Can non-REM sleep be depressogenic?

Domien G.M. Beersma and Rutger H. van den Hoofdakker

Department of Biological Psychiatry, University Hospital, Groningen, Netherlands

(Received 20 March 1991)
(Revision received 25 October 1991)
(Accepted 12 November 1991)

Summary

Sleep and mood are clearly interrelated in major depression, as shown by the antidepressive effects of various experiments, such as total sleep deprivation, partial sleep deprivation, REM sleep deprivation, and temporal shifts of the sleep period. The prevailing hypotheses explaining these effects concern the antidepressant potency of the suppression of either REM sleep or non-REM sleep. This issue is discussed in the light of present knowledge of the kinetics of non-REM sleep intensity, REM sleep production, and their interaction. Recent findings have led us to suggest that the suppression of non-REM sleep intensity is the common pathway in the set of experimental data on the antidepressant effects of sleep manipulations.

Key words: Sleep; Major depression; Non-REM sleep

Introduction

There is ample evidence indicating that sleep and mood are strongly interrelated in patients suffering from major depressive disorder. In these patients, sleep latency is generally long, REM sleep (REMS) latency is shortened, the first REMS episode is relatively long and shows increased frequencies of rapid eye movements, sleep is frequently interrupted by wakefulness and is usually terminated early, and the amount of time spent in slow wave sleep (SWS, i.e., stages 3 and 4) is reduced (Gillin et al., 1979; Feinberg et al., 1982). These sleep disturbances decrease in remission. Some authors suggest that SWS production is not completely normalized (Hauri et al., 1974; Schulz et al., 1978). In addition, various manipulations of the sleep-wake cycle result in a remarkable improvement of mood in a considerable number of cases. Total sleep deprivation (TSD) is among the best studied sleep-wake manipulations in this respect. About 60% of patients with a major depressive disorder respond to TSD with an immediate and substantial reduction of depressive symptomatology. Unfortunately, however, subsequent recovery sleep usually leads to a
relapse (Pflug and Tolle, 1971; Van den Burg and Van den Hoofdakker, 1975; Gillin, 1983; Kuhs and Tolle, 1986; Elsenga and Van den Hoofdakker, 1987). Other sleep-wake manipulations which have been studied include partial sleep deprivation (PSD), during which sleep is allowed only during the first or the last hours of the night; sleep stage deprivation, during which REMS, electroencephalographically determined, is selectively deprived (RSD) or during which non-REMS sleep is selectively interrupted (NRI), and temporal shifts of the sleep-wake rhythm (TS), during which patients are instructed to sleep at unusual clock times. The results of such manipulations are: that depressive mood is more likely to improve when sleep is restricted to the beginning of the night (Schilgen and Tiille, 1980; Sack et al., 1988; Elsenga et al., 1990); that RSD is a better treatment strategy than NRI (Vogel et al., 1975); and that shifting sleep to advanced clock times may sometimes have considerable antidepressive effects (Wehr et al., 1979; Sack et al., 1985; Souètre et al., 1987; Van den Hoofdakker and Beersma, 1988).

Although various hypotheses have been proposed to explain the experimental results (Kripke et al., 1978; Wehr and Wirz-Justice, 1981; Borbély and Wirz-Justice, 1982; Wehr and Goodwin, 1983; Vogel et al., 1990; Wu and Bunney, 1990), there are no sleep-physiological hypotheses in the literature which are compatible with the total data set.

The RSD experiment carried out by Vogel and coworkers (Vogel et al., 1975) suggests that increased REM pressure is a sufficient prerequisite for an antidepressive response to sleep-wake manipulations. This hypothesis does not, however, provide an explanation of the results of other interventions in the sleep-wake cycle. The most important manifestation of this inconsistency is that the antidepressant response to RSD develops in the course of 3 weeks of REMS deprivation, whereas the effects of TSD and PSD are immediate. The differences in the response kinetics cannot be explained by differences in REMS pressure, as both TSD and RSD result in virtually complete REMS suppression. Furthermore, PSD procedures usually result in the occurrence of more REMS than selective REMS deprivation does, and yet PSD leads to immediate improvement whereas REMS deprivation does not. Finally, allowing patients to take a nap during the day following a successful sleep deprivation may sometimes result in a relapse towards a depressive state, irrespective of whether the nap contained REM sleep or not (Riemann et al., 1990).

On the other hand, it has been suggested that the pressure for non-REMS is crucial for an antidepressive response (the so-called S-deficiency hypothesis (Borbély and Wirz-Justice, 1982; Borbély, 1987)). According to this hypothesis, low pressures for non-REMS facilitate depression. Although this hypothesis is qualitatively compatible with the results of TSD, PSD and TS, it seems to be incompatible with the results of the RSD study. The control group in that experiment, which was subjected to NRI, did not show an antidepressant response (Vogel et al., 1975).

Since the publication of the REM pressure hypothesis in 1975 (Vogel et al., 1975) and the S-deficiency hypothesis in 1982 (Borbély and Wirz-Justice, 1982), there have been major developments in the knowledge of sleep regulatory mechanisms. The present paper presents a theoretical discussion of the data and hypotheses in the context of the current knowledge of such mechanisms. Arguments will be provided to show that non-REM sleep may be depressogenic. It will be argued that this hypothesis is consistent with all sleep manipulation studies in depressives, including the RSD experiment carried out by Vogel et al. (1975).

Non-REMS sleep intensity

The idea that the efficiency of recovery processes during sleep may vary is an old concept which is in agreement with the results of many studies (Patrick and Gilbert, 1896; Feinberg, 1974; Borbély, 1982; Daan et al., 1984; Dijk and Beersma, 1989). For non-REMS it has been shown that the threshold for awakening when a subject is subjected to acoustic stimuli is directly related to the magnitude of slow waves in the EEG, that is, the larger the amplitude, the higher the threshold (Rechtschaffen et al., 1966). Fur-
thermore, EEG amplitude is strongly determined by the history of sleep and wakefulness. The longer the duration of prior wakefulness, the higher the EEG power density in non-REM sleep (Borbély et al., 1981; Dijk et al., 1987), and therefore the higher the non-REMS EEG amplitude (EEG power density is proportional to the square of EEG amplitude). It has been concluded that non-REMS becomes more intense in response to a longer duration of wakefulness.

During undisturbed sleep, the opposite effect is observed. Here, EEG amplitude and EEG power density decrease in the course of the sleep episode (Feinberg, 1974; Borbély et al., 1981; Dijk et al., 1989). Apparently, non-REMS intensity gradually decreases when the need for non-REMS is being satisfied. Interference with non-REMS intensity, as, for instance, by means of acoustic stimulation, appears to postpone high non-REMS EEG power density to a later time (Dijk et al., 1987; Dijk and Beersma, 1989). Taking EEG power density as a measure of non-REMS intensity, it is even possible to quantitatively predict the non-REMS recovery response to experimentally induced suppression of non-REMS (Dijk et al., 1987). Clearly, non-REMS intensity is subject to homeostatic control.

These data have important consequences for the interpretation of sleep manipulation studies in depressed patients. A reduction in the duration of stage 4 sleep, for example, may be partially or totally compensated by an increase in the intensity and/or duration of the other non-REM sleep stages, including stage 2. The existence of this regulatory mechanism is particularly important in the interpretation of the results of the control group in the RSD experiment carried out by Vogel et al. (1975). In this group non-REMS was interrupted by inducing epochs of wakefulness in order to control for the sleep interruptions and intermittent wakefulness caused by the RSD procedure in the experimental group. This NRI did not lead to a significant improvement of mood. However, because of the homeostatic control of non-REMS intensity, the NRI would probably not have led to a substantial non-REMS intensity deficit. Instead, it must be expected that the NRI-induced loss of non-REMS time was compensated for by increased non-REMS intensity during the undisturbed sleep epochs.

Admittedly, this argument is based on extrapolation from data on healthy subjects. The question must be raised as to whether the homeostatic processes regulating non-REMS intensity in normals are intact in depressives. This is still not completely decided. Two types of data are relevant to this question. First, the analysis of EEG power density (1–15 Hz) before and after TSD shows the same increase in depressives as in healthy subjects (Beersma and Van den Hoofdakker, 1988). In other words, extension of the duration of prior wakefulness leads to similarly increased sleep intensity in both normals and depressives. So the response of EEG power density to manipulation of prior wakefulness seems similar in depressives and controls. The controversy concerns evaluations of normal nights of sleep. Some studies indicate that depressed patients and age matched controls do not differ when EEG power density is summed over total sleep time (Mendelson et al., 1987 (power density 0.23–25 Hz); Van den Hoofdakker and Beersma, 1988 (power density 1.0–15 Hz)). In spite of the fact that the patients in both studies had reduced amounts of slow wave sleep (60% and 49% of control values, respectively), integrated power density was slightly higher than control values in both studies (101% and 104%, respectively). In other studies, however, reduced power densities were found in non-REM sleep of depressed patients (Borbély et al., 1984; Kupfer et al., 1984), suggesting impaired homeostatic control of non-REM sleep. Apparently, depression interferes with stages 3 and 4, leading to reduced amounts of SWS. In some patients, the loss of SWS time is, however, compensated for by increased non-REMS intensity during the rest of the sleep episode (predominantly in stage 2), just as in experimental sleep disruption in healthy subjects. In other patients such compensatory reaction is not observed. So, although some patients may show reduced values of non-REM EEG intensity while others do not, the reaction of non-REM EEG intensity to sleep deprivation seems normal. Consequently, there is no reason to presume that non-REMS regulatory processes generally differ between depressives and healthy subjects. There-
fore, it is unlikely that the non-REMS interruptions in Vogel et al.'s control study yielded a substantial non-REMS intensity deficit. This conclusion is even more warranted when the duration of sleep is taken into account. It must be noted that the NRI group slept 40 min longer than the RSD group. The longer the opportunity for non-REM sleep, the more a possible intensity deficit can be compensated for.

Pressure for REM sleep

Knowledge of the mechanisms regulating the occurrence of REM sleep is less detailed. It has been noted that the propensity for REMS varies with the time of the day, with a maximum in the early morning (Czeisler et al., 1980). Furthermore, it is obvious that there is some kind of homeostatic control, because RSD leads to increased duration of REMS during recovery sleep. This is true of RSD carried out using REMS interruption by awakening (Dement et al., 1966; Agnew et al., 1967; Beersma et al., 1990) as well as of pharmacologically induced RSD (Oswald, 1973; Cadilhac, 1976; Chen, 1979). However, an intensity dimension of REM sleep is not known. It has been argued that the frequency of rapid eye movements during REMS cannot be considered a measure of REMS intensity (Borbély and Wirz-Justice, 1982; Beersma et al., 1990). In the absence of an adequate measure for REMS intensity, it can still be concluded that the need for REMS increases with the duration of REM sleep suppression in both healthy subjects and depressed patients, since an increasing number of awakenings is needed to prevent REMS from occurring. The only reason to presume that there are differences in REMS regulation between healthy subjects and depressed patients is the phenomenon that, in depressives (unlike in controls), the REMS latency is occasionally less than 40 min, while the first REMS period is abnormally long, showing increased frequencies of rapid eye movements. Detailed analysis (Beersma et al., 1983; Van den Hoofdakker and Beersma, 1985) has led to the conclusion that these abnormalities merely represent a sleep onset phenomenon. After 1 h of sleep, the further accumulation of REMS is normal. It is highly conceivable that this sleep onset abnormality results from a disturbance in arousal mechanisms rather than from the disturbed regulation of REMS. According to this assumption, there is no reason to presume that there are differences between healthy subjects and depressed patients in the regulation of REM sleep per se.

Interaction between non-REM sleep and REM sleep

The regulatory mechanisms of non-REMS intensity and of REMS duration are not completely independent. A recent study in healthy subjects (Beersma et al., 1990) showed a reduction of non-REMS intensity in response to the deprivation of REMS during the first 5 h of sleep. The control study in which the same subjects were awoken from non-REMS did not show a comparable non-REMS intensity reduction. It was concluded that increased pressure for REMS causes a reduction of non-REMS intensity. A similar conclusion resulted from another study in which sleep duration was restricted to 4 h per night for two successive nights (Brunner et al., 1990). These data have consequences for the interpretation of sleep stage deprivation data in depressives. Since regulatory mechanisms of non-REMS and REMS are presumably very similar in depressed patients and healthy controls, it may be expected that the interaction between the two sleep stages is also similarly regulated. Thus, RSD is not as selective as was previously thought. Vogel et al.'s RSD experiment in depressives probably has led to reduced non-REMS intensity, whereas the control condition may not have been very efficient in creating a non-REMS intensity deficit. This reasoning leads to a totally different explanation of the antidepressive effects of RSD in depressed patients.

A non-REMS intensity hypothesis for depression

On the basis of these facts and considerations, it seems to be justified to postulate that, in depressed patients, non-REMS is a pathogenetic factor. The core of this postulate is the hypothesis that suppression of non-REM sleep has an antidepressive effect. This hypothesis is qualita-
tively compatible with the results of all sleep-wake manipulations, that is, TSD, PSD, RSD, NRI and TS. Moreover, it may explain effects of antidepressive medication. In the subsequent paragraphs, this compatibility will be explained.

**Total sleep deprivation**

If suppression of non-REMS intensity has an antidepressive effect, TSD should obviously be antidepressive because this intervention completely suppressed non-REMS intensity. Recovery sleep after TSD is more intense than baseline sleep. Therefore, the relapse usually observed after recovery sleep from TSD is consistent with the hypothesis.

**Partial sleep deprivation**

The antidepressive effects of PSD, as well as the impact of PSD on non-REMS intensity, depend strongly on the time interval during which sleep is allowed (Schilgen and Tolle, 1980; Sack et al., 1988; Elsenga et al., 1990). If sleep is confined to the beginning of the night, non-REMS shows its usual characteristic of the first part of the night in depressives, that is, it is relatively shallow. Therefore, a significant fraction of non-REMS will be blocked by the early termination of sleep. In contrast, if sleep is allowed in the second part of the night, the need for sleep will be increased. Therefore, sleep efficiency is greatly increased (Sack et al., 1988) and (because of the regulatory mechanisms of non-REMS) non-REMS will be more intense (Dijk et al., 1987). Thus, the therapeutic superiority of PSD in which sleep is restricted to the beginning of the night is in accordance with the hypothesis.

**Selective** sleep stage deprivation

As has been argued, the interruption of non-REMS will not be very effective in creating a non-REMS deficit due to the homeostatic control of non-REMS intensity. Therefore, the absence of an antidepressive response to NRI is in accordance with the hypothesis.

In contrast, the deprivation of REMS is likely to yield a suppression of non-REMS intensity. It is conceivable that the non-REMS intensity suppression will last as long as RSD is continued. The steadily increasing pressure for REMS will prevent normal homeostatic control of non-REMS intensity. The continued suppression of non-REMS over days is qualitatively compatible with the gradually developing improvement of mood in response to RSD. Obviously, the non-REMS intensity hypothesis for depression explains the results of Vogel et al.'s RSD experiment.

**Shifting the sleep-wake schedule**

Advancing the time interval allowed for sleep to earlier clock times may, in some cases, lead to an improvement in mood over the course of 3–4 days. This is usually followed by a relapse after 1–2 weeks (Wehr et al., 1979; Sack et al., 1985; Souëtre et al., 1987; Van den Hoofdakker and Beersma, 1988). Again a reduction of non-REMS intensity may arise. This may be due to two factors. First, as an immediate response to the advance of bedtime, the homeostatic control of non-REMS intensity yields a reduced need for non-REMS (Daan et al., 1984). This effect occurs mainly during the first night of sleep. Secondly, it can be shown mathematically (according to the two-process model of sleep regulation (Daan et al., 1984)) that advancing the timing of sleep must lead to a lower average intensity of non-REMS. This is due to the fact that at the new phase of sleep, the thresholds to the S-process are at a lower level. The latter factor becomes less effective when the circadian system begins to adapt to the time-shifted situation. In summary, therefore, the results obtained with phase advances of the sleep-wake rhythm are consistent with the non-REM intensity hypothesis for depression. In studies on the clinical effects of phase delays, the course of mood has not been described in sufficient detail to be discussed here.

**Antidepressive medication**

Most antidepressive drugs reduce the amount of REM sleep. Spectral analysis has, however, seldom been applied as a method of quantifying the impact of these drugs on non-REMS intensity. In one recent study (Kupfer et al., 1989), the first 100 min of baseline sleep were compared with the same interval of sleep after a 'loading dose' of clomipramine. It was concluded that
non-REMS intensity, summed over the 100-min interval, was significantly increased in response to clomipramine. However the duration of REMS and wakefulness within the 100-min intervals would be expected to be very different under the two conditions. No data regarding these critical issues are provided. Indirect evidence for the existence of such differences can be derived from REMS latency data. Average REM latency was 45 min in baseline versus 310 min after the loading dose of clomipramine. These differences have not been taken into account. Furthermore, comparisons should include all non-REMS episodes for the entire sleep period because the temporal distribution of non-REMS intensity may have been altered by the use of the drug. Since the authors did not perform this comparison, it cannot be concluded that clomipramine increases non-REMS intensity.

In the context of the non-REMS intensity hypothesis for depression, we would speculate that antidepressant drugs are effective because they suppress non-REMS intensity. This suppression is not necessarily immediate, but may develop in the course of drug treatment. Such suppression may be a direct result of the influence of antidepressive medication on non-REMS or may be an indirect consequence of the suppression of REMS. In fact, recent animal data indicate that clomipramine has a strong suppressive effect on non-REMS intensity (Dijk et al., 1991).

In spite of this apparent consistency, it must be noted that in drug studies the interpretation of the non-REMS EEG in terms of non-REMS intensity may be difficult. Various drugs may modify the EEG power spectrum in different ways (Acherman and Borbély, 1987; Dijk et al., 1989). It has been hypothesized that benzodiazepines, for instance, reduce the amplitude of the non-REMS EEG, without having much influence on the underlying non-REMS regulatory mechanisms (Acherman and Borbély, 1987; Borbély et al., 1991). Apart from the fact that this interpretation in the context of the present hypothesis may explain why benzodiazepines are not generally used as antidepressants, this interpretation also renders it very difficult to interpret the results of any drug study in terms of its influence on non-REMS mechanisms.

Comparison with the S-deficiency hypothesis

According to the S-deficiency hypothesis (Borbély and Wirz-Justice, 1982), process S (i.e., the pressure for non-REMS) is the link between sleep and depression. In addition, the build up of S during waking is assumed to be impaired. Under this hypothesis, depressed mood is related to low levels of S, and increased levels of S are needed to normalize mood. Since suppression of non-REMS intensity prohibits the decrease of S, non-REMS suppression results in a relatively high value of S, which also according to the S-deficiency hypothesis would result in improved mood. So, although this has not been claimed by the authors of the S-deficiency hypothesis, this hypothesis does qualitatively explain the effects of REMS deprivation just as well as does the present non-REMS hypothesis. There is, however, one essential difference between the two hypotheses. Under the non-REMS intensity hypothesis of depression, S is not supposed to be deficient. It is only the suppression of non-REMS that is considered to influence the improvement of mood. This distinction is crucial, as, in several studies, no differences in accumulated sleep EEG energy (i.e., non-REMS intensity, summed over non-REM sleep time) between depressives and controls (Reynolds et al., 1985; Mendelson et al., 1987; Beersma and Van den Hoofdakker, 1988) were found. In spite of the fact that some other studies (Borbély et al., 1984; Kupfer et al., 1984) did report reduced values of accumulated sleep EEG energy in depressives, apparently subgroups of depressed patients can be found without a deficient process S. As a result, the present non-REMS intensity hypothesis can be considered a modification of the S-deficiency hypothesis, where high pressures for non-REMS still are considered to be antidepressive, but where a deficiency of process S is not required.

Perspective

The present hypothesis offers an alternative to the frequently quoted hypothesis that increased REMS pressure leads to reduced depressive symptomatology (Vogel et al., 1975, 1980). The increasingly large body of literature which is com-
compatible with the REMS pressure hypothesis has recently been reviewed (Vogel et al., 1990). In our view, possible influences of drugs on non-REMS sleep are underestimated and have not been adequately quantified. The hypothesis that reduced non-REMS intensity may have an antidepressive effect seems to be totally compatible with the data.

Finally, we would like to point out that the present hypothesis is more than a mere framework upon which experimental results can be integrated. It may also be used to predict the outcome of future experiments. A variety of experimental interventions in non-REMS can be performed. The impact of these interventions on non-REMS intensity can easily be quantified by spectral analysis. The present hypothesis predicts that all designs which result in a decrease of non-REMS intensity summed over time will reduce depressive symptomatology.

References

Feinberg, M., Gillin, J.C., Carroll, B.J., Greden, J.F. and Zis,


