Defining and determining the properties of the human sleep homeostat
Zavada, Andrei

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1

General Introduction

Humans are diurnal mammals. Like all higher primate species they are active in daytime and have adapted to sleep at night. Other mammalian species, such as mice and most other rodents, have evolved to be nocturnal or they are cathemeral, which means that they obtain their sleep at any time of the day–night cycle (e.g., cats). There is no universal benefit in having sleep to occur specifically at night, but every species is adapted to obtain sleep in its own orderly fashion. It will revert to this temporal organization if disturbed.

So, sleep in humans occurs at night and in young and healthy subjects it commonly occurs in one consolidated bout. The timing of the sleep episode within the night varies, both between subjects and between nights within subjects. The variations include variation in sleep duration. Apart from showing variations in timing, sleep also varies in intensity. Episodes of sleep deprivation are followed by episodes of more intense sleep, which is expressed in larger slow waves in the EEG signal (Borbély, 1981) and giving rise to lower response rates to acoustic signals (Dijk et al., 1987b). During the sleep episode this indicator of sleep intensity varies systematically as well. High EEG slow waves are characteristic for the beginning of the sleep episode, while slow wave sleep is of reduced amplitude at the end of the night. These phenomena of sleep timing and intensity are not independent.

Spontaneous sleep timing is not only controlled by these endogenous processes. The possibilities of electric light together with social demands often make humans extend their wakefulness in the evening and terminate sleep prematurely in the morning. This occurs at the expense of sleep time. The reduction of sleep time puts the mecha-
nisms regulating sleep under strain, and encroaches upon the presumed recuperative function of sleep.

These two topics, the preference in timing of sleep and properties of the sleep homeostat, will form the subject matter of this thesis.

1.1 Human sleep timing

1.1.1 Behaviour

People with a natural propensity to go to bed early in the evening and to wake up early, are commonly referred to as ‘larks’. By analogy, those who can hardly resist to go to sleep late and do not get up until late in the day are known as ‘owls’. In the current scientific literature these subgroups are called Early and Late chronotypes. Early chronotypes are relatively less common than late chronotypes. In a human population, this is reflected by a skewed distribution of the midsleep on free days, with a steeper rising slope on the early side and a shallower drop on the late side (Roenneberg et al., 2003; Paine et al., 2006, see also Chapter 2).

We conducted a survey of chronotypes in The Netherlands (Chapter 2) using the recently developed Munich ChronoType Questionnaire (MCTQ; Roenneberg et al., 2003). This survey fully confirmed the age dependence of midsleep on free days reported by Roenneberg et al. (2004a). We compared the results of this chronotype questionnaire with the morningness–eveningness preference questionnaire (MEQ) of Horne and Östberg (1976). This is of interest because the Morningness-Eveningness Questionnaire, often used in older literature, asks for preferences in a qualitative manner, while the Munich ChronoType Questionnaire refers to actual sleep timing. The statistical comparison of the results of both questionnaires allows us to estimate to what extent the preferences influence actual behaviour. It also serves as a ‘validation’ of the more recent MCTQ against the widely accepted MEQ.

Extreme chronotypes suffer from their chronotypical trait to varying degree. Extreme early types have little difficulty with rising in the morning, but they experience it as hard to stay awake in the evening, when often opportunities to engage in social activities prevail. Late types can easily participate in social events up to late hours in the night, but they have great difficulty to rise in the morning and be awake and alert at their work.
1.1. HUMAN SLEEP TIMING

The established view on the causes defining the spontaneous sleep phase includes two distinct aspects: a circadian aspect regulated by the circadian clock in the brain, and a homeostatic aspect relating the current sleep propensity to the prior history of sleep and wake alternation.

The circadian clock is located in the suprachiasmatic nucleus (SCN), the master pacemaker in the mammalian circadian system. The SCN generates a rhythm with a cycle length of about 24 hours by itself, which is entrained to the environmental light-dark cycle (Pittendrigh and Daan, 1976a). The SCN coordinates different rhythms in the body including the activity rest cycle. Thereby organisms adjust their behavioural patterns to the relevant 24-h changes in the environment. Two characteristics of the circadian pacemaker are relevant to sleep timing: the intrinsic period of the endogenous oscillation and the sensitivity to light of the entrainment process. If the period of the pacemaker is long, the pacemaker will have the tendency to lag behind the periodicity of the zeitgeber. It is commonly assumed that late chronotypes are likely to have such long intrinsic period of the pacemaker (Duffy et al., 1999; Czeisler et al., 1999; Duffy et al., 2001). Yet, the timing of the pacemaker signal relative to the light-dark cycle is not solely based on intrinsic period. It is also affected by the sensitivity to the environmental zeitgeber, i.e., light (Daan and Pittendrigh, 1976; Dijk et al., 1995; Terman et al., 1995). Circadian pacemakers respond to light in the morning by advancing their phase, effectively shifting the rhythm to earlier clock times. A light pulse applied in the evening does the opposite. The actual phasing of the pacemaker signal against clock time thus results from the balance between effects of intrinsic period and responsiveness to light. As a consequence, it is not only intrinsic period that determines chronotype, but also the sensitivity to light. One might thus presume that a subject very sensitive to light in the morning would be subject to a large advance of its pacemaker, and be prone to become an early type. The systematic information on sleep timing collected by means of the Munich Chronotype Questionnaire in a large population sample allowed us to recruit subjects to an experiment (Chapter 3) to investigate these two aspects of their circadian organization: the endogenous circadian period (τ) and the susceptibility to shift phase under the effect of bright light in the subjective evening and subjective morning.

The other aspect of sleep regulation to be investigated in the context of chronotypes is the homeostatic aspect. As most of us know
Figure 1.1 Two-process model simulations showing the effective displacement of sleep phase through variation of the rise rate of Process S. The rise rate determines the steepness of S build-up. Each panel has two instances of S: solid line representing a 'normal' subject, and dashed line representing a subject with higher (A) or lower (B) rise rate. The time constants (τ_r for Wake and τ_d for Sleep) are inversely proportional, respectively, to the rise and decay rates of Process S. Solid bars below zero line represent sleep time (upper for 'normal' rise rate, lower for deviating rise rate). The phase, amplitude and skewness of Process C (two thin skewed sine waves limiting the oscillation of Process S) are identical in (A) and (B), and so is the decay rate. Simulation data: τ_d = 4.2 h; 'normal' subject, τ_r = 18.2 h; stable sleep start 0:17, midpoint 4:19, end 8:21, length 8.1 h; 'early/late' subjects: (A) τ_r = 14.0 h, stable sleep start 22:03, midpoint 2:50, end 7:37, length 9.6 h; (B) τ_r = 23.0 h, stable sleep start 2:15, midpoint 5:36, end 8:58, length 6.7 h.

From own experience, sleep can easily be delayed beyond the preferential timing. Such disturbance will elicit a rebound in sleep intensity and/or duration in the following night (Berger and Oswald, 1962; Williams et al., 1964a; Borbély, 1981). The response suggests the presence of a recuperative, or homeostatic, component involved in sleep regulation. It has been demonstrated that the high-amplitude
1.1. HUMAN SLEEP TIMING

Slow waves in the NREM sleep EEG yield a good indicator of this homeostatic process: slow wave activity (SWA, spectral power in the \(~0.75–4.5\text{ Hz range}\) is high in the beginning of sleep when the need for sleep is high; SWA is low at the end of the sleep episode when sleep need is satiated. After sleep deprivation, SWA is increased above the normal levels (Borbély, 1981), and suppression of SWA in the beginning of the night leads to predictable rebounds in the later part of the night (Dijk et al., 1987b). In the two-process model of sleep regulation (Daan et al., 1984), SWA has been taken as an indicator both of the need for sleep, designated ‘Process S’, and of the rate of its dissipation. Process S increases during wakefulness and decreases during sleep. In the model it is proposed that the alternations between sleep and wake occur because of Process S reaching thresholds that are set by Process C, which in turn is controlled by the circadian pacemaker. Process C rises in the course of the day, and thereby runs parallel to the rise in S during wakefulness. The rise in Process C prevents the subject from increasing fatigue and reduced alertness due to the rise in S. For this reason C and S have also been termed “opponent processes” by Edgar et al. (1993), who demonstrated by SCN lesions in squirrel monkeys that the alerting effects of C in daytime are directly controlled by the SCN.

Figure 1.1 presents a schematic representation of the two-process model. It is evident from the model that shifts in phase of the circadian Process C should lead to similar shifts in sleep timing because the inversions (sleep onset and end) of the course of Process S are regulated by Process C. There are yet other options in the model to yield shifts in sleep timing. Faster time constants of Process S, for instance, will lead to earlier sleep timing relative to Process C. Slower time constants will result in later timing (Fig. 1.1). Whether a connection exists between the time constant of Process S build-up and the chronotype, is investigated in Chapter 4.

1.1.2 Properties of the human sleep homeostat

In the original formulation, the decay rate of Process S during sleep was estimated from average SWA values of each NREM–REM cycle in EEG recordings (Borbély, 1981) by fitting an exponential breakdown curve to them (Daan and Beersma, 1983). Estimation of the rise rate (the inverse of time constant) was done by fitting an exponentially saturating curve to three values derived from EEG recordings of baseline
sleep and recovery sleep after sleep deprivation: the end of baseline sleep, the beginning of baseline sleep, and the beginning of recovery sleep (Daan and Beersma, 1983). Implicit in the exponential decay of $S$ during sleep is that the instantaneous decay rate at any time is proportional to the level of $S$ at that time. Experiments applying selective deprivation of SW A and investigating subsequent recovery (Dijk et al., 1987b) indeed proved that the amount of SW A present in the NREM EEG is proportional to the rate of decay of Process S. Suppression of SW A for three hours resulted after that interval in SW A that was more intense than in baseline for the time of the night, but only marginally less intense than observed at the onset of an undisturbed baseline night. This finding quantitatively matched predictions from the S-C model (Daan et al., 1988). It led Achermann et al. (1993) to develop a comprehensive model in which (a) fine details of SW A dynamics are taken into account (such as, periodic intrusions of REM sleep and latencies on vigilance state changes), and (b) Process S is made clearly distinct from SW A with a new parameter, called ‘gain constant’, mediating the relationship between these variables, thus: $dS \propto -gc \cdot \text{SW A}$. An example of such profile is presented in Fig. 1.2.

In Chapters 5 and 6 of this thesis I focus on SW A patterns for various positions on the skull and discuss local differences in the regulation of SW A. For that purpose I rely heavily on Achermann’s modeling work. The concepts of this approach are explained here.

Achermann and coworkers started from the model’s prediction and its experimental confirmation by Dijk et al. (1987b) that Process S declines during sleep in linear proportion to average current values of SW A. Achermann et al also quantified SW A dynamics throughout the sleep episode in great detail and incorporated the results into the model. In this version of the model, the decay rate of $S$ is thus determined directly by SW A and only indirectly by the level of $S$. Furthermore, the model uses the hypnogram (i.e., the observed sequence of sleep stages and intervening wakefulness) to simulate the SW A pattern. The simulation starts from the empirical level of SW A ($SW A_0$), and, aided by the hypnogram, iterates along the entire timeline reconstructing the course of SW A. During NREM sleep, SW A has the tendency to increase in value, at a rate dependent on the current value of $S$. The magnitude of this increase is a parameter (SW A rise constant, $rc$) of the model. When wakefulness occurs, SW A goes down rapidly with a time constant provided by another parameter (SW A fall constant triggered by Wake, $fcW$) of the model. The values do not drop to zero,
1.1. HUMAN SLEEP TIMING

**Figure 1.2** Two cases of simulation of a frontal derivation SWA and Process S using Achermann’s model, from a simple sleep deprivation experiment (baseline night, night awake, recovery night; detailed in Chapter 6). Original SWA profile shown in solid grey, simulated SWA course as thick black line; thin black line running on top of SWA is Process S. SWA and $S$ are codimensional, expressed in percent of the average SWA in Baseline. Below the zero-line is a simplified hypnogram, showing NREM sleep in full-height solid bars and REM sleep in half-height bar. The ticks on $y$-scale are the starting values of $S$ (the lower value of the two), and the upper asymptote of $S$ (the higher of the two; out of range in case A). Note that the peak SWA in Recovery in case A is clearly higher than in Baseline, whereas these values are rather similar in case B. This is interpreted as an indication of very weak homeostatic response in case B.

but level off at a value set by another model parameter ($SWA_L$). In subsequent NREM sleep, the increase of SWA follows its course in a similar manner as at sleep onset. In anticipation of REM sleep ($t_a$ min before scored REM), SWA begins to decrease towards the level during REM sleep (95% of the lowest SWA observed in REM is $SWA_L$). Similarly, the subsequent rise of SWA after the REM episode is inhibited for a short period ($t_p$ min). During wakefulness, $S$ increases towards an upper asymptote ($S_U$), With a certain rise rate ($r_s$). In contrast to the original proposition for $S$, the increase of $S$ in Achermann’s model is not cancelled during sleep; however, with sufficiently intense SWA, the decline of $S$ outweighs the tendency to increase. This aspect is a subtlety of the model, incorporated by Achermann because the accuracy of simulations was enhanced by it.

The model of Achermann et al. (1993) was not designed to predict sleep timing; it was made to simulate the fine structure of SWA profiles. Therefore, the parameters of Achermann’s model are not identical to the parameters of the two-process model, even though some of them refer to similar aspects of regulation. The rise rate of Process S, for instance, is interpreted similarly in both models, but not
identically. In the two-process model, the rise of $S$ is postulated to only occur during waking, not during sleep, whilst in the SWA model, the rise of $S$ is explicitly allowed to take effect during waking as well as REM sleep and even NREM sleep (although in NREM, its effect is by and large overruled by the SWA-controlled decline of $S$). Hence, the exact values of the rise rate may be different for the two models. The other parameters that play similar (but not identical) roles in the two models are the decay rate of $S$ in the two-process model and the “gain constant” of the SWA model. While the decay rate refers to an exponential decline of $S$ during sleep in the two-process model, the gain constant concerns the reduction in $S$ per unit of SWA in the SWA model. A fundamental difference between the two parameters is that the decay rate represents a global characteristic of the overnight SWA profile, while the gain constant describes the direct relationship between SWA and $S$ without reference at all to the overnight SWA profile.

Achermann’s SWA model was taken as the working model in the second part of this thesis. In Chapter 4 it was applied on the sleep data of the early and late chronotypes who also contributed to the data of Chapter 3.

1.1.3 Topography of sleep: Multiple sleep homeostats or variable reflections of a single homeostat?

Given an EEG profile, Achermann’s method infers a set of descriptive parameters that indicate, in particular, the sleep debt ‘repaid’ by SWA generated during the sleep episode. As a result, one can obtain the starting levels of $S$ at sleep onset and the final levels at sleep termination. Given multiple EEG profiles, recorded simultaneously from different locations, the question arises whether there is any uniformity to be anticipated in the starting and ending values of $S$? Considering the accumulating evidence for local aspects of most processes occurring in the brain, this is not extremely plausible. If there is no uniformity, then the two process model can no longer predict a single timepoint for sleep initiation and termination, as a single $S$ is required to enter into interaction with Process C and trigger the state change.

Indeed, the answer is negative. In Chapter 5 we report on a multiple-location EEG study aimed to estimate the distribution of the gain constant over 26 locations on the scalp. There is no uniformity in SWA dynamics across the brain, and consequently, the original
1.1. HUMAN SLEEP TIMING

definition of Process S of Borbély and Daan should be distinguished from its close descendant, Process S of Achermann. We denote the latter as ‘Process Z’ in the remainder of this thesis. Process Z describes the local cortical changes in the rise and fall of NREM slow wave activity across the sleep wake cycle. Process S describes, as was originally conceived in the S–C model, the global role this homeostat plays in controlling the spontaneous human behavioural decisions of sleep onset and sleep end.

1.1.4 Challenging sleep homeostat

In Chapter 6, an attempt is made to further probe the properties of the sleep homeostat using Achermann’s method, but now under conditions of sleep deprivation. The advantage of a sleep deprivation protocol is that the experiment includes substantial variation in wake intervals, i.e., about 16 hours of wakefulness prior to baseline sleep as compared to about 40 hours of wakefulness prior to recovery sleep. This variation allows a better estimation of the rise rate of Process Z than under normal (baseline) sleep conditions, in which wakefulness is limited. In this study, we confirm the conclusion of Chapter 5, namely, that SWA regulation is not uniform across the scalp.

In Chapter 7, I summarize and discuss the role of the different elements of the Circadian system and Sleep homeostat in the timing of sleep in different chronotypes on the basis of the results of this thesis.