The Sun is an average small ('dwarf') star. In terms of its position half-way along a galactic spiral arm and its temperature (and, consequently, its luminosity) it is quite unremarkable. The Solar System with which it evolved is, in our view, remarkable, particularly since conditions were right for life to begin on the third planet some 4 billion years ago and to evolve from there.

The Sun is remarkable to us as it has a large enough disk for us to observe and the radiation we receive is intense compared with that from distant stars. Consequently, by studying the Sun we can try to understand how average stars work.

The present model of the solar interior is that hydrogen fusion reactions burn in a hot (15 million K), dense (over 100 000 $kg\,m^{-3}$, or several times the density of lead), gaseous core. The core is depleted in hydrogen abundance compared with the remainder of the interior due to fusion reactions. Outwards from the core, the density and temperature decrease rapidly. Heat is radiated outwards up to *ca.* 80% of the solar radius until the convection zone begins. The visible surface or 'photosphere' has a temperature of 6000 K.

The photosphere, and above it the chromosphere and corona, have energetic large- and small-scale structures organized by the magnetic field. Large-scale structures include the coronal holes, whose extent wanes with the increasing solar cycle, and which are the source of the high-speed solar wind that escapes along the magnetic field. Smaller-scale structures include coronal loops, sunspots, prominences and filaments. Short-lived features causing flares and coronal mass ejections are also important in determining the electromagnetic environment.

By comparison with other similar stars, we know that when the Sun runs out of hydrogen fuel in about 5 billion years the core will start to support other nuclear reactions involving carbon, nitrogen and oxygen. The core will contract but the envelope will expand to beyond the Earth's orbit. At that point the outer Solar System will be warmer than its current temperature and the situation may prevail long enough for life to develop on Europa or Titan.

Much of our present picture of the Sun, including models of its structure, has emerged from space missions. Most of the energy of sunlight, 1370 $W\,m^{-2}$ at Earth's orbit, and the spectral information it contains is only visible to space-borne detectors. However, there are several areas for which there are significant outstanding questions, as follows. How are transient events such as coronal mass ejections triggered? Will we ever be able to forecast their onset? Why is the temperature of the corona, at a million kelvins, so much hotter than the visible surface? Are our models of the internal structure correct? Why are there fewer neutrinos from the burning core than expected? How is the solar magnetic field produced? We await the space missions of the next millennium to answer these questions, and, inevitably, to pose more.

4. The inner planets

Due to the temperature variation in the collapsing primordial nebula, the inner regions of the Solar System contain rocky planets (Mercury, Venus, Earth and Mars). The condensation temperatures of the minerals forming these planets were higher than the icy material in the outer Solar System. While it is fair to treat the inner planets as a group, the diversity of the planets and of their atmospheres is remarkable. The three processes of impacts, volcanism and tectonics are vital ingredients in the

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF

evolution of the planets. Our understanding also depends critically on the sources and dissipation of heat.

The origin of the atmospheres of the inner planets is an important topic in itself. It is now thought that the origin is due to outgassing of primitive material from which the planets were made. The subsequent evolution of the atmospheres depends on two major factors: the distance from the Sun (which controls radiation input) and the mass of the body (controlling, firstly, heat-loss rate from the initial increase due to accretion, secondly, heating rate due to radioactivity, and thirdly, atmospheric escape speed). The presence of life on Earth has also played an important role in determining atmospheric composition here.

Mercury is a hot, heavily cratered planet that is difficult to observe because of its proximity to the Sun. Only one spacecraft, Mariner 10, has performed three fast fly-bys of the planet. Mercury is remarkable because of its high density, second only to the Earth. Another remarkable and unexpected discovery from Mariner 10 was the presence of a strong magnetic field and a magnetosphere. It is likely that the planet has a larger iron-rich core in relation to its radius than the others. Mercury has an 'exosphere' rather than an atmosphere, since the pressure is so low that escape is as likely as a gas collision. An atmosphere was not retained because of the low mass (size) and high temperature (proximity to the Sun). The planet has an eccentric inclined orbit. The rotation period is in a 2:3 ratio with its rotation. This indicates that the slightly non-spherical shape of the planet was important during its formation. The surface of Mercury contains a well-preserved little-disturbed cratering history. During the early bombardment in the first 0.8 billion years after formation, ending with a large impact that produced the 1300 km Caloris Basin feature, there was important tectonic activity. There is also evidence of significant shrinkage of the planet due to cooling, causing 'lobate scarp' structures; this shrinkage was also early in the planet's history.

The Venus–Earth–Mars trio is particularly important to our understanding because we know that life evolved at least on the Earth. In the case of Venus, the planet, although similar in size to the Earth, was closer to the Sun; water evaporated and caused a greenhouse effect, causing further evaporation of water and, eventually, a runaway greenhouse effect. Lack of surface water meant that fixing of evolved carbon dioxide in the rocks via absorption in the oceans as on Earth was impossible. Carbon dioxide that has evolved into the atmosphere over the 4 billion years since formation continues the greenhouse effect after the hydrogen (from water dissociation) has escaped to space and the oxygen has oxidized rocks. The surface temperature is now 750 K, the atmosphere is thick and supports sulphuric-acid clouds, and our sister planet is completely inhospitable to life. The clouds are observed to rotate much faster (*ca.* 4 days) in the equatorial regions than the planetary rotation rate (once per year); this 'super rotation' is one of the aspects of the Venusian atmosphere that is not yet well understood.

In terms of the space programme, there have been missions to study the atmosphere (Mariner, Pioneer, Venera), map below the clouds using radar (Magellan), and landers (Venera) have transmitted pictures from the surface (see figure 2).

Mars, on the other hand, is smaller than the Earth and further from the Sun. Isotopic ratios between radiogenic ^{40}Ar and primordial ^{36}Ar indicate that only *ca.* 20% of the gas in the rocks has been evolved into the atmosphere because of reduced tectonic activity (small size). Also, due to the size, much of the early atmosphere was lost and the present atmospheric pressure is less than 1% of that on Earth. Substantial carbon-dioxide ice deposits are present at the poles, but there is not enough in the atmosphere to cause a greenhouse effect. Oxygen isotopic ratios show that there must be a source of oxygen, perhaps frozen sub-surface water, that increases the proportion of the lighter isotope, which would otherwise preferentially be lost to space, to the observed level. On the other hand, images from Viking and Mars Global Surveyor contain evidence that liquid water flowed on the Martian surface about 3.8 billion years ago. Mars is presently much colder, dryer and much less hospitable to life than it once was.

Some serious scientists propose that greenhouse gases could be introduced into the Martian atmosphere to warm the planet and release some of the trapped water and carbon dioxide, ultimately giving a hospitable environment for humans. This is an interesting idea in theory and a good target for computer simulations. In the opinion of this author, terraforming would be the ultimate in cosmic vandalism if implemented.

The Earth was at the right place in the Solar System and was the right size for life to evolve. The presence of liquid water on the surface meant that carbon dioxide could be dissolved in rocks as carbonates; some of this is recycled due to volcanism. As life developed, photosynthesis became important, leading to the production of oxygen and the fixing of some of the carbon in the biomass. Enough oxygen in the atmosphere led to production of stratospheric ozone, which allowed the protection of land-based life forms from harmful EUV radiation.

The Earth is also the planet we know the most about. Looking at the Earth from space gave us a new perspective: an enhanced feeling that the Earth is special and indeed fragile. The average temperature of the Earth's surface is close to the triple point of water where solid, liquid and vapour may all exist. That is part of how we came to be here.

Our Moon is the first planetary satellite in terms of proximity to the Sun. Its density is much lower than the Earth's and there is effectively no atmosphere. The cratering record is, therefore, well preserved, but the maria show that volcanic activity was important after the early bombardment and up to about 3.2 billion years ago. Despite intensive study by spacecraft (Luna, Ranger, Surveyor, Apollo, Clementine, Lunar Prospector), the origin of the Moon has not yet been determined from the competing theories (simultaneous formation, catastrophic impact on early Earth, capture).

The satellites of Mars, namely Phobos and Deimos, may be captured asteroids based on their physical characteristics. However, the understanding of the dynamics of their capture is by no means solved.

Our planetary neighbours

Despite the many space missions that have explored our planetary neighbours, many important questions remain. Why is Mercury's core so large? Might a catastrophic collision early in its life explain this and its orbital eccentricity and inclination? How is seismic activity affected? Is there ice at Mercury's poles? Why is there super rotation in the Venus atmosphere? What is the surface composition of Venus and Mercury? What is the geological history? How oxidized is the Venusian surface and what is the history of water in the Venusian atmosphere? What changes

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

is humankind making to the Earth's climate, and do these need to be ameliorated? What is the origin of the Moon? Do the Martian atmospheric loss rates to space support the models? Where is the water on Mars now? What is the history of other volatiles? Was there life on Mars? Could and should we terraform Mars?

5. The asteroids

In some sense, the asteroids belong with the inner planets. Many of the asteroids occur in the main belt in between Mars and Jupiter. Some are in other orbits, including orbits that cross the Earth's path. A wide variation of eccentricities and inclinations of the orbits are also present. Spectral studies allow the classification of asteroids into several types: C-type, dark, rich in silicates and carbon, mainly outer main belt; S-type, rocky bodies, mainly inner main belt and Earth crossing; M-type, iron and nickel. A few other asteroids do not fit this scheme. It seems likely that asteroids are the remains of inner-Solar System planetesimals as opposed to being the result of the destruction of a larger body. However, there have been collisions between some bodies since the early bombardment leading to fragmentation and other processing. Collisions with the Earth may have been important. In future, commercial mining of asteroids for minerals may become economically feasible.

Asteroid missions

Why are asteroid types diverse? What is the composition; which, if any, are planetesimals? How pristine? Do they contain interstellar grains from before the Solar System? Is there any water? What is their origin? Which asteroids do meteorites come from? Might they be a future source of raw materials?

6. The outer planets

The outer-planets group contains the gas giants (Jupiter, Saturn, Uranus and Neptune) and the icy object Pluto. The gas giants are heavy enough, and were cold enough when they were formed, to retain the light gases hydrogen and helium from the solar nebula, and these constituents form most of the mass of the planets reflecting the early composition. The visible disk for telescopes and space probes is ammonia and water-based clouds in the atmosphere. At Jupiter, the largest planet and closest gas giant to the Sun, the cloud structure shows a banded and colourful structure caused by atmospheric circulation. The detailed cloud colours are not fully understood. There is no solid surface as such, but models of the internal structure of the gas giants show increasing pressure below the cloud tops, ultimately the pressure becomes so high that a metallic hydrogen layer forms at *ca.* 80% and 50% of the radius, respectively. Dynamo motions in this layer, assuming it must be liquid, power the powerful planetary magnetic fields. A rocky–icy core is thought to be present at *ca.* 25% of the planetary radius.

Jupiter rotates rapidly, providing some energy via the Coriolis force for atmospheric circulation. However, both Jupiter and Saturn have internal heat sources, which mean that they emit 67% and 78% more energy than they receive from the

Sun, respectively. This gives most of the energy for the atmosphere, but the origin of the internal heat source is not fully understood. Models indicate that helium precipitation within the metallic hydrogen core, in which helium is insoluble, may be responsible. There are also strong zonal (east–west) winds near the equator on Jupiter and Saturn, stronger on Saturn where they reach two-thirds of the speed of sound. Their origin is not fully understood. The planets also have important long- and short-lived atmospheric features, of which the most prominent is the great red spot on Jupiter. This long-lived feature, seen for at least 300 years, is surprisingly stable, and so far there is no adequate model to describe it. A similar spot feature appears on Neptune.

In situ results at Jupiter have recently been enhanced by data from the Galileo orbiter and probe. While the orbiter has discovered unexpected dipole magnetic fields in some of the Galilean satellites, the probe has sampled the atmospheric composition, winds, lightning and cloud structures at only one point, which turned out to be a non-typical location in the Jovian atmosphere. One of the discoveries was a lower-than-solar helium abundance, which provides some support for the idea of helium precipitation in the metallic hydrogen layer; a similar conclusion was arrived at based on Voyager data at Saturn. Also, there is less water than expected.

The gas giants each have important and fascinating moons. At Jupiter, Io has the only known active volcanoes other than Earth, providing sulphur-based gases for the Jovian magnetosphere; Europa may have a liquid-water ocean under its icy crust; Ganymede has its own 'magnetosphere within a magnetosphere'; and Callisto has a very old cratered surface. Our knowledge of these has been revolutionized by *in situ* observation; before this, only the albedos and orbital periods were known. At Saturn, Titan is a tantalizing object, planet like in structure and the only moon with a significant atmosphere: 1.5 times the Earth's pressure at the surface. However, its face was shrouded from Voyager's view by organic haze in its thick nitrogen–methane atmosphere. The atmosphere may hold clues about Earth's early atmosphere; there may be methane- or ethane-based precipitation systems; and the ionosphere forms a significant source for Saturn's magnetosphere. Cassini–Huygens will study Titan and some of Saturn's 20 or so icy satellites in detail starting in 2004. At Uranus, the moon Miranda graphically indicates the accretion theories as it appears to be made up of several different types of structure seen elsewhere. Also, the moon system is out of the ecliptic because the spin axis of Uranus, at 98° inclination, is almost in the ecliptic itself. At Neptune, Triton is an icy satellite with a very thin atmosphere but it is in a retrograde orbit and is spiralling closer to Neptune; in tens of millions of years it may break up to produce spectacular rings. It may be similar in characteristics to Pluto and Charon.

Ring systems are present at all the gas giants but spectacularly so at Saturn. Saturn's rings were discovered by Galileo, found to be separate from the planet by Huygens and found to have gaps by Cassini, who also suggested that they were composed of separate particles; this idea was mathematically proved two centuries later by James Clark Maxwell. Detailed exploration was begun by the fly-by missions Pioneer and Voyager, which found remarkable structures including warps, grooves, braids, clumps, spokes, kinks, splits, resonances and gaps. Whole new areas of Solar System dynamics were opened up, including the study of electromagnetic forces that may be important in spoke formation. The rings are less than a kilometre thick, as low as tens of metres in places, and composed of billions of chunks of ice and dust

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

ranging from microns to tens of metres in size. But the main question has not yet been satisfactorily answered: how did the rings form? Was it break-up of a smaller satellite or cometary capture?

And then there is Pluto with its moon Charon. Following an elliptical inclined orbit and currently the furthest planet, Pluto is an icy body rather than a gas giant (Stern 1992). It may be closely related to, but larger than, the icy Kuiper belt objects, the outer Solar System planetesimals, and it may also be related to Triton. Much will be learned by the first spacecraft reconnaissance of the Pluto–Charon system. However, as Pluto goes towards aphelion, its tenuous and interesting methane-based atmosphere will condense and become much less dense. In 2010–2020 it is expected that a rapid atmospheric collapse occurs. There is a good case for getting to Pluto as soon as possible and another excellent case for a visit near the next perihelion in 2237.

Questions for the next millennium on outer planets and their satellites

What causes the cloud colours in the gas giants? Are the internal structure models correct? What causes the internal heat source? Why are the zonal winds so high on Jupiter and, particularly, Saturn? Why are atmospheric features, such as the Jovian great red spot, so stable? Does Europa have water oceans and is life a possibility there? Will the Titan atmosphere evolve further when the Sun becomes a red giant? Why is the Uranian spin axis so tilted and what are the effects? Can the study of Saturn's rings give us more information about the radiation–plasma–dust mixture in the early Solar System? What are the basic characteristics of Pluto–Charon?

7. The comets

Comets are the building blocks of the outer Solar System. Their formation was at the low temperatures prevalent there in the primordial nebula, and they retain volatile material from the early Solar System. They are relatively pristine bodies, making their study important. From the orbits of comets we find that some were formed in the Uranus–Neptune region but were expelled to form the spherical (and theoretical) Oort cloud, with a radius of *ca.* 50 000 AU. The orbits of these distant members of the Solar System may be disturbed by passing stars. They may then plunge into the inner Solar System, where their orbits have random inclinations. Others form the Kuiper belt just beyond Pluto's orbit; their inclinations are close to the ecliptic plane. Whatever their origin, comet nuclei are dirty snowballs, which, when they near the sun, emit gas and dust that form the plasma and dust tails. Halley's comet, for example, loses about a metre of material per orbit and has orbited the Sun about 1000 times, but activity varies significantly from comet to comet.

The space missions in the mid-1980s confirmed that comets have a distinct nucleus, measured gas, plasma and dust composition and led to an understanding of tail formation. The main surprises were the darkness of the nucleus and the jet activity rather than uniform gas and dust emission.

Cometary collisions played a key role in the inner Solar System. Collisions can still occur now, as shown by the Shoemaker–Levy 9 impact with Jupiter in 1994.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Cometary missions

Given their importance in the early Solar System, what is the detailed composition of several comets? Can we bring an icy sample back to Earth for analysis? Is there an Oort cloud? What is the relation to planets? Might comets have brought volatiles to the inner planets?

8. The interplanetary medium and solar wind interactions

The aurora and cometary tails, the only visible clues to the solar wind, have been observed for centuries. Occasionally, the aurora is seen as far south as Britain. Near the dawn of this millennium in 979 AD when Ethelred (the Unready) was pronounced King, a possible occurrence was recorded in the Anglo-Saxon chronicle:

> In this year Ethelred was hallowed king at Kingston (Surrey) on the Sunday, fourteen nights after Easter... In the same year was seen oftentimes a bloody cloud, in the likeness of fire; and that was most apparent at midnight; and was coloured in various ways. Then when it was about to dawn, it glided away ...

Although it sounds ominous he reigned until 1061. As with many other scientific phenomena, the beginning of understanding of the aurora had to wait until the present era. Gilbert's ideas (in 1600) of the Earth as a magnet, Halley's (1698 and 1700) magnetic maps of the Earth from ships and his idea of the aurora being associated with the magnetic field, and George Graham's observation in 1722 of the motion of compass needles, were important early contributions. More recently, in Birkeland's terrella experiments in the early 1900s he fired electrons ('cathode rays') at a magnetized sphere in a vacuum, Appleton and others studied the ionosphere in the 1920s and Chapman laid the foundations of modern solar–terrestrial theory starting in 1930; these all contributed to our current understanding of the solar–terrestrial relationship. However, at the dawn of the space age, the first and completely unexpected scientific discovery from the space programme by Van Allen's team was of the radiation belts of energetic charged particles trapped in the Earth's magnetic field.

Observations of comets in the 1950s had led to the idea of a constantly blowing but gusty solar wind. The solar wind was confirmed by early space probes but it was not until the mid 1980s that the comet–solar wind interaction was understood and backed up by *in situ* data. Solar–terrestrial research, and solar wind interaction with other bodies, remain active areas of research as we approach the millennium. One of the drivers for understanding is the effect that violent solar activity can have on the electromagnetic environment of the Earth and on humankind's technological systems. I have been privileged to be a part of some of the exciting exploratory space missions in these fields.

In its present state, the Sun emits about a million tonnes of material, in the form of plasma, per second. At this rate, it would take 10^{14} years to disintegrate; well before then, in about 5 billion years, the hydrogen fusion fuel will be exhausted and the Sun will become a red giant. In the meantime, the Earth, other planets and comets are all bathed in the solar wind (see figure 3).

The solar wind is highly conducting such that, to a good approximation, the magnetic field is frozen into the flow. Plasma in the solar corona is hot enough to escape

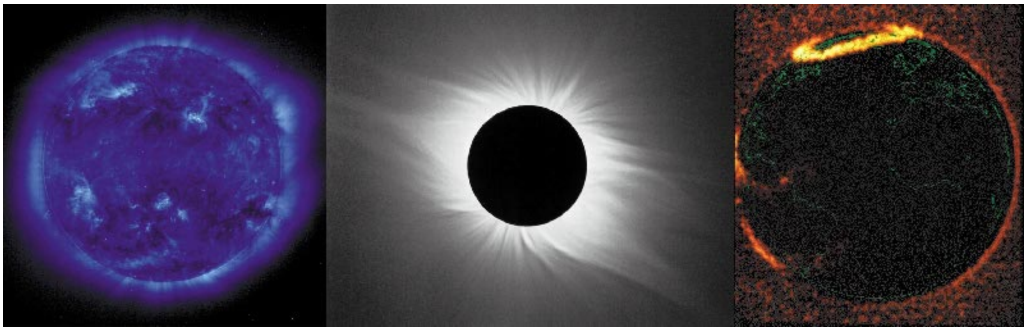

Figure 3. The solar source and terrestrial effect of the solar wind. Left, a SOHO image from 26 February 1998, in the light of eight and nine times ionized iron at 17.1 nm, reveals the magnetic structure in the plasma near to the Sun at a million degrees (Courtesy of EIT team/ESA/NASA). Centre, an eclipse image from the same date showing the corona (copyright Fred Espenak). The eclipse occurred soon after sunspot minimum; coronal holes, characterized by open magnetic field lines, are visible. Right, an ultraviolet image of the Earth from the Dynamics Explorer satellite shows the aurora from space (courtesy of University of Iowa).

the Sun's gravity along the magnetic field, where motions are unrestrained, and it can be shown that the solar wind becomes supersonic within a few solar radii. Beyond this, the speed is almost constant for a particular element of solar wind but the value can vary between *ca.* 300 and 800 km s^{-1}. As plasma expands radially through the Solar System, it drags the solar magnetic field along, but this forms a spiral pattern in space due to the solar rotation. By Earth's orbit the density is *ca.* 5 cm^{-3} but variable and disturbed by coronal mass ejections, shocks and discontinuities. As the solar cycle waxes and wanes, the source region of the solar wind changes; in particular, the coronal holes get smaller (increased magnetic complexity) and larger (decreased complexity), respectively. The electrodynamic environment of Solar System bodies is extremely variable on time-scales of seconds (ion rotation around the magnetic field) up to 11–22 years (solar cycle).

The extent of the solar wind reaches well beyond the planets. Ultimately, a heliopause is required where the solar-wind pressure balances that of the local interstellar medium, at *ca.* 150 AU on the upstream side. Before that, at *ca.* 100 AU, a terminal shock slows the solar wind from supersonic inside to subsonic outside, and upstream of the heliopause a bow shock may form if the LISM motion is supersonic. The terminal shock, heliopause and bow shock are all hypothetical, as the Voyager spacecraft, the most distant human-made object, has not yet crossed these boundaries. However, the inner heliosphere is becoming better understood from several spacecraft, notably Ulysses, which is measuring the structure out of the ecliptic plane in the Sun–Jupiter region for the first time.

The interaction of the solar wind with an obstacle depends, critically, on the obstacle itself: its state of magnetization, its conductivity, and whether it has an atmosphere. We will consider two main types of object: a magnetized planet, such as Earth; and an unmagnetized object, such as a comet (see figure 4).

Other objects include interesting features of both extremes. Mars and Venus are unmagnetized but have some cometary features. Io is conducting, is within the subsonic Jovian magnetosphere, produces a plasma torus, drives huge field-aligned currents to Jovian auroral regions causing light emission there, and supports Alfvèn

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

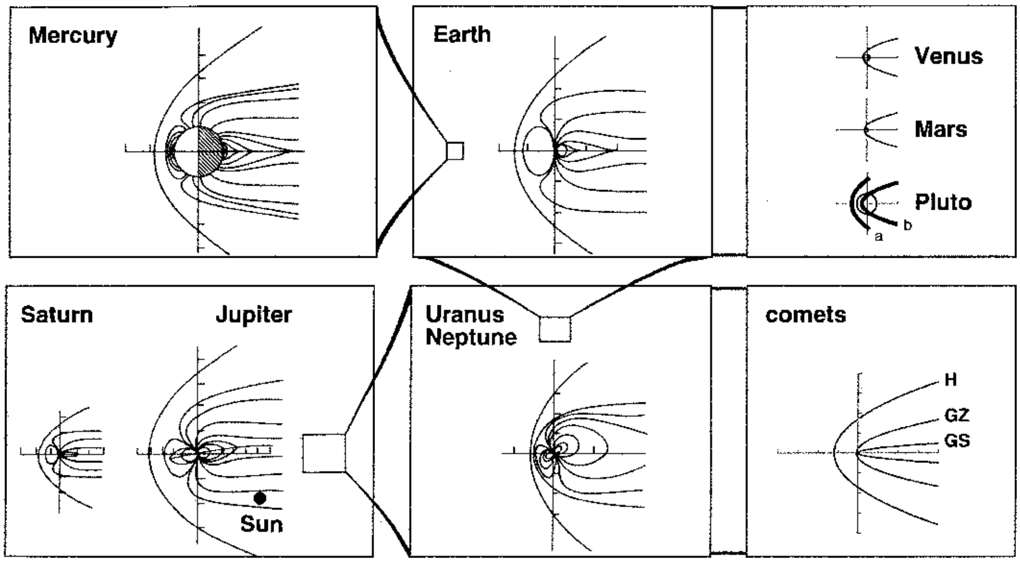

Figure 4. Comparison of the sizes of magnetospheres and non-magnetic interactions with the solar wind (adapted from Kivelson & Russell 1995). The two panels on the right illustrate the bow shocks of non-magnetic objects. In the Pluto panel, Charon's orbit is shown as a circle, and the anticipated shock locations are shown when Pluto is closest to the Sun (a) and furthest from the Sun (b), so that the atmosphere is more tenuous. Bow-shock locations of the three comets visited so far by spacecraft are shown as H (Halley, Giotto), GZ (Giacobini–Zinner, ICE) and GS (Grigg–Skjellerup, Giotto).

wings. Titan has a dense ionosphere, it is usually in the subsonic magnetosphere of Saturn but it is sometimes in the solar wind, and field line draping seems to occur, as at comets.

(*a*) *Magnetized planet interaction*

The discussion concentrates on the Earth but is also relevant to Mercury, Jupiter, Saturn, Uranus and Neptune; differences are caused by different magnetic dipole strengths and orientations, spin rates, and particle sources such as moons and ionospheres and sinks such as rings. In the case of Mercury, there is no ionosphere, so that, at present, we do not understand how the electrical currents close.

Magnetized plasmas do not mix. As the solar wind approaches the Earth, a current sheet, the magnetopause, is set up to separate the regions of solar and planetary magnetic field. The solar-wind particle pressure outside balances the magnetic pressure inside (these are the dominant pressure components), so that the magnetopause is compressed on the day side and extended like a comet tail on the night side. This simple model was formulated by Chapman and Ferraro earlier this century and gives a good prediction for the magnetopause location. Outside the magnetopause, a bow shock stands in the supersonic solar-wind flow. The nature of this collisionless shock changes with magnetic-field orientation as, to some extent, the magnetic field plays the role that collisions play in a fluid shock.

However, the real situation is not this simple. Shocked solar-wind particles can penetrate the magnetopause directly via the funnel-shaped cusp regions on the dayside to cause dayside aurora. Also, if the solar-wind magnetic field and the northward directed terrestrial magnetic field (the magnetic dipole of Earth is northward in the Southern Hemisphere at present) are oppositely directed, then it was realized by Dungey that our nice fluid model is not enough. The explosive process of magnetic reconnection takes place at just the distance scale where the fluid approximation breaks down. This causes solar-wind field lines to be connected to terrestrial ones and these are dragged over the polar cap like peeling a banana. Ultimately, this leads to a build-up of magnetic energy in the tail, further reconnection in the deep tail, and, perhaps, important trigger processes nearer to Earth. The effect is that some plasma is shot down the tail and some towards Earth, causing night-side aurora. The reconnection process, and the electric field across the tail caused by the motion of the solar wind relative to the magnetized Earth, sets up a convection system.

Another convection system is caused by plasma corotating with the planet. The atmosphere and ionosphere corotate with the Earth, as does the inner part of the magnetosphere (the plasmasphere). The overall circulation in the Earth's magnetosphere is given by the sum of the corotation electric field and the convection electric field. As we can see, the convection part of this is extremely dynamic and is still the focus of intense research.

Measurements from ground and space are used together to diagnose the near-Earth environment. Despite the success of early satellites in mapping the various regions, one of the major difficulties of space measurements has been the use of single or dual satellites and the consequent space–time ambiguity. If one or two satellites see a signal due to a boundary, how can we know if the boundary has moved over the satellite or vice versa? The only way to resolve this is with more satellites. At present, this is possible on the large scale with the ISTP (International Solar-Terrestrial Physics) fleet. Satellites upstream of the Earth monitor the Sun and solar wind (SOHO, ACE, Wind), while satellites in the magnetosphere (Polar, Interball, FAST and others) and in the tail (Geotail) monitor the overall effects.

This combination has proved to be extremely useful in following coronal mass ejections from the Sun, through the interplanetary medium and to the Earth. In one case, the consequent increase of radiation belt particles (via an as-yet unknown process) may have caused a commercial telecommunications satellite to fail. This was the first time that 'space weather' has been measured while causing catastrophic effects. As electronics integration and our dependence on technology increases these effects will become more important still. Effects are felt on satellites and in electricity cables and oil pipelines where currents are induced.

On a small scale, the Cluster mission will be the first to make progress here. This is a group of four spacecraft for launch in 2000 following an abortive attempt on the first Ariane 5 in 1996. They will fly in tight formation, starting a few hundred kilometres apart, in a polar orbit through the magnetosphere. At one part of the orbit the spacecraft will be at the corners of a tetrahedron. With two spacecraft we can measure along a line in one dimension, with three we measure a two-dimensional plane, but with four spacecraft we can measure in three dimensions. For the first time, the space–time ambiguity will be resolved and it will be possible to measure vector gradients. This gives the exciting prospect of being able to measure the parameters in Maxwell's equations of electromagnetism directly and to do some real plasma physics with the results. The reconnection process, cusp entry processes and tail dynamics will be directly measured.

At Saturn, the Cassini mission will orbit the planet for four years. This will give us more information about the similarities to and differences from Earth's magnetosphere. Saturn's magnetosphere is controlled by the solar wind but corotation is more important, the magnetosphere is bigger so time-scales are longer, and the rings and moons make important differences. This will be one mission that will contribute to studies of comparative magnetospheres and will help to tell us more about our own.

For the future, even the large-scale ISTP constellation and the small-scale three-dimensional Cluster measurements will still give a significant undersampling in near-Earth space. This is now starting to be addressed by two techniques: magnetospheric imaging and multi-spacecraft (30+) missions. These will be important techniques in the coming decades.

Questions on interplanetary medium and magnetized planet interaction

How do solar and interplanetary events relate? Will it be possible to forecast events at Earth, particularly the sign of the interplanetary magnetic field? Is there a terminal shock, heliopause and bow shock far outside the planets? What triggers reconnection? What causes energy release in the tail? What do different time-scales and other differences at the planets teach us about Earth's magnetosphere? How are radiation belt particles accelerated? Can we develop better protection techniques for our satellites? How do field-aligned currents in Mercury's magnetosphere close? Is it important that the planet occupies much of the inner magnetosphere? How does this magnetosphere work? How do the ecliptic plane orientation of the Uranus geographic axis, the 60° tilt of the magnetic axis with respect to this, and the rapid rotation rate affect the Uranian magnetosphere? How is the planetary magnetic field produced at Mercury, Uranus and Neptune? How important to the survival of life on Earth was the protection from solar wind and cosmic radiation afforded by Earth's magnetic field?

(*b*) *Comet–solar wind interaction*

A comet interacts with the solar wind in quite a different way to a magnetized planet. The cometary nucleus is not magnetized and is a 'dirty snowball'; in the case of comet Halley the nucleus is some 15 km by 8 km, as determined by the Giotto spacecraft. When the nucleus warms near to the Sun, gas (mainly water vapour) sublimes away at *ca.* $1\ \mathrm{km\ s^{-1}}$, carrying dust with it. Comet Halley, for example, produces *ca.* $20\ \mathrm{t\ s^{-1}}$ while Grigg–Skjellerup, Giotto's weaker target, produced only $200\ \mathrm{kg\ s^{-1}}$ at the encounter time. The gas may ionize in sunlight or by charge exchange with the solar wind on a time-scale of a week or so. When ionized, the new-born ion interacts with the electric and magnetic field in the solar wind.

The new ion is first accelerated along the electric field and then gyrates around the magnetic field. The motion of the particle is a cycloid, the same path taken by the valve on a bicycle wheel as it moves. The motion is well known to plasma physicists as '$\boldsymbol{E} \times \boldsymbol{B}$ drift', and the speed of the centre of the motion is given by the ratio of the electric to magnetic fields. The motion in real space is equivalent to a ring in velocity space, centred on the drift speed in a direction along the magnetic field.

The ring causes plasma instabilities which excite Alfvèn waves moving predominantly upstream along the magnetic field. We were able to show both experimentally and theoretically that due to energy conservation, energy from the particles is given to the waves, causing the particle distributions to follow bispherical shells in velocity space, centred on the upstream and downstream waves. This can cause particle acceleration as well as deceleration and, contrary to expectation, comets are good particle accelerators.

The solar wind is slowed due to the added mass (it is 'mass loaded'). This leads to draping of the magnetic field around the comet, as predicted by Alfvèn, as the plasma is frozen to the flow. If the slowing proceeds rapidly enough, a bow shock is formed in the flow. This was observed at the three comets visited by spacecraft so far, although, in some cases, the boundary was a 'wave' rather than a shock. Due to the cometary ions whose gyration radius is much larger than that of the solar-wind particles, cometary bow shocks are the most complex in the Solar System.

The regions inside the bow shock gave several surprises and boundaries not predicted by models. Nearest to the nucleus, however, a predicted boundary appeared where the magnetic field plummeted to zero: the cavity boundary. The number of cometary ions was so high here that the magnetic field was excluded completely.

Following the spacecraft encounters, we have a detailed understanding of some of the physics at work in this interaction. The obstacle to the solar wind is very diffuse and dependent on the outgassing rate of the nucleus and its position in the Solar System. At small comets, we even discovered non-gyrotropic ring distributions.

Since the encounters, an exciting discovery was made of X-rays produced by comets, and it appears that the best explanation is due to the decay of excited states formed due to charge exchange between heavy solar-wind ions and the cometary ions.

Questions on solar wind–comet interaction

How permanent are the various plasma boundaries? Exactly how is the solar wind slowed by the mass loading? Is particle acceleration from the diffusing cometary ions enough or are other mechanisms needed? How does the cometary tail form, how does the magnetic cavity connect to it and what is the importance of tail rays seen in remote observations? Could we fly along a comet's tail to better understand it? What can comets tell us about instabilities in Earth-bound fusion machines or about astrophysical phenomena such as supernova explosions?

9. Future exploration

During our review, we identified some important questions to be answered for each class of body or region in the Solar System. In order to answer these questions, and others that will arise as some are answered, it will be necessary to continue space exploration far into the next millennium. Remote sensing techniques from the ground or from Earth orbit are unlikely to have sufficient resolution, the ability to penetrate clouds at the target, or the ability to see the far side of objects. In addition, the *in situ* measurements of plasma, dust, composition and direct sampling cannot be done remotely at all. Most of the questions and studies highlighted here play an important part in answering why humankind has evolved here. Some are directly related to the possible existence of life elsewhere. Answering the questions is, thus, of important cultural value as well as purely scientific curiosity.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 1. *In situ Solar System exploration*

object	past missions	stage	future missions (approved)
Mercury	Mariner 10	1	ESA cornerstone, Messenger
Venus	Mariner, Pioneer Venus, Venera, Vega, Magellan	3	—
Earth	many	n/a	many
Moon	Luna, Ranger, Surveyor, Zond, Apollo, Clementine, Lunar Prospector	4	Lunar-A, Selene, SMART-1
Mars	Mars, Mariner, Viking, Phobos, Pathfinder, Global Surveyor	3	Surveyor 98, 01, 03, 05, Nozomi, Mars Express
Jupiter	Pioneer, Voyager, Galileo, Ulysses	3	Europa Orbiter
Saturn	Pioneer, Voyager	1	Cassini–Huygens
Uranus	Voyager	1	—
Neptune	Voyager	1	—
Pluto	—	0	Pluto–Kuiper Express
asteroids	Galileo, NEAR	1	DS 1, Muses-C
comets	ICE, Sakigake, Suisei, VEGA, Giotto	1	Stardust, Rosetta, CONTOUR, DS4, Deep Impact
Sun & i/p medium	WIND, ACE, ISEE, AMPTE, SMM, Yohkoh, SOHO, TRACE	n/a	Cluster, Genesis, Solar-B, Solar Probe

Table 1 shows a list of the past and future missions approved over the next decade or so. The natural sequence of Solar System missions involves four stages: (1) initial reconnaissance by fly-by; (2) detailed study by orbiter; (3) direct measurement of atmosphere or surface via entry probe; and (4) sample return. The stage we have reached for each body is also shown in table 1. The approved programme includes a mission to pursue the exploration of Mercury; an important series of missions to Mars culminating in *in situ* searches for life and sample returns to Earth; the exploration of two possible future sites for natural life, namely Europa and Titan; an in-depth exploration of Saturn's system; the first reconnaissance of the Pluto system; asteroid and comet landers and sample return missions; solar-wind sample return; and the first near-Sun (four solar radii) fly-by. This is an exciting and vibrant programme. But is the current drive, forced by budgetary necessity, for 'small, fast, cheap' robotic missions, working? One is tempted to answer yes, as every mission in the table is answering important questions and new missions can be proposed rapidly in response to new discoveries. The payload carried by each of the missions is limited, however, and only a few questions will be studied by each.

However, some of the missions (Cassini in particular) are not of the 'smaller, faster, cheaper' type. The strength of the Cassini mission is in its multidisciplinary approach: we will only understand the complexities of Titan's atmosphere by using several techniques, the *in situ* Huygens probe and several measurements from the orbiter. The probe will only measure at one place, and, as seen at Jupiter, it may not be representative. We must be careful, therefore, to include some larger missions with a balanced payload in the future programme. For possible missions to the outer planets in particular it would be better to send larger, multidisciplinary missions. Another approach could be multi-spacecraft, multi-purpose missions. The smaller, faster, cheaper philosophy only really seems relevant in the inner Solar System.

One problem in Solar System exploration is the time taken to get to the target. Cassini took eight years to be built and the flight time to Saturn is another seven. While some opportunistic science is possible—in this case at Venus, Earth, Jupiter and in the distant solar wind—it will take a significant proportion of the careers of the scientists involved to arrive, those that are not retired first. To get a reasonable payload to Pluto would take much longer than the seven years foreseen for the small mission with severely limited payload that is the Pluto–Kuiper Express.

If missions within the Solar System take a long time to reach the destination, is it realistic to consider missions beyond this? Using current spacecraft technology, assuming the same speed as Voyager (3.3 AU per year), missions to the heliopause at 150 AU would take over 50 years, the Oort cloud at 50 000 AU would take at least 15 000 years, and to Proxima Centauri, our nearest star at 4.2 light years away, would take *ca.* 80 000 years. It is clear that better trajectories and advanced propulsion systems would be needed for the more accessible missions, and remote sensing is a better technique for the stars. Within and closer to the Solar System, ion propulsion or inner-Solar System light-driven designs may be appropriate. Another problem for these remote missions will be the provision of electrical power so far away from the Sun. The only solution appears to be nuclear power.

As we find out more about each object or region, further questions will be raised. It will also be necessary to explore further afield. Even if firm evidence is found for past life on Mars we will need to understand how common the occurrence of life is in the universe, hence the search for extra-solar planets. We need to understand our own Solar System properly before extrapolation is possible.

We are making good use of space for peaceful meteorological, communications and positioning reasons on Earth. Uses of space may also reach further into the Solar System in the new millennium: for example, there may be economic sense in asteroid mining. Another example is mining the moon for helium. Because the moon is unmagnetized, the solar wind can impact it directly. The magnetic field diffuses relatively rapidly but the particles are buried in the regolith. Over billions of years, there may be enough ^{3}He buried to make mining of this isotope worthwhile for use in future fusion reactors on Earth. Another possible, but in my view unpalatable, application is planetary engineering or terraforming.

One controversial issue is the direct involvement of humankind in *in situ* exploration. So far, only a handful of people have travelled beyond 400 km and only a few hundred beyond 100 km from the Earth's surface. However impressive the missions and brave the people involved, the reasons for sending them are to do with politics and public relations. These were associated with national pride during the cold war and science was an afterthought. Lunar samples were returned by Russian robots as well as by the Apollo programme, although, admittedly, much less massive. Humankind as a whole can be proud of the achievements but it seems unlikely that the same political will or economic priority will exist to propel people far beyond low Earth orbit.

On the other hand, robots have successfully explored all the planets except Pluto, and the furthest human-made object, Voyager 1, is over 70 AU away and counting. Do we need the encumbrance of life-support systems? Can we accept human error in space? Would space travellers be adequately protected from radiation away from the protection of our magnetic field except in an extremely thick-walled spacecraft? Is there the economic or political will or necessity to support expensive crewed exploration? If computer and virtual reality techniques continue to accelerate in development do we need to send people? From a purely scientific perspective the answer is no.

Another obstacle for any kind of mission is launch cost. This is the prime reason that crewed exploration is likely to stay in Earth's vicinity, and is also a severe limitation on robotic exploration. Launches more frequent than the present rate will have to await technological developments on reusable rockets, aerospike engines and perhaps air-breathing engines. Studies are currently underway in the United States with the X-33 demonstrator being built, and it seems likely that this technology will become usable in the next decade or so.

Robotic space exploration should be continued as rapidly as possible. This tool, which has become available within the last half century, has already proved its worth and should be used for important scientific purposes. Robotic exploration is much better value for money.

We may speculate on possible targets for robotic missions for the new millennium as follows: constant monitoring of the Sun and solar wind as part of an integrated space weather forecasting system; constant monitoring of weather on Mars, Venus and Jupiter to improve models; detailed explorations of Mercury and Pluto; return to sites of earlier exploration with better instruments, new ideas and atmospheric probes in several places; sample return from nearby Solar System bodies following detailed mapping and *in situ* composition measurements; exploration of Uranus and its extraordinary magnetosphere; explorations of outer planetary moons; investigation of the feasibility of asteroid mining; terminal shock, heliopause, heliospheric bow-shock exploration; Oort-cloud exploration; investigation of the feasibility of sending spacecraft to nearby stars and planetary systems.

In summary, there are many exciting and challenging possibilities for future Solar System exploration. The answers to be gained are fundamental to a better understanding of our place in the universe.

I thank the Royal Society for support as a University Research Fellow. I am grateful to Alan Elliot for providing the quote from the Anglo-Saxon Chronicle, and to Martin dela Nougerede at MSSL for help with the illustrations. I thank the Editor for his patience in awaiting the manuscript.

References

Beatty, J. K., Petersen, C. C. & Chaikin, A. 1999 *The new Solar System*. Cambridge University Press.

Kivelson, M. G. & Russell, C. T. (eds) 1995 *Introduction to space physics*. Cambridge University Press.

Lewis, J. S. 1997 *Physics and chemistry of the Solar System*. Academic.

Stern, S. A. 1992 The Pluto–Charon system. *A. Rev. Astron. Astrophys.* **30**, 185–233.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

AUTHOR PROFILE

A. J. Coates

Andrew Coates, born at Heswall, Cheshire, gained a BSc (first) in physics from UMIST in 1978, and MSc (1979) and DPhil (1982) in plasma physics from Oxford University. He has been at UCL's Mullard Space Science Laboratory since. He was a Royal Society University Research Fellow and is now a Reader in Physics. He was guest scientist at MPAe, Germany, and the University of Delaware, USA, and was a media fellow at the BBC World Service. Space mission involvements include AMPTE, Giotto, Meteosat, Cluster and Cassini, where he led the team providing the electron spectrometer (shown in photograph). Aged 42, he is on several PPARC committees and the ESA Solar System Working Group. Scientific interests include the solar wind interaction with planets and comets and space instrumentation; he has over 100 publications. Interests include popularizing science, skiing and helping to bring up twin daughters.