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

On Possible Origins of Relatively Short-Term Variations in the Solar Structure

D. O. Gough

Phil. Trans. R. Soc. Lond. A 1990 **330**, 627-640
doi: 10.1098/rsta.1990.0043

Phil. Trans. R. Soc. Lond. A **330**, 627–640 (1990) 627
Printed in Great Britain

On possible origins of relatively short-term variations in the solar structure

By D. O. Gough

Institute of Astronomy, and Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, U.K., and Joint Institute for Laboratory Astrophysics, University of Colorado and National Institute of Standards and Technology, Boulder, Colorado 80309-0440, *U.S.A.*

The general large-scale redistribution of magnetic field over the solar cycle is possibly associated with an overall variation of thermal structure of the convection zone, which modulates not only the total luminosity but also the latitudinal distribution of radiative flux, thereby modifying the irradiance of the Earth. Whether the cause of this variation lies within the convection zone or is more deeply seated is still an open question.

1. Introduction

There are three obvious natural timescales on which one would expect the Sun to vary: the nuclear evolution time, about 10^{10} years, the thermal diffusion time characteristic of the entire Sun, about 3×10^7 years, and the acoustic travel time from centre to surface, about 1 h. The last of these timescales is a dynamical time; it is of the order of the free-fall time from the surface to the centre of the Sun, and it is approximately equal to the period of the gravest mode of acoustic oscillation.

There is strong astronomical evidence that the Sun varies on the nuclear timescale. Moreover, it is also an inevitable consequence of hydrostatic balance under newtonian gravity, granted that the fusion reactions cause the mean molecular mass (the stellar physicists' jargon for the mean mass of the particles that constitute stellar material, measured in units of the mass of the hydrogen atom; the interior of the Sun is too hot for there to be any molecules) in the core of the Sun to increase with time. Moreover, it has been observed that the Sun varies dynamically; in particular, acoustic oscillations have been well studied. Their amplitude is extremely low, and it is unlikely that the oscillations that have been observed have a significant influence either on the hydrostatic structure of the Sun or on its temporal variation on timescales substantially greater than hours. However, the oscillations have important diagnostic power, and I shall therefore return to them later.

There is no direct evidence that the Sun varies on the thermal timescale. Linearized stability analyses have found the Sun to be thermally stable, so one would expect variation on the thermal scale only as a relaxation to thermal equilibrium if the structure of the Sun were to have been perturbed by some other process. I shall not discuss such possibilities here, because the timescale is not of interest to this meeting. Indeed, nor are the nuclear and the dynamical timescales. I have introduced them simply to point out that the phenomena discussed in this meeting do not find a ready explanation as a simple natural solar variation.

There are two other, somewhat less natural timescales that come to mind: the characteristic turnover time of turbulent eddies in the Sun's convective envelope, and the thermal relaxation time of the convective envelope. The former varies from about 15 min at the top of the convection zone to about a month in the almost adiabatically stratified layers beneath. The characteristic rotation period of the Sun is also about a month (the Sun's angular velocity is not uniform), which means that rotation neither dominates the dynamics of the convection zone nor acts as a small perturbation. Therefore, the interaction is quite complicated, and probably leads to a complicated flow field. The motion can twist and stretch magnetic fields, producing the rich variety of magnetic activity observed in the surface layers that is normally attributed to a dynamo. The thermal relaxation time of the convection zone is about 10^5 years, and may very well be of some significance to climatic issues discussed in this meeting, even though the timescale is greater than those of direct concern. As is the case with the thermal relaxation of the entire Sun, however, it can be manifest only as a transient response to some other stimulus, or via an interaction with some other process. I should perhaps also mention at this juncture the thermal relaxation timescale of the core of the Sun within which energy is being generated: it is *ca.* 10^6 years.

I do not include the rotation period in the same category as the five natural timescales I have just enumerated, even though it is a very obvious property of the Sun. The reason is that at our present state of understanding it is hardly natural. It is a consequence of the particular value of the angular momentum the Sun possesses at the moment. To be sure, a systematic variation of angular velocity Ω of main-sequence stars has been observed: Ω depends both on the mass and age of a star. Moreover, there are theoretical rationalizations of this variation. However, the theories have been tuned, and even the most modern ideas would have been unlikely to have successfully predicted the observations had the observations not preceded them.

Another obvious solar timescale is the characteristic period of the solar cycle, *ca.* 22 years. This is very relevant to this meeting. But it is not an obviously natural timescale of the Sun. It can, however, be constructed from natural timescales; it is the geometric mean of the characteristic dynamical time of the Sun and the thermal relaxation time of the energy-generating core. Whether this coincidence has any physical significance is not known.

It is evident from this prologue that any theoretical discussion of temporal solar variations on timescales relevant to this meeting must necessarily be either subtle or uncertain. It is probably true to say that the subtle discussions always rest on severe idealizations, and although they may be correct within the restricted context defined by those idealizations, their application to the Sun lies outside the realm of validity of the arguments, and therefore they too are uncertain. I have dynamo models particularly in mind now. Therefore, notwithstanding the intricacies of the mathematical models and the many hours of computer time that have been devoted to studying those models, either of which can sometimes leave unwarranted impressions of reliability in the mind of an unwary bystander, I feel obliged to regard all theories that address solar variation on the timescales relevant to this meeting as being quite uncertain. I hasten to add that this is not to say that those theories are of no value. I am quite sure that some of their components well describe processes that do go on, and without an understanding of those processes we would never understand the Sun.

With this uncertainty in mind, I shall now discuss, quite briefly, some of the ideas that have been aired. The purpose is to provide some idea of the kinds of processes that may be taking place inside the Sun, and which cry for further study. Most of my discussion will be directed towards the solar cycle, for that is the only solar temporal variation directly relevant to this meeting that we are really sure takes place.

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

2. The standard solar model

The so-called standard theoretical solar model is important because it provides a basis from which to discuss the processes that might lead to interesting temporal variations. The model is a highly idealized representation of the Sun. Nevertheless, astrophysicists believe it contains the most important ingredients of a more realistic description.

The model assumes the Sun to be hydrostatic and essentially spherically symmetric; rotation and magnetic fields are ignored, which is certainly a good first approximation. There is no motion other than a gradual expansion or contraction, except where the stratification is convectively unstable. This occurs only in the outer 30%, by radius, of the envelope, where the only role the turbulent fluctuations are assumed to play is in providing contributions to the flux of energy and, sometimes, momentum, via a turbulent pressure tensor. Elsewhere, energy is presumed to be transported solely by radiation. There is no significant loss of mass from the surface (the mass-loss timescale due to the present solar wind is about 10^{13} years) nor accretion of mass onto the surface.

The star changes its structure because nuclear fusion reactions in the core increase the mean molecular mass, which leads to contraction and a rise in temperature in the radiative interior. According to the simple mixing-length prescriptions used to model the outer boundary layer of the convection zone, the upper layers of the Sun expand. Concomitant with there being no motion in the radiative interior, the products of the nuclear reactions are assumed to remain *in situ*. Notice that by assuming no motion outside the convection zone the nonlinear development of any instability to which the radiative interior might be subject is ignored.

The rise in the density and temperature of the core leads to a rise in the rate of thermonuclear energy generation, and an increase of entropy. Because evolution is slow compared with the rate at which energy is transported from the core to the surface, there is a close balance between the energy generation rate and the solar luminosity L, the net rate at which energy is radiated from the surface. According to the theory, L rose gradually and monotonically from the initial value of about 70% of its value today, the initial value being the luminosity after the Sun had settled down to a state of thermal balance. Roughly speaking, that was the epoch at which the earth was formed, about 4.6×10^9 years before present (BP). The rise in L is a robust conclusion, and is not dependent on any subtle details of the theory.

Temporal variation on the nuclear timescale is not the subject of this meeting. Nevertheless, I have taken the space to mention it because, so far as I am aware, all self-contained quantitative climatic models, which are relevant to our discussions, cannot reconcile the past low luminosity of the Sun with the Earth not having been almost entirely glaciated for most if not all of its existence. Therefore a healthy scepticism must always be maintained when discussing long-term solar–terrestrial relations.

Some scepticism must also be maintained when discussing the standard solar model. Aside from the adoption of what might seem highly idealized assumptions, we are faced with the fact that the model that otherwise best fits the astronomical observations yields a neutrino flux on Earth some three times greater than the value measured by Davis and his colleagues. This too appears to be an astrophysically robust result, provided the physics used to determine the nuclear reaction rates is correct. I shall mention this discrepancy again later.

Unfortunately, despite their name, standard solar models have not yet been standardized. The models are continually being updated as new modifications to the microscopic physics

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

determining the opacity, the nuclear reaction rates and the equation of state are suggested. Fortunately, the differences between the models are much too small to be of any concern to my discussion here.

3. Competing hypotheses about the solar cycle

An early suggestion for explaining the solar cycle of magnetic activity was that of Walèn (1946). The solar interior was thought to be undergoing oscillations for which magnetic stresses provided the major component of the restoring force; the oscillation is essentially a resonant Alfvén mode. Evidently the motion must be essentially horizontal, to avoid substantial generation of buoyancy forces that would lead to periods of hours rather than 22 years. The oscillations are modes that become the zero-frequency torsional modes as the magnetic field strength approaches zero. They could therefore be called magnetic torsional oscillations. Because the Sun rotates with a period much less than the cycle period, the motion must also be such that the Lorentz force largely balances the Coriolis force, otherwise the oscillation period would once again be too short.

The oscillation period depends on the strength of the magnetic field. A value of a few hundred milliteslas yields 22 years. That is typical of the fields seen in sunspots. Therefore, if one believes that there is a global field in the interior of the Sun, loops of which sometimes rise through the convection zone and break through the photosphere to form sunspot pairs, 22 years does not seem too unnatural a value to expect for the period of the cycle. At least the hypothesis is economical in assumptions. The detailed implications of Walèn's idea appear not to have been worked out.

There is another reason for not being too surprised at a large-scale interior magnetic field with magnitude of a few hundred milliteslas. It is natural to presume that the Sun condensed from a typical interstellar gas cloud, as have most other stars. The cloud was presumably intrinsically rotating, thereby having vorticity with respect to an inertial frame, and was pervaded by a magnetic field. Both vorticity and magnetic field in the interstellar turbulence are random; for want of better data I shall adopt typical values of $2 \times 10^{-8}\ \mathrm{a}^{-1}$, corresponding to the galactic rotation, and 3 μG†, respectively. We know that neither angular momentum nor magnetic flux were conserved by the material that eventually formed the stars. For that to have been so, the collapse would have had to be restricted to parts of the gas cloud that happened to have essentially no vorticity and no magnetic field; how would similar neighbouring parts of the cloud, with somewhat greater vorticity and magnetic field, know not to collapse because it would present them with serious dynamical problems in the distant future? It is more likely that these problems were faced by all protostars, and solved. A plausible and commonly held opinion is that much of the solution occurred in a turbulent phase of contraction in which magnetic field was expelled. In an ideal homentropic fluid, both potential vorticity $\boldsymbol{\omega}$ and $\boldsymbol{b} = \boldsymbol{B}/\rho$, where $\boldsymbol{B}$ is the magnetic field and ρ is the fluid velocity, behave like a material line element: a streak of dye that has been placed in the fluid. Even if dissipation is considered, the equations for $\boldsymbol{\omega}$ and $\boldsymbol{b}$ are similar; so one might expect $\boldsymbol{\omega}$ and $\boldsymbol{b}$ to behave similarly, particularly in a turbulent fluid where any differences in initial conditions have been forgotten. This argument was first used by Batchelor (1950) in a discussion of hydromagnetic turbulence in an incompressible fluid, and although particular cases are known

† 1 G = 10^{-4} T.

where the analogy between $\boldsymbol{\omega}$ and $\boldsymbol{b}$ breaks down, it is an open issue whether it holds under the circumstances I am discussing here. Nevertheless, if one accepts it, even though the collapsing gas is not strictly homentropic, and even though one does not understand the mechanism of vorticity and magnetic field expulsion well enough to predict how much is lost, one can deduce that the ratios of the final to the initial values of both the mean vorticity (and therefore intrinsic angular velocity) and the mean magnetic field are the same. The present solar rotation period is about 27 d, corresponding to a mean vorticity of 160 a^{-1}, an increase over the initial value by a factor of about 10^{10}. Thus a magnetic field of $10^{10} \times 3 \times 10^{-6} = 30$ kG seems not unreasonable. Some further credence has been given to a value of this order by Mestel & Weiss (1987), who point out that fields of about 500 G have been observed at the surfaces of T Tauri stars. These stars are at the end of the extensively turbulent phase of their evolution, and the surface fields may be in the process of being expelled. If they are representative of the interior field, and flux is conserved throughout the subsequent evolution to the main sequence, an increase in field strength by a factor of 10 occurs, yielding a zero-age main-sequence field of 5 kG. The difference by a factor 2π between these two estimates is hardly material for such a rough argument. Of course, in the absence of dynamo action in the radiative interior the field will decay on the main sequence; moreover, there is also strong evidence for substantial angular momentum loss during the main-sequence lifetime. There is no reason to believe that the different processes causing these decays would keep in step, so the angular velocity to field ratio might have changed by a factor 10 or so since the Sun arrived on the main sequence. Nevertheless, this factor is quite insignificant in the light of the wide range of values (1 μG–10 MG) proffered by various authors for the present field in the solar interior.

The major competing hypothesis to explain the solar cycle is that the convection zone of the Sun is a turbulent dynamo. Theories of dynamos in rotating, roughly spherical electrically conducting fluid bodies were developed initially to explain the existence of the Earth's magnetic field. The natural decay time of the field is much less than the age of the Earth. Therefore, if the terrestrial field were not regenerated, any primeval field would by now have decayed to an imperceptible level; a dynamo is therefore essential. As in the Sun, the dipole component of the terrestrial field reverses sign, roughly but apparently not exactly, in a periodic manner. Thus it is perhaps no surprise that having laboured long to develop the mathematical armoury to attack the terrestrial problem, the temptation to turn it towards the Sun was irresistible, even though the natural decay time of the largest-scale component of the solar magnetic field is greater than the age of the Sun.

The characteristic cycle period of the supposed dynamo is not an easy quantity to predict. It depends on subtle properties of the structure of the turbulence in the convection zone, which are difficult to ascertain. It is the case, however, that the observed period is not in obvious conflict with the theoretical estimates.

The earliest dynamo theories assumed the dynamics to be confined entirely within the convection zone. The major difficulty encountered by even the most elaborate studies was that consistent dynamical theories could not reproduce simultaneously the two most pertinent observations: that the photospheric angular velocity is greater at the equator than it is at the poles, and that immediately after field reversal the predominant magnetic activity occurs at mid-latitudes and subsequently migrates towards the equator. To reproduce the more rapidly rotating equator the uncertain parameters of the theory must be adjusted in such a way as to tend to an angular velocity that increases with depth, at least in the equatorial regions; whereas to reproduce the equatorial migration of magnetic field the opposite radial rotation gradient is required.

In recent years, consideration of the dynamics has been extended a little beneath the base of the convection zone. It has been suggested that the essence of the dynamo action is confined to a relatively thin interfacial layer between the true convection zone and the radiative interior (Spiegel & Weiss 1980). Basically, magnetic field is forced downwards through the convection zone as a result of a topological asymmetry in the flow (Drobyshevski & Yuferev 1974). It accumulates as a thin toroidal belt near the base of the convection zone until it becomes buoyant enough to overcome the convective downthrust. Then it surges to the surface to start a new cycle. It was hoped that with such an approach the apparently insuperable difficulty encountered by previous theories of reproducing both the rapid equatorial rotation of the photosphere and the equatorial flux migration at low latitudes might be surmounted. The attention of dynamo theorists has not yet penetrated fully into the radiative interior of the Sun.

How might one decide between the two hypotheses? Before discussing that it would be wise to make it quite plain what the fundamental difference between them is. To be sure, if dynamo calculations were extended deep into the radiative interior, some response would be found, but because some 98% of the inertia of the Sun resides in the radiative region, the amplitude of that response would be relatively small. The existence of the radiative region is not essential to the operation of the dynamo, and presumably its presence modifies the cycle only slightly. The basic mechanism is a loss of magnetic field either by dissipation in the convection zone or loss of flux through the photosphere, followed by amplification of the residual field due to stretching by shearing motion; field amplification to cancel losses is an essential feature of a dynamo. The torsional oscillation, on the other hand, does not require the dissipated field to be replaced. The dominant controlling dynamics of the cycle is in the radiative interior; it is there that the characteristic cycle period is determined. Of course, the processes that are presumed to take place in the turbulent dynamo would be operative in the convection zone, and no doubt would have a very strong influence on the observable features of the oscillation in the photosphere. Therefore, there may be similarities in some of the superficial consequences of the two hypotheses.

Perhaps the principal superficial observed property of the magnetic field is the polarity reversal of the dipole component. This feature is often considered to be an overwhelming objection to torsional oscillations, it being implicitly assumed that superficial behaviour would also be required to take place throughout the Sun, and that therefore the entire radiative interior would have to turn over. That is not the case. If the oscillation were one predominantly in rotational shear, for example, the interior field would be stretched azimuthally, presenting the convection zone with a field that is hardly discernible from that of the interfacial dynamo. The difference between the two hypotheses is not essentially in the field configuration at the base of the convection zone, but in how it is produced.

Another argument that is often tendered in favour of the solar dynamo is that some of the more subtle features of the photospheric flow field have been reproduced by suitably tuned dynamo models, whereas that is not so of a global torsional oscillation. The torsional oscillations discovered by LaBonte & Howard (1982) constitute an example of such flow. When considering this argument one should remember that, as I have just pointed out, many of the processes in the convection zone that might constitute an integral part of a turbulent dynamo are also likely to be operative, albeit incidently, if the cycle is controlled from within the radiative interior. More importantly, it should be realized that one reason certain phenomena have not been reproduced by a theory of global torsional oscillations is that nobody has attempted to do so.

4. Temporal coherence of the solar cycle

Because the convection zone is turbulent, any model of the cycle that depends in any way on the dynamics of the convection must be subject to some stochastic element. There may be other sources of variation, such as the deterministic chaos discussed by Professor Weiss (this Symposium). It is not out of the question that some information could be gained about the dynamics controlling the cycle by studying the properties of the variations from cycle to cycle. This is quite an old idea, and there have been many studies since the pioneering paper by Yule (1927) on the behaviour of a randomly perturbed pendulum.

The issue I address here is that of phase maintenance. My interest was raised in this subject by Dicke (1970), who remarked that because the early arrivals of the sunspot maxima of 1778 and 1788 were followed immediately by a compensating long cycle, thereby restoring the phase, the cycle must be controlled by some mechanism that keeps time. This could be so of a coherent laminar torsional oscillation of the relatively quiescent radiative interior of the Sun (whose signature at the surface is delayed by some random process), but it would hardly be true of a process that is controlled largely by the random motions in the turbulent convection zone, unless, of course, the dominant motion in the convection zone were not really random.

Of course the occurrence of but a single pair of exceptional cycles whose mean period is similar to that of the entire record is only weak evidence for phase maintenance. The entire record should be examined. Thus with the aim of distinguishing between the two hypotheses discussed in the previous section in mind, two simple stochastic models have been tested against the sunspot record. The first, which is a simple yet extreme representation of a turbulent dynamo, is a process with no long-term memory. Cycle n is considered to be of duration $P_n = P+\sigma_n$, where P is the characteristic period of the process that controls the dynamics of the oscillation and σ_n is a random variable resulting from the stochastic turbulent perturbation which is considered to be uncorrelated with the perturbations $\sigma_m (m \neq n)$ to the periods of the other cycles. The time of occurrence of some well defined epoch (say sunspot maximum or sunspot minimum) signifying the 'beginning' of cycle $N+1$, measured from the same epoch in cycle 1, is thus $T_N = NP+\sum\sigma_n$, where the sum is over all cycles from 1 to N. I am assuming that there is nothing special about cycle 1; it is simply the first cycle of the record available for analysis. The second model is intended to represent a periodic torsional oscillation of the radiative interior. The oscillation keeps perfect time, but because information must pass through the turbulent convection zone to produce a visible change in the photosphere there is again a stochastic element to the observable signal. In this case the time of some epoch signifying the 'beginning' of cycle $n+1$ relative to the same epoch in cycle 1 is $T_n = nT+\tau_n-\tau_1$, where T is the period of the torsional oscillation and τ_n is a random variable representing the turbulent perturbation which, as in the first model, is uncorrelated with the perturbations to the other cycles. Thus in this model the period of cycle n is $T+\tau_n-\tau_{n-1}$. The idea is to compare the properties of the two models with the sunspot record.

No statement has yet been made about how close to being periodic the sunspot record actually is. That would depend on the standard deviations σ and τ of σ_n and τ_n, which depend on how much the turbulence in the convection zone influences the sunspot signal, and about which it is assumed we have no *a priori* information. What must therefore be tested is a statistic that is independent of σ or τ. The obvious choice is the variance σ_T^2 of T_n from the perfectly regular time sequence whose period and phase best match the data, measured in units of the variance σ_P^2 of P_n^2. This is then compared with the value of the ratio $\Phi = \mathrm{E}(\sigma_T^2)/\mathrm{E}(\sigma_P^2)$ predicted by the two models, where E denotes expectation value. When N is large, Φ increases linearly with N for the turbulent dynamo model, whereas it tends to $\frac{1}{2}$ for the torsional oscillation model. For $N = 2$, the smallest value for which the analysis can formally be carried out, Φ is the same for both models. The crucial question is whether the sunspot record is long enough for the predictions of the two models to be distinguishable.

The analysis was first carried out by using separately the epochs of sunspot maxima and sunspot minima, dating from immediately after the end of the Maunder Minimum (Gough 1978). The result was inconclusive: the time series of sunspot maxima is slightly closer to the prediction of the periodic oscillation model, whereas the series of sunspot minima is slightly closer to the model representing the turbulent dynamo. It appeared that the sunspot record is too short. Shortly afterwards, Dicke (1978) analysed a modified time series based on the epochs of sunspot maxima alone, and concluded that phase has been maintained. The analyses have been rediscussed, and extended by dividing the record into shorter segments and comparing the dependence of Φ on N with the predictions of the models (see, for example, Gough 1988). The results depend on whether 11-year records or 22-year (derectified) cycles are used. In the former case, Φ follows the trend of the stochastic dynamo model except for the greatest value of N, at which, as I have already mentioned, it lies inconclusively between the two models. Because it should be the longest record that is the most reliable, one is bound to be left in an uncertain state. The derectified cycles are similar, except that the undivided record is closer to the dynamo model than it is for the 11-year rectified cycle.

An obvious drawback to taking the epochs of either sunspot maxima or sunspot minima is that in a noisy record they are subject to error due to local fluctuations, particularly because the first derivative is zero at extrema. It is surely preferable to use all the data in the record. Recently, Bracewell (1985) has carried out a Hilbert transform of derectified sunspot data, representing the record as a mildly nonlinear function of a sinusoid that is displaced from zero by a variable amount, having varying amplitude and phase offset. The variance of the phase offset, which represents the deviation of the oscillator from a perfect clock, is very much less than the value of σ_T^2 computed from the raw sunspot data. Indeed, Bracewell (1988) has regarded so small a variance as evidence that phase is maintained, and proposed an explanation in terms of wave motion that is not very dissimilar from the torsional oscillator that I have discussed here. However, as I have already argued, a small value of σ_T^2 does not by itself distinguish between the two models; for that a statistic must be used that does not depend on the magnitude of the stochastic fluctuations. Accordingly, I have defined two time sequences of a given epoch in the 22-year magnetic cycle (representing odd- and even-numbered sunspot maxima) from the phase offset in Bracewell's Hilbert transform, and have subjected them to the same analysis as the raw sunspot data. The result is illustrated in figure 1, together with the values of Φ predicted by the two statistical models. Once again the results are inconclusive: although the N dependence of the ratios obtained from the subsequences looks more similar to the predictions of the turbulent dynamo model, the standard deviations are large enough for

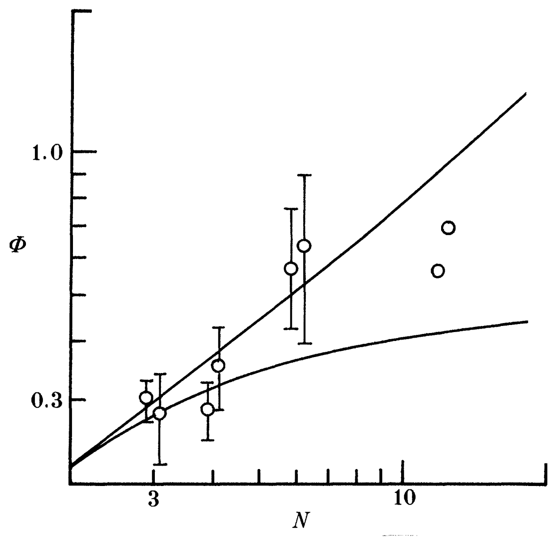

FIGURE 1. Ratios Φ of phase deviation to period deviation from a perfect clock of the solar cycle for two time sequences of constant cycle phase computed from Bracewell's (1985) Hilbert transform of the $12\frac{1}{2}$ cycles of derectified sunspot record between 1705 and 1979. The sequences start at 1705 and 1716, and represent approximate sunspot maxima. (To compute the final date of the second sequence it was necessary to extrapolate Bracewell's phase to 1980.) They have each been divided into contiguous subsequences of length N, from each of which Φ has been computed separately. The points plotted are the means of the values from each subsequence, and the error bars denote ± one standard deviation. To avoid overlapping, points computed from the sequence starting in 1705 are displaced slightly to the right; those from the sequence starting in 1716 to the left. The upper continuous curve is $N(N+3)/15(N+1)$ and represents the predictions of the turbulent dynamo model; the lower curve is $N^2/2(N+1)^2$ and represents the prediction of the periodic internal oscillator model.

either model to be possible; and the longest records yield ratios lying between the two theoretical curves.

There thus seems to be little evidence in the sunspot record to support the idea that the cycle is controlled by a periodic oscillator. But that does not imply that the Sun is not undergoing torsional oscillations in its radiative interior. If the oscillations are nonlinear, they could be chaotic, in a manner similar to the oscillators discussed by Professor Weiss (this Symposium) in connection with deterministic dynamo models. In that case phase would not be well maintained, and might be consistent with the data plotted in figure 1.

5. The seat of the cycle

There are other diagnostics that can shed light on the mechanism of the cycle. To interpret them, the structure of the entire Sun must be taken into account.

Any perturbation to the balance of forces in a star causes a global readjustment of its structure. Just how that adjustment takes place depends on the nature of the perturbation and the timescale on which it occurs. Here, interest is on a timescale of order 10 years, which is much less than the thermal relaxation times of either the solar core or the convection zone. However, it is much greater than the internal thermal readjustment timescale of the convection zone, which is the time it takes for a convective wave to propagate through the extent of the zone: about one month for a spherically symmetrical readjustment. Therefore any perturbation on the timescale of the solar cycle would cause an adiabatic hydrostatic readjustment of the radiative interior, leaving the convection zone in internal thermal balance. Of course, the convection zone would not be in balance with the radiative interior; a thermal boundary layer would develop, of characteristic thickness 0.1 % of the solar radius (1 % of the pressure scale height), if intermittent convective overshooting were unimportant.

The hydrostatic readjustment of the convection generally changes the radius R of the photosphere and the photospheric temperature. Consequently, there is a change in the luminosity L. This changes the solar irradiance S on Earth by the same relative amount. Broadly speaking, the magnitude of the ratio W of the relative perturbations to L and R tends to increase with the depth of the perturbation (Gough 1981; Däppen 1983), and so provides a diagnostic for the location of the perturbation. A recent paper by Willson & Hudson (1988) reports that S varies with the rectified 11-year cycle (figure 2), with a relative amplitude of 0.04 %, maximum S corresponding to maximum sunspot number. Unfortunately, unambiguous variations in the photospheric radius have not yet been measured. There are many problems associated with the determination of R, among which is the variation of the limb-darkening function at the very limb resulting from magnetic activity in the solar atmosphere, which varies with the solar cycle. Therefore, this potential diagnostic has not yet given us any secure information.

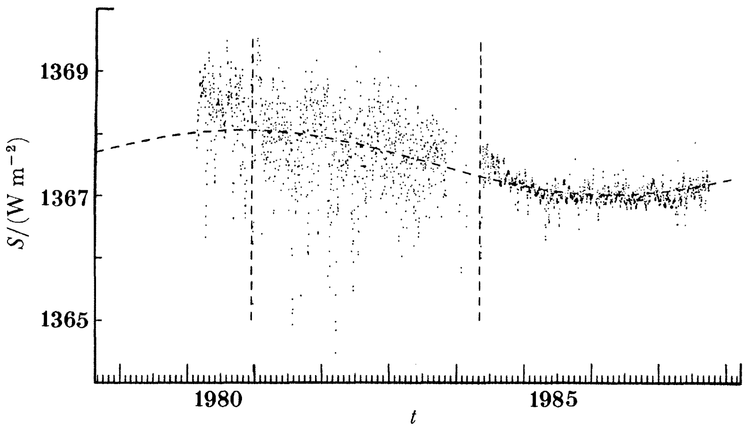

FIGURE 2. Solar irradiance S from the *Solar Maximum Mission* Active Cavity Radiometer Irradiance Monitor reported by Hudson & Willson (1988). Data in the interval from 1980 day 346, when the satellite's attitude control system failed, until the repair in March 1984 (delimited by the vertical dashed lines) are of lower quality than the rest. The dashed curve is the least-squares cosine function

$$S = S_0\{1+0.00039 \cos [2\pi(t-1980.82)/10.95]\},$$

where S_0 is constant and t is time in years from the beginning of the century. Fluctuations are lower at sunspot minimum because there are fewer sunspots appearing and disappearing to modulate the flux.

If one does accept at face value the semi-diameter measurements plotted in figure 3, combines them with the irradiance measurements of figure 2, and assumes that $\delta S/S \approx \delta L/L$, one obtains $W \approx -0.7$. This is larger in magnitude than the theoretical values that have been reported for even the most deeply seated perturbations that occur on a solar-cycle timescale. It is interesting to note, however, that this value agrees almost exactly with the value one can infer from the calculations reported by Gough & Thompson (1990), who considered the response of the Sun's structure due to a broad equatorial belt of toroidal magnetic field at the base of the convection zone. The vertical extent of the belt was about 10 % of the solar radius; and the characteristic field intensity required to obtain variations of L and R of the observed

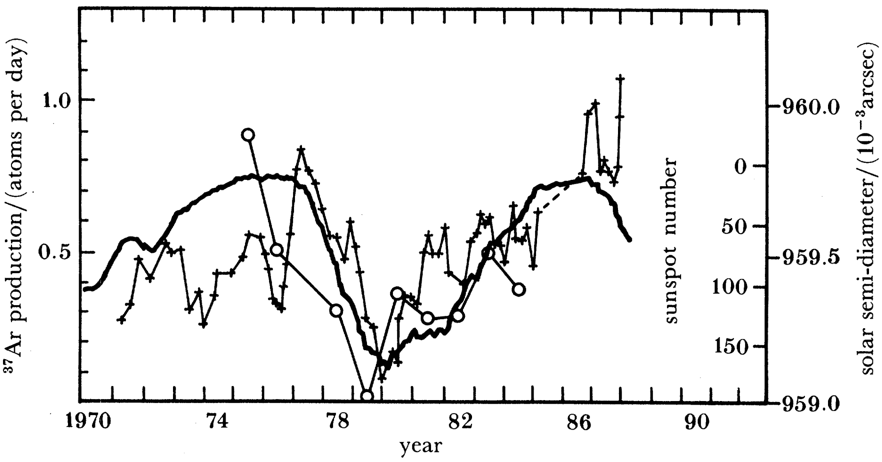

FIGURE 3. Five-monthly averages of sunspot number (thick line), solar neutrino flux recorded by Davis and his collaborators (crosses) and apparent semi-diameter of the Sun from astrolabe measurements (circles), perhaps suggesting a causal connection (after Davis *et al.* 1988).

magnitude was about 10 MG, which is rather greater than the values I have been discussing explicitly in this paper. The computations were not strictly applicable to the solar cycle, however, because the radiative interior was maintained in thermal balance rather than being perturbed adiabatically as would essentially be the case for a variation on a timescale of 10 years. (The purpose of the computations was not to describe solar-cycle variations.) Therefore, the numerical coincidence might just be fortuitous.

Magnetic fields cannot be everywhere spherically symmetrical. Therefore, any magnetic field that perturbs the solar structure must distort the Sun from its otherwise almost spherical configuration. This distortion results, in general, in an aspherical variation of any property of the Sun. In particular, it causes a variation of the photospheric temperature T_e, leading to a redistribution of radiant flux over the surface. This provides another contribution to the variation in the irradiance S on Earth. Recently, Kuhn *et al.* (1988) have shown that the latitudinal variation δT_e of T_e is correlated with the sunspot distribution, a result that is hardly surprising. But what is surprising, however, is Kuhn's (1988) observation that if δT_e is assumed to translate directly into a variation δc of sound speed c, whose relative value $\delta c/c$ is then assumed to be independent of depth in the convection zone, then a certain degeneracy splitting observed in the frequencies of acoustic oscillations of the Sun could be reproduced theoretically. This would imply that the latitudinal relative variation $\delta T/T$ of temperature is independent of depth throughout the convection zone. No consistent calculation of the Sun's convective envelope, perturbed in some plausible way by a magnetic field, has yet yielded such a property. Kuhn's remarkable result therefore offers an interesting challenge to theoretical heliophysicists, which is bound eventually to yield a better appreciation of solar-cycle dynamics.

It should be appreciated that a naive identification of acoustic wave propagation speed with the sound speed of the solar gas may not be correct. Small-scale fibril magnetic fields, temperature inhomogeneities and convective velocities all modify the speed of propagation of large-scale acoustic waves, and thereby modify the resonant frequencies of global oscillations. Therefore it does not necessarily follow from the observations that $\delta T/T$ is independent of depth. It must also be borne in mind that the measurements of Kuhn *et al.* (1988) might themselves have been influenced by fibril fields, and therefore do not reflect true temperature variations.

Solar oscillations have not yet provided a clear indication of whether the asphericity extends into the radiative interior. Degeneracy splitting in low-degree acoustic modes, which penetrate to the energy-generating core of the Sun, is difficult to discern; no convincing measurement has yet been made. Absolute frequency shifts, correlated with the solar cycle, of groups of nearly degenerate modes have been reported by several observers. The shifts are of such a magnitude as to require a substantial contribution to come from an aspherical component of the structure of the radiative interior (Gough 1988), should they be real. However, one must always be wary of measurements of groups of unresolved modes, because one runs the risk of the data being biased by interference phenomena.

There is another direct measurement I wish to mention that might be pertinent to the issue: the solar neutrino flux. In figure 3 is plotted the smoothed neutrino counts of Davis and his collaborators (1988) on the same diagram as smoothed sunspot numbers and measurements of the apparent solar diameter by Laclare and his collaborators (cf. Delache *et al.* 1985). There is not an obviously convincing correlation. I know that because I submitted some of the data to a critical test: I took the solar diameter measurements (sequence A) and the neutrino data over the same interval (sequence B) and supplemented them with a series of random numbers (sequence C). I took the liberty of removing just one (outlying) point from each sequence, and offered then to a well-known expert in correlative science for his advice. Sequences B and C were unmarked, but sequence A was labelled in Laclare's handwriting, and therefore identifiable. My adviser found an interesting correlation between sequences A and C (subsequently admitting that he had believed sequence C to be T. M. Brown's unpublished solar-diameter measurements), and recommended that I devote some time to work more on the subject. Notwithstanding the outcome of this isolated experiment, it may seem for some that there is at least a hint of a causal connection, and indeed there have been serious claims of a significant correlation (Davis 1988), in addition to claims to the contrary (Bahcall *et al.* 1987). If future data support that hint, then something very important will be learned; indeed, it is because it is so important that I have taken the trouble to discuss what at present is such weak evidence.

A detectable flux of solar neutrinos can come only from the solar core. Therefore, if a variation in that flux is convincingly found, and if it is associated with the solar cycle, one can conclude only that either the source of neutrinos is modulated or that the neutrinos are changed as they pass through the Sun. If the former were the case, then the very core of the Sun would be partaking in the cycle; the cycle would not be caused by a dynamo confined to the outer convective envelope and its immediate environs. In this context it is perhaps worth mentioning that the cycle could result from a 22-year modulation of a relatively short-period nonlinear oscillator, such as an internal gravity mode; as Roxburgh (1986) has pointed out, because of its spatial structure such an oscillator is likely to have the added attraction of reducing the neutrino flux of theoretical models to the observed value without having recourse to modifications to nuclear or particle physics. If the latter were the case, it would follow that neutrino transitions take place, the most likely cause being a dipole interaction with the magnetic field inducing neutrino helicity flipping. This would require the neutrino to have a magnetic moment. Current bounds on the value of that moment, coupled with the uncertainty in the interior solar magnetic field, certainly do not rule out this interesting possibility (see, for

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

example, Leurer & Liu 1989), particularly if the transition were a resonance phenomenon (Akhmedov 1988; Minakata & Nunokawa 1989).

Finally, I return to the problems encountered by dynamo theorists in simultaneously explaining both the decrease in the photospheric angular velocity with latitude and the equatorward migration of low-latitude magnetic activity. As I have already mentioned, dynamical calculations require parameters to be adjusted such that the angular velocity Ω increases with depth to produce the observed latitudinal dependence of Ω; yet Ω needs to decrease with depth to reproduce the observed migration of magnetic activity. Recent analyses of rotational degeneracy-splitting of the Sun's acoustic oscillations (Christensen-Dalsgaard & Schou 1988; Brown *et al.* 1989; Dziembowski *et al.* 1989) have revealed in some detail how Ω really varies in the convection zone. The result is illustrated in figure 4. Throughout most of the convection zone Ω increases with depth. But in a possibly thin transition layer at the base of the convection zone Ω decreases somewhat with depth at low latitudes. (The separation of the contours in figure 4 is too great for this small decrease to be visible.) In a discussion of the

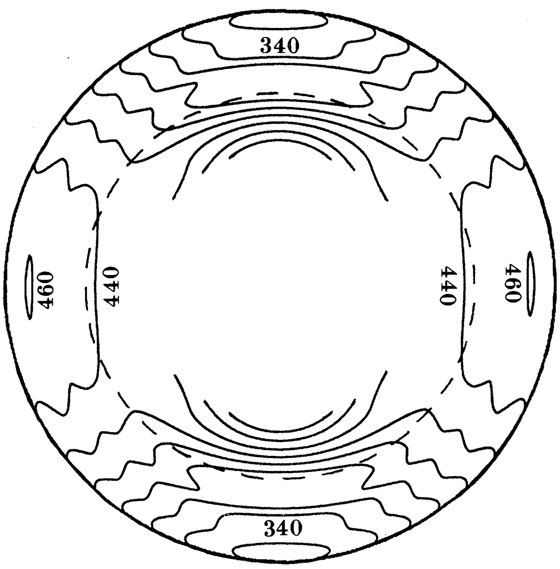

FIGURE 4. Contours of constant angular velocity Ω in the Sun, obtained by averaging the results of the inversions by Christensen-Dalsgaard & Schou (1988), Brown *et al.* (1989) and by Dziembowski *et al.* (1989) of measurements of rotational splitting of solar five-minute acoustic modes. The contours are drawn in a meridional plane through the Sun. They are labelled in nanohertz, and are separated by 20 nHz. They are drawn only in the outer half, by radius, of the Sun; the data are unreliable nearer the centre. The dashed circle indicates the base of the convection zone. The axis of rotation is vertical.

dynamical implications of this discovery, Morrow *et al.* (1988) have recently pointed out that this leads the way to a dynamo theory that might explain the observations: the negative value of $d\Omega/dr$ throughout most of the convection zone is consistent with the overall dynamics of the convection zone in which there is a decline of Ω with latitude, whereas the dynamo is situated in the thin transition zone in which the possibly time-dependent value of $d\Omega/dr$ is positive at low latitudes, leading to low-latitude dynamo waves that propagate equatorward. The ideas also suggest explanations to other properties of the solar cycle, which I have not discussed in this paper.

6. Conclusion

After carefully weighing the evidence, and discounting the numbers of scientists who would vote for each of the principal types of proposed explanation of the solar cycle, the conclusion must be that we do not yet know whether the Sun is an oscillating dynamo or whether the solar magnetic field is decaying from its primeval state. Nor do we know whether the mechanisms that cause the polarity of the dipole component of the surface field to reverse are confined to the convection zone or whether they are controlled by a deep-seated oscillator in the solar core.

I am grateful to R. N. Bracewell for giving me the phases of his Hilbert transform of the sunspot record, and to W. J. Merryfield for his assistance in producing figure 4. I gratefully acknowledge support from NASA grant NSG-7511.

References

Akhmedov, E. K. 1988 *Phys. Lett.* B **213**, 64–68.
Bahcall, J. N., Field, G. B. & Press, W. H. 1987 *Astrophys. J.* **320**, L69–L73.
Batchelor, G. K. 1950 *Proc. R. Soc. Lond.* A **201**, 405–416.
Bracewell, R. N. 1985 *Aust. J. Phys.* **38**, 1009–1025.
Bracewell, R. N. 1988 *Q. Jl R. astr. Soc.* **29**, 119–128.
Brown, T. M., Christensen-Dalsgaard, J., Dziembowski, W. A., Goode, P., Gough, D. O. & Marrow, C. A. 1989 *Astrophys. J.* **343**, 526–546.
Christensen-Dalsgaard, J. Schou, J. 1988 *Seismology of the Sun and Sun-like stars* (ed. E. J. Rolfe), pp. 149–153. ESA SP-286, Noordwijk.
Däppen, W. 1983 *Astron. Astrophys.* **124**, 11–22.
Davis, R. 1988 Seventh workshop on grand unification/Icoban '86 (ed. J. Arafune), pp. 237–276. World Scientific.
Delache, P., Laclare, F. & Sadsaoud, H. 1985 *Nature, Lond.* **317**, 416–418.
Dicke, R. H. 1970 *Stellar rotation* (ed. A. Stetteback), pp. 289–317. Dordrecht: Reidel.
Dicke, R. H. 1978 *Nature, Lond.* **279**, 676–680.
Drobyshevski, E. M. & Yuferev, V. S. 1974 *J. Fluid Mech.* **65**, 33–44.
Dziembowski, W. A., Goode, P. & Libbrecht, K. G. 1989 *Astrophys. J. Lett.* **337**, L53–L57.
Gough, D. O. 1978 *Pleins feux sur la physique solaire* (ed. J. Rösch), pp. 81–103. Paris: CNRS.
Gough, D. O. 1981 *Variations in the solar constant* (ed. S. Sofia), pp. 185–206, NASA Conf. Publ. 2191. Washington, D.C.: U.S. Govt. Printing Office.
Gough, D. O. 1988 *Solar–terrestrial relationships and the Earth environment in the last millenia* (ed. G. C. Castagnoli), pp. 90–132. Bologna: Soc. Italiana di Fisica.
Gough, D. O. 1988 *Seismology of the Sun and Sun-like stars* (ed. E. J. Rolfe), pp. 679–683. ESA SP-286, Noordwijk.
Gough, D. O. & Thompson, M. J. 1990 *Mon. Not. R. astr. Soc.* (In the press.)
Kuhn, J. R. 1988 *Astrophys. J. Lett.* **331**, L131–L134.
Kuhn, J. R., Libbrecht, K. G. & Dicke, R. H. 1988 *Science, Wash.* **242**, 908–911.
LaBonte, B. & Howard, R. F. 1982 *Sol. Phys.* **80**, 373–378.
Leurer, M. & Liu, J. 1989 *Phys. Lett.* B **219**, 304–308.
Mestel, L. & Weiss, N. O. 1987 *Mon. Not. R. astr. Soc.* **226**, 123–135.
Minakata, H. & Nunokawa, H. 1989 *Phys. Rev. Lett.* **63**, 121–124.
Morrow, C. A., Gilman, P. A. & DeLuca, E. E. 1988 *Seismology of the Sun and Sun-like stars* (ed. E. J. Rolfe), pp. 109–115. ESA SP-286, Noordwijk.
Roxburgh, I. W. 1986 *The internal solar angular velocity* (ed. B. R. Durney & S. Sofia), pp. 1–5. Dordrecht: Reidel.
Spiegel, E. A. & Weiss, N. O. 1980 *Nature, Lond.* **287**, 616–617.
Walèn, C. 1946 *Ark. Mat. astr. Fys.* A **33** (18).
Willson, R. C. & Hudson, H. S. 1988 *Nature, Lond.* **332**, 810–812.
Yule, G. V. 1927 *Phil. Trans. R. Soc. Lond.* A **226**, 267–298.

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