THE CIRCADIAN SYSTEM AND THE THERAPEUTICS OF THE AFFECTIVE DISORDERS

D. HEALY* and J. M. WATERHOUSE†

*Academic Sub-Department of Psychological Medicine, Hergest Unit, Ysbyty Gwyneud, Gwyneud LL57 2PW, U.K.
†Department of Physiological Sciences, Manchester University, Manchester M13 9PT, U.K.

Abstract—Establishing that a circadian rhythm is abnormal tells us little about the cause, which can arise from changes in the patient’s lifestyle, irregularities of the body clock or a malfunction in the process of entrainment of the clock. In a clinical context, such a range of possible explanations implies differences in the most appropriate mode of treatment. Against this background, the conventional view that the underlying abnormality in endogenous depression is due to a disorder of the body clock is challenged. The challenge is based on difficulties of interpretation of the clinical data and the results of studies on circadian rhythms in patients. It is suggested that the state of the circadian system in depression resembles its state in healthy individuals after time-zone transitions or in shift work maladaptation syndrome and that this disturbance should be seen as resulting from changes in the phasing of external zeitgebers rather than from an abnormality in the clock itself.

Keywords—Depression, social zeitgebers, shift work maladaptation syndrome.

CONTENTS

1. Abnormal Circadian Rhythms
   1.1. Estimating the endogenous component in field conditions
   1.2. Melatonin
2. How Regular are Circadian Rhythms in Healthy Subjects?
3. Abnormally-phased Rhythms
   3.1. Time-zone transitions
   3.2. Long-haul aircrew
   3.3. Night workers
4. Abnormal Clock or Environment?
5. The Circadian System and the Affective Disorders
   5.1. Phase advance hypothesis
   5.2. The endogenomorphic features of depression
   5.3. Seasonality and periodicity
6. Therapeutics of the Affective Disorders
   6.1. What are antidepressants?
   6.2. Sleep and light therapies
   6.3. Cognitive and behavioural therapies
   6.4. Chronopharmacokinetics
7. Mania and Its Therapy
   7.1. Neuroleptics and mania
   7.2. Lithium and mania
8. A Shift Work Model of Depression
   8.1. A research programme
   8.2. Implications for the therapeutics of the affective disorders
References

Abbreviations—DLMP, dim light melatonin onset; ECT, electroconvulsive therapy; 5-HT, 5-hydroxytryptamine; SAD, seasonal affective disorder
1. ABNORMAL CIRCADIAN RHYTHMS

There have been a number of studies of affective disorders that have found abnormalities in the profiles of circadian rhythms (Souetre et al., 1991; Tsujimoto et al., 1990). As outlined in a previous article (Redfern et al., 1991), these abnormalities can be described mathematically as altered parameters of a cosine curve — the period, mean value (mesor), amplitude and the time of peak (acrophase). Whilst this approach has the obvious merit that the abnormality has been quantified, it gives no information on the closely related issue of how the abnormality arises. Such knowledge would be valuable, however, since it would enable patient treatment to be rationalised.

Since rhythms have endogenous and exogenous components, it is important to establish which of these is responsible for the abnormality. The standard way to measure the endogenous component of a circadian rhythm is by placing the subject on a constant routine (Minors and Waterhouse, 1984). Abnormalities in the rhythm measured under these circumstances would strongly suggest that the internal timing system — the endogenous component — was responsible.

1.1. Estimating the Endogenous Component in Field Conditions

Unfortunately, the demands of the constant routine protocol — no sleep, constant low levels of activity and equi-spaced snacks — are rather severe and so can be unsuitable for patients, for investigators in field conditions and for studies of, say, depression or mental performance, in which the sleep loss and/or abnormal lifestyle might affect the depression itself.

Alternative methods, which attempt to estimate the phase of the endogenous component from data obtained under conditions when masking factors are present, are being developed (Minors and Waterhouse, 1989, 1992, 1993). They tend to make two assumptions. The first is that the masking effect, due to activity (in the case of body temperature) or time spent awake (alertness or mental performance tests) is independent of the phase of the endogenous component. For example, the fall in temperature produced by a sleep or the rise produced by activity will be the same whether they take place coincidentally with the trough or peak of the endogenous component. In humans, the evidence is that this assumption is approximately correct. The second assumption is that the shape of the endogenous component remains constant (and very close to that of a cosine curve). This has not been tested, but there is no reason to believe that the assumption is false under most circumstances. Methods exist that do not make the first assumption (see Spencer, 1989, for example), but they require more frequent data collection than can normally be achieved in the field.

One of the simplest methods using the above concepts is based upon the ‘purification’ principle of Wever (1985) and can be used with masked temperature data collected in the field (Minors and Waterhouse, 1993). It requires a log to be kept of the subject’s sleep times and types and times of activity (sitting, walking, etc.). The method then purifies the masked rhythm by calculating the most appropriate amount by which the temperature rhythm would have to be decreased (purified) as a result of each type of activity and how much it would have to be increased as a result of sleep. ‘Most appropriate’ would be those changes that produced a ‘purified’ rhythm that approximated most closely to the shape of ‘normative endogenous data’. These are the average endogenous rhythm from a large number of individuals who have been studied in constant routine conditions. Any shift of the normative endogenous data that was needed to produce the closest approximation to the purified rhythm would then be a measure of the shift of the internal timing system, the circadian oscillator.

This method has been used with nurses doing night work (Minors and Waterhouse, 1993). It indicated that the shift of the body clock was less than that assessed conventionally from masked data. (Since nurses on night duty tend to delay their lifestyle — and hence the masking effects — the masked data would be predicted to overestimate the rate of adjustment of the body clock to night work.)

It might be possible to use a variant of such a purification model with individuals with affective disorders, provided that an accurate activity log could be kept. (An alternative to this would be the automatic monitoring of movement, say by a device strapped onto the non-dominant wrist.) There is some evidence that the amplitude of the endogenous component decreases in such patients (Rosenthal et al., 1990), but it should be possible to incorporate the possibility of a decrease in amplitude of this component into the model.
It has not been investigated if the method could be adapted to purify other rhythms, for example, those from hormones or urinary metabolites of neurotransmitters. In cases other than body temperature, the nature of the masking factors would need to be established. These factors might be activity and sleep for other variables also, but mealtimes and exposure to stressors or bright light (see melatonin, below) might be important in other cases: it is these factors, of course, that would have to be recorded in the patient's log and used in the 'purification' method.

1.2. Melatonin

There is one variable, melatonin, and its urinary metabolite 6-OH melatonin sulphate, that appears not to be masked, at least if light intensity is low (such as would occur if a subject remains indoors in a dimly lit room). Accordingly, melatonin has been used as a robust marker of the internal clock. In particular, the time at which the concentration of plasma melatonin rises above a threshold value (this being arbitrarily set) in a subject in dim lighting is taken as a phase marker of the internal clock (Lewy and Sack, 1989). (It has been abbreviated to DLMO, dim light melatonin onset.) While this appears to be a most useful measure, there are some points that deserve comment.

First, the maximum concentration of plasma melatonin, which occurs during the night, varies between individuals. Setting the same threshold for each as the phase marker might not be ideal, particularly if the peak values vary day by day in the same individual. Setting a threshold at a constant proportion of the nocturnal maximum might have some advantages.

Second, there is an upper limit to the frequency at which blood and, particularly, urine can be sampled, and this limits the resolution of the method for estimating DLMO. In practice, however, this limit is less than the phase changes that have been found in some disorders.

In summary, with the possible exception of melatonin, we conclude that in the absence of data obtained from constant routines, from some form of purification of masked data, or from some other means to separate the endogenous and exogenous components of a rhythm, there must be difficulty in being able to establish if a rhythm abnormality is caused by the internal clock, the external factors or some interaction between the two. These interpretative caveats will apply to nearly all published literature on altered circadian rhythms in affective disorders.

2. HOW REGULAR ARE CIRCADIAN RHYTHMS IN HEALTHY SUBJECTS?

A huge body of data has been collected from healthy subjects to assess how regular circadian rhythms normally are. Such data have been obtained from many variables and from people living in many parts of the world (Halberg et al., 1969). In addition, data are now available to compare rhythms in subjects of different ages; and longitudinal studies have been carried out that compare the same individual on successive days or in successive decades of life (Brock, 1991; Monk, 1989). What these results indicate is the similarity of the amplitude and, particularly, the phase of a particular rhythm, especially if the phase of the rhythm is expressed with reference to mid-sleep. This attests to the reliability of circadian rhythms, but it does not allow us to decide to what extent this is a reflection of the regularity of exogenous influences (especially when sleep is used as the reference point for expressing the phase) rather than the stability of the endogenous component (stabilised by regular exogenous influences acting as zeitgebers).

Repeated estimation of the phase of the endogenous component via repeated constant routines is not feasible, of course, but there are two pieces of information that are relevant. First, we have performed constant routines in well over 100 subjects who have been living a normally timed lifestyle for the previous 3 or more days. Estimates of phase (of the endogenous component) during these constant routines have established that the population as a whole shows rhythms of deep body temperature and urinary constituents that are phased close together (Minors and Waterhouse, 1990). Inspection of the scientific literature produces confirmation of this result for deep body temperature (Moog and Hildebrandt, 1989) and enables the variety of variables to be expanded to include melatonin (assessed by the DLMO method, above, so exogenous effects are eliminated) (Lewy et al., 1987) and several hormones (assessed in circumstances of controlled meals and bed rest when exogenous influences are minimised) (Van Cauter et al., 1991).
The second approach is to consider the phase of a rhythm that has been purified by the method described above. The example of deep body temperature in two subjects living a normal lifestyle is shown in Fig. 1. This shows that daily changes in mid-sleep are quite small, that is, the routine appears to be quite stable. The endogenous component of the rhythm, assessed by the purification method (Minors and Waterhouse, 1992, 1993), shows a daily variation within a range of 2 hr either side of the mean value. Assessment of the phase of the masked rhythm showed intermediate shifts, as is to be expected, since it results from both of the other components.

The results show that in both subjects, there is clear evidence that the rhythm has been entrained, though perhaps not as accurately as is sometimes assumed. The usefulness to a healthy individual of a stable and entrained circadian rhythm has been described by Redfern et al. (1991). Here we stress that in normal circumstances, there is a daily reproducibility in the phasing of environmental zeitgebers, the body and circadian rhythms.

3. ABNORMALLY-PHASED RHYTHMS

Several laboratory-based experiments have shown that the body clock is slow to adjust to a change in timing of zeitgebers. This inertia is important for it means that neither a daytime nap nor a night-time foray to the larder will change the clock. Under normal circumstances, this inertia to change can be seen as being of adaptive worth; but the lifestyle of humans has meant that this slow adjustment can be disadvantageous when it leads to an individual’s lifestyle and body clock being out of phase with each other. What are the causes of these abnormalities and what effects do they have upon the individual concerned?

3.1. Time-zone Transitions

After long-distance flights to the east or west, travellers experience ‘jet lag’ (Graeber, 1982). This is an assortment of physiological and psychological abnormalities, amongst which are fatigue and yet inability to sleep during the local night, loss of appetite and indigestion, decreased mental and physical performance during the local daytime, loss of concentration, generally feeling ‘below par’ and even mildly depressed. Such symptoms worsen with the number of time zones that have been crossed, and if the journey is to the east rather than the west. Investigators have shown that the symptoms are present when the circadian rhythms have not adjusted to the new time zone. Such

---

Fig. 1. Two subjects who kept a daily record of rectal temperature and activity: •–•, midsleep; o–o, phase of temperature rhythm estimated by the ‘purification’ method. For more details, see text.
findings can readily account for the abnormal sleep pattern and consequent fatigue, and the combination of abnormally phased rhythms and fatigue can account for poor performance and loss of concentration. Even so, a strict correlation between the severity of symptoms and the mismatching between the environment and body clock problem is unlikely to exist.

The explanation for symptoms of jet lag being worse after an eastward flight is probably that adjustment requires the body clock to advance, and its free-running period (in the absence of zeitgebers) is greater than 24 hr, about 25 hr. Indeed, after a simulated flight to the antipodes, adjustment almost invariably takes place by a delay, rather than an advance, of the internal clock (Mills et al., 1978).

Attempts to ameliorate these problems have centred on combatting the symptoms of jet lag or attempting to promote adjustment of the body clock (Arendt et al., 1987; Redfern, 1989, 1992). Inability to sleep can be countered by the use of a short-acting hypnotic of the benzodiazepine group, and melatonin capsules taken in the evening of new local time decrease subsequent daytime fatigue. Whether such substances act on the body clock itself is not clearly established, although it is certainly a possibility.

In hamsters, there is good evidence to indicate that activity can adjust the clock, the amount and direction of which are determined by the phase of the clock when activity is taken (Mrosovsky et al., 1989). There are also claims that alterations to an individual's diet, via a sequence of changes involving plasma amino acid concentrations, their uptake into the brain and incorporation into neurotransmitters, can promote adjustment, but the scientific evidence so far is poor (Leathwood, 1989).

By contrast, recent work has shown that bright artificial light — equal to that found outdoors about 1 hour after sunrise — is an effective means to shift the body clock (Czeisler et al., 1989; Minors et al., 1991). Briefly, bright light at a time soon after the minimum of the body temperature rhythm can advance the clock and just before the minimum can delay it. In practice, a combination of appropriately timed exposure to light and adjustment of meals, sleep and activity times is a practical and effective set of measures.

The duration of the symptoms of jet lag is a few days only and so it is unlikely that individuals could suffer from anything more than the most mild and temporary form of anxiety or depression. However, in two groups of people, the potential exists for any such symptoms to develop further. These are long-haul air crew and night workers.

### 3.2. Long-haul Aircrew

Long-haul aircrew continually undergo time-zone transitions, in addition to having to work irregular duty hours. Potentially, chronic fatigue and a decrement of mental and physical performance might arise and a chronic version of jet lag might exist, which could lead to anxiety and depressive states in some. The few studies that have been carried out have concentrated on the quality and quantity of sleep (Preston, 1973; Nicholson, 1987). It is becoming evident that senior aircrew are particularly likely to lose sleep and have difficulty with adjustment, but that all are concerned to lose as little sleep as possible. To achieve this aim, many make use of more than one sleep period per day and, whenever duty schedules permit, arrange for some sleep to be taken during the night on home time (Graeber et al., 1986).

This 'anchor sleep', in combination with the associated regularity of meals, light exposure and activity, is likely to act as a zeitgeber and so, stabilise circadian rhythms with a phasing that is appropriate when the aircrew have time off at home (Minors and Waterhouse, 1983). It also means that a nap at night (body time) will be easy to initiate and maintain. Information on performance decrement is not available. There is no evidence that aircrew are prone to anxiety or depressive states. Whereas this might be part of a 'macho image', there are other factors to take into account. First, there is now recognition by management and personnel of difficulties with sleep loss and the need to plan leisure time; second, there is legislation designed to prevent excessive work hours and provide opportunity for recuperation; and third, there is the possibility that any staff with severe problems due to jet lag can choose to leave.
3.3. Night Workers

Night workers also work and sleep at abnormal hours, and this can occur intermittently for 40 years or so. Difficulties with poorer daytime sleep and performance at night (due to the effects of working at the circadian trough, as well as sleep loss) are widely reported; chronic fatigue, chronic indigestion and an increased frequency of gastrointestinal disorders are widely reported also; and even cardiovascular disorders are more common than amongst day-working control groups (Waterhouse et al., 1992). Symptoms similar to jet lag are experienced, and they are called ‘shift work maladaptation syndrome’. Not surprisingly, surveys indicate that night work is not often liked, but that most put up with it because of the extra money and days off, for example. With increasing years spent on night work, there is a tendency for the perceived disadvantages to increase at the expense of the advantages, but there is also the finding that groups who have worked for many years at night appear to be anomalously healthy. This anomaly exists because those who remain are unrepresentatively tolerant of night work, those who experience much difficulty choosing to leave it.

Adjustment to night work is slower than to a time-zone transition involving a similar change in sleep habits because the zeitgebers acting upon a night worker are not all promoting entrainment of circadian rhythms to night work. Any advice to the night worker, therefore (Redfern, 1989; Waterhouse et al., 1992), tends to stress the use of zeitgebers timed appropriately for night work (meals and sleep, for example) and to suggest a distancing of the individual from those zeitgebers that remain wrongly timed (many social influences and natural lighting, for example).

4. ABNORMAL CLOCK OR ENVIRONMENT?

For long-distance travellers and night workers, the unpleasant consequences of ignoring the body clock are common and well known, influencing as they do many physiological and psychological systems. Perhaps it is because the difficulties are widely experienced, they have some readily apparent explanation, and because they can be avoided if they become too marked, that severe anxiety and depressive states are less frequent than might have been predicted. It is important also to realise that the problem has not arisen because of a faulty body clock, but rather because of a change in the phasing of environmental time cues. On the contrary, given the abnormality of a mismatch between the body clock and an abnormally phased zeitgeber, the body clock has behaved perfectly normally. Furthermore, advice that is given to travellers and to night workers — advice that attempts to provide adjustment of the body clock, or else to stabilise it at a particular phase, by the appropriate use of zeitgeber — makes the assumption that the body clock and its means of entrainment are normal.

The position is similar in many ways in blind subjects. It has been shown on several occasions that their circadian rhythms are more irregular than those of sighted controls. Some rhythms even appear to lose entrainment to the 24-hr day (Lewy and Sack, 1989). Symptoms include fatigue and insomnia when the circadian rhythms are phased inappropriately for a normal lifestyle. If the loss of light as a zeitgeber can be substituted for by regular exercise, meals and times of retiring and rising, or by regularly taking melatonin, then improvements in sleep and alertness result (Arendt et al., 1988; Sack et al., 1991). Again, there is no evidence that the body clock is abnormal, only that it is receiving inadequate zeitgeber information and so, predictably, it is unadjusted, or only poorly so, to the solar day.

But what would be the result if there were no obvious reason for abnormal circadian rhythms; or if the environment appeared to be normal and if the individual was not obviously unable to respond to normal zeitgebers?

5. THE CIRCADIAN SYSTEM AND THE AFFECTIVE DISORDERS

In the case of the depressive disorders, there have been a number of studies that have found abnormalities in the profiles of circadian rhythms (Souetre et al., 1991; Tsujimoto et al., 1990). As outlined in Section 4, however, it is not necessarily clear from such altered rhythmic profiles what the cause of the abnormality is and hence, what their significance is. Applying the principles outlined
in Sections 1 and 4 to the affective disorders, a number of questions arise regarding depression and antidepressants.

Regarding depression, the question arises as to whether altered rhythmic profiles arise from a disturbance of a circadian clock or from a disorganisation of the circadian system; this latter might stem from either a failure of entrainment to zeitgebers, because of altered sensitivity to these zeitgebers, or a disorganisation in external zeitgebers. An attempt to answer this question will require some consideration of the nature of the affective disorders.

Regarding antidepressants, there appear to be findings of alterations in rhythmic profiles of hormonal outputs, as well as in the rhythms of receptor densities and affinities (Wirz-Justice, 1987), but there is no consensus as to whether the period of the underlying circadian rhythm is altered by antidepressant therapies and/or whether there is a consistent phase shift induced by antidepressants. Do antidepressants act on the circadian clock or do they affect the profile of circadian rhythms? A consideration of this question will involve an attempt to define what an antidepressant is. This will include a consideration of the recently introduced sleep and light therapies for depression. It will also necessitate a consideration of chronopharmacokinetetic aspects of antidepressants.

5.1. Phase Advance Hypothesis

In the early 1980s, the phase advance hypothesis of depression was articulated by Wehr, Goodwin, Wirz-Justice and colleagues (Wehr and Goodwin, 1981; Wehr and Wirz-Justice, 1982). It established a basis for thinking about the role of circadian rhythms in depressive disorders (Healy, 1987a). The hypothesis drew support from evidence of altered rhythmic profiles in depression and from evidence that antidepressant agents have effects on rhythmicity.

However, there are a number of problems with the experimental designs employed in studies of this hypothesis. Firstly, many different rhythms have been measured and compared with no explicit reference being made to the relative importance of exogenous and endogenous components to the rhythms in question. Second, in general, these studies have not considered whether the results reported could only be explained in terms of an abnormal clock—changes in lifestyle, for instance, have been poorly reported.

For example, in a recent study, Souetre and colleagues (1991) subjected five depressed individuals to a phase-shift experiment, which involved advancing by 5 hr the timing of major zeitgebers. They reported that this was of benefit clinically and that it led to an enhancement of the amplitude of a number of biological variables they were looking at. They concluded that their findings indicated that the entrainment of internal clocks by environmental information may be impaired in depression and that the imposition of a time zone shift had helped to correct this defect. However, at no point do they discriminate between faulty entrainment of the clock owing to a poverty in environmental rhythms on the one hand and a loss of entrainment owing to a clock abnormality on the other. They do not advert to the fact that the imposition of a time zone shift also imposes a structured day on individuals and that, in its own right, this may be therapeutic.

It appears to have been assumed generally that features of depression, such as its periodic and its bipolar nature, in addition to its seasonal incidence, its response to sleep deprivation and light therapies, as well as clinical features, such as diurnal variation mood and early morning wakening, all pointed to a basis for the disorder in a disturbance of circadian rhythms.

These aspects of depression have also often served to label a particular form of depression as endogenous as opposed to reactive. The endogenous nature of such symptoms appeared to indicate, in turn, that the disturbance must be endogenous rather than exogenous/reactive in origin. It was natural, as a consequence, to assume that the disorder must have its basis in some fault in the functioning of a clearly endogenous unit, such as the clock, rather than that it might arise from a disturbed interaction between an individual and their environment (Healy and Waterhouse, 1990, 1991).

Such an assumption was consistent with the prevailing assumptions regarding the depressive disorders themselves in the late 1970s and early 1980s. At that time, it was held that endogenous and reactive forms of depression were two distinct entities, with the endogenous form arising 'out of the blue', in contrast to the reactive form, which was seen as a neurotic state arising as a response to psychosocial stressors. An implication of this formulation was that the endogenous form of the
illness involved a biological disturbance of some sort that was most appropriately corrected by biological (pharmacological) treatments, in contrast to the reactive form of the disturbance, which could be managed by psychotherapeutic or hygienic manoeuvres (Healy, 1990a).

The dichotomy of these assumptions has been breaking down in recent years. A number of studies have shown that the 'endogenous' form of the illness is endogenomorphic in the profile of clinical symptoms that it demonstrates rather than endogenous in the sense of arising from a primary endogenous disturbance (Healy and Williams, 1988). It has been clearly shown that environmental disruptions (life events) precipitate the endogenomorphic form of the illness, at least as frequently as they do the 'reactive' form of depression, which is characterised by misery and unhappiness without any evidence of biological disturbance (Hirschfeld, 1981; Paykel, 1985).

Furthermore, it seems that a good proportion of 'biological' depressions (all but the most severe) respond to a variety of psychotherapeutic measures. The psychotherapies in question, however, are not interpretative psychotherapies, of the kind that were in favour during the 1960s and 1970s and that appeared to be ineffective for depressive disorders. Responses have come to novel therapies based on the cognitive model of Beck (1967) or interpersonal therapy as developed by Klerman et al. (1984). A common feature of such treatments is their recourse to programmes that involve behaviour activation, or motivated activity (Healy and Waterhouse, 1991). In this, they resemble, to some extent, the therapy or hygiene advice given to individuals likely to be exposed to jet lag or shift work maladaptation syndrome.

5.2. The Endogenomorphic Features of Depression

There is a traditional notion that individuals who are depressed feel worse in the morning and that their mood improves during the day (Healy, 1993). This appears to be thought of as a relatively stable feature of endogenomorphic depressions, and is often cited as evidence in favour of there being a disturbance of the circadian clock in depression. However, the evidence on this point is ambiguous. (As well as the longstanding association between morning dysphoria and endogenomorphic depression, there has been an association between evening dysphoria and neurotic depression.)

In a recent prospective study of this issue, however, Carpenter et al. (1986) were unable to confirm this finding. They specifically found that individuals who had an evening dysphoria had a higher incidence of decreased appetite than those with morning dysphoria. They also found that individuals with clear diurnal variations of mood, with either morning or evening dysphoria, on the one hand tended to have longer episodes of depression, but on the other, experienced a more rapid clinical response to treatment than those with no marked diurnal variation at all. Finally, Carpenter et al. (1986) found that morning dysphoria appeared to be associated particularly with retardation and an absence of agitation.

There is some debate about the stage of a depressive episode in which diurnal variation is most typically found. Beck (1967) has suggested that it occurs more frequently in severe depression than in minor stages of depression, but there is no consensus on this (Healy, 1993). A study by Waldmann (1972) indicated that diurnal variation may be lost in the most severe stages of a depressive episode, which throws up the possibility that diurnal variation may be absent initially, appear at moderate levels of intensity and disappear again at the severest levels of intensity.

In favour of the suggestion that diurnal variation of mood is likely to present in mild depressions and to be lost as the disorder becomes more severe are studies by Middelhoff (1967) and Waldmann (1972), which found that diurnal variation of mood was present in a substantial proportion of individuals during symptom-free intervals. It did not appear, therefore, only with the onset of depression, although the number of people affected with diurnal variation increased with the onset of a depressive episode.

However, all of these proposals are compromised by a number of studies that suggest that this prototypical feature of depression has no typical presentation at all. Stallone et al. (1973) looked intensively at a small group of patients and found that diurnal variation is not a stable pattern for individuals. In some cases, it may be present at one point of an illness and absent later on, for no obvious reason. When present, it is so for variable lengths of time during the day, even in individuals who believe themselves to have a typical and consistent pattern of mood variation.

In another study using a similar intensive design, Tolle and Goetze (1987) found that the typical
diurnal variation of morning dysphoria is irregular and intraindividually unstable. They questioned 63 depressed subjects, among which were 35 endogenomorphic depressives, over a total of 273 days and found an unexpected lack of pattern to the daily course of the depressive state. They concluded that daily rhythms in mood are polymorphic. Reliable regularities cannot be established either intraindividually or in subgroups of subjects. In fact, they found that a constant state of mood during the day was the most frequent finding, with typical diurnal variation being noted in only about one-third of the patients. However, within this relatively constant state, they noted that there appeared to be a number of ultradian rhythms. Their opinion was that there must be considerable doubt as to whether these ever could be investigated properly, in view of the numerous situational variables likely to influence mood during the course of the day.

As regards early morning wakening in depression, a similar ambiguity prevails. While early morning wakening is thought of as having some specificity for depressive disorder and as being in some way characteristic of it, the disturbances of sleep in individuals who are depressed are as likely to include difficulties in falling asleep, as well as broken sleep during the night, as they do early morning wakening. Furthermore, as with diurnal variation of mood, there is no evidence to suggest that individuals who wake early at a particular time one morning awaken regularly at that time. Not only that, but a proportion of individuals far from having early morning wakening actually have hypersomnia.

Thus, there is little basis in terms of the clinical features of depression to point toward a reliable disturbance of circadian clock functioning. In particular, there is no clear evidence for an alteration of period or for a phase advance or phase delay. Attempts to establish clearly the existence of clock-derived features of depression would require a careful monitoring of ongoing behaviour and social influences and an effort to discriminate between the effects of such influences and changes stemming from the clock. In contrast, the picture that emerges from the present state of the evidence is rather one of shifting clinical symptoms that are neither constant to the individual or constant across individuals. Such a disturbance is potentially consistent with a disorganisation of circadian rhythms, such as occurs in jet lag or shift work—a disorganisation that does not involve a disturbance of clock functioning, but that stems from alterations in lifestyle.

5.3. Seasonality and Periodicity

Since the first articulation of the phase advance hypothesis, a subgroup of affective disorders has been delineated, the seasonal affective disorders (SAD). Attention to this disorder was stimulated by the case of a particular individual whose disorder showed an apparent seasonal pattern and who also had a dramatic response to light therapy (Lewy et al., 1982). The syndrome was first described by Rosenthal and colleagues in the early 1980s (Rosenthal et al., 1984). SADs are illnesses that supposedly start regularly in the autumn or winter and remit in the spring or summer. From the start, therefore, there has been a close connection between SAD and phototherapy. However, there are problems with this. The initial claim for the efficacy of phototherapy was in terms of phototherapy extending the photoskeleton (the period of the day during which the clock might be exposed to light). This was done by giving bright light therapy early in the morning and also in the evening. However, at present, there is no evidence that the effects of light therapy depend on an administration that increases the photoperiod (Terman, 1988).

There is a further problem in that if a decreased duration of the photoperiod were an important precipitant of depressive disorders, one would expect the peak of SAD to occur in the winter and for remissions to occur at some point during the spring. In fact, the peak time of incidence for affective disorders is in the spring, with a second peak somewhere in the autumn (Eastwood and Peter, 1988). It would appear, therefore, that there is a seasonal pattern to the incidence of affective disorders, but that there is no obvious relation between this and photoperiodicity. It must also be stressed that while there are peaks of the affective disorders in spring and perhaps a smaller peak in the autumn, affective disorders occur at any point in the year. Thus, the epidemiological evidence suggests that if there is a SAD of the type proposed by Rosenthal and colleagues, it must account for a very small proportion of affective disorders in general (Eastwood and Peter, 1988).

Regarding periodicity, the picture is somewhat similar. Observers of psychopathology, from as early as Aristotle, Aretaeus and Hippocrates, noted that some patients had periodic as opposed to
simply episodic illnesses (Sampson and Jenner, 1975; Jenner, 1988; Wehr and Rosenthal, 1989). The issue was pursued in an increasingly systematic way from the 18th century onwards by Gasimirs (1764), Testa (1787), Kirn (1878), Koster (1882) and Pilez (1901). This tradition culminated in the work of Rolf Gjessing (1932), who described and provided considerable empirical evidence in favour of the existence of a number of periodic catatonias (Gjessing, 1974).

Amidst this interest in periodicity, in the 1860s Falret and Baillarger independently described folie circulaire or folie de deux periodes (Healy and Waterhouse, 1991). This was later to form the basis of Kraepelin's manic-depressive disorder. Despite recognising and having an interest in the periodicity (Sampson and Jenner, 1975) and seasonality (Wehr and Rosenthal, 1989) of mental illness, Kraepelin felt unable to agree that periodicity was a key feature of the affective disorders, arguing that attempts to determine any underlying periodicity 'must of necessity wreck on the irregularity of the disease' (Kraepelin, 1921).

In an attempt to avoid such a conclusion, one recourse since Gjessing has been to attempt to delineate subtypes of affective disorder, one of which might be more clearly periodic. In general, such attempts have distinguished between severe and mild depressive disorders (Checkley, 1989). The former are then taken to represent an endogenous or biologically based disorder. In contrast, the mild quality of the latter has been taken to imply that it is a neurotic disorder — more a matter of a psychological problem than a clear-cut illness. The latter are often said to have no biological stigmata. In the case of the endogenous depressions, it has been argued that there is a modest amount of evidence in favour of an association of disturbed periodicity with these disorders (Checkley, 1989).

The focus on periodicity began before the physiological basis for rhythmicity was formally identified as being 'circadian'. A circadian or chronobiological paradigm only began to emerge in 1937 with the founding of the International Society for the study of Biological Rhythms (Brown, 1982). In 1971, this was rechristened the International Society for Chronobiology to distinguish circadian research from research on biorhythms. The term circadian itself was only coined in 1959 by Franz Halberg.

The distinction between chronobiology and biorhythms is of central importance. The latter study, in which Freud's colleague Wilhelm Fliess played a prominent part (Sulloway, 1980), was predicated on a basis of rigid periodicity. Cycles of 23, 28 and 33 days were postulated and considerable evidence was marshalled in favour of each of these. The cycles were represented as fitting a set of three cosine curves. The effects of these cycles and the conjunction of the various nadirs and peaks of different cycles on human behaviour was held to be important (Sulloway, 1980; Brown, 1982).

While the study of biorhythms has degenerated today into a party game, its significance lies in the rigidity of the periodicities involved. Whether years, months, days or hours, it was assumed that these periods were being marked off by some internal clock that might be liable to pathological dysfunction in a manner that would yield psychiatric illness. This led to both biorhythms and attempts to find regular intervals between the onset of affective episodes, or regularity in the cycling between manic and depressive episodes or distinct lengths to depressive episodes.

To date, however, despite moving away from a biorhythmic paradigm, work on a possible circadian pathophysiology in affective disorders has been dominated by the notion of an aberrant clock. As mentioned in Section 5.1, the phase advance hypothesis has been one such hypothesis. Another was the internal desynchronisation hypothesis (Halberg, 1968; Kripke et al., 1978). This proposed that there were two internal oscillators that ran with different periods. At least one supposedly could not be entrained, and so free-ran with its natural period. The consequence of this would be that beat phenomena should arise as the two oscillators continuously come into phase and antiphase. These beat phenomena, it was argued, might trigger off affective illnesses. Furthermore, the symptoms were supposed to relate to whether the oscillators were in or out of phase.

The only evidence in support of this has been one patient whose temperature rhythm, which was taken to reflect one of the oscillators, appeared to advance continuously and whose clinical course appeared explicable in terms of a pathogenic interaction between this free-running and a normally entrained rhythm (Kripke, 1983; see also Checkley, 1989). As outlined in Section 5.2, the variability of circadian changes between individuals and within individuals between days would argue against any such notion. So also would the continuing lack of evidence for any fixed relationships between the durations of affective disorders and the durations of illness-free periods or relationships between the frequency of manic and depressive episodes in the case of bipolar disturbances (see Section 7).
The circadian system and affective disorders

6. THERAPEUTICS OF THE AFFECTIVE DISORDERS

In addition to the above endogenous features of depression and its supposed endogenous origin in a circadian clock disturbance, the effects of antidepressant drugs on circadian rhythms and the rather mysterious nature of the action of antidepressants, in general, have been offered in favour of the hypothesis that affective disorders involve disturbances of the circadian clock.

There are a number of features of both the clinical and physiological effects of antidepressants that might suggest an action upon the circadian system. The first of these is that antidepressants, quite unlike any other psychotrophic agents, appear to have no common acute behavioural effects. In particular, they are not stimulants. While a number of antidepressants may produce anxiolytic effects relatively rapidly, their major effect on depressed mood is generally held to take longer than 10 days or 2 weeks to appear. This could be seen as consistent with an action to realign a disordered circadian clock or set of rhythms.

Furthermore, antidepressants appear to have no common acute biochemical effects. As regards chronic biochemical effects, the picture is confusing. In the early 1980s, there was a focus on down-regulation of $\beta$-adrenoceptors (Sulser et al., 1978; Healy, 1987b). However, there are a number of puzzling aspects to this in that the capacity of an antidepressant to down-regulate $\beta$-receptors does not correlate with its affinity for these receptors (Willner, 1984).

More recently, there has been a focus on the 5-hydroxytryptamine (5-HT) system and in particular 5-HT$_2$ receptors, with the suggestion that antidepressants act to bring about a delayed onset sensitization of postsynaptic 5-HT$_1$ receptor function (De Montigny et al., 1988; Healy, 1991). However, at present, the balance of evidence suggests that while many antidepressants do lead to a sensitisation of postsynaptic 5-HT$_2$ receptors, this occurs independently of any change in mood (De Montigny et al., 1988; Healy, 1991). There is, furthermore, just as with antidepressants and $\beta$-adrenoreceptors, no correlation between the affinity of antidepressant drugs for 5-HT$_2$ receptors and their ability to modulate the functioning of these receptors.

This lack of correlation between affinity for particular receptors and either the ability to modulate these receptors or therapeutic efficacy suggests that changes in these receptors may come about indirectly. A possible mechanism by which this could happen would be if antidepressants shift the circadian period or alter its phase such that fluctuations in receptor density or affinity are altered as a consequence. Consistent with this are findings that when taken chronically, antidepressants also appear to alter a number of other receptors, apart from the $\beta$ and 5-HT$_2$ receptors, for which they have no intrinsic activity (Healy and Paykel, 1989).

An alternative is that antidepressants alter either the sensitivity of physiological systems, including the clock, to environmental influences or that they modulate the sensitivity of physiological systems to clock outputs. Are these effects of antidepressants underpinned by any demonstrable effects on either rhythms or clocks? There is no coherent answer to this question at present.

The evidence has been reviewed recently by Duncan and Wehr (1988) and by Kripke et al. (1986). What emerges is that very few of the wide range of antidepressants have actually been tested at all. What work that has been done has been performed exclusively on animals. From the animal work, it would appear that there is no consistent effect of antidepressants on either the circadian period or its phase.

The experimental design of the work that has been performed gives no clear indication that serious consideration has been given to potentially differing effects across species, to the impact of time of day of administration of the drugs, to the effect of the half-life of the various drugs being tested or to the dose range being employed (Turek, 1988).

Experiments on antidepressants stand in marked contrast, for example, to work done by Turek and colleagues (see Turek, 1988, 1989, for reviews). They have shown that, using a properly designed experiment, drugs such as triazolam and carbachol can be shown to have clear effects on the circadian clock. From their experiments, it is clear that carbachol can mimic the effects of light. Given at a light-sensitive time, carbachol induces phase shifts in circadian rhythms that can be specifically blocked by cholinergic antagonists such as mecamylamine.

These workers have also shown that triazolam can induce both advances and delays of the circadian phase, depending on the time of its administration (Turek, 1989). Triazolam can also induce changes in the period of the clock. Such effects can be demonstrated when experimental methods
are sufficiently sensitive. In contrast, work on antidepressants has revealed no clear effects of this nature. It is not clear to what extent such a negative result must be attributed to inadequate experimental design.

6.1. What are Antidepressants?

There is a further problem when it comes to considering the potential effects of antidepressants on circadian rhythmicity and the role of such effects in the treatment of depressive disorders and the implications of that role for the aetiology of the disorder. This is the fact that there is no agreement as to what constitutes an antidepressant.

It is generally conceded that electroconvulsive therapy (ECT) is the most effective antidepressant treatment, although it is used relatively infrequently. Based on the delay in the onset of clinical improvement following ECT and the earliest tricyclic antidepressants amitriptyline and imipramine, it appears to have been assumed that these drugs were in some way ‘ECT equivalents’.

The situation has grown more complex since the 1960s with the development of agents acting on the 5-HT system for which the evidence of ‘ECT equivalence’ is lacking (Healy, 1991) and also the production of evidence that a great number of depressive disorders respond to agents that would appear not to be ECT equivalent (Healy, 1990b).

There is considerable evidence that all neuroleptics (Robertson and Trimble, 1982) and all minor tranquilizers (Healy, 1990b) are ‘antidepressant’. If by antidepressant what is meant is the capacity to improve the clinical state of individuals who rate as moderately depressed on instruments such as the Hamilton Rating Scale for Depression, then these agents are antidepressant. The problem with this criterion is that the Hamilton Rating Scale, in particular, codes for anxiety as well as depression and hence, a good anxiolytic may lead to a reduction in scores. However, it is probably also the case that these compounds, by reducing anxiety, lead to a resolution of the disorder itself.

6.2. Sleep and Light Therapies

One difference between these nonspecific compounds and the conventional or canonical antidepressants is that these atypical agents produce rapid improvements in mood and clinical state. The significance of this for the present review is that similarly rapid responses to sleep and light therapies have been taken to indicate that in some way, these therapies are more specific for the affective disorders than are the tricyclic antidepressants, for example (Wu and Bunney, 1990). These therapies, in turn, have also been seen generally as chronotherapies and as such, the rapidity and drama of the response to them has been taken to indicate general support for a hypothesis of clock pathology in the affective disorders.

There are a number of problems with this nexus of assumptions and inferences, that are rarely commented on. The first problem has been alluded to above. It is that sleep and light therapies are not unique in producing rapid responses in the overall level of distress of depressed subjects. A variety of agents, including neuroleptics and minor tranquilizers, as well as psychostimulants, will similarly do so. Although there has been no formal comparison between these atypical antidepressants and sleep and light therapies, regarding the proportion of individuals who respond to treatment or the rapidity of response, there is no reason to believe that such a comparison would yield significant differences.

In contrast to what has been demonstrated for neuroleptics, it would appear that there are specific effects that treatments such as phototherapy may have on the circadian system. Judiciously timed phototherapy may lead to a phase advance or phase delay of circadian rhythms (Czeisler et al., 1989). However, it may be a mistake to assume that this physiological response to phototherapy and the apparent behavioural effects of phototherapy are causally correlated (Terman, 1988).

One argument against such a causal correlation lies in the evidence presented above that a great number of agents lead to relatively rapid improvements in the clinical states of individuals with a depressive disorder.

Quite apart from this, at present in the case of depression, the studies that have been undertaken have not clearly excluded the exogenous masking effect of the treatment process itself. Light energises: thus, while there appears to be a slightly higher response rate to light delivered in the morning, both
midday and evening light appear superior to dim light, for example (Terman et al., 1989). Furthermore, improvements in mood ratings appear to occur with light treatment even in those who do not have a depressive disorder (Terman, 1988).

Except for the studies of Lewy et al. (1987), in the case of the studies of phototherapy that have been presented hitherto, it has been difficult to assess the true phase of circadian rhythms from the material that has been presented. This is important if a hypothesis that phototherapy works by phase advancing (Lewy et al., 1987) is to be sustained. In this case, it would also seem important to establish whether phase delaying treatments — evening light — bring about a deterioration in clinical state, with, ideally, midday treatment making little difference. Current findings in the case of the therapy of the depressive disorders are not easily reconciled with either a specific phase shifting mechanism of action for phototherapy or a straightforward action to extend the photoperiod.

6.3. Cognitive and Behavioural Therapies

In recent years, a number of cognitive and behavioural therapies have been developed that appear effective not only for neurotic or reactive forms of depression, but also for depressions with endogenomorphic features (Healy and Williams, 1988). A common feature of such cognitive and behavioural treatments is their recourse to programmes that involve behavioural activation and motivated activity (Teasdale, 1988). In this they resemble, to some extent, the therapy or hygiene advice given to individuals likely to be exposed to jet lag or shift work maladaptation syndromes (Waterhouse et al., 1987).

In terms of actions on the circadian system, such therapies can be conceived of as acting on the disturbance by altering an individual's lifestyle in a manner that strengthens environmental rhythms. A similar mechanism of action may play a part in the effectiveness of both phototherapy and sleep therapies. While bright light may have clear phase shifting effects and may benefit jet lag states by acting to shift the circadian phase, present evidence, as reviewed above, does not convincingly indicate that this is the mechanism whereby it brings about beneficial effects in affective disorders.

Both phototherapy and sleep therapies, as delivered clinically, involve considerable changes to lifestyle (environmental rhythms) that amount to programmes of behavioural activation, on the lines of those currently employed as part of the cognitive/behavioural arsenal. At the very least, experimental designs aimed at isolating what is specifically therapeutic about light and sleep therapies should take this possibility into account.

6.4. Chronopharmacokinetics

There are a number of chronopharmacokinetic issues raised by antidepressant treatments that have not been adequately studied as of yet. As mentioned in Section 6, current reviews of the effects of antidepressants on circadian parameters have indicated that studies hitherto have paid no clear attention to the timing of psychotropic drug administration to experimental animals. Neither is it clear that any thought has been given to the implications of the half-life of the antidepressant being investigated for experimental outcomes.

There are a number of related chronopharmacokinetic questions, arising from current clinical practice with antidepressants, that have not been addressed and, indeed, have hardly been raised. Antidepressants are unusual compounds in that in clinical practice, it would seem that there is no need to prescribe them on a basis that would take into account their half-life — that is on a divided dose basis, such that steady-state levels of these compounds are reached in plasma. As of yet, there are no indications that there is any correlation between particular steady-state plasma levels and therapeutic recovery.

In the early years of the antidepressant era, in line with the administration of other psychotropic compounds, it was common to have antidepressants prescribed in regimes that involved taking medication three or four times a day. Increasingly, however, the clinical norm appears to be changing to one of prescribing antidepressants in single daily dosing regimes. In the case of the tricyclics, the recent tradition has been to give these last thing at night. Individuals, therefore, get a single dose during the day, usually just before going to bed — although there is no evidence that any one time
is particularly efficacious. Indeed, there are also some indications that giving an antidepressant every few days may be as therapeucic, if not more so, than giving it several times a day (Pollock et al., 1989).

In the background to the question of antidepressant regimes, there is the paradigmatic example of ECT. This is a treatment that is given in single doses and, moreover, is given once every few days rather than three times per day. There is no evidence that increasing the frequency of its administration promotes more rapid response.

In the case of lithium, it was customary to give this treatment in divided doses, owing to a belief that this would minimise the plasma levels at any one point in time and as a consequence, reduce the potential damage to the kidney. Again, this tradition has been overturned recently, and there is some evidence to indicate that single daily dosing, even though this may produce higher acute levels than divided dosing, does less harm to the kidney rather than more (Schou et al., 1982). Indeed, there is also some evidence to suggest that lithium may be as efficacious and safer if given every few days rather than being given every day (Abou-Saleh, 1987).

A number of questions emerge from such considerations. One is whether antidepressants are actually more effective when given in a pulsatile fashion rather than when given regularly or continuously, as is done with other psychotropic compounds. A second question is whether there is a specifically sensitive time when giving antidepressants may produce a more efficacious outcome.

Neither of these questions has been addressed. Such regimes do not seem appropriate to adjusting the clock, but they might play a part in making it more responsive, particularly if time of day of administration is unimportant.

7. MANIA AND ITS THERAPY

It has been estimated that up to one-half of individuals with a depressive disorder who are seen in clinical settings will have an episode of mania at some point during their life. Where individuals with affective disorders have a history of having suffered from both episodes of mania and depression, the disorder is termed a bipolar one. The clinical realities of mania provide a counterpoint to a number of the issues considered in regard to depression. In general, it seems reasonable to suggest that hypotheses regarding the nature of the biological disturbance in depression can only be correct if in some plausible way, they can account for the occurrence and clinical features of mania.

Mania, in contrast to depression, is characterised by a sense of well-being with increased activity levels, increased appetites and reduced need for sleep. Although not uncommon, in general, it has not been well researched in any of its particulars.

Traditionally, bipolar disorders have been seen as being among those affective disorders that are most ‘endogenous’. These are the affective disorders for which researchers have particularly sought for some periodic relationship between the onset of manic and depressive episodes or some periodic aspect to the length of both manic and depressive episodes. None has been forthcoming.

Whatever the precipitation of depressive disorders by life events, until recently, mania generally had been held to arise autonomously. However, the evidence on this latter point is somewhat contradictory at present. Recent research suggests that mania, like depression, may be precipitated by psycho-social disruption. This relationship between the onset of mania and life’s events appears to hold most convincingly for first or early episodes of mania (Ambelas, 1987). Thereafter, the disorder appears to develop some autonomy and to be liable to arise apparently out of the blue or in response to quite minor stresses.

In the limiting case, there appears to be a variant of bipolar affective disorder, termed rapidly cycling affective disorders, in which an individual may cycle with great frequency between the depressed and manic poles of the disorder. To be diagnosed as having a rapidly cycling disorder, an individual must have at least four episodes of depression or mania per year. In a great number of cases, however, the frequency of episodes may be much greater than this. In some rare cases, there may be up to 20 or 30 episodes per year as individuals cycle continuously from depression to mania and back again, with perhaps only brief spells of normality in between. A proportion
of individuals also switch abruptly from depression to mania, or vice versa, in the course of a few hours.

These rapidly cycling disorders are obviously of great theoretical interest in any consideration of a role of the circadian system in affective disorders. There is some evidence that abnormalities of the hypothalamo–pituitary–thyroid axis are associated with rapid cycling (Alarcon, 1985). However, there has been very little systematic work on this group of disorders and none of this work has been replicated.

7.1. Neuroleptics and Mania

The relevance of mania to any consideration of the role of the circadian system in depression becomes clear if one asks whether a treatment for mania could be devised that would be complementary to phototherapy for depression. There are several options. One could enforce darkness and inactivity on someone — this could be done either first thing in the mornings or all day, or particularly in the evening. If phototherapy acts in depression by a specific phase shifting mechanism, then some complementary mechanism must apply in mania and some prediction should be possible regarding what would work.

As regards to whether light itself or light given in the context of motivated activity is more important in depression, clinical practice in cases of mania may hold some pointers. There is at present no treatment for mania that involves putting affected individuals into a darkened and unstimulating environment. It can be argued, however, that the first line of treatment for mania (neuroleptics) does just this.

No clear theoretical rationale for the use of neuroleptics in mania has ever been developed. Their use has developed empirically. The first requirement in a case of mania is behavioural control. Ordinarily, affected individuals are extremely overactive and disinhibited, which puts both themselves and others at potential risk. This leads to the widespread use of neuroleptic drugs for mania, in regimes that produce Parkinsonian-like immobilising effects and a demotivated state.

This demotivation and immobilisation appear conducive to recovery. Whether this is simply a matter of containing disruptive behaviour, until the underlying disorder resolves spontaneously, or whether the induction of immobility and the demotivating effects of neuroleptics have a more specifically therapeutic effect is uncertain. What is clear, however, is that a great number of episodes of mania resolve on treatment with neuroleptics alone.

7.2. Lithium and Mania

While neuroleptics are the first line of treatment for mania, lithium has been held by many to be a more specific treatment. Its effects are not as immediate in their onset as those of neuroleptics. For this reason, it is not as useful for the containment of disruptive behaviour in the early stages of hospitalisation. However, it has been claimed that treatment with lithium brings about a much clearer resolution of manic episodes than does treatment with neuroleptics (Shopsin et al., 1979). It has also been claimed that lithium has specific phase-delaying effects on circadian rhythms in animals (Kripke et al., 1986). This could be taken to indicate an underlying clock pathology in the case of mania.

However, in the case of lithium, all the chronopharmacokinetic questions raised in Section 6.4 regarding antidepressants apply — in a way that they do not for the neuroleptics. This would argue against a clock disorder. A further point of note here is that there is also some evidence that ECT may be an even more effective treatment for mania than lithium (Small et al., 1988).

At this point, a further paradox in the treatment of the affective disorders should be apparent. Both ECT and lithium appear to be of benefit in the treatment of both mania and depression. The conventional antidepressants such as the tricyclic antidepressants have typically been held to be antidepressants only and are liable to cause mania rather than to cure it. However, the evidence for this is scanty. There have been a number of studies of tricyclics given to individuals with mania, which have suggested that they also may have antimanic, as well as antidepressant, properties (Arnone et al.,
If this were the case, then we would be left with a situation in which treatments for depression and mania would perhaps more appropriately be described as therapeutic for affective disorders in general (thymoleptic), rather than specifically antidepressant or antimanic. Given that there are such basic uncertainties about the nature of these treatments, there is further cause to doubt arguments for a clock disturbance in depression that might claim legitimacy from the supposed actions of antidepressants.

8. A SHIFT WORK MODEL OF DEPRESSION

There is a considerable phenomenological overlap between the depressive disorders and both jet lag and shift work maladaptation syndrome. All of these conditions are characterised by dysphoria, anergia, apathy, listlessness, sleep disturbances, as well as disturbances of concentration, along with increased irritability, anxiety and instances of psychosomatic disturbances (Healy and Waterhouse, 1991).

Put in psychiatric language, jet lag and shift work maladaptation syndrome have endogenomorphic features that are comparable to those found in endogenous depression. However, these endogenomorphic features arise in the presence of a disturbance that is not endogenous in origin — a disturbance that arises in response to environmental dislocation (life events).

In addition to a considerable overlap in the phenomenology of depression and shift work maladaptation syndrome, there is a good deal of evidence that similar personality profiles are predisposed to having difficulty managing shift work and depression and that the quality of social support offered to both sets of sufferers has comparable effects on the evolution of the respective syndromes (Healy and Waterhouse, 1991).

Putting these phenomenological aspects of depression together with its apparent response to cognitive/behavioural approaches, it is possible to propose a ‘shift work’ model of depressive disorders. The outlines of this would be as follows: depressive disorders are triggered by environmental disruption, that is by altered social routines and, consequently, altered zeitgebers; they involve, at their heart, disturbances that resemble jet lag and shift work, and they respond to the kind of measures that have also proved useful in the management of jet lag and shift work.

There are a number of problems with this model. One is that the depressive disorders typically involve cognitive features, such as hopelessness, guilt and suicidal ideation, which are not typical of jet lag or shift work. The other point is that the depressive disorders, in general, tend to last for a longer period of time than jet lag or shift work. Their duration and severity has been such that pharmacological management has often appeared necessary to correct some specific endogenous disturbance.

Regarding the first problem, the cognitive features of depression not found in jet lag and shift work maladaptation syndrome, such as hopelessness, guilt and suicidal ideation, studies from general practice suggest that these cognitive features are not a necessary part of a depressive disorder either. Some individuals have them, others do not (Blacker and Clare, 1987; Angst and Dobler-Mikkola, 1984). Their presence typically aids the general practitioner in the diagnosis of the disorder and appears to have some association with the severity of the disorder. Healy and Williams (1988) have argued that such features can be understood as arising within a depressive episode, leading to an increase in severity and potentially to a maintaining of the disorder. Evidence in favour of this has been provided by Williams et al. (1990), who found that dysfunctional attitudes predicted the duration of a depressive disorder. Regarding the second of these points, the duration of the depressive disorders, in recent years, there has been some change in the estimates of the duration and intensity.
of depressive disorders. The older stereotype of depression in the early psychopharmacological era saw it as a disorder of endogenous and/or periodic onset, liable to last for upwards of 1 year and requiring hospital admission during this period and possible treatment with ECT. It is now thought that this stereotype was based on an unrepresentative minority of sufferers of depressive disorders — those who required hospitalisation. Studies during the past 2 decades have focused instead on individuals with depressive disorders presenting to general practitioners. These far outnumber those individuals who are referred to psychiatrists (in a ratio of the order of 20:1) (Goldberg and Huxley, 1980). Based on studies of such populations, it now appears that the majority of depressive disorders, in all likelihood, are much briefer and less intense than was formerly thought (Healy and Waterhouse, 1991).

Current estimates of the duration of the depressive disorders suggest a duration of something around 14 weeks (Blacker and Clare, 1987). There are a number of reasons why even this is likely to be an overestimate. In the first instance, it excludes individuals with brief depressive episodes. The reason for this exclusion is that studies in general practice have sought to identify individuals who meet current research criteria for depressive disorders. Hitherto, these have required affected individuals to have a disorder that lasts at least 2 and sometimes 4 weeks. More recently, Angst et al. (1990) and Montgomery (1991) have drawn attention to the prevalence of recurrent brief depressive episodes. These are depressive episodes that last less than 2 weeks. It has been suggested that these may outnumber those major depressive episodes that have formed the subject matter of research studies hitherto, by a ratio of up to 3:1 (Angst et al., 1990).

A further finding from general practice depression studies has been that the intensity of most depressive disorders is very much less severe than that found in hospitalised depressives. There is no evidence that this milder form of depression is in any way a different disorder to the kind of depression that leads to some individuals being hospitalised. Rather, the evidence at present is that far from having a mild ‘psychological’ problem in contrast to the serious physical disorder that leads to people being hospitalised, individuals who meet criteria for depression and who present to general practitioners complain of disturbances such as lassitude, aches and pains and a general dysphoria rather than any misery, unhappiness or existential despair. The severity of their condition, however, and the lack of cognitive concomitants, appears to lead to general practitioners missing the diagnosis in up to 50% of cases (see Blacker and Clare, 1987, for a review).

8.1. A Research Programme

A number of consequences follow from this formulation. First, there should be some evidence that all depressions involve a core dysrhythmia-related experience of the type encountered in jet lag or a shift work maladaptation syndrome, some of which later develop the cognitive features of hopelessness, helplessness, guilt and suicidality. The studies from general practice, cited in Section 8, would appear to offer a considerable amount of evidence in favour of this possibility. Taking this view, the apparently endogenous origin of endogenomorphic depressions would be more apparent than real.

A second consequence is that depressions should start as relatively mild disorders and become more severe and potentially chronic with the addition of cognitive distortions. Indeed, for cognitive distortion to be a significant factor in the pathogenesis of the affective disorders, the initial experience, almost of necessity, must be noticeable, but not disabling (Healy and Williams, 1988; Healy and Waterhouse, 1991).

A further implication of this proposal regarding the duration of an affective disorder is as follows. It is being postulated that disturbances of rhythm lead to dysphoria, lethargy and apathy, and that demoralisation arises as a consequence of this initial disturbance. It would seem likely that this demoralisation might persist for some time after the provoking disturbance has cleared up and persisting would be rated on depression rating scales and be interpreted as a persistence of the full-blown illness. In favour of this argument are findings from pattern analysis of the response to antidepressants that features of depression, such as sleep and appetite disturbances, typically clear up within 2–3 weeks of antidepressant treatment, whereas impairments of self-esteem and lack of self-confidence take longer (Quitkin et al., 1984; Kravitz et al., 1989). On this basis, the actual
duration of the core of a typical depressive disorder, even without including brief depressive episodes, is likely to be somewhat less than Blacker and Clare’s estimate of 14 weeks (Blacker and Clare, 1987).

A final implication is that cognitive reactions, rather than any immutable clock pathology, may be what determines the chronicity of an affective disorder, as well as its severity. It has been shown that depressed mood differentially activates global self-devaluative concepts in subjects who previously have been depressed (Teasdale and Dent, 1987). This, Teasdale (1988) has argued, is the kind of cognitive processing that might transform a mild and transient depressive episode into a more severe and persistent one. A question that arises from this formulation is whether comparable cognitive distortions have similar effects in shift work maladaptation syndrome.

The central aspect of the proposal being offered here is that there is a mismatch between exogenous and endogenous influences on rhythmicity, and that this mismatch gives rise to the core experiences of dysphoria, lethargy and listlessness found in depression and after shift work. This mismatch has been demonstrated for shift work (Minors and Waterhouse, 1981) and in the affective disorders, in the latter giving rise to the phase advance hypothesis (Wehr and Goodwin, 1981). The correlation of altered rhythms with subjective states has yet to be conclusively demonstrated in either. However, there is some evidence in favour of such a correlation. Recently, both Souetre et al. (1989) and Tsujimoto et al. (1990) have reported negative correlations between Hamilton Rating Scale scores in depressed subjects and the circadian amplitudes of rhythms in temperature and a variety of hormones. Reinberg et al. (1983) have also shown that non-tolerant shift workers have lower amplitude circadian rhythms than tolerant ones.

The task of correlating altered rhythms with subjective states will require a comparison of shift work-induced disturbances and depression, aimed at establishing the initial physiological and cognitive responses to desynchronisation. This will entail prospective, rather than cross-sectional, studies of shift work, to include those subjects who may have difficulties tolerating the change, and a detection of early episodes of hypomelancholia in younger subjects, to minimise contamination by factors making for severity or chronicity.

Both populations would need to be rated on a common scale. There are certain difficulties in devising this, in that what needs measuring is as much the experience of purposiveness, expectancies and incentives as it is of sleep and appetite disturbances or aches and pains. This follows as, for our purposes, what matters in shift work and hypomelancholia is not so much a change in routines as the match between the subject and new routines. In so far as shift workers, for example, are prepared to accommodate their lifestyle to the new pattern required, they do well. Where they attempt to work shift work and live a normal diurnal lifestyle, they do poorly. Therefore, the questions needed will involve some measure of how stressful the changed routines are perceived to be and some assessment of the motivation to overcome it. Some assessment of how life is perceived to be developing, rather than just how much it has changed — ‘Is my life meeting up to the criteria I have set? Questions such as, ‘How much effort would it now take to do . . . a variety of daily tasks’. There would also have to be open questions, in which the response to the subject was judged in terms of the strategy they saw themselves employing to overcome the problem. For example, ‘How will you tackle your sleep difficulties?’ — the options being either to reorganise one’s life or to focus on sleep hygiene, with the former being predicted to indicate a good prognosis and the latter a poorer one. (Our version of such a scale is available on request.)

There are a number of other projects that might critically test the model being offered. A central prediction of this proposal is that subjects who become depressed after life events do so by virtue of the socially disorganising consequences of these events. It follows that there should be evidence of increased disruption of social routines in subjects who become depressed following life events, compared with those also exposed to such events but who do not become depressed. There is some preliminary evidence in favour of this position (Ehlers et al., 1988), but this area needs considerably more research. Accordingly, some form of diary keeping would seem indicated. A number of predictions can be made in addition to the proposal that disruption of social routines will correlate with the development of affective disturbances. One is that this correlation will be strongest for initial affective episodes, as the initial establishment of cognitive distortions will create a vulnerability that will lead to a more rapid emergence of distress of clinical intensity in subsequent episodes.

A second prediction is that ongoing social disruption will correlate with chronicity of affective disorders. A third is that the rate at which subjects otherwise being treated equally with
The circadian system and affective disorders

antidepressants and in comparable therapeutic milieu response to treatment will correlate with the regularity or perceived regularity of routines in their environment. Given two groups of subjects, one of whom are instructed to keep to fixed routines and the other exposed to varying routines, we would predict that the former would respond to treatment earlier.

At a biochemical and physiological level, comparable measurements could be made with shift workers and subjects with hypomelancholia. The hormones cortisol, prolactin and growth hormone could usefully be measured, as these have a varying amount of endogenous and exogenous inputs from primarily endogenous, in the case of cortisol, to largely exogenous, in the case of growth hormone (Minors and Waterhouse, 1981). A study of their various profiles and their relationship to each other in shift work and hypomelancholia would be of interest. Given recent developments in our understanding of endocrine rhythmicity (Rasmussen, 1986), such a study would need to collect samples of each hormone every 1–2 min for a 24-hr period. It is not clear how many subjects would need to be tested. This figure would depend on the actual data produced and the requirements of a mathematical model that would describe the findings adequately. There are a number of parameters that might yield important information, namely the number and frequency of hormonal pulses, as well as the overall rhythmic profile (Veldhuis et al., 1988).

Such a project would be very costly in time, energy and money. A further option would be to use a putative marker of depression, such as the Dexamethasone Suppression test, platelet 5-HT uptake, rapid eye movement sleep latency or the concentration difficulties reported in depressed subjects (Watts and Sharock, 1985). These could then be compared in depressed subjects, in shift workers while working daytime and then while doing night work. In the latter case, one would need to look at the various parameters in the normal diurnal way, but also phase shifted. For example, dexamethasone suppression testing would be performed as normal while the worker worked a night shift. It would then be performed giving dexamethasone at noon and taking blood samples at 0400 hr. A drawback to both these neuropsychological and physiological programmes is that it is not known to what extent any of these altered functions really mark depression and, in particular, early depression rather than later developments.

With many of these and other putative markers of the body clock, there is a lack of data from healthy individuals, which can act as a standard. In the case of deep body temperature, for which the circadian rhythm in health has been adequately described, a recent development has been the idea of using an insulated axillary skin probe. This method is unobtrusive and non-invasive, and early results (Motohashi et al., 1987) suggest that it enables reliable estimates of the phasing of circadian rhythms in shift workers to be made.

Methods also exist to correct the measured temperature rhythm for exogenous influences, so that the phase of the internal clock may be inferred (Minors and Waterhouse, 1989). Correction factors are based on current lifestyle — time of sleep, exercise, leisure, etc. Such data could be derived from the same diary used to assess social routines, as proposed in Section 8.1. Based on this, it would be possible to look at correlations between the phase of the internal clock and changes in activity and sense of well-being.

Another testable implication of this proposal is that subjects who have had an affective disorder should be less tolerant of shift work and jet lag than subjects who have not been affected. This follows as the proposed similarity of the core phenomenological experiences with the affective disorders might be expected to overcome, to some extent, the influence of attributions as to the cause of the disorder. There is some evidence in favour of this with regard to plane travel (Jauhar and Weller, 1982) and in response to sleeplessness (Wehr et al., 1987).

8.2. Implications for the Therapeutics of the Affective Disorders

An implication of this model for the therapeutics of the affective disorders is that a range of treatments of varying degrees of specificity might be expected to influence a 'rhythms' disorders of this type. Accordingly, as regards drug treatment, the apparent current lack of specificity of antidepressants, in their chemical and therapeutic properties, is consistent with the model being proposed. If the affective disorders are something similar to the disorders that are shift work maladaptation syndrome and jet lag, then one might predict that a variety of interventions from
simple anxiolytics through to ‘ECT equivalents’ and on to psychosocial manoeuvres might be of benefit.

A further prediction of the model is that programmes of structured and motivated activity might be expected to help resolve the disorder as they do in jet lag and shift work. Increasingly, this appears to be true. Given this, one can even argue that perhaps the mechanism whereby sleep and light therapies work is on the lines of providing such a framework rather than in terms of any more specific effects on the circadian system that such therapies may also have.

REFERENCES


