HOW IS SOLAR ACTIVITY INFLUENCING THE STRUCTURE OF THE SUN?

D. O. Gough
Institute of Astronomy, Madingley Road, Cambridge CB3 0HA and Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW

To Roger Bonnet

ABSTRACT

To what extent does solar activity change the structure of the Sun? And what is solar activity, for that matter? Some would say it is the gamut of violent events seen in and above the upper atmosphere. If that is all it is, then its effect on the Sun itself, that is the region beneath the photosphere where almost all the matter resides, must be utterly negligible. But if one includes the cause of these violent events as part of the activity, then associated structure changes in the Sun must be more profound.

1. INTRODUCTION

It was more than twenty years ago that I first thought seriously about those putative structural changes in the Sun with which solar activity might be associated. Trying to explain the immediate cause of the activity and its variation was in those days mainly the realm of solar-dynamo theorists, as is still the case today; but the dynamo theorists confined their attention to only kinemal and dynamical issues, paying little or no attention to associated structural variation, as is essentially still the case today. I was therefore moved to address the question of how one might relate potentially observable quantities to putative structural changes, and, in particular, whether one can learn from observation where in the Sun those changes take place. The results were summarized in an article entitled: ‘On the seat of the solar cycle’ (Gough, 1981) which was delivered to a workshop in 1980 on variations in the solar constant organized by Sabatino Sofia. Much of what I say here simply brings part of that article up to date.

There has been some theoretical advance since that time, but mostly in the details of the processes taking place in the Sun’s superficial layers. There has been little work on more profound processes. Most of the real progress that has been made is observational: principally, the detailed accurate measurement of radiance and irradiance, and its interpretation in terms of observable surface features, much of which has been carried out by the VIRGO investigation on SOHO, led by Claus Fröhlich, this symposium’s principal scientific organizer. Although the results are of immense importance, their value, in my opinion, would be enhanced enormously for a theorist interested in structural change if they were complemented by contemporary measurements of the solar radius. The reason for this opinion will be elucidated in the next section. But such measurements are difficult to make from the ground, because the changes are so small that telluric atmospheric fluctuations mask them (Brown and Christensen-Dalsgaard, 1998). In principle the measurements should be relatively straightforward from a purpose-built spaceborne instrument such as the solar sextant proposed by Sabatino (cf Sofia, Heaps and Twigg, 1994).

It was therefore a source of great delight to me when Emilio et al. (2000) announced a positive measurement of possibly solar-cycle-related variation in the Sun’s radius, using SOI/MDI data (Scherrer et al., 1995), and a glorious opportunity to express that delight when I was invited to write News and Views about it for Nature. My tentative inference, as I explain below, was that, if correct, the measurements implied that the principal structural change responsible for those variations was located near to, but perhaps not extremely near to, the solar photosphere. But as is so often the case when one expresses one’s views, there were repercussions. In this case they started apparently mildly, with a message from my friend Peter Foukal welcoming me from the heretical school of profundists to the ubiquitous school of superficialists. My immediate reaction was one of mild appreciation. But it subsequently transpired that Peter’s welcome was misconceived: in a letter he wrote in response to my article in Nature he made it clear that his view of superficiality encompassed not merely the regions in the Sun in which he believes that all solar-cycle-related structural changes take place, but also a domain of intellectual pursuit beyond which it is unproductive to tread; he regards it to be ‘unnecessary’ to consider processes that may operate more profoundly to draw on the potential and thermal energy of the convection zone to pro-
duce the plainly visible dark spots and bright faculae which modify the photospheric thermal impedance and thereby induce the observed luminosity changes. I disagree with this view profoundly, and therefore I cannot number amongst Peter’s superficialists. And in accepting the invitation by the organizers of this symposium to address the question that entitles this presentation I trust that it will become clear why. Since Peter Foukal is not present at this meeting, I shall take care to refer to his contrary views only by quoting verbatim from his letter, trusting that the many of his advocates who must surely be present will ensure that I do not mislead by quoting out of context.

Thus I dedicate this presentation not only to Roger Bonnet, but also to profundity, the quest of which drove Roger first to argue for and then to support SOHO, one of the most, if not the most successful of the missions that ESA has ever conceived.

2. THE STORY OF $W$

In the absence of a comprehensive theory of the solar cycle, we are not in a position to predict the absolute magnitude of any structural change, and hence of the amplitude of any observable consequence. We are thus led to consider amplitude ratios, with the expectation that because the absolute values of the perturbations are observed to be small, linearized theory may not be too far from being representative of possibly pertinent changes that one might contemplate, and that therefore amplitude ratios are robust diagnostics.

The amplitude ratio that I have in mind is

$$W = \frac{\Delta \ln R}{\Delta \ln L},$$

(1)

where $R$ and $L$ are respectively the radius and luminosity of the Sun, and $\Delta$ donates a difference. There was a flurry of activity in the late 1970s and early 1980s in trying to compute $W$ theoretically, typically by tinkering with numerical computations of stellar evolution. At first, the only kind of change contemplated was an adjustment to the efficacy of convection, via a (typically instantaneous) modification to the mixing-length parameter $\alpha$. One of the principal motivations for carrying out such calculations was for estimating variations in $L$, and hence variations in the solar irradiance, to study implications for the terrestrial climate. One was acutely aware in those days that small yet possibly significant irradiance variations could not be measured from the ground, yet if armed with a value for $W$ one ought to be able to infer those variations from measurements of the variations in $R$, which at the time were considered to be almost within reach (e.g. Sofia, 1981). Little did one know how differently the subject was to evolve: two decades later we are uncertain of $R$ and we don’t know $W$; the only reasonably reliable values we have are those of $L$, which we infer from the superb irradiance measurements plotted in Figure 1.

In the early days there was a very broad dispersion in the theoretical estimates of $W$ (e.g. Dearborn and Newman, 1978; Sofia et al., 1979; Dearborn and Blake, 1980; Gilliland, 1980; Sofia and Endal, 1980; Twigg and Endal, 1981). One of the culprits was a variation amongst the time steps used in the stellar evolution computations, few, if any, of which were short enough to resolve the true response of a solar model to an instantaneous perturbation. Thus in some sense the computations inadvertently approximated crudely the relaxation of the theoretical models on different timescales, and we know today that the Sun’s response to perturbations is strongly time dependent (see Figure 2). Indeed, that was realized at the time of Sabatino’s conference (Gough, 1981; Sweigart, 1981), so the dispersion amongst the published results was hardly surprising. But, in addition to the variation of effective timescales, there was probably also substantial numerical error in representing the spatial stratification in many of the computations (as was noted by Sweigart, 1981), because typical stellar evolution computations that were not designed for studying helioseismology were not accurate enough to represent faithfully the typically tiny structural modifications that were being sought.

All these deficiencies were really relatively minor, because they were correctable. What could not be accounted for was our lack of knowledge of how and where the modifications to $R$ and $L$ are induced.

At Sabatino’s conference a variety of putative perturbations to the structure of the Sun were considered. Some of the results are summarized here in Table 1. Although the observable responses at the surface of the Sun were always all small, the perturbations
Table 1. Summary of the responses of the Sun to various disturbances. The characteristic radius $r_0$ at which the disturbance occurs was estimated for the first two cases from the first moment of the depth weighted by $\nabla = \nabla_{id}$ and $\rho_t/p$ respectively, where $\nabla = \text{dln}T/\text{dln}p$, $\nabla_{id} = (\text{dln}T/\text{dln}p)_{id}$, $\rho_t$ is turbulent pressure, $p$ is gas pressure and $T$ is temperature (from Gough, 1981 and Balmforth, Merryfield and Gough, 1996).

<table>
<thead>
<tr>
<th>Disturbance</th>
<th>$\Delta \ln L$</th>
<th>$\Delta \ln R$</th>
<th>$1 - \frac{r_0}{R}$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibition of convection by reducing the mixing-length parameter $\alpha$ by $6.3 \times 10^{-4} \alpha$</td>
<td>$-4.5 \times 10^{-4}$</td>
<td>$-8.6 \times 10^{-7}$</td>
<td>$1.1 \times 10^{-3}$</td>
<td>0.002</td>
</tr>
<tr>
<td>Introduction of a tangled magnetic field in energy equipartition with the convective motion</td>
<td>$2.6 \times 10^{-2}$</td>
<td>$1.2 \times 10^{-4}$</td>
<td>$1.4 \times 10^{-3}$</td>
<td>0.005</td>
</tr>
<tr>
<td>Modifying the temperature gradient in the bottom 0.1 pressure scale heights of the convection zone by increasing the mixing-length parameter $\alpha$ by a factor 1000</td>
<td>$-9 \times 10^{-6}$</td>
<td>$-1.6 \times 10^{-6}$</td>
<td>0.3</td>
<td>0.18</td>
</tr>
<tr>
<td>Any disturbance that is confined to the energy-generating core</td>
<td>$\geq 0.8$</td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
</tbody>
</table>

that produced them were not, and therefore the computations were performed by piecing together time-independent portions of the model in such a manner as to accommodate the nonlinearity, calculating changes in the radiative interior by adiabatic hydrostatic readjustment of a normal solar model, supporting a convective envelope with uniform luminosity $L(r)$, potentially different from the original, but with unchanged specific entropy deep in the zone. Thus the readjustments were meant to represent the response on timescales much less than the gross characteristic thermal relaxation times of the radiative interior ($3 \times 10^8$ years) and the convection zone ($10^9$ years) yet long compared with the internal thermal readjustment time of the convection zone (1 year). From these scant data it was conjectured that the obvious properties exhibited in the table, namely

$$W(r_0) > 0$$

and

$$\frac{dW}{dr_0} < 0,$$

where $r_0$ is the radius about which the (rather localized) perturbing agent operates, hold universally for all perturbations, at least on solar-cycle timescales. This raised the possibility of being able to say something about the location of the seat of the solar cycle from a measurement of $W$. (Of course, if the perturbing agent were not localized in radius, or if it were localized about more than a single radius, the value of $W$ represents only some intermediate location.) Subsequently, an extensive linear perturbation analysis was undertaken by Däppen (1983), who considered a complete set of elementary perturbations at a variety of locations $r_0$ operating over a wide range of timescales, the responses to which more-or-less confirmed the conjecture. His results for thermal perturbations are reproduced in Figure 2.

We come now to the radius variations reported by Emilio et al. (2000). After correcting for an error (of a factor 2) in what the authors mean by amplitude, and combining the result with the irradiance variations illustrated in Figure 1 (which one interprets as being proportional to variations in $L$), the result implies that $W \approx 0.04$ (Gough, 2001). On referring this value to Table 1 and Figure 2, one would infer that the structural perturbation responsible for the variations might be located in the outer layers of the Sun, perhaps near $r \approx 0.98 R$ ($q_e \approx -5$), which, perhaps fortuitously, is the location of the zone of the second ionization of helium. It was this conclusion that prompted Peter Foukal to attempt to welcome me to superficialism. I must emphasize, however, that $r = 0.98 R$ is too deep for the structural modification to have been produced solely by a uniform modification to the efficacy of convection (see Table 1) the cause of the photospheric thermal impedance variations considered by Foukal (e.g. Foukal, Fowler and Livshits, 1983), or even by the introduction of a tangled magnetic field; an adoption at face value of the result derived from the paper of Emilio et al. (2000) implies that there is a significant perturbation located more deeply.
It behoves me to point out that the conclusion of Emilio et al. is not uncontroversial. Indeed, one might even be tempted to entertain some doubt simply from noticing that at the end of the paper Emilio et al. offer the inferred radius variation merely as an upper bound. Nevertheless, Jeff Kuhn at this meeting has advised us to take the value seriously. I shall not enter a discussion of the wisdom of either taking or rejecting this advice. Instead, I simply point out that there appears to be evidence from the temporal variation of high-degree f-mode frequencies that the variation of $\bar{R}$ is actually in the direction opposite to that reported by Emilio et al. (implying that $W < 0$), and of a somewhat smaller magnitude (Antia et al., 2000; Dziembowski, Goode and Schou, 2001).

In the absence of horizontal inhomogeneity, convective motion and magnetic fields, the frequencies $\omega$ of high-degree $f$ modes are extremely insensitive to the stratification of the star, and depend only on the mass, $M$, and radius, $R$, of the Sun (e.g. Gough, 1980). Since $M$ is conserved, variations of $\omega$ might therefore be a robust indicator of variations in $R$. But the Sun is not smooth and quiescent in its outer layers, and although Antia et al. argue that the frequency variation does not exhibit the degree-dependence that one would expect from perturbations produced by convective inhomogeneity, it is probably fair to say that an adequately careful account of the possible interactions of convection and magnetic fields with the $f$ modes has not yet been undertaken (cf Murawski and Roberts, 1993a,b; Mydrek, Murawski and Roberts, 1999). So perhaps one should be cautious, and simply regard the value of $W$ as an open issue.

3. SUPERFICIAL AND PROFOUND PERTURBATIONS

In order to be specific, I restrict my discussion to consideration of only thermally induced perturbations, although much of what I have to say applies more generally. So far as the reaction of the convection zone is concerned, the pertinent point to realize is that the internal thermal equilibration time (namely, the time $\tau_\eta$ — about 1 y — that it takes to reestablish local thermal balance within the convection zone after a perturbation) is very much shorter than the cooling time (the time $\tau_c$ — about $10^5$ y — that it takes for an overall readjustment to a perturbation in the rate at which the zone radiates heat into the universe, or in the rate at which it receives heat through its base). Roughly speaking, though adequately closely for my immediate purpose, I can liken the situation to a thick block of silver surmounting a block of ebony heated from below, and radiating from its upper surface into its (cool) environment (Figure 3). Because the thermal conductivity of silver is very high, the silver block is almost isothermal, and because that block is thick, the heat capacity, and hence the thermal inertia, is high — the time $\tau_c$ it takes for the silver block to cool is much greater than the time $\tau_\eta$ it takes a temperature perturbation to diffuse throughout its interior. Thus the situation is similar to the solar convection zone, although the details are different, but not materially so: the thermal inertia of the convection zone is high, and so is the rate at which thermal perturbations are redistributed through the medium, although the mechanism of the transfer is different from that in a block of silver. (It is by convection, rather than conduction, and satisfies an equation of hyperbolic character — because the temperature perturbation that is transported modifies the buoyancy and hence influences the driving of the convective flow — which is more similar to a wave equation than it is to a (parabolic) diffusion equation; moreover, the steady state to which the system evolves is essentially isentropic rather than isothermal.)

Consider now the upper surface of the silver block suddenly to be painted with a thin uniform coating of a poorly conducting lacquer. As a result of the immediate decrease of the conductivity of the outermost layer, a temperature gradient is set up in that layer, causing the surface of the lacquer to cool. Therefore the rate $L$ at which heat is radiated from the surface suddenly declines. But there is no immediate change to the heat content of the block; that change occurs slowly, on the timescale $\tau_\eta$. If, on the other hand, it were the lower surface that were painted, there is no immediate change to conditions at the upper surface; only on a timescale $\tau_\eta$ does the heat content of the block slowly fall, and with it the temperature, and in particular the temperature of the upper surface (because $\tau_c \ll \tau_\eta$), which causes $L$ to fall in step. These statements remain true if instead of regarding the painting to be instantaneous the lacquer were imagined to come and go cyclically with a period $\tau_\eta$ that satisfies $\tau_\eta \ll \tau_\eta \ll \tau_c$. The statements are basically true of the solar convection zone too, except that in
that circumstance the compressibility of the material of the convection zone must also be taken into account; the convection zone contracts as the heat content of the zone decreases. Thus, if conditions at the base of the convection zone were changed in such a manner as to decrease the heat flux into the zone, both L and R would fall together, although not by very much, and one would expect \( 0 < W = O(1) \). But when it is near the upper surface that heat transport is inhibited, the diminution of L is much greater than the change (actually an augmentation) in the heat content of the zone, and \( |W| \) is small. One might be tempted to infer from this argument that \( W < 0 \), but that is not actually the case, because the cooling of the outer poorly transporting superadiabatic layer contributes to a decrease in radius that on a timescale \( \tau_\xi \) exceeds the minute expansion of the rest of the zone. Thus \( 0 < W \ll 1 \).

I have advanced these arguments to provide some appreciation of the dominant physical consequences of imagining a thermal perturbation. And I have chosen a thermal perturbation, in fact a perturbation to the efficacy of thermal transport, rather than a dynamical perturbation, because the convective transport is easily imagined to be modified without the expenditure of a substantial amount of (internal) work whose consequences on the balance of forces determining the hydrostatic stratification of the zone would otherwise need to be taken into account. But, of course, the horizontally uniform perturbations that I have considered do not represent what we believe to occur in the Sun. In particular, heat transport near the surface is known to be blocked locally by sunspots and enhanced locally by faculae and bright points in the magnetic network. This causes almost instantaneous modifications to L, as indeed does applying lacquer only in spots to the upper surface of the silver block in Figure 3. So once again \( W \) is small in magnitude, although its sign is not immediately evident, partly because one first has to decide how, or perhaps I should say where, to define R. If one uses a quiet-sun value, which is probably mainly what the \( f \) modes are sensing, then by the arguments that I advanced above one might expect \( W < 0 \). It is perhaps prudent for me not to presume what radius Emilio et al. (2000) have measured.

Substantial effort has been devoted to trying to understand how heat is redistributed after the appearance of a sunspot. The diminution of the efficacy of heat transport through the spot brought about by the magnetic inhibition of convection is believed to be the dominant agent. This causes a cooling of the uppermost optically thick layers and, on the timescale \( \tau_\xi \) of the spot lifetime (which is less than \( \tau_\xi \)) a damming up of the heat beneath the spot, and a consequent heating. The telechronoseismological analysis of Kosovichev, Duvall and Scherrer (2000) and Zhao, Kosovichev and Duvall (2001) confirms this picture. Spruit (1982), Foukal, Fowler and Livshits (1983) and Chiang and Foukal (1985) have tried to estimate how much of the dammed-up heat is transported around the spot to the surface, to produce a bright ring. They adopted a (parabolic) diffusion approximation for the transport, which, as I pointed out above, and have pointed out before (e.g. Gough, 1981), is fundamentally wrong. Nevertheless, it does not follow that their conclusion is wrong too.

The first simple models produced little change in the energy emitted from sunspots, at least on timescales comparable or less than \( \tau_\xi \), although Chi\-ang and Foukal (1985) demonstrated that excess emission from a surrounding bright ring can be engineered to cancel, more or less, the blocking by a sunspot, at least on a timescale greater than \( \tau_{fp} \). However, although the observable aspects of the global energy balance can be described well in a parametrized way (e.g. Fröhlich, 1994), a sound theoretical basis for the variation seems still to be lacking (e.g. Spruit, 1994). Models based on thermal diffusion alone appear to be inadequate. Moreover, the seismological analysis reveals coherent outflows beneath the spot, no doubt driven by the temperature perturbation produced by the spot, which is bound to transport some of the heat in a nondiffusive manner.

Less attention has been paid to laterally inhomogeneous perturbations near the base of the convection zone. Kuhn (1994) has carried out an interesting numerical simulation, however, and has found, not surprisingly, that the outcome is not in accord with the predictions of simple diffusion. Heat transport is predominantly vertical so that a 'shadow' of a profound disturbance is imprinted on the superficial layers of the convection zone, and is not washed out by as much as one who believes in simple convective diffusion of an active scalar might suppose.

4. HELIOSEISMOLOGICAL DIAGNOSIS

The distinction between the reaction of the convection zone to superficial and profound perturbations is manifest in the perturbation to the acoustic radius (namely, the perturbation to the sound travel time \( \tau = \int c^{-1} \, dr \) from the centre of the Sun, where \( c(r) \) is the adiabatic sound speed at radius \( r \)), which is illustrated in Figure 4 for the reaction to perturbations to the efficacy of convection (produced by changing the mixing-length-to-pressure-scale-height ratio \( \alpha \)) that are (i) near the bottom of the convection zone, which induces a disturbance throughout the entire convection zone, and (ii) uniform in space, creating a structural disturbance principally in the outer layers of the convection zone where the convective-adiabatic balance is relatively delicate.

The acoustic radius is the principal determinant of the propagation time of acoustic modes, and the relative perturbation \( \delta r / r \) is the pertinent quantity that determines frequency changes. (There are modifications due to wave refraction and the inhibition of propagation by density stratification.) It should be appreciated, however, that it is really the acoustic radius difference between the lower and upper boundaries of the acoustic cavity that matters, for that determines the propagation time across that cavity, and consequently the resonant frequencies.
concentrate my discussion on the upper reflecting boundary, whose location depends predominantly on the frequency of the mode, and not so much on the degree. Reflection occurs when a wave encounters a region in which it cannot propagate, which, for the case of an acoustic wave, occurs when the background state no longer varies slowly on the natural lengthscale $\lambda$ of the wave: actually, when, roughly speaking, $\lambda \equiv k^{-1} \gg 2H$, where $k_\nu$ is the vertical component of the wavenumber $k$ of the wave and $H$ is a scale height (I shall not discuss here the details of which scale height, but for reflection near a stellar surface one may safely regard it as a scale height of pressure — Gough, 1993). Since $H$ is roughly proportional to depth below some (appropriate) fiducial level in the atmosphere, waves with higher frequency $\omega \approx (\omega_0)$ are reflected nearer to the surface.

We can now discuss the frequency dependence of the p-mode frequency perturbation resulting from various kinds of perturbation to the background state. Other similar discussions have been presented by Gough and Thompson (1988), Gough (1990a, 1994), Goldreich et al. (1991) and Balmanfort et al. (1996). It is evident from Figure 4 that an increase in the efficacy of convection solely at the base of the convection zone (case i) on a solar-cycle timescale decreases $\tau$ throughout almost all of the convection zone, yet leaves it essentially unchanged beneath; it therefore increases the frequencies of all p modes (the negative contribution from the lowermost layers of the convection zone is always dominated by the influence of the more extensive frequency-increasing layers above), by an amount which increases gradually with increasing frequency because the higher-frequency modes propagate up to higher levels in the envelope where the magnitude of $\delta \tau/\tau$ is greater. The frequency shift $\delta \omega$ is a gradually increasing function of $\omega$, as is illustrated in Figure 5a for low-degree modes. That variation is not in accord with the observations. A uniform change to the efficacy of convection (case ii), on the other hand, induces a perturbation to $\delta \tau$ that is concentrated only in the very uppermost region of the convection zone, much of which lies above the level at which the acoustic waves are reflected. In this case the frequency dependence of $\delta \omega$ is dominated by the degree of evanescence in the perturbed layers, which is proportional to the inverse mode inertia $I^{-1}$ (Gough, 1993b) and is a steeply increasing function of frequency. The outcome is more-or-less in agreement with observation, at least for modes whose cyclic frequencies are below about 4 mHz. So is not the issue settled?

![Figure 4](image_url)

**Figure 4.** In-phase component of relative Lagrangian variation $\delta \tau/\tau$ of the acoustic radius $\tau$ plotted against $\tau$ resulting from a 22-year periodic perturbation in the mixing-length parameter $\alpha$ with relative amplitude $\delta\alpha/\alpha = 1000$ in the bottom 0.1 pressure scale heights of the convection zone, and $\delta\alpha/\alpha = 3.3 \times 10^{-3}$ everywhere; the out-of-phase components are relatively small (from the calculations of Balmanfort, Gough and Merrifield, 1996).

![Figure 5](image_url)

**Figure 5.** The squares are the cyclic frequency increases from 1986 to 1988 of low-degree modes of the Sun, averaged over degree in the range $5 \leq l \leq 60$, with 1-$\sigma$ error bars in (b) (from Libbrecht and Woodard, 1990a). The solid lines in (a) join the theoretical frequency differences resulting from the disturbance illustrated in Figure 4a; the theoretical differences have been scaled up by a factor $4$ to allow clearer comparison with the observational data. The solid curve in (b) is proportional to the inverse mode inertia $I^{-1}$ for $l = 20$ normalized at an optical depth of 0.05, and is close to the theoretical frequency differences (without scaling) resulting from the perturbation illustrated in Figure 4b. The dashed curve differs from the solid curve by an oscillatory function with 'period' 700µHz with amplitude proportional to $I^{-1}$, chosen to accentuate a putative property of the data.

The answer to that question is: It is not. Firstly, as was stressed by Gough and Thompson (1988b), Goldreich et al. (1991) and Balmanfort et al. (1996),

© European Space Agency • Provided by the NASA Astrophysics Data System
the perturbation to a required to reproduce the observed frequency changes produces a grossly incompatible value of $\Delta \ln \nu$; too great in magnitude by a factor of about 45 and of the wrong sign. So the frequency changes must be produced by something else. Secondly, observations subsequent to those illustrated in Figure 5 showed that there is a fairly rapid, although not precipitous, decline in the magnitude of the frequency perturbation in the vicinity of 4 mHz, although possibly, according to some measurements (not depicted in Figure 5) even a reversal of sign, followed by return towards the theoretical curve at yet higher frequencies (Woodard and Libbrecht, 1991). This is not explained by a thermal perturbation of the kind produced by convection. (A very narrow excursion to near-zero values could result from a sharp chromospheric resonance, of the kind exhibited in smooth spherically symmetrical solar models, that does not change substantially with the solar cycle (Gough, 1990a); perhaps a broader resonance might be more realistic, because the chromosphere is horizontally inhomogeneous, but if it were as broad as the observations suggest, it is questionable whether it would be able to produce an adequate reduction in the frequency perturbation without producing other observable effects.)

Gough and Thompson (1988), who addressed spatial, latitudinal variation, suggested that such a perturbation is magnetic, and due to fibril fields of a magnitude that appeared to be not incompatible with fields that had recently been measured by the photosphere might be responsible instead (Evans and Roberts, 1999). Such fields might also be responsible for the observed reduction in the frequency shift at cycle frequencies higher than about 4 mHz, although thermal changes above the photosphere might be responsible instead (Evans and Roberts, 1990; Goldreich et al., 1991; Jain and Roberts, 1994). It seems not to be possible at present to distinguish between different kinds of superficial structural variation by seismological means. Therefore, if the goal is to determine what it is that causes the structural perturbation, as it is mine, rather than merely to produce a model that is not ruled out by observation, one is forced to seek other diagnostics. That is why $W$ is not uninteresting.

Before leaving this subject, permit me to recall an additional feature of the solar-cycle frequency variations plotted in Figure 5b. Not surprisingly, there are deviations from the smooth theoretical curve. Indeed, it can hardly be surprising, given that I have argued that the theoretical curve may be inapplicable. But one might expect the outcome of an appropriate distribution of magnetic field to be able to produce a variation similar to the theoretical curve in the figure. Could any remaining discrepancy below 4 mHz simply be observational error, notwithstanding the estimated errors? Many may argue that this is likely, but rather than being so dismissive I wish at least to rekindle an idea that there is evidence, ad

mittedly rather weak, for a residual signal. With the eye of faith the credulous might discern an oscillatory component to that residual, and I have included in the figure an oscillatory dashed curve to guide (or, perhaps, misguide) the eye.

The oscillation was noticed also by Goldreich et al. (1991). And, if real, would be an indicator of a temporally varying structural perturbation localized about an acoustic depth of $(2P_0)^{-1}$, where $P_0$ is the period of the oscillation with respect of cyclic frequency $\nu$ (Gough and Thompson, 1988; Vorontsov, 1988; Gough, 1990a). The period of the dashed curve in Figure 5b is about 700 $\mu$Hz, which corresponds to a depth of about 0.03 $R_\odot$ beneath the photosphere, in the He II ionization zone where about 80 per cent of the helium is completely ionized; this raises the issue of whether the equation of state might be particularly susceptible to changes in the strength of the magnetic field when the plasma as in a state of partial ionization. But, of course, the putative oscillation may not be real. Indeed, Sarbani Basu has announced at this meeting that in response to my earlier discussions of the matter (Gough, 1991, 1990a) she has sought the feature in more recent MDI data and has failed to find it. I should point out also that if the BBSO data of Libbrecht and Woodard (1990a) are binned differently, the oscillation can be suppressed. So perhaps the signal is revealed only by appropriate binning; or perhaps it is actually an artefact of inappropriate binning. Or - and I regard as being much less likely - perhaps the perturbation producing such an oscillatory component to the frequency variation is present only in alternate sunspot cycles; are there no other 22-year features in the cycle aside from the polarity of the dipole component of the magnetic field? Certainly some have been claimed.

Of course one must be careful not to overinterpret the data. But it is also unwise to infer from someone's failure to see a signal that that signal is not there. It is unadventurous not at least to speculate on possible signs of interesting phenomena; it is the pursuit of ways of testing such speculation that is likely to lead us most readily towards an eventual understanding of the mechanism of the solar cycle.

5. ROTATION AND THE VARIATION OF MAGNETIC ACTIVITY

I do not presume to summarize the enormous amount of research that has been devoted to explaining the variation of the Sun's magnetic field. Suffice it to say here that there is apparently universal belief that winding of magnetic field by differential rotation coupled with field ascension, probably associated with further rotation (this time about a different axis) are essential kinematical ingredients. There can be little doubt that that belief is correct. But what is somewhat less clear, at least to me, is whether the magnetic field plays a dynamical role that is essential to the operation of the cycle. And what is even less
clear is the separate issue of whether the process is a genuine dynamo, maintaining the global magnetism, or whether it is merely a mechanism to distort and possibly amplify the decaying relic of the Sun's primordial field, producing activity that is decaying with that relic.

Given the kinematical considerations it is natural to suspect that there may be solar-cycle related variations in the Sun's angular velocity. There were indications of variations in the rotational splitting of p modes measured by Libbrecht and Woodard during Hale cycle 22 (Libbrecht, 1989; Libbrecht and Woodard, 1990b; Goode and Dziembowski, 1993; Gough et al., 1993; Gough and Stark 1993), but the epochs during which the data were acquired were too sparse for a coherent variation to be discerned. The almost continuous data acquired in the current cycle have changed the situation dramatically. Recently, Howe et al. (2000) have studied inversions of p-mode rotational splitting frequencies from GONG and SOI/MDI, in search of a long-term variation. There was no convincing indication of an 11-year variation (nor of a 22-year variation, but the data set is not long enough for that). However, a marked oscillatory variation in the angular velocity with a period of about 1.3y was found immediately above and below the base of the convection zone in the equatorial regions (within the sunspot belt, although the spatial resolution of the inversions was inadequate to define the latitudinal extent of the oscillation to a precision significantly better than the extent of the belt). To have found this period, which might perhaps be the fifth harmonic of the 22-year magnetic variation, with no evidence of a lower harmonic, was a surprise, although subsequently this period (amongst others) was reported to be present in variations in the solar wind and the Earth's magnetosphere (e.g. Lockwood, 2001).

I have looked at the splitting data, and I am reasonably convinced that the inferred oscillation in the angular velocity is a credible consequence of those data, provided that the splitting really is produced by rotation; and there is even evidence that the oscillation extends beneath the levels reported by Howe and her colleagues. But one must be cautious, and entertain the possibility that the splitting is produced by some form of contamination of the data by surface activity on the Sun, and not by rotation at all. I should also point out that Antia and Basu (2000) have failed to find a similar oscillation in their analysis of splitting data, and have concluded thereby that the inferences of Howe et al. are flawed. As in the case of the apparently oscillatory feature in the data plotted in Figure 5b, one should therefore be cautious; but equally one should not immediately jump to the conclusion that a failure to find something implies that that something is not there.

6. WAVE-INDUCED OSCILLATIONS

For half a century or so there have been discussions of how the structure and the angular velocity of the Sun might be influenced by internal waves. We now know that wave transport plays a crucial role in the dynamics of the earth's atmosphere, and it is hardly unlikely that the same is true of the solar interior. Gravity waves generated by the turbulence in the convection zone propagate into the radiative interior, and those that dissipate deposit energy and angular momentum. Most estimates of the spectrum of the waves that are generated imply an energy flux that is rather small compared with the radiative flux, but perhaps the flux of angular momentum could be great enough to have a significant impact on the distribution of angular velocity in the Sun's radiative interior.

The interesting property of gravity waves to which I draw attention here is that, unlike acoustic waves, they have a tendency, if generated randomly, to enhance (rather than suppress) shear. They act in a manner that some have likened to 'negative viscosity', but I refrain from that expression because, unlike viscous transfer, the process of (angular) momentum transfer by waves is not local. Indeed, it is the very property of nonlocality that is responsible for the richness in the behaviour of sheared stratified flows. Most of the discussions of gravity-wave angular-momentum transport in the Sun have been motivated by the problem of solar spin-down. To what extent has the solar-wind torque which is slowing down the outer layers of the Sun been transmitted to the core? Is the core rotating at a rate similar to the surface, or is it rotating much more rapidly? These were issues of extensive debate in the 1960s, before helioseismology, and were triggered largely by Dicke and Goldenberg's (1967, 1974) attempt to measure the oblateness of the Sun's gravitational equipotentials and their claim that a rapid rotation of the core implied that the observed precession of the perihelion of the orbit of Mercury was not in accord with General Relativity. Perhaps the most commonly held view at the time was that macroscopic fluid instabilities led to some kind of turbulent mixing which tended to oppose rotational shear, but there was no consensus about the distribution of angular velocity that resulted. Only if the radiative zone were pervaded by a large-scale magnetic field, which to me seems to be inevitable (Gough, 1990b), can a tight coupling between the core and the convection zone be assured (Gough and McIntyre, 1998).

Gravity waves are generated at the base of the convection zone by the turbulent velocity and pressure fluctuations. One can make a rough estimate of the generation rate from the Eulerian pressure fluctuations $\rho'$ induced beneath the convection zone which deflect downward moving convective flow, inhibiting penetration into the stably stratified layers beneath. Roughly speaking, $|\rho'| \approx \rho \omega^2$, where $\rho$ is
the fluid density and $\omega$ is a typical vertical velocity of a convective eddy at the base of the zone. This pressure fluctuation then propagates into the radiative interior as a gravity wave; I shall ignore other forms of wave transport in this simple discussion, although there must be others present — inertial waves, Rossby waves, Alfvén waves, acoustic waves — but I presume them, perhaps incorrectly, to be less effective. To add to the simplicity I shall replace the broad spectrum of gravity waves that are generated by a single representative wave of frequency $\omega$ (which I take to be positive) and a single horizontal wavenumber $k$.

With the help of the horizontal momentum equation for the waves we can set $\beta \rho_0 v_x = |\mathbf{v}| = \rho_0 k^{-1} |\mathbf{g}|$ for the representative gravity wave with horizontal displacement amplitude $\xi$, where $\beta$ is a numerical factor of order, but probably somewhat less than unity. I need now to estimate the flux of momentum carried by the waves, and the manner in which it is transferred to the mean flow.

Before proceeding with the estimate let us make a simplification by noticing that the characteristic growth rate $\mu$ of convective eddies at the base of the convection zone, about $10^{-4} \text{yr}^{-1}$, is only about $10^{-3}$ of the typical buoyancy frequency $N$ in the tachocline, so the most strongly generated gravity waves, which must have frequencies comparable with $\mu$ (indeed, I shall take $\omega = \nu \mu$, where $\nu$ is another numerical factor of order unity, which should not be confused with the cyclic frequency $\omega/2\pi$ introduced in section 4), have vertical wavelengths only $10^{-3} \text{yr}$ of the horizontal wavelengths, the latter presumably being comparable with the horizontal scale $\lambda_0$ of the eddies. Thus I set $k = \pi/\lambda_0$, where $\lambda$ is another numerical factor of order unity. For such waves the magnitude of the vertical component $k_v$ of the wavenumber is much greater than the horizontal component $k_h$ (since $k_v \approx 2\omega^{-1} N k$) and the waves dissipate rapidly within a vertical distance $d = v_r r_v \approx \omega^2 k^2 N^2 \nu$, where $v_r$ is the vertical component of the group velocity and $r_v = (\kappa k_v^2)^{-1}$ is the characteristic dissipation time of the waves, $\kappa$ being the thermal diffusivity of the medium through which the waves propagate. The group velocity $(u_g, v_g)$ is approximately $\omega (k^{-1}, -k_v^{-1})$, and the vertical component of the flux of horizontal momentum (measured in the positive $x$ direction) carried by the waves is $F = \omega^{-1} k F_g$, where $F_g = 3\kappa k_v^2 \nu$ is the energy flux. Thus, using the estimate of $\xi$ from the convection-coupling argument, the positively directed momentum flux carried by a representative wave is $F \approx \frac{1}{2} \pi^{-1} \beta \omega N^{-1} k^2 \rho_0^2$, the additional factor $\pi^{-1}$ arising from the assumption that the waves are generated isotropically in the horizontal. Note that because $\lambda_0$ is presumed to be of the order of a scale height, and therefore substantially less than the radius $r_0 \approx 0.7 R$ of the base of the convection zone, and the vertical wave scale is even smaller, it is quite unnecessary to take the details of spherical geometry into account. Therefore it was adequate to use the planar approximation in the estimates of the properties of the waves.

The phenomenon to which I wish to draw attention here is the difference between the dissipation rates of waves propagating into or against a shear. This situation is illustrated in Figure 6: Waves propagate downwards, in the $x-z$ plane, say, in a medium moving with horizontal velocity $U(z)$, where $z$ is depth beneath the convection zone. Since the background state in the radiative zone is independent of both time $t$ and the horizontal coordinate $x$, $\omega$ and $k$ are constants. However, the relations I have quoted between the other properties of the waves and $\omega$ and $k$ are valid in a frame of reference moving with the fluid, and therefore, in particular, the frequency must be Doppler shifted to $\omega' = \omega - kU$. Consequently $k_v \approx -kN/(\omega - kU)$ for downward propagating waves. It is now clear that the gravity waves propagating into the shear ($u_g \approx \omega/k > 0$) have a vertical wavenumber whose magnitude increases with depth — and therefore they are refracted towards the horizontal, along raypath A in Figure 6 — whereas waves propagating against the shear experience a decreasing $|k_v|$ and are refracted towards the vertical (raypath B). Moreover, the dissipation rate of the positively directed waves exceeds that of the negatively directed waves, and, under the assumption that the two classes of waves have been generated with approximately the same amplitude, net positively directed momentum is transferred to the mean flow, and the shear is thereby enhanced. The ray path of the forward waves approaches the horizontal as $U \to k^{-1} \omega$ (provided $U$ indeed attains such a value); this occurs at the critical level, $z = z_c$. As $z \to z_c$, $|k_v| \to \infty$ and the dissipation rate diverges. Thus all the momentum of the waves is absorbed by the mean flow before the critical level is reached, although much of the absorption takes place very close to the critical level if the waves are otherwise damped slowly. There is no critical level for waves propagating against the shear.

I come now to an interesting consequence of the dif-
differential dissipation: namely that it can generate a mean flow $U$ which oscillates in time. This phenomenon was first demonstrated both theoretically and experimentally by Plumb and McEwan (1978), and is generally believed to be the cause of the quasi-bienniel oscillation (QBO) of the Earth’s atmosphere. A simplified representation of Plumb and McEwan’s apparatus is depicted in Figure 7. An annular tank of water, 50cm high and about 10cm thick, stably stratified with dissolved salt and illustrated from the side in the figure, has a rubber bottom supported by pistons which are caused to oscillate and thereby excite gravity waves in the fluid. The whole system starts from rest. What happens? As a result of spontaneous symmetry breaking, there arises a random net excess of angular momentum, either positive or negative, in the waves, some of which is imparted to the mean flow, generating a shear. Once the initial shear is produced, it is enhanced by the differential dissipation of the waves, causing after an hour or two a more rapid flow to, say, the right, as is represented by the lower single horizontal solid arrow in the figure. At a greater height, where substantially more of the prograde wave (angular) momentum than retrograde momentum has been dissipated, the wave motion is dominated by retrograde waves, whose dissipation produces negative mean flow (represented by the dashed arrow). As time progresses, the shear is enhanced yet further, causing the maximum in the positive flow to descend (indicated by the thick vertical arrow in the figure) until it is absorbed at the lower boundary; the negative mean flow is enhanced and descends to replace the positive flow, and a new positive mean flow is generated above, producing a situation like that in the figure but with the directions of the horizontal arrows reversed. Thus a large-scale nonlinear wave of rotation propagates downwards. The characteristic period $P$ of the mean flow at any fixed height is proportional to the characteristic time $T = \omega d / k F$ associated with accelerating a layer of fluid of height $d$; Plumb and McEwan found that over a fairly wide range of parameters, $P \approx 3T$, and that the vertical scale between positive and negative rotation is approximately $d$, although it is not immediately obvious that the factor 3, and the implicit factor unity multiplying $d$, are relevant to solar conditions. Combining the estimates yields $P \approx 10 \beta^{-2} (\omega / F)^{2} N^{-2} k^{-1} w^{-1}$, in which I have absorbed into the factor $\beta$ the uncertainty in the ratio of $P$ to $T$.

I have been fascinated for more than two decades by the possibility that such an oscillation might arise in the Sun. To be sure there is angular momentum enough in the gravity waves to be interesting: the flux in just the prograde modes is sufficient to transfer through the base of the convection zone the entire angular momentum of the radiative interior of the Sun in roughly $10^{6}$ yr. So with differential dissipation one might expect that perhaps there is the (perhaps remote) possibility of an oscillating shear in the tachocline (whose moment of inertia is 2 per cent of that of the radiative interior), on a solar-cycle timescale. The crucial issue is whether the angular momentum transported by the waves is transferred to the mean rotation in an appropriate location. I once looked into this issue (Gough, 1978), and concluded that formally $d$ is likely to be only about 1km; most of the waves would dissipate before they escaped from the turbulent interface between the convection zone and the radiative interior, whose thickness, although too small yet to be resolved by seismology, must surely be much greater than $d$. In making that estimate I used one of the standard prescriptions of mixing-length theory as proposed by Böhm-Vitense (1958), which implies that $w \approx (\alpha F_{c} / 10 \rho)^{1/2}$, where $F_{c} \approx L_{0} / 4 \pi^{2}$ is the convective heat flux near the base of the convection zone, and $\mu = 2w/aH$, in which I have taken a mixing length of $\alpha$ times the pressure scale height $H$; the solar calibration obtained basically from requiring that the radius of a standard solar model agrees with observation yields $\alpha \approx 2$. I also assumed that $\omega = \mu \approx 1 \times 10^{-4} s^{-1}$ (i.e. $\nu = 1$), that $N \approx 10^{-8} s^{-1}$ and that $k = 2 \pi / H$, which implies that $\lambda \approx 1/4$ if one adopts the standard assumption $i_{0} = aH$. Fress (1981) subsequently adopted the same assumptions. It might be noted that if one adopts what I would now regard as the more natural assumption: $\lambda = 1$, then $d$ is increased by a factor $d^{2} \approx 1/5$, although still the dissipation occurs in a very thin layer — within about 1 per cent of the thickness $\Delta$ of the tachocline.
The corresponding QBO period $P$ is about a month. It might be noted that Press admitted that there are uncertainties of order $\pi/2$ in factors such as $\lambda$ and $\nu$ in the estimates, but he regarded it as being unlikely that they are as great as $2\pi$. I tend to agree. I had earlier argued that all the factors in both the wave-generation formulae and the relations between the variables in the convection formalism could be in error by about a factor $2$ (Gough, 1978), which leads to an uncertainty in the acoustic emissivity by the turbulence of at least a factor $100$ after the constraint of the solar calibration of the heat flux has been imposed. The uncertainty in the factors in the gravity-wave estimates must be similar. However, the factors are no longer obviously constrained by the solar radius calibration, which concerns only the upper superadiabatic boundary layer of the convection zone; that renders the uncertainty yet greater.

It should be noted that the estimates for both $d$ and $P$ contain products of quite high powers of the scaling factors; even ignoring the uncertainty in the relation between $w$ and $F_0$ and in the assumed shapes of the convective eddies upon which the mixing-length formalism depends, we have $d = \left(\frac{1}{2}\alpha\right)^{1/3} \lambda_0^2 \nu_0$ and $P = \left(\frac{1}{2}\alpha\right)^{2/3} \lambda_0^2 \nu_0^2 \beta^{-2} P_0$, where $d_0$ and $P_0$ are what one might call the 'standard' values (i.e. the values of $d$ and $P$ when $\alpha = 2$ and $\beta, \lambda, \nu$ are all unity; here, noting that near the base of the convection zone $H \simeq 5 \times 10^8$ cm, $\kappa \simeq 1.5 \times 10^8$ cm$^2$s$^{-1}$ and $\rho_0 \simeq 0.2 \rho$ cm$^{-3}$, I take $d_0 \simeq 1.6 \times 10^{-4}$R$_\odot$ and $P_0 \simeq 0.4$ months). So there is substantial room for manoeuvre. Indeed, if one accepts possible doubling or halving of all the factors independently, $d$ could be multiplied or divided by as much as $2^{10} \simeq 10^3$, and $P$ by $2^{20} \simeq 10^6$, which are values that might appear to encompass the possibility of having the tachocline, in which the essential dynamics of the solar cycle is commonly believed to play a role, oscillate with a period of 22y. Actually it is not quite so easy, because the joint requirements on $d$ and $P$ constrain the factors quite severely. The requirement that $d \simeq \Delta \simeq 0.02$R$_\odot$, in fact $\simeq \alpha \lambda_0^2 \nu_0^2 \simeq 125$, from which it follows that $\beta \nu \simeq 125 (P_0/P)^{1/2} \simeq 15$ if $P \simeq 22y$. If one accepts that $\beta$ is unlikely to be less than unity, then the value of $\nu$ must be at least about $5\pi$, some of which might be accounted for. It is even greater than Press's 2y limit. However, one must not forget that there are additional uncertainties in the convection theory (e.g. Gough, 1978) which I am not discussing explicitly here and upon which $d$ and $P$ also depend, which might ease the constraint on $\beta \nu$. One might note that if $\beta = 1$ and $\nu = 5\pi$, then $\lambda \simeq 0.13$ if $\alpha = 2$, which implies that the horizontal scale of the representative gravity wave is substantially smaller than the horizontal dimension of a typical convective eddy.

It was the discomfort with a value of $\nu$ as great as $5\pi$, coupled with a value of $\lambda$ as small as $2/5\pi$, which has inhibited serious consideration of the QBO mechanism as a possible mechanism to control the solar cycle. Of course, it is not unlikely — most would say it is extremely likely (and others entertain no other possibility) — that the solar cycle is driven by a turbulent dynamo, situated probably in the tachocline. But a devil's advocate like me might wonder otherwise. The possibility of a torsional azimuthal Alfvén oscillation, suggested originally by Walen (1946), has never been convincingly ruled out; if such oscillations are to be sustained they must survive the danger of rapid dissipation associated with phase mixing, and although there is a chance that they might (e.g. McIntyre, 1994; Gough, 1995b) it has not actually been demonstrated to be possible. McIntyre (1994), who has provided the most comprehensive review of the QBO mechanism in the astrophysical literature, suggested that such an Alfvén mode might be driven by the QBO mechanism operating on the gravity waves generated by convection — and Gough (1991) by the QBO mechanism operating on gravity modes generated by $^3$He burning in the core, although the latter now appears to be unlikely because such waves are likely to be heavily damped by non-linear coupling to pairs of modes of very high degree (Jordison and Gough, 2000; Jordison, 2002; see also Dziembowski, 1983) — although recently McIntyre (2002) has rejected the idea on the ground that the coherence of the oscillation would be destroyed by the magnetorotational instability of Balbus and Hawley (1991). If confined to the tachocline, for example, the magnetorotational instability would occur provided that the intensity $B_0$ of the component of the magnetic field parallel to the axis of rotation does not exceed 1T and to the idea on the ground that there is no helioseismic evidence for such a large-scale 22-year modulation of the rotation of the radiative interior.

The reader might be wondering why it is that I have devoted so much space to rotational dynamics in a discussion of the hydrostatic structure of the Sun. The reason is firstly that it is extremely likely that the source of magnetic activity, which has some influence on the structure, lies deep in the Sun, possibly in the tachocline, and secondly, as I shall discuss in the next section, the secondary flows associated with the dynamical processes that transport angular momentum beneath the convection zone redistribute the chemical composition, and thereby induce structural changes predominantly through changes in the equation of state.

7. ON THE UNIFORM ROTATION OF THE RADIATIVE INTERIOR

That most of the radiative interior of the Sun appears, according to helioseismological evidence, to rotate uniformly, is a startling property which demands explanation. The issue was grasped by Spiegel and Zahn (1992), who suggested that a high turbulent viscosity acting in horizontal directions produced by layerwise two-dimensional turbulence
in the tachocline holds the base of the tachocline rigid against the convection-zone shear. Thus, they argued, provided any angular-velocity gradient imparted by the solar-wind braking can somehow be removed, the resulting steady state would be one of uniform rotation beneath the tachocline. Even leaving aside the issue of spin-down, it is difficult, if not impossible, to see how such a mechanism could work. The reason is that layerwise two-dimensional turbulence tends not to act in such a manner as to mimic viscosity. The tendency of two-dimensional flow to conserve potential vorticity has the effect of making potential vorticity, not angular momentum, more uniform in well mixed regions. There is a wealth of meteorological evidence to support this tenet, and also to suggest that the mixing process is unlikely to homogenize an entire spherical surface; instead, finite regions of well mixed fluid develop, separated by regions of more quiescent fluid. The transport of angular-momentum is not by small-scale turbulence alone, but is accompanied by waves, and by an induced large-scale circulation. The transport process is not wholly local, as it is by viscosity. McIntyre (1994), and Fritts, Vadas and Andreassen (1998), have discussed the processes that operate. The processes tend to move the mean flow away from, rather than towards, uniform rotation.

In two recent papers, Kumar and Quataert (1997) and Zahn, Talon and Matias (1997) simultaneously claimed that the angular-momentum transported by gravity waves generated near the base of the convection zone holds the radiative interior rigid. Although the claim itself is quite obviously wrong, as was immediately pointed out at the time (Gough, 1997; Ringot, 1998) and as is evident from my discussion here in the previous section, the papers had the merit of raising the possibility that a significant amount of angular momentum might be carried by the waves to interestingly great depths. For this to be so, it would be necessary for the representative wave discussed in the previous section to have a frequency substantially greater than the growth rate of the convective eddies or a horizontal wavelength much smaller than the scale of the convective eddies. Indeed, it would require \( \lambda^2 r^4 \approx 10^3 \). A crucial issue, therefore, is whether or not the bulk of the waves are generated apparently so far off resonance. Of course, in reality a spectrum of waves is generated, and both Kumar and Quataert (1997) and Zahn, Talon and Matias (1997) took cognizance of that. However, from the gross energetics of the wave generation, it seems to me to be unlikely that there is sufficient power in the high-frequency tail of the wave spectrum to transport a substantial amount of angular momentum well into the radiative interior. I hasten to add that Zahn, Talon and Matias do not regard the waves that are required for such transport to be in the high-frequency tail; they postulate for the characteristic frequency and wavenumber of resonating waves: \( \omega = 2\pi \nu \omega / H \) and \( k = 2\pi / H \), where effectively \( \omega \approx \alpha^{-1/3} \nu \), which leads to \( \omega \approx \nu \alpha^{2/3} \mu \), corresponding to \( \nu = 1.6 \pi \) when \( \alpha \approx 2 \), and \( \lambda = 1/4 \). However, with these values, the value of \( \lambda^2 r^4 \) is only 10. Such waves have a period of about 10-15 days. The resonating frequencies of Kumar and Quataert (1997), however, appear to correspond to \( \nu \approx 1 \), yet Kumar and Quataert appear to have found similar momentum fluxes to Zahn, Talon and Matias. This highlights the degree of uncertainty in the estimates, which translates to even greater uncertainty in the estimated gravity wave fluxes of energy and momentum, as is the case also of the acoustic wave fluxes (Gough, 1978).

In view of the tendency of all fluid-dynamical processes to enhance shear, McIntyre and I (1998) concluded that the only way to account for the uniform rotation of the radiative interior is for it to be held rigid by a large-scale magnetic field. It had long been known that uniform rotation could be maintained if even quite a weak poloidal magnetic field were present (e.g. Mestel, 1953); what McIntyre and I concluded is that the field must be present. In a simple model introduced to expound the relevant dynamics, the interior is separated from the convection zone by the thin tachocline in which baroclinicity drives a meridional flow. That flow downwells in extensive regions surrounding the poles and in a broad equatorial belt, and upwells in the vicinity of the latitudes of no tachocline shear, which corresponds roughly to the latitudes of the emergent sunspots at the start of a new cycle. The downwelling fluid confines the internal field, and the essentially horizontal field at the interface with the tachocline deflects the downwelling tachocline flow to the horizontal, the base of the tachocline being determined by a balance between downward advection and outward diffusion. Except in the upwelling regions at mid latitudes, the tachocline is therefore field free, aside for any field that has been transported downwards from within the convection zone. This simple model ignores the influence of gravity waves, and was taken as the first instance to be steady. An important structural consequence of the model is brought about by a homogenization by the meridional flow of the tachocline material with that of the convection zone, which produces an observable anomaly in the sound speed relative to standard solar models, in which helium settling has left a composition gradient. The sound speed in the tachocline is too thin to be resolved directly by helioseismology. But by calibrating a model of the sound-speed anomaly seismologically, it has been possible to estimate the mean thickness of the tachocline to be about 0.02R⊙ (Elliott and Gough, 1998).

It should be appreciated that the thickness of the tachocline cannot be uniform. At locations where the magnetic field is nearly vertical, and the horizontal component of the field is therefore small, the tachocline is much thicker. This can arise near the axis of a dipole field, for example. There the downwelling is deep, and may even be deep enough to burn a significant amount of lithium and maybe some beryllium. The axis of the field need not be aligned with the axis of rotation; indeed it may even lie near the equatorial plane, rendering the structure of the Sun nonaxisymmetric (e.g. Mestel, 1999). Such a field imbedded in the uniformly rotating radiative
interior might then be responsible for the persistent longitudes of activity observed in the solar atmosphere.

Gravity waves can modify the dynamics of the tachocline via the transport of momentum and, probably to a lesser extent, energy. Fritts, Vidas and Andraeassen (1998) have described how wave transport can contribute not only to a modification of the angular velocity but also to the driving of the meridional flow. The characteristic ventilation time of the wave-driven flow was estimated to be between $2 \times 10^4$ and $2 \times 10^5$ years, substantially shorter than the $10^5$-year timescale of a flow driven solely by the steady baroclinicity discussed by McIntyre and myself (1998). The flow they discuss is unlikely to be correct in detail, however, for the overall penetration of the tachocline sound-speed anomaly inferred by helioseismology. Nevertheless, the physical processes are no doubt operating.

It is interesting to consider further whether the QBO mechanism can operate in the tachocline. Kumar, Talon and Zahn (1991) have recently addressed this question, correcting their earlier discussions and extending their scope to include the influence of rotation and a large-scale magnetic field on the propagation of the waves. They used the formalism of Goldreich, Murray and Kumar (1994) to estimate the power in the gravity waves, as had Kumar and Quataert (1997) before them, although now they apparently expected the resonant waves to have periods of about 10 days (and consequently $\nu \approx 2\pi$). Yet the estimated total energy flux in the waves agrees numerically with the estimate for the single representative wave given in the previous section having a frequency corresponding to $\nu = 1$. Presumably the horizontal wavenumber of the representative wave has been adjusted to compensate. Kumar, Talon and Zahn did not comment on this apparent adjustment. With the corrected momentum flux they succeeded in generating a strong QBO-like shear layer beneath the convection zone with a characteristic vertical scale of order $10^{-1} R_\odot$, although they failed to produce a temporal oscillation. However, they did argue that in reality the shear layer might be destroyed by merging with the convection zone, and thereby permit a cyclic behaviour, analogous to that found in Plumb and McEwan's (1978) experiment. The timescale for generating the shear layer was 20-30 years which, they asserted, should be considered to be of the same order as the solar cycle period. Surely the uncertainty in the theory should be considered to be far too great to justify being so confident. In the light of the substantial variation amongst the wave-frequency estimates, the associated estimates of the characteristic wave number, not to mention the magnitude of the wave momentum flux, could one not even contemplate the QBO mechanism to be responsible for the 1.3-year oscillation in the rotational shear in the vicinity of the tachocline that was discovered recently from GONG and MDI data by Howe et al. (2000)? In this regard it is interesting to note that the strongest evidence for a coherent oscillation was found in the equatorial regions, where $\Omega$ increases away from the axis of rotation and therefore cannot drive a magnetorotational instability that might destroy the coherence. However, we should not forget that Antia and Basu (2000), who reanalysed MDI data, failed to find the 1.3-year oscillation, and claim that it is not present.

8. STRUCTURAL VARIATION DEEP IN THE CONVECTION ZONE

Is there any? We cannot be sure. But it is perhaps not out of place to be reminded here that evidence is accumulating for some kind of interesting phenomenon occurring a few per cent of the solar radius beneath the photosphere. Over the years there have been many allusions to a local maximum of angular velocity at a radius $r \approx 0.95 R_\odot$ (e.g. Schou et al., 1998). It is discernible even in the original inversions (Duvall et al., 1984), and there is evidence for it in earlier rotational-splitting data (Deubner, Ulrich and Rhodes, 1979). Although there is yet no cogent simple explanation, it has been postulated that it is associated with magnetic activity. Indeed, Komm, Harvey and Howard (1993) have suggested that surface magnetic features are anchored by the field to fluid beneath and rotate with it, and they have thus attempted to map the subsurface rotation rate by assigning successively greater anchoring depths to supergranulation, small magnetic features and the large-scale field pattern, comparing the outcome with the helioseismically determined rotation, and suggesting a connexion with the solar cycle. Indeed, one might even wonder whether the solar cycle is driven in a layer near the angular-velocity maximum, perhaps by a QBO-like mechanism, although not, of course, with gravity waves.

In a recent analysis of the even component of degeneracy splitting of seismic mode frequencies, Antia, Chitre and Thompson (2000) found a structural anomaly in the convection zone which they interpreted as being caused by a magnetized layer with field intensity of about 20 kG located around $r = 0.96 R_\odot$. Although one cannot determine whether any anomaly found from seismic frequency analysis alone is due to a magnetic field (Zweibel and Gough, 1995), the hypothesis is certainly plausible, and links interestingly with Komm, Harvey and Howard's conjecture. Moreover, it is interesting that the location of the anomaly is not far removed from the depth (0.03R_\odot) at which there would not be a time-dependent anomaly to account for a 700 microHz oscillatory component in the solar-cycle changes to the (m-averaged) multiplet frequencies plotted in Figure 35. It is also, perhaps fortuitously, more-or-less consistent with $W$ having a value of about 0.04, which is what motivated the manner in which I have attempted to address the question posed by the title of this contribution.

Thus there does appear to be meagre but mounting evidence that there is some dynamical process associ-
ated with the solar cycle that causes profound structural changes. It is quite evident, therefore, that it is important to anyone profoundly interested in the solar cycle that further investigation into the matter should be undertaken.

9. CONCLUDING REMARKS

There is evidently considerable uncertainty in the mechanisms that control the solar cycle. Although there is almost a consensus amongst solar physicists that the Sun harbours an oscillatory dynamo, and it seems to be most commonly believed at the moment that that dynamo is situated, at least in part, in the tachocline, the manner in which that putative dynamo operates is not agreed upon. I have not discussed that matter in this review. Instead, I have brought attention to the fact that the mere existence of a near-consensus amongst scientists does not necessarily imply that that consensus is the truth. Indeed, there remains the possibility that the surface magnetic activity is a vestige of a primordial field permeating the radiative interior, which is modulated cyclically by fluid motion. It is probably true to say that most of the few solar physicists who are prepared to contemplate that possibility would maintain that the cyclic behaviour is at least a result of magnetohydrodynamical processes operating deep in the convection zone or in the tachocline. Those processes are commonly called dynamo action (by dynamo theorists). That they are genuinely dynamical is never questioned; and that they are profoundly situated is hardly in doubt. Indeed, I think that actually the chances are that the view that the Sun really is a dynamo, a view which has emerged from the cumulative wisdom of those who work in the subject, is essentially correct. But until that view has been demonstrated, it is healthy occasionally to contemplate alternatives. It would be interesting indeed if the cycle were actually the product of a purely fluid dynamical QBO mechanism in which the magnetic field is a mere slave to the cyclic process, and which plays no essential dynamical role.

Associated with the cycle dynamics must be some structural response. The principal issue with which my discussion here has been concerned is whether any part of that response is profound. There can be no doubt that there is a cyclic variation of the superficial layers of the Sun. We observe that directly. And indeed, the evidence may be 'already overwhelming', as Peter Foulal has claimed, 'that measured luminosity variations arise chiefly from modulation of photospheric heat flow by plainly visible dark spots and bright faculae'. But I cannot agree with Peter's view that to contemplate a deeper root of these processes is 'unnecessary'. Our objective is not merely to account for the gross behaviour of the variation in a manner that is not obviously contradicted by observation: it is to understand how the Sun really works. And for that one must look for clues beyond the chief contributor. The hint I re-

called in the previous section of a tiny contribution to the solar-cycle seismic frequency variations being an oscillatory function of frequency is one such example. The solar diameter variation is another. Both observations are marginal at best. But if either is correct, it points to the existence of processes affecting the Sun’s structure that operate more deeply than those that we plainly see. To be sure it is not productive to make unrealistic claims for such observations. But it is not without utility to pursue their consequences, even though they may in the past have contributed little to our relatively mature understanding of solar luminosity variation, and (may) have even less prospect of competing against more promising diagnostics in the future.'

I am very grateful to D. Bayston for preparing the \LaTeX file, and to G. Houdek and R.W. Sword for help with the preparation of the table and diagrams.

REFERENCES

Antia, H. M., Chitre, S. M. and Thompson, M., 2000, Astron. and Astrophys., 360, 335-344


Kumar, P and Quataert, E. J. 1997 Astrophys. J. Lett., 475, L143-


Quinn, T. J. and Fröhlich, C., 1999, *Nature*, 401, 841-842
Twigg, L. W. and Endal, A. S., 1981 *Workshop on variations of the solar constant*, NASA CP-2191