The ori
evolutio
solar s

M Woolfson discusses theories of

No entirely satisfactory theory has been proposed for the origin of the Solar System, a problem of increasingly wider interest with the discovery of other planetary

The heliocentric nature of the solar system with its major components – the Sun, planets and satellites – was firmly established well before the end of the 17th century. After the publication of Newton's *Principia* in 1687 it became possible to apply scientific principles to the problem of its origin.

Most theories that have been advanced in the last 300 years are obviously untenable, but some contain the germs of what might be part of a viable theory. It would not be practical to attempt to deal with all theories in detail in a short review article. Here we shall mention five theories, recently developed or still in the process of development, that have a reasonable scientific basis. Two of them, the Solar Nebula Theory and the Capture Theory, will be described in more detail, emphasizing what they have and have not explained and what their remaining difficulties are. Two early theories will be described first, chosen because they relate closely to the extant ones and illustrate the major problems for theories.

Early theories

Based on ideas and observations by Descartes, Kant and Herschel, Pierre Laplace (1796) put forward the first really scientific theory (summarized in figure 1). A slowly spinning cloud of gas and dust cooled and collapsed under gravity. As it collapsed, so it spun faster and flattened along the spin axis. It eventually took on a lenticular form with equatorial material in free orbit around the central mass. Thereafter material was left behind as a set of rings within which clumping occurred. Clumps orbiting at slightly different rates combined to give a protoplanet in each ring. A smaller version of the scenario, based on the collapse of protoplanets, produced satellite systems. The central bulk of the original cloud collapsed to form the Sun.

This *monistic* theory, that produced the Sun and the planets in a single process, has an attractive simplicity but a fatal flaw. It suggests that most of the angular momentum of the system is in the Sun – which is not so. The Sun with 99.86% of the mass of the system has only 0.5% of the total angular momentum contained in its spin; the remainder is in the planetary orbits. All 19th century attempts to rescue the theory were unsuccessful. The theory, although based on scientific principles, did not agree with observation and so had to be abandoned.

Some time later James Jeans (1917) suggested a *dualistic* theory, one for which the Sun and planets were produced by different mechanisms. A massive star passed by the Sun, drawing from it a tidal filament (shown in figure 2). The gravitationally unstable filament broke up with each condensation forming a protoplanet. The protoplanets, attracted by the retreating star, were retained in heliocentric orbits. At first perihelion passage a small-scale version of the same mechanism led to a filament being drawn from a protoplanet within which protosatellites formed.

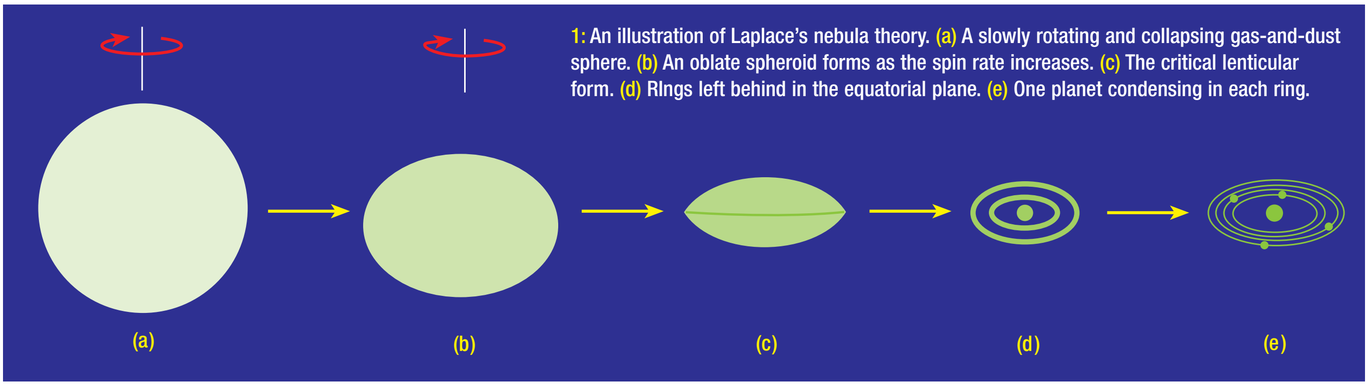
The theory had a good reception – especially as it was supported by some elegant analysis. Jeans found how a tidally affected star would distort and eventually lose a filament of material from the tidal tip. He showed that the filament would fragment through gravitational instability and he also derived a condition for the minimum mass of a filament clump that could collapse. Despite the initial enthusiastic acceptance of the theory, it soon ran into trouble. Harold Jeffreys (1929), by a mathematical argument involving the concept of circulation, suggested that Jupiter, which has the same mean density as the Sun, should have a similar spin period. The periods differ by a factor of 70. Other simpler, and hence more readily accepted, objections followed. Henry Norris Russell (1935) showed that material pulled from the Sun could not go into orbit at more than four solar radii – well within Mercury's orbit. This was another type of angular momentum problem. Then Lyman Spitzer (1939) calculated that a Jupiter mass of solar material would have a temperature of about 10^6 K and would explode into space rather than collapse. Later, other objections were raised concerning the presence of lithium, beryllium and boron in the Earth's crust, light elements that are readily consumed by nuclear reactions in the Sun.

Jeans tried to rescue his theory by having a cool extended Sun with the radius of Neptune's orbit, but this created new problems – not least that the newly formed planets in a diffuse form would be ploughing through the Sun. He finally conceded that "the theory is beset with difficulties and in some respects appears to be definitely unsatisfactory".

The Laplace and Jeans theories were scientifically based but finally succumbed to scientific criticism. They both had angular momentum problems although of different kinds. Nevertheless all the modern theories described here involve ideas that they introduced. They also illustrate important problems that theories must address to be considered as plausible.

What is a good theory?

Those producing cosmogonic theories usually provide lists of "facts to be explained" but, as the scientific historian Stephen Brush concluded, such lists often emphasize those facts that the individual's theory best deals with. This could well be true. To avoid that possibility, I



1: An illustration of Laplace's nebula theory. (a) A slowly rotating and collapsing gas-and-dust sphere. (b) An oblate spheroid forms as the spin rate increases. (c) The critical lenticular form. (d) RIngs left behind in the equatorial plane. (e) One planet condensing in each ring.

gin and
n of the
ystem

how the Sun and the planets began.

systems. Here I give an account of the development of ideas in the field of solar system evolution, matching the evolution of models to the discoveries of the key evidence.

give below the union of *all* “facts” suggested by various workers. They are separated into groups according to whether they are gross features or relate to details of the system.

Gross features:

- the distribution of angular momentum between Sun and planets
- a planet-forming mechanism
- planets to form from “cold” material
- direct and almost coplanar orbits
- the division into terrestrial and giant planets
- the existence of regular satellites.

Secondary features:

- the existence of irregular satellites
- the 7° tilt of the solar spin axis to the normal to the mean plane of the system
- the existence of other planetary systems.

Finer details of the solar system:

- departures from planarity of the system
- the Earth–Moon system
- variable directions of planetary spin axes
- Bode’s law or commensurabilities linking planetary and satellite orbits
- asteroids: origin, compositions and strutures
- comets: origin, compositions and structures
- the formation of the Oort cloud
- the physical and chemical characteristics of meteorites
- isotopic anomalies in meteorites
- Pluto and its satellite, Charon
- Kuiper-belt objects.

The least that a theory should deliver is convincing explanations of the gross features. A theory without a slowly spinning Sun and a planar system of planets with regular satellite systems for some is, at best, implausible.

If alternative plausible theories are available then one may resort to the principle first enunciated by the English philosopher William of Occam (1285–1349), known generally as Occam’s razor. Loosely translated from the Latin this implies that “if alternative theories are available that explain the observations equally well then the simpler is to be preferred”.

The goal then is to find a simple theory based on well-established scientific principles, that explains what is known and that cannot be refuted by scientific arguments. We shall now look at the ideas that have been put forward over the last half century, roughly in their date order of presentation.

The accretion theory

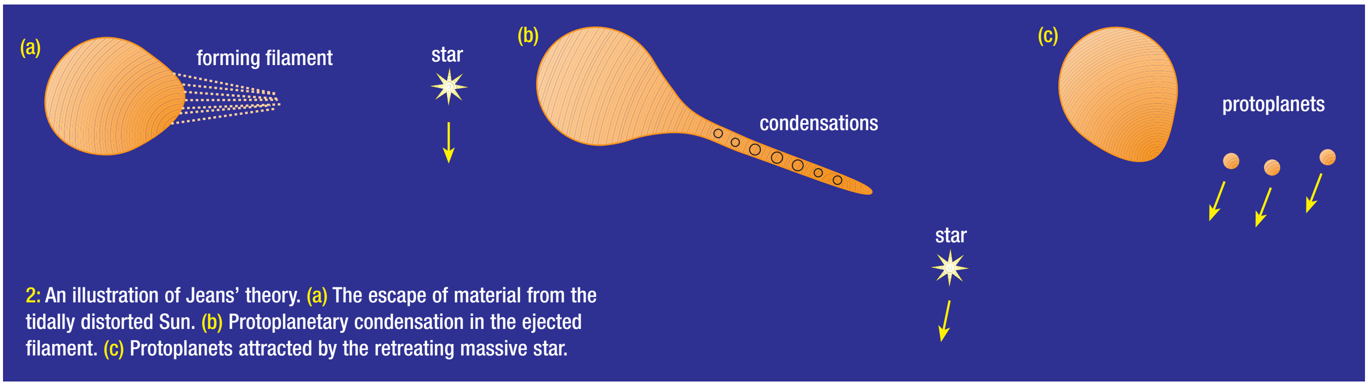
In 1944, Soviet planetary scientist Otto Schmidt suggested a new kind of dualistic theory. It was known from telescopic observations that cool dense clouds occur in the galaxy and Schmidt argued that a star passing through one of these clouds would acquire a dusty-gas envelope. Schmidt believed from energy considerations that, for two isolated bodies, material from one body could not be captured by the other and so he introduced a third body nearby, another star, to remove some energy. The need for a third body made the model rather implausible but, as Lyttleton showed in 1961, Schmidt’s argument was invalid since the cloud was of large extent and the star-plus-cloud behaved like a many-body system. Lyttleton proposed capture of material by an accretion mechanism first suggested by Bondi and Hoyle (1944) and illustrated in figure 3. The cloud material moves relative to the star at speed *V*, greater than the escape speed. Deflected interacting streams, such as at point G, lose their component of velocity perpendicular to the original direction of motion and the residual speed can then be less than the escape speed.

Lyttleton used parameters for the model that gave the mass and angular momentum of captured material compatible with that of the planets, although no process was suggested for producing planets from the diffuse envelope. However, Lyttleton’s parameters were implausible. The temperature of the cloud was 3.18 K, in equilibrium with galactic radiation, and the relative velocity of cloud and star was 0.2 km s^{-1}. A cloud temperature of 10–20 K or even greater is more consistent with observation, and the relative speed is more likely to be of order 20 km s^{-1}. The proposed mechanism does no more than suggest a source of planetary material. It cannot be regarded as a convincing theory, especially as planet formation from diffuse material presents additional difficulties, as we shall see later.

The floccule/protoplanet theory

In 1960, McCrea suggested a theory that linked planetary formation with the production of a stellar cluster and also explained the slow rotation of the Sun. McCrea’s starting point was a cloud of gas and dust that was to form a galactic cluster. Due to turbulence, gas streams collided and produced regions of higher-than-average density. The high-density regions, referred to as “floccules”, moved through the cloud and combined whenever they collided. When a large aggregation formed, it attracted other floccules in its region so producing a protostar. Since floccules joined the accreting protostar from random directions, the net angular momentum of the protostar was small; for a particular set of parameters it would be only a few times the present angular momentum of the Sun and the excess can be removed after formation by various physical processes.

It was assumed that star-forming regions were isolated and McCrea showed that the angular



2: An illustration of Jeans’ theory. (a) The escape of material from the tidally distorted Sun. (b) Protoplanetary condensation in the ejected filament. (c) Protoplanets attracted by the retreating massive star.

momentum contained in a region due to the original floccules was much greater than that residing in the protostar. The missing angular momentum was assumed to be taken by smaller aggregations of floccules that were captured by the protostar to form a set of planets.

In the original form of the theory, each floccule had about three times the mass of the Earth so many of them had to combine to form the giant planets. The resultant planetary aggregations contained much more angular momentum than the present planets. McCrea turned this apparent problem into an asset. As the protoplanet collapsed it would have become rotationally unstable and behaved as described by Lyttleton (1960) and shown in figure 4. The protoplanet would have broken into two parts with a mass ratio of about 8:1. The smaller part, moving faster relative to the centre of mass, could escape from the solar system, with most of the angular momentum. In a neck between the two separating parts, small condensations would form and be retained by the larger part as a satellite family. To explain the terrestrial planets, McCrea had to assume that the fission process took place in a dense core of the protoplanet. In the inner part of the solar system, with higher escape speeds, both parts were retained and formed the pairs Earth–Mars and Venus–Mercury.

With some parameters deduced from the present solar system and others chosen to give the best possible results, the Sun plus planets and satellites system could be explained. Nevertheless the theory has severe problems. First, the floccules were unstable, with lifetimes much less than the time between floccule collisions. In response, McCrea (1988) produced a modified form of the theory where the initial condensations, now called "protoplanets", were of Saturn's mass and stable. The initial system would not have been coplanar and indeed there could have been retrograde orbits although, with motion in a resisting medium and collisions to remove a minority population of retrograde objects, the system could have evolved to the present state. However, what is highly suspect is the idea that the angular momentum not present in the protostar must necessarily reside in a planetary system. It is much more likely that the "missing" angular momentum would reside in relative motions of protostars than in planetary systems.

The Solar Nebula Theory

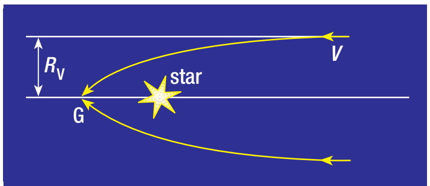
Over the past 30 years a paradigm has arisen – a model that has wide acceptance and is the basis of thinking about contingent matters. This is the Solar Nebula Theory (SNT).

In the 1960s it became clear that many features of meteorites were interpretable in terms of condensation from a hot vapour, encouraging the view that early solar system material had been in a hot gaseous form. In addition, in the 1960s Victor Safronov was working on planet formation from diffuse material and in a seminal paper translated into English (Safronov 1972) he summarized this work. Driven by these twin developments a new Solar Nebula Theory (SNT) quickly took off as a major research activity. It was believed that new knowledge and approaches should enable the original problems of Laplace's nebula theory to be solved.

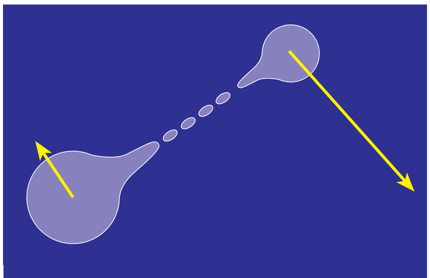
An early worker on the SNT concluded quite quickly: "At no time, anywhere in the solar nebula, anywhere outwards from the orbit of Mercury, is the temperature in the unperturbed solar nebula ever high enough to evaporate completely the solid materials contained in interstellar grains," (Cameron 1978). Although this undermined an important *raison d'être* for the revival of nebula ideas, by this time the work was in full flow and proceeded without interruption.

Work on the redistribution of angular momentum has been central in the development of the SNT. Lynden-Bell and Pringle (1974) described a mechanism in which, given turbulence and energy dissipation in a disk, the disk would evolve to conserve angular momentum by inner material moving inwards while outer material moved outwards. This is tantamount to the outward transfer of angular momentum. However, it does not solve the basic angular momentum problem. Material joining the central condensation gradually spirals inwards so that it is always in a near-Keplerian orbit around the central mass. A useful way of thinking about the spin angular momentum of the Sun is to equate it to one-quarter of a Jupiter mass orbiting at the Sun's equator. If the Sun could form in its present condensed configuration by material spiralling inwards, which it could not, then it would still have hundreds of times its present angular momentum. Realistically, without having much less angular momentum it could not form at all. Various mechanisms have been suggested for transferring angular momentum (Larson 1989). An example is by gravitational torques due to spiral arms in the disk (figure 5). To be effective this requires a massive nebula, which is undesirable for other reasons, but *any* mechanism giving a spiralling motion for material does not solve the problem.

An effective mechanism for removing angular momentum from a pre-existing star involves a loss of ionized material from the star plus a strong stellar magnetic field, both likely in a young active star. Ionized material moves outwards locked to a magnetic field line. The field rotates with the star so the ionized matter moves outwards with constant angular speed; the increased angular momentum it acquires is removed from the star. It remains attached to the field line until the kinetic pressure of the ion flow exceeds the magnetic pressure that, in the case of a dipole field, varies as r^{-6}. Analysis shows that, with plausible stellar winds and fields, some 90% or so of the original angular momentum can be removed in this way.



3: Streams of material arriving at point G cancel their velocity components perpendicular to the axis.



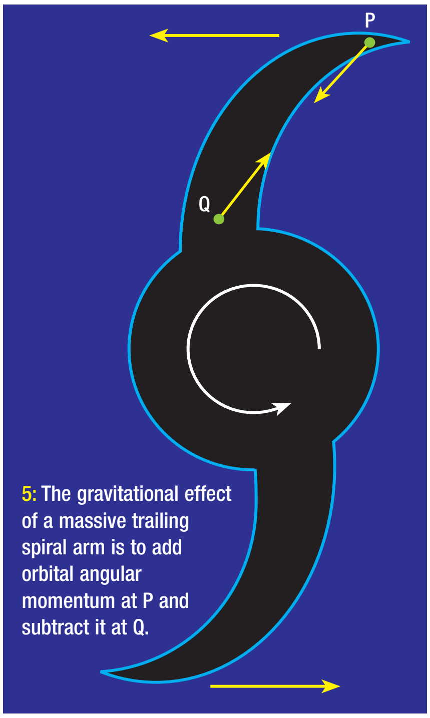
4: The fission of a rapidly spinning protoplanet with the formation of protosatellite droplets.

T-Tauri emission, at the deduced rate of $10^{-7}\ M_\odot\ \text{year}^{-1}$ for a period of 10^6 years, is often cited as a model for mass loss. However, spectroscopic evidence shows that T-Tauri emitted material is only lightly ionized and hence would be feebly coupled to the field. In addition, low-mass stars, for which no T-Tauri emission occurs, also spin slowly so a second mechanism would be needed for these stars.

Forming the Sun requires *inward* movement of material while the magnetic field mechanism for removing angular momentum requires *outward* movement. If a way could be found whereby the nebula core would grow and simultaneously lose highly ionized material which coupled to a strong stellar magnetic field (~10^5 times as strong as the present solar field) then the angular momentum problem would be solved. For example, one could envisage a bipolar inflow of neutral material adding to the mass of the star with an equatorial loss of ionized material to remove angular momentum – although it seems unlikely that such a pattern would arise naturally. To summarize, while it is not possible to say that the angular momentum problem *cannot* be solved, it has certainly not been convincingly solved as yet although general papers on the evolution of disks appear from time to time (e.g. Pickett and Durisen 1997).

Forming planets from a diffuse medium

There are two possible planet-forming scenarios for the SNT. In the first, the nebula disk had about a solar mass and a density and temperature such that regions of it contained a



5: The gravitational effect of a massive trailing spiral arm is to add orbital angular momentum at P and subtract it at Q.

Jeans critical mass and spontaneously collapsed to produce planets. This gives planets, but so many that there is a challenging disposal problem. SNT theorists no longer seriously consider this possibility.

The other scenario is with a disk of mass between 0.01 $M_\odot$ and 0.1 $M_\odot$, similar to that considered by Safronov (1972) whose work has been developed by others. Recent observed infrared excess radiation from young stars is almost certainly due to the presence of dusty disks. These observations, taken as supporting the SNT, also impose a constraint; stars older than a few million years do not show infrared excess radiation. It has been inferred, and generally accepted by the SNT community, that planet production has to be completed within 10 million years of disk formation.

What emerges is a multi-stage process:

(i) Dust within the disk settles into the mean plane. For dust grains as small as normal ISM grains this process would take too long. Weidenschilling *et al.* (1989) suggested that grains were sticky so that large dust particles formed, thus drastically shortening the settling time. There is controversy about the need for sticky dust but general agreement that the dust disk must form in a reasonably short time.

(ii) The dust disk is gravitationally unstable and fragments to form kilometre-size bodies, called “planetesimals”. The early nebula might have had to be turbulent to allow transfer of angular momentum but a quieter nebula is now required to allow the planetesimals to form.

(iii) Planetesimals accumulate to form planets. This is the awkward part of the process. Planets would form in the terrestrial region within 10^7 years but, according to Safronov’s theory, it would take 1.5×10^8 years to produce a Jupiter core and 10^{10} years or more to produce Neptune – more than twice the age of the solar system.

There are conflicting requirements here. Short formation times require a turbulent environment to bring planetesimals together quickly while, for planetesimals to amalgamate, approach speeds must be low. Stewart and Wetherill (1988) suggested conditions that would lead to runaway growth. These include local density enhancements in the disk, viscous forces to slow down planetesimals and the application of an energy equipartition principle so that larger bodies would move more slowly and hence be able to combine more readily. These are *ad hoc* assumptions but reduce formation times to within the allowed period – except for Uranus and Neptune. In the first programme of a recent BBC television series *The Planets*, an SNT theorist said, “according to our theories, Uranus and Neptune do not exist”!

(iv) Planetary cores accrete gaseous envelopes. This would take about 10^5 years for Jupiter.

Satellite formation is taken as a miniature version of planet formation although angular momentum transfer is not such a serious problem in this case. The ratio (intrinsic orbital angular momentum of the secondary body)/(intrinsic equatorial spin angular momentum of the primary body) is 7800 for Jupiter–Sun and 17 for Callisto–Jupiter so that only a modest outwards transfer of angular momentum is required.

Comments and residual difficulties

The difficulties of angular momentum transfer and planet formation have not been convincingly resolved after 30 years of concentrated effort so the SNT *per se* has not progressed beyond these basic problems.

Papers are produced from time to time on planet formation, usually involving special assumptions that are not justified other than that they lead to a desired outcome. For example Pollack *et al.* (1996), by numerical simulations involving the simultaneous accretion of solid planetesimals and gas, gave the formation times of Jupiter, Saturn and Uranus as a few million years. The major assumption they made was that the growing planet was in a disk of gas and planetesimals with uniform surface density and that planetesimals had to remain within the feeding zone of the planet. More recently Chambers and Wetherill (1998) have simulated the formation of terrestrial planets on the assumption of a pre-existing Jupiter and Saturn but, even then, the period covered by the simulation is an unacceptable 3×10^8 years. There is no model for planet formation that has commanded general support from the SNT community which describes a progression from a believable initial condition through a series of well-founded physical processes to planetary formation.

The division of planets into terrestrial and giant categories is related to the temperature of their formation. Mercury is formed where only iron and silicate grains can survive and the Mercury region would have been iron-rich. However, there is no simple explanation for the seemingly erratic pattern of densities of the terrestrial planets. Beyond the orbit of Mars, ice grains would have been stable, so allowing massive planetary cores to form that attracted extensive atmospheres.

On the question of angular momentum transfer the situation is perhaps less favourable than for planet formation. Again papers appear giving rather general results which are not, and cannot be, directly related to the problem of a slowly spinning Sun.

The SNT should yield the solar spin axis strictly perpendicular to the mean plane. An explanation for the 7° tilt could be perturbation by a passing star that disturbed the orbital planes of the planets subsequent to their formation. There are some tricky problems with this explanation. Neptune’s orbit is almost perfectly circular and any perturbation that significantly changed its inclination would also have greatly changed its eccentricity. There is, however, a ready explanation for the tilts of the planetary spin axes. Planetesimals, or larger aggregations, will build up planets by collisions from random directions and spin axes could be in almost any direction, although the preponderance of direct planetary spins may require explanation.

The capture theory

The Capture Theory (CT) (Woolfson 1964) actually predated the advent of the SNT by several years but its arrival was largely unnoticed. The basis of the CT, as first presented, is illustrated in figure 6 which shows a point-mass model, an early one of its kind, in which interpoint forces simulated the effects of gravity, gas pressure and viscosity. It depicts a tidal interaction between the Sun and a diffuse cool protostar, of mass 0.15 $M_\odot$ and radius 15 AU. As Jeans had deduced, the protostar distorts and eventually a filament of material escapes from the tidal tip. The model was too coarse to show filament fragmentation, but individual mass points were captured by the Sun. This model, which involved mechanisms analysed by Jeans, was free of all the criticism that had been raised against the original tidal model. The angular momentum of the planetary orbits comes from the protostar–Sun orbit and the range of perihelia given by the model, up to 38 AU, matches that of planetary orbital radii. Since the material is cold it satisfies the chemical constraints. The orbital planes are close to the Sun–protostar orbital plane although, due to protostar spin throwing material slightly out of the plane, there would be some variation of inclinations.

It was seven years before the next CT paper was published. This paper (Dormand and Woolfson 1971) improved the original model by exploiting the dramatic increase in available computer power. The paper confirmed the validity of the capture process and showed, from several simulations, that the calculated radial distributions of planetary material agreed reasonably well with that in the solar system (figure 7). From the properties of the filament it seemed that six or so protoplanet condensations would be expected. Much later, by the use of a smoothed particle hydrodynamics (SPH) approach, Dormand and Woolfson (1988) modelled filament fragmentation that was found to take place much as Jeans had described.

The modelling showed that the protoplanets began moving towards the aphelia of very eccentric orbits. If the collapse time of a protoplanet was substantially less than its orbital periods (>100 years) then this would enable it to condense before being subjected to disruptive tidal forces at perihelion. The collapse of a Jupiter-like protoplanet, under the conditions of CT formation, was modelled in detail by Schofield and Woolfson (1982). This indicated planetary collapse time as short as 20 years with reasonable model parameters.

Satellite formation

While the planets could survive, they were subjected to considerable tidal forces during their first orbit. Consequently they would go into their final collapse stage in a distorted form that included a tidal protuberance. The characteristic of a nearly free-fall collapse is to amplify any distortion so that what began as a tidal bulge turned into a tongue or filament. Condensations within this filament would give a family of regular satellites. Williams and Woolfson (1983) found good quantitative agreement between predictions based on this model and the properties of the regular satellite families of Jupiter, Saturn and Uranus. Actually, this mechanism is similar to that suggested by Jeans for satellite formation – a small-scale version of his planet-forming process. The Jeans tidal theory had insuperable angular momentum problems for planets but not for satellite formation.

Orbital evolution

Dormand and Woolfson (1974), investigating the effect of a resisting medium around the Sun, found that protoplanet orbits quickly round off. In one simulation, with a medium with five times Jupiter's mass, it was found that Jupiter rounded-off in 10^5 years, Saturn in 3×10^5 years and Uranus and Neptune in 2×10^6 years. The times depend on the density of the medium and were also approximately proportional to the inverse of the planet's mass. They are comfortably less than the inferred lifetimes of disks around young stars if, indeed, the resisting medium acts as a disk.

Table 1: Planets in the early solar system according to the Capture Theory

Planet	Mass ($M_{\odot}$)	Radius (10^3 km)	Semi-major axis (AU)	Eccentricity	Inclination (°)	Rounding time (years)
Neptune	18	28	62.3	0.720	3	2×10^6
Uranus	15	26	35.6	0.690	2.5	2×10^6
Saturn	100	66	18.6	0.680	1.5	3×10^5
Jupiter	330	78	14.6	0.800	2	1×10^5
A	17	27	12.2	0.874	1	2×10^6
B	5.5	21	9.1	0.908	1	6×10^6

The periods of Jupiter and Saturn and those of Uranus, Neptune and Pluto are close to being commensurate. Melita and Woolfson (1996) showed that orbital evolution in a resisting medium leads to resonance locking between pairs of planets. During the evolution of the orbits with energy loss, the periods reach some commensurability. Thereafter an automatic feedback mechanism ensures a difference between the energy lost by the outer planet and its gain of energy from the inner planet such that the resonance is maintained. This does not give Bode's law – but it does explain commensurabilities that have a firmer physical foundation.

Spin axes

The original solar spin axis could have been in any direction. However, during the dispersal of the resisting medium – mostly by being pushed outwards by radiation pressure and the solar wind – larger solid grains would have spiralled inwards due to the Poynting–Robertson effect. As they joined the Sun, their angular momentum contribution pulled the solar spin axis towards the normal to the mean plane. Absorption of a fraction of a Jupiter mass in this way would give the spin axis nearly, but not quite, normal to the mean plane – not a problem for, but a natural consequence of, this model.

The basic CT provides an explanation of the tilts of the planetary spin axes as due to strong tidal interactions between planets that approached closely while their orbits were still highly eccentric. Woolfson (2000) describes a point-mass model of a proto-Uranus with a radius of 0.25 AU in an orbit of semi-major axis 35.6 AU and eccentricity 0.69 interacting with a model Jupiter on an orbit with semi-major axis 14.8 AU and eccentricity 0.826. Jupiter passes over Uranus with nearest approach 1.15 AU and the spin axis of Uranus changes from being normal to the original orbit to being at an angle of 98.7° to the almost-unchanged new orbit. Other planetary spin-axis inclinations are readily explained in this way.

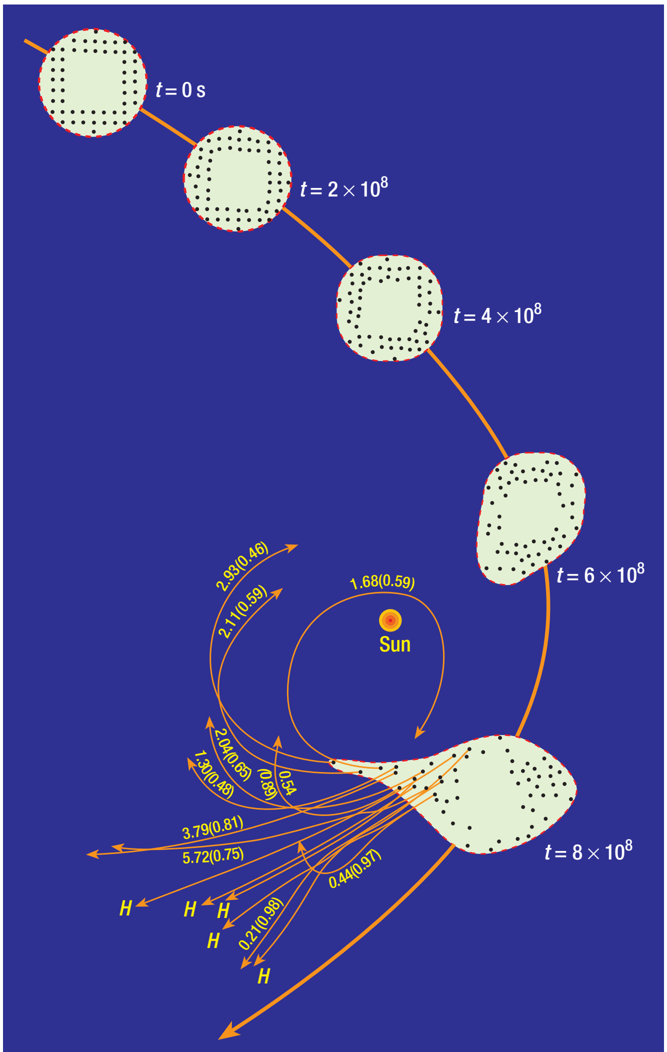
Star formation

The CT is a dualistic one and offers no explanation for the slow solar spin, something that must always be of concern to the cosmogonist. To address this concern, Woolfson (1979) described a model for star formation within a galactic cluster and similar ideas have been investigated by Pongracic *et al.* (1991). The model followed the evolution of a collapsing dark cool cloud within which turbulent energy steadily increased. The collision of turbulent gas elements gave compressed hot regions that cooled much faster than they re-expanded. If the free-fall time of the cool dense region was less than the coherence time for the whole cloud, during which matter was completely redistributed within it, then a star could form. Producing stars this way, with subsequent accretion to form more massive stars, gave spin rates for different classes of stars similar to those observed. Additionally, the rate of star formation and the variation of the masses of formed stars with time agreed with observations from young clusters. The predicted mass index of stars, that gives the stellar mass distribution, also agreed with observation. Given at least one star-forming model that explains solar spin in the context of the spin characteristics of all stars, it is reasonable for a dualistic theory to confine itself to the problem of planetary orbital angular momenta.

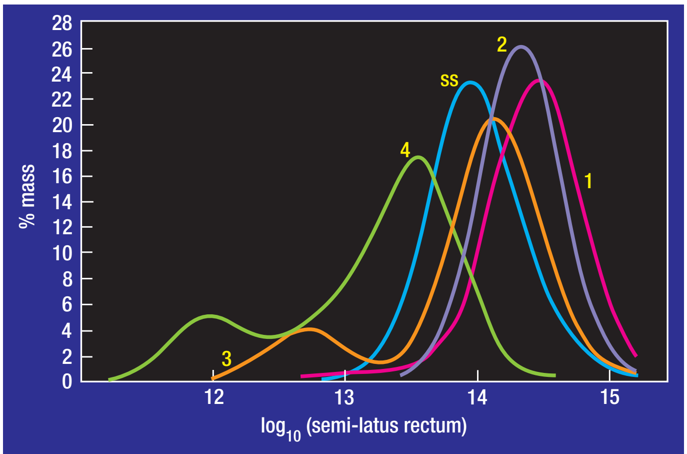
Planetary collision and terrestrial planets

The basic CT gives planets formed from cold material, in direct almost coplanar orbits of the right dimensions and accompanied by natural satellites. However, there were problems with the original model. Dormand and Woolfson (1971) reported that, according to their model, terrestrial planets would have gone too close to the Sun and so have been disrupted.

The first orbital round-off calculations by Dormand and Woolfson (1974) were two-dimensional but later they explored a three-dimensional scenario. They found, as expected, reducing orbital inclinations but they also found other, unexpected, orbital behaviour. Due to the medium's gravitational influence the eccentric orbits precessed in a complex way. The original inclined orbits did not intersect in space but, because of differential precession, pairs of orbits did occasionally intersect. Strong interactions could occur if planets arrived together near a point of intersection. A



6: The disruption of a model protostar. Captured points are marked with their orbital perihelion distances (10^{12} m) and eccentricities (in brackets). Escaping points are marked H (hyperbolic orbits).



7: The mass distribution from four Sun–protostar encounters together with the smoothed-out distribution for the solar system.

Table 2: Characteristic times for (a) planet 1 to be expelled from the system, (b) planet 2 to be expelled and (c) a collision

Planets		Time in millions of years			
1	2	(a)	(b)	(c)	(a), (b) or (c)
B	A	2.41	1790	33.3	2.24
B	Jupiter	0.11	∞	13.1	0.11
B	Saturn	2.53	∞	41.2	2.38
A	Jupiter	0.09	∞	5.91	0.09
A	Saturn	2.56	339	28.1	2.33
A	Uranus	∞	183	465	131
Jupiter	Saturn	453	0.22	139	0.21
Jupiter	Uranus	∞	0.34	111	0.34
Jupiter	Neptune	∞	0.95	327	0.94
Saturn	Uranus	∞	4.40	314	4.32
Saturn	Neptune	∞	17.8	1150	17.5
Uranus	Neptune	3420	2190	5240	1060

tidal interaction between a proto-Uranus and proto-Jupiter was previously described, but Dormand and Woolfson (1977) considered much stronger interactions where either one or other of the planets was ejected from the solar system or where there was a direct collision. Straightforward calculations showed that characteristic times for strong interactions were similar to those for orbital round-off.

Dormand and Woolfson took an initial system with six major planets, the present four plus two others denoted by A and B in table 1. The characteristics of A and B are speculative but the conclusions that follow are insensitive to the parameters chosen. From table 2, it appears that at least one major event was more likely than not in the early solar system.

Dormand and Woolfson (1977) modelled a collision between protoplanets A and B and showed that A could be expelled from the solar system while B was sheared into two parts that would have rounded off to the present orbits of the Earth and Venus. The largest terrestrial planets were interpreted as two non-volatile residues of a disrupted major planet.

The possible outcomes for the planetary satellites were that they could leave the solar system, go into independent heliocentric orbits, or be retained or captured by one or other of the B fragments. Thus, in one computational model the Earth fragment captured a satellite of A into a very stable orbit with an eccentricity of 0.4. The capture readily occurred in the presence of other bodies that removed energy from the Earth–satellite (Moon) system.

The Moon, Mars and Mercury

This scenario explains a curious feature of the Moon. The Moon's far side lacks large mare features, so characteristic of the near side. Since altimetry from lunar orbiters shows the presence of large basins on the far side, the usually accepted and sensible conclusion is that the solid crust was thicker on the far side so that magma was unable to reach the surface. Complicated explanations for this have been advanced yet simple tidal effects should lead to a thicker crust on the *near side*. Planetary collision is a straightforward explanation. Collision debris, travelling at more than 100 km s^{-1}, would have bombarded the satellites and abraded their surfaces. A thickness of a few tens of kilometres of the Moon's original surface could have been removed in this way – but only from the planet-facing hemisphere.

Protoplanets A and B would have had small perihelia and, because of large solar tidal forces, families of large satellites. A satellite origin for Mars explains its hemispherical asymmetry. The surface features of Mars, and their relationship to its spin axis, were explained by Connell and Woolfson (1983) who also considered the early water-rich evolution of that planet. Mercury too could be an escaped satellite, originally of similar mass to Mars but so heavily abraded that its surface completely reformed and it was left with a high density (Woolfson 2000).

Smaller bodies

The CT model does not predict large satellites for the outer planets. Neptune's large satellite, Triton, is also anomalous in its retrograde orbit. Woolfson (1999) described a computational model in which Triton was an escaped satellite from the collision. This collided with an existing regular satellite of Neptune, Pluto, which was expelled into a heliocentric orbit like its present one while Triton was captured by Neptune. The collision sheared off a portion of Pluto to give its satellite, Charon.

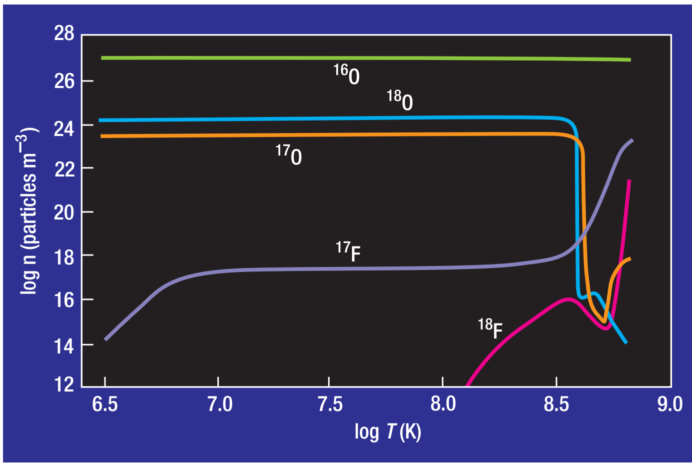
Debris from the planetary collision would have had the greatest concentration in the inner part of the system. Near-surface volatile-rich material from the colliding planets would have moved out furthest and, interacting with protoplanets near the aphelia of their original elliptical orbits, have provided a comet reservoir beyond the present planetary region. Inner larger members of this reservoir form Kuiper belt objects. Others, perturbed outwards by occasional close passages of stars or giant molecular clouds, formed the Oort cloud. Perturbations now remove Oort cloud comets and replenish them from the inner reservoir.

Debris closer in provided the early heavy bombardment within the solar system for which there is so much evidence. Those bodies that were in "safe" orbits remain today as asteroids or as captured irregular satellites.

Isotopic anomalies in meteorites

Models of a planetary collision (Woolfson 2000) show a collision-interface temperature in excess of 3×10^6 K. With a wide range of temperatures available there would have been an abundance of molten and vaporized material to explain chondrule formation and rapid cooling to give unequilibrated mineral assemblages within chondrules. There are interesting isotopic anomalies in meteorites including important ones for oxygen, magnesium, neon, silicon, carbon and nitrogen. An intriguing anomaly in some meteorites is neon-E, almost pure ^{22}Ne, assumed to be the daughter product of ^{22}Na with a half-life of 2.6 years. This sodium isotope was produced by nucleosynthesis and trapped in a cold rock within a few years.

Most explanations of isotopic anomalies deal with them individually on an *ad hoc* basis. The excess ^{16}O in some meteorites is ascribed to formation from ^{12}C in some far region of the galaxy, then transport in grains to the solar system and then exchange with normal oxygen.



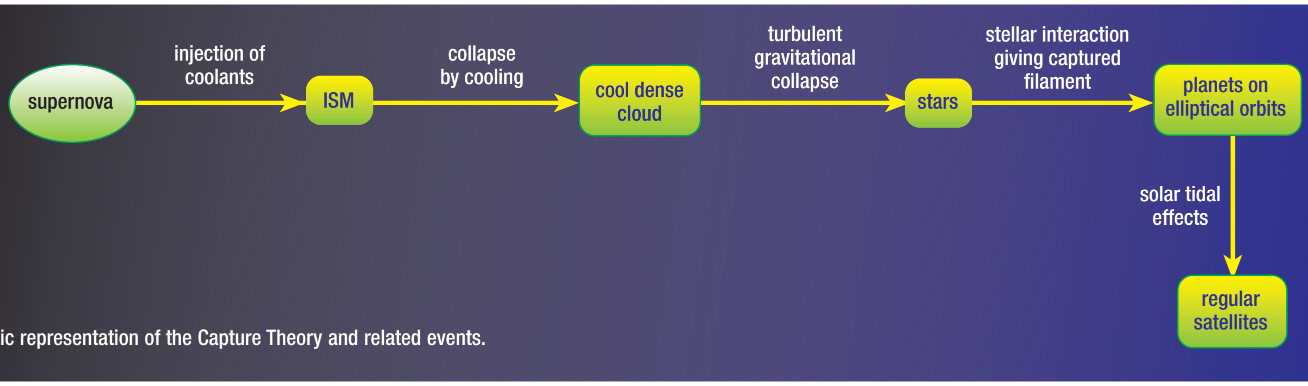
8: The variation of the concentration of stable oxygen isotopes and radioactive fluorine isotopes with temperature. (Holden and Woolfson 1995.)

One widespread anomaly within the solar system is the D/H ratio – 2×10^{-5} for Jupiter, 1.6×10^{-4} for the Earth, a few times the Earth value for some meteorites and 100 times the Earth value on Venus. Michael (1990) showed that the early evolution of intermediate-mass protoplanets could lead to differential loss of D and H and a D/H ratio as high as that of Venus. The consequence of a colliding planet having such a high D/H ratio was quantitatively examined by Holden and Woolfson (1995). A triggering temperature of 3×10^6 K sets off a nuclear reaction chain, at first involving D but later other nuclei as the temperature rises. All the isotopic anomalies referred to above can be well explained as mixtures of processed and unprocessed material; there is no need for *ad hoc* explanations. For example, figure 8 shows the variation of the concentration of oxygen isotopes (and ^{17}F and ^{18}F that decay quickly to ^{17}O and ^{18}O) with temperature during the nuclear reaction. At $\sim5\times10^8$ K the system explodes, the collision region expands and cools and reactions virtually cease. The oxygen content of processed material is almost pure ^{16}O; mixing it with unprocessed material explains the anomaly.

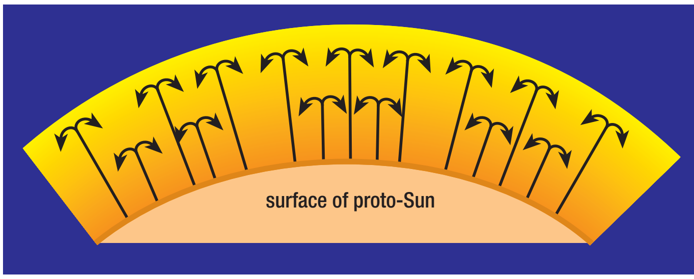
The modern Laplacian theory

The Solar Nebula Theory is clearly related to the original Laplace model but the Modern Laplacian Theory (Prentice 1974) follows the Laplace scenario much more closely.

To solve the problem of a slowly spinning Sun, Prentice followed a suggestion of Reddish and Wickramasinghe (1969) and assumed that the Sun formed from grains of solid molecular hydrogen settling within a dense cool cloud to which they were strongly coupled. The gravitational energy of the collapse vaporized the grains to give a cloud of hydrogen of radius $10^4\ R_\odot$ with a dense core formed by faster-falling CNO grains. By the time the radius of the cloud equalled that of Neptune's orbit, the boundary material was in free orbit. At this stage Prentice introduced turbulent stress. Supersonic turbulence within the cloud gave density variations and less dense regions were propelled outwards from the surface by buoyancy effects in the form of needle-like elements. Motion outwards would have been fast but inward motion slower, giving a higher density in the surface region (figure 9). Prentice showed that an instability would occur from time to time at the cloud equator so that material would be lost in the equatorial plane in the form of rings, much as Laplace postulated. All the rings had a similar mass, about $10^3\ M_\oplus$, with temperatures falling off with increasing ring radius. Prentice postulated that the several rings within the orbit of Mercury were vaporized, for a terrestrial ring there would have been silicate and metal grains with total mass $4\,M_\oplus$ and in major planet regions there would have been additional ice grains giving a total ring mass of 11–13 $M_\oplus$.



10: A schematic representation of the Capture Theory and related events.



9: Needle-like elements due to supersonic turbulence. Material in the shaded region slowly falls back to the surface of the proto-Sun.

Prentice presented an analysis in which solid material fell towards the axis of each ring and then came together to form a single planet or planetary core. In the major planet region the cores were sufficiently massive to accrete gas. While this gas contracted, a smaller scale version of the process, including supersonic turbulence, was taken to produce planetary systems.

This theory is by far the most complex of the current theories but despite its attention to the fine details of the system it does have severe drawbacks. The several rings within Mercury would have had an angular momentum several hundred times that of the Sun so they would not fall into the Sun. It can be shown that the rings would not have been stable and have had lifetimes much shorter than the time required for material within them to aggregate. The process by which material falls towards a ring axis is based on rather dubious mechanics requiring quite large solid bodies to be strongly coupled to a very diffuse gas. Finally, the system produced by this model would be highly coplanar and could not explain the tilt of the solar spin axis.

Conclusion

The current paradigm, the SNT, has not yet been successful in explaining the structure of the solar system at a very basic level. The observation that young stars are accompanied by dusty disks does not necessarily confirm the validity of the SNT because it predicts and depends upon a disk. Indeed, it is difficult to envisage a star-forming process that would not provide extraneous material that would form a disk. The important thing is not the disk but whether or not it gives planets. Nevertheless all observations are interpreted in terms of the SNT. For example, the nebula concept naturally suggests that radioactive isotopes were uniformly distributed in the early solar system. Hence, by looking for daughter products of particular decays in various types of object one can get relative times for when they became closed systems. The timings thus deduced are confusing and inconsistent – although the measurements are of good quality. Conformity reigns supreme and there is reluctance to consider that the SNT may not be valid. A more fruitful approach would be to find out what the experiments and observations are indicating rather than trying to force them into a theoretical strait-jacket. To quote Richard Feynman: "The test of all knowledge is experiment. Experiment is the sole judge of scientific 'truth'." This is applicable to cosmogony where "experiment" is usually observation.

By contrast the CT provides a coherent self-consistent model where single events explain many observations and events occur in causally related sequences. Figure 10 shows a schematic flow diagram for the CT including a planetary collision. Explanations have been given for all but one of the 20 features referred to previously in this article – the existence of other planetary systems. It turns out that CT interactions would probably be common in an evolving stellar cluster. Recently there has been much discussion of the *embedded* phase in the evolution of a galactic cluster (Gaidos 1995) where stellar density can be of order 10^5 pc^{-3}. Recent work, as yet unpublished, has not only realistically modelled planetary formation in great detail, showing the formation of single-planet or multiple-planet systems, but also indicated that the predicted frequency of planetary systems is consistent with recent observations. ●

Prof. M Woolfson is at the University of York.

References

Bondi H and Hoyle F 1944 *Mon. Not. R. Astr. Soc.* **112** 273–282.

Cameron A G W 1978 *The Origin of the Solar System* (ed. S F Dermott) pp49–74 Wiley, Chichester.

Chambers J E and Wetherill G W 1998 *Icarus* **136** 304–327.

Connell A J and Woolfson M M 1983 *Mon. Not. R. Astr. Soc.* **204** 1221–1230.

Dormand J R and Woolfson M M 1971 *Mon. Not. R. Astr. Soc.* **151** 307–331.

Dormand J R and Woolfson M M 1974 *Proc. R. Soc. Lond.* **A340** 349–365.

Dormand J R and Woolfson M M 1977 *Mon. Not. R. Astr. Soc.* **180** 243–279.

Dormand J R and Woolfson M M 1988 *The Physics of the Planets* (ed. S K Runcorn) pp371–383. Wiley, Chichester.

Gaidos E J 1995 *Icarus* **114** 258–268.

Holden P and Woolfson M M 1995 *Earth, Moon & Planets* **69** 201–236.

Jeans J H 1917 *Mon. Not. R. Astr. Soc.* **77** 186–199.

Jeffreys H 1929 *Mon. Not. R. Astr. Soc.* **89** 636–641.

Laplace P S de 1796 *Exposition du Système du Monde* Paris, Imprimerie Cercle-Social.

Larson R B 1989 *The Formation and Evolution of Planetary Systems* (eds H A Weaver & L Danley) pp31–48 Cambridge University Press Cambridge, England.

Lynden-Bell D and Pringle J E 1974 *Mon. Not. R. Astr. Soc.* **168** 603–637.

Lyttleton R A 1960 *Mon. Not. R. Astr. Soc.* **121** 551–569.

Lyttleton R A 1961 *Mon. Not. R. Astr. Soc.* **122** 399–407.

McCrea W H 1960 *Proc. Roy. Soc. Lond.* **256** 245–266.

McCrea W H 1988 *The Physics of the Planets* (ed. S K Runcorn) pp421–439 Wiley, Chichester.

Melita D M and Woolfson M M 1996 *Mon. Not. R. Astr. Soc.* **280** 854–862.

Pickett B K and Durisen R H 1997 *Icarus* **126** 243–260.

Pollack J B *et al.* 1996 *Icarus* **124** 62–85.

Pongracic H *et al.*1991 *Mem. S.A. It.* **62** 851–858.

Prentice A J R 1974 *In the beginning…* (ed. J P Wild) Australian Academy of Science, Canberra.

Reddish V C and Wickramasinghe N C 1969 *Mon. Not. R. Astr. Soc.* **143** 189–208.

Russell H N 1935 *The Solar System and its Origin* MacMillan, New York.

Safronov V S 1972 *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* Israel Program for Scientific Translations, Jerusalem.

Schmidt O Y 1944 *Dokl. Acad. Nauk. USSR* **45** 229–233.

Schofield N and Woolfson M M 1982 *Mon. Not. R. Astr. Soc.* **198** 947–961.

Spitzer L 1939 *Astrophys. J.* **90** 675–688.

Stewart G R and Wetherill G W 1988 *Icarus* **74** 542–553.

Weidenschilling S J, Donn B and Meakin P 1989 *The Formation and Evolution of Planetary Systems* (eds H A Weaver & L Danley) pp131–150 Cambridge University Press, Cambridge, England.

Williams S and Woolfson M M 1983 *Mon. Not. R. Astr. Soc.* **204** 853–863.

Woolfson M M 1964 *Proc. R. Soc. Lond.* **A282** 485–507.

Woolfson M M 1979 *Phil. Trans. R. Soc. Lond.* **A291** 219–252.

Woolfson M M 1999 *Mon. Not. R. Astr. Soc.* **304** 195–198.

Woolfson M M 2000 *The Origin and Evolution of the Solar System.*