Chapter 1

What Is Sleep?

Introduction

Our patients are engaged in a struggle for sleep, and they pit their wits against wakefulness. They find sleep hard to initiate, and hard to sustain. Their quality of sleep is typically poor, and they suffer also in their waking lives from the effects of insomnia. Yet, in truth, all they crave is normality. They want to sleep well. So, what is this thing that eludes them so steadfastly, yet seems inadvertent to the good sleeper? What is sleep? In this chapter we will explore what the Israeli scientist Dr Peretz Lavie, in his semi-autobiographical book, called *The Enchanted World of Sleep* (1996). For me his book title captures the nature of sleep. Sleep is rich, diverse, and precious, as well as fascinatingly mysterious – not something you would want to miss out on! So, as you set out to help your patient with their insomnia, it will be useful to start your journey with an understanding of what sleep is.

Sleep and Wakefulness

What do you think of when you hear the word 'sleep'? What does it mean in everyday language? I was curious to see what the dictionary had to say, and although a professor at the University of Oxford, perhaps I felt some loyalty to our 'rival' institution because Cambridge University Press is the publisher of this book! Anyway, I looked at the Cambridge Dictionary Online (2024) and found the noun 'sleep', succinctly defined as 'not awake'. Consistent with this I found that 'wakefulness' is defined as 'the state of being wakeful . . . not asleep or (not) able to sleep'. I think this confirms what most people would expect: sleep and wakefulness are pretty much opposites of each other. You are either asleep or awake, are you not?

Well, scientific understanding of sleep and wakefulness is rather more complex than this, and much more intriguing. Sleep is not simply the absence of wakefulness. Falling asleep is not like having your light switch in the 'off' position, compared with wakefulness being like a light that is switched 'on'. The on/off idea would suggest that we live our lives either in one category or the other. This is not correct because there is variation within sleep, just as there is variation within wakefulness. You are not always 'wide awake', are you? Similarly, you are not always 'fast asleep'. Let me give you some examples that illustrate the non-binary nature of sleep and wakefulness.

Let's consider *sleepwalking*. The sleepwalker gets out of bed, walks around the house, with eyes open, but is still asleep. The sleepwalker can see, can navigate, can perform routine activities, and of course can walk. A classic single-photon emission computed tomography (SPECT) study by the Swiss neurologist Claudio Bassetti demonstrated

convincingly that sleepwalking arises from activation of thalamocingulate pathways while there is simultaneous and persisting deactivation of other thalamocortical arousal systems (Bassetti et al., 2000). This represents a true dissociation between what this team called 'body sleep' (where the motor system is active) and 'mind sleep' (where the frontal executive of the brain is offline). These findings have been confirmed more recently and point to blurred boundaries between wakefulness and the deepest portions of sleep (Januszko et al., 2016). The sleepwalker is manifestly awake and asleep at the same time.

Another example would be 'hearing' while you are asleep. Think of responding to your baby crying at night. This can cause disputes at home, as Mum always seems to be first to react! Indeed, there appears to be a basis to gender-dependent modulation of brain responses to infant cries (De Pisapia et al., 2013), and basic science suggests that the neuropeptide oxytocin increases the salience of acoustic stimuli in maternal mice (Marlin et al., 2015). However, I digress slightly. My point is that the brain may wind down when asleep, but it by no means loses all responsiveness. We remain attentive to information that is personally relevant to us while we are asleep.

Event-related potentials (ERPs) have been used to investigate the ability of the sleeping nervous system to monitor and respond to the external environment (Colrain and Campbell, 2007). ERPs quantify the neurophysiological activities elicited by the presentation of auditory stimulation (e.g., noises, words) and do not require any behavioural responses or conscious awareness. They have proven to be a useful technique to study information processing during sleep, showing that there is a stronger brain response when we hear something that we would normally react to, for example our name, an infant cry, or anything that might reflect a sense of danger. Isn't it fascinating to realise that you don't need to be awake for your brain to monitor and filter crucial information?

Sleepwalking and selective hearing during sleep are examples of two typical wake behaviours (walking and listening) overlapping with sleep, but the inverse is also true: sleep may also occur while we remain awake. One of my colleagues at Oxford, Vlad Vyazovskiy, has conducted studies in laboratory animals demonstrating the presence of *local sleep* during periods of wakefulness when the animals were sleep-deprived (Vyazovskiy et al., 2011). His research team found that subsets of neurons, even neurons within the same cortical area, entered periods of sleep while overall brain activity remained typical of wakefulness. The animals acted as if they were awake. They had their eyes open and were responsive to stimuli, but they were also partly asleep at the same time. These findings that local sleep may occur in a sleep-deprived but awake brain may account for the cognitive impairments that we see in people who are sleep-deprived (e.g., Van Dongen et al., 2003; Tononi and Cirelli, 2006; Hudson et al., 2020). It is essentially synaptic overload.

So, it seems that sleep and wakefulness sometimes blend, and at times this wake–sleep or sleep–wake dissociation can be adaptive. Mothers (especially) appear to have a built-in baby monitor that remains switched on while they get some sleep, and in the animal kingdom, creatures including dolphins can swim and birds can fly when parts of their brain get their sleep (Pigarev et al., 1997). Of course, the sleepwalker's dissociation is more commonly maladaptive, and this is one reason why it can become a disorder of sleep.

Blurred Boundaries between Sleep and Wakefulness in Insomnia

Surely people with insomnia couldn't be awake and asleep at the same time?! Well, we have known for 40 years that insomnia patients report that they are awake even though

laboratory measurement indicates they are asleep (Borkovec et al., 1981). What is going on? Are they exaggerating their complaints because of their frequency, severity, persistence, and impact? Are they misperceiving their sleep and making attributional errors about their sleep and its consequences? Is insomnia primarily a psychological disorder, or is it a psychophysiological disorder? A lot of time and research effort has been put into understanding what is often called the *subjective–objective discrepancy* in insomnia. Those with insomnia do have poorer objective sleep. They take longer to get to sleep, spend longer awake during the night, and sleep less than good-sleeper controls, but the differences are relatively small and do not fully account for the complaints and distress (Edinger et al., 2000). Perhaps they are also awake while they are asleep?

Investigations of various ERP component signals suggest that compared with good sleepers, those with insomnia have difficulty both disengaging from wake processes and in initiating sleep processes (Yang and Lo, 2007; Bastien et al., 2008), which suggests greater ongoing information processing while asleep. Moreover, hyperactivation and inhibition deficits have been directly associated with poorer reported sleep quality (Devoto et al., 2005; Turcotte and Bastien, 2009). Studies using *power spectral analysis (PSA)* of the *electroencephalogram (EEG)* (Perlis et al., 2001b; Krystal et al., 2002; Perrier et al., 2015) and a brain imaging technique called positive emission tomography (PET) (Nofzinger et al., 2004; Kay et al., 2017; Hsiao et al., 2018) similarly suggest that people with insomnia have reduced down-regulation of their brain arousal level during sleep compared with good sleepers. In other words, combined with ERP data, these fine-grain measures of brain electrical activity (PSA) focusing on high-frequency EEG activity and of brain glucose metabolism (PET) indicate that while they are asleep, people with insomnia remain relatively responsive, hyperaroused, and hypermetabolised. This is consistent with a neurocognitive perspective that suggests that the patient with insomnia maintains a level of processing that blurs the distinction between sleep and wakefulness and influences their experience and reporting of sleep (Perlis et al., 1997).

The Process of Falling Asleep

Can you tell if someone is asleep? At the observable level a particular body posture (whether lying down or slouched) and persisting quiescence might suggest that a person has fallen asleep. Add to that an elevated arousal threshold, for example when you speak to them and they don't respond, coupled with a change in their behaviour on awakening when you raise your voice or nudge them might confirm they were in prior sleep. I'm sure these descriptions are familiar. Sleep, then, is a behavioural state that is associated with a relative lack of movement and response compared with the awake state. When we talk about falling asleep we are referring to a transition from wakefulness, not simply an event. If we examine that process in a little detail we will see how behavioural and physiological processes mirror one another.

Have a look at Figure 1.1. In this experiment participants were instructed to squeeze a ball in time with their breathing as they fell asleep. In the left panel you can see the task and the *electromyography (EMG)* system recording muscle activity from the hand and forearm (Prerau et al., 2014). In the upper part of the right panel (marked A) you see the amplitude of the EMG gradually reducing over time, until there is persistent response failure. Immediately below the EMG there are measures taken from an EEG recording system where electrodes on the scalp are used to quantify the brain's electrical activity.

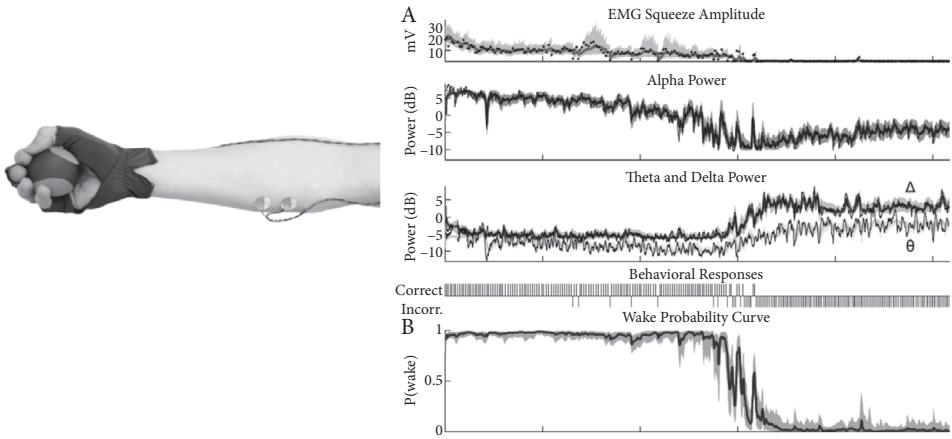


Figure 1.1 Changes in EMG and EEG activity associated with sleep onset (from Prerau et al., 2014, pages 4 and 8)

You can see the reduction in *alpha* (α) power, an EEG index of the waking brain when the eyes are closed, and an associated increase in the sleep EEG signals associated with *theta* (Θ) and *delta* (Δ). If you look now at B in the lower part of the right panel you can see that the investigators were able to use these dynamic brain and body changes to plot the participants' wake probability curve. Can you see the moment of sleep onset?

You probably noticed from this illustration that although complex physiology was measured, the presence and then the absence of the behavioural responses (the ball squeezes) corresponded quite reliably with the onset of sleep. I'm sure you can relate to this. Can you imagine someone falling asleep while they are sitting reading a book? As they fall asleep their head bobs or nods, and they lose their grip on the book. Very commonly, at the point of falling asleep the book falls only to wake them up momentarily until they decide to give up trying to read! Later, when we are talking about cognitive behavioural therapy (CBT), we will come back to this observation. It is a useful analogy.

The Stages of Sleep

We have been considering how sleep differs from wakefulness, although one sometimes intrudes upon the other; and we have been thinking through and visualising the process of falling asleep. I have already mentioned a few technical procedures and terms, so I think it will be helpful now to take a closer look at how scientists and clinicians study sleep in the laboratory. There are three principal measurements that help us to define the stages of sleep.

Electrical activity in the brain is measured by means of electroencephalography (EEG). In the sleep EEG, surface electrodes are attached to the scalp in central, frontal, and occipital positions according to the international 10–20 system of electrode placement (Jasper, 1958). Reference electrodes are sited on the mastoid bones located behind each ear, and the EEG is the primary variable to document wakefulness, arousals, and sleep stages during the sleep study. I have also referred to electromyography (EMG). The recording of EMG activity under the chin is used for determining the level of muscle tone and provides supplemental information regarding patient movements and arousals. EMG may also be helpful in distinguishing artefact in other channels. In clinical studies EMG measurement applied to the legs can help evaluate the presence of limb movements

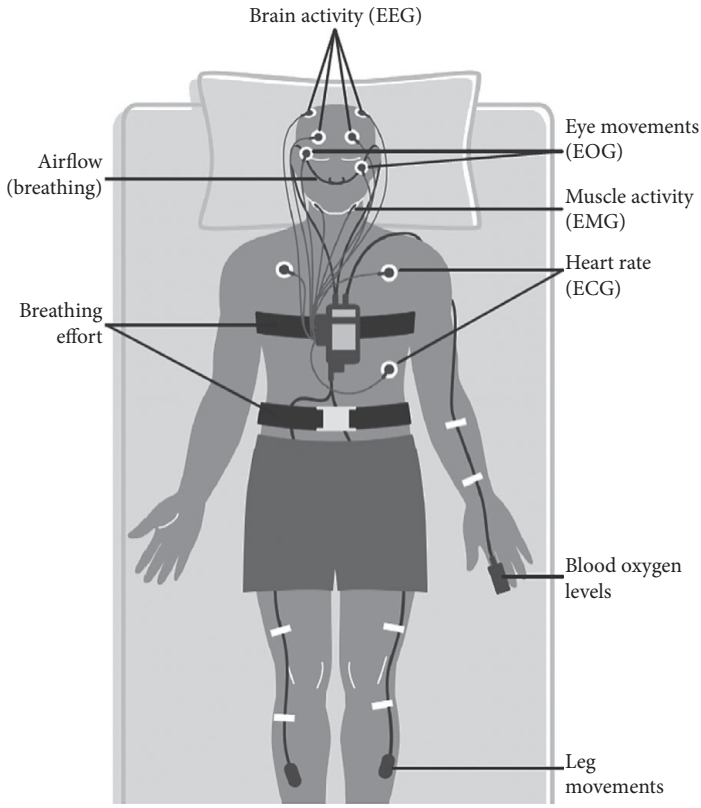


Figure 1.2 A typical PSG set-up
(reprinted from Espie, 2021) Copyright © Liane Payne

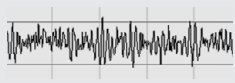
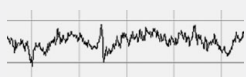
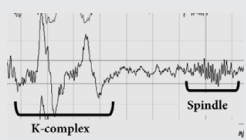
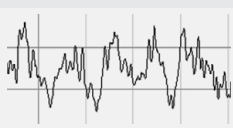
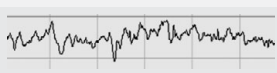
and thereby sleep-related movement disorder. The third standard sleep-recording measure is *electro-oculography (EOG)* where electrodes are positioned 1 cm above or below the outer canthi, that is, adjacent to where the eyelids meet. EOG recording aids the identification of sleep onset by monitoring for slow, rolling eye movements that occur with transition from wakefulness into early sleep, but also helps with identification of the rapid eye movements that we shall see shortly are characteristic of REM sleep. Finally, an evaluation of respiratory airflow and respiratory effort and pulse *oximetry* measurement (of blood oxygen levels) provide information on breathing during sleep and help detect sleep-related breathing disorders (SRBDs). Figure 1.2 shows a typical *polysomnographic (PSG)* set-up comprising these many (poly) channels of information.

Non-Rapid Eye Movement (NREM) Sleep

When we fall asleep we enter *non-REM sleep*, typically abbreviated to NREM sleep; so we will consider NREM first and then turn to thinking about REM sleep.

Table 1.1 starts with a representation of the waking EEG. Brain activity in the range of 13–30 cycles per second (or Hertz (Hz)), known as beta waves, was first described by

Table 1.1 PSG waves, characteristics, and features of each stage of sleep (courtesy of Alasdair L. Henry; EEG waveform images provided courtesy of Rachel Sharman)

State	EEG waveform	Frequency (Hz)	Proportion of night (approx.)*	Characteristics
Attentive wakefulness		13–30 (beta)	5%	Open eyes/ voluntary rapid eye movements
Restful wakefulness		8–12 (alpha)		Blinking/slow eye movements, voluntary muscle movement and tonic activity
Stage 1 (N1)		3–7 (theta)	5%	Decrease in tonic activity and slow rolling eye movements, hypnic jerks
Stage 2 (N2)		3–7 (theta) with k-complexes and sleep spindles (rapid bursts of 12–14 Hz)	50%	Low-level tonic activity
Stage 3/ SWS (N3)		0.5–2 (delta)	15%	No eye movements, low-level tonic activity
REM		13–30 (beta)	25%	Rapid eye movements, muscle atonia, dreaming mentation

* *Note*: these proportions are based on a healthy, middle-aged adult.

Hans Berger and was associated with focused attention when the eyes are open. Beta rhythms (β) are sometimes subdivided into beta1 (13–20 Hz) and beta2 (21–30 Hz), and the even higher frequency *gamma activity* (γ: 30–60 Hz) is associated with deep concentration. For our purposes here, my point is that alert wakefulness is associated with waves

that are low amplitude (height) and generated in close proximity (high frequency). Sometimes we refer to this as fast EEG activity.

Next in Table 1.1 we see the alpha rhythm (α), the EEG signature of 8–12 Hz that we have already discovered is characteristic of contemplative or quiet wakefulness when the eyes are closed. In Figure 1.1 this was presented as alpha power to demonstrate the dynamics of how wake signals gradually give way to sleep signals as sleep onset approaches. The typical progression is for the alpha waves to become more intermittent, then for the EEG to flatten and ripple before the emergence of stage 1 sleep (Hori et al., 1994). NREM stage 1 (or N1) sleep is not only a transitional phase between wakefulness and sleep but also a transitional stage before progressing into the second stage of sleep. In N1 sleep, *vertex sharp waves* and incomplete sleep spindles emerge, and EEG activity slows to around 3–7 Hz (see the N1 panel in Figure 1.1). These are the previously mentioned theta (Θ) waves. Muscles begin to lose their EMG tone (that is, they relax) in comparison with wakefulness, and the EOG begins to show slow rolling eye movements.

I trust you are getting the sense of the dynamic interplay between sleep and wakefulness? It is useful to keep in mind a balance between the utility of *sleep-staging*, where we allocate a piece of recording to a given stage (Table 1.1), and the more quantitative EEG approach, illustrated by PSA (Figure 1.1) where signals may coexist within the EEG. Before moving on let me give you a further example that I'm sure you will understand from your own experience. It is well established that under conditions of *sleep deprivation* while individuals are attempting to remain awake, EEG power density in the theta/alpha range increases significantly and is correlated with self-rated fatigue (e.g., Cajochen et al., 1995). That sense of zoning in and out of wake tasks and drifting in and out of sleep is very familiar to us when we lack sleep.

Looking again at Table 1.1 you will see that in N2 (stage 2) sleep the EEG varies quite considerably. These are what we call mixed-frequency EEG waves (some fast, some slow, some high amplitude, some low). However, there are two characteristic formations that occur repeatedly, and these are the defining features of N2 sleep. The *K-complex* takes its name from the shape of an initially descending and then ascending sharp change in voltage. If you are trying to make sense of this by following the graph, you will need to know that by tradition upward deflections of EEG are 'descending' and downward deflections are 'ascending' – just to confuse you! The other features of N2 sleep are known as *sleep spindles*. This is a descriptive term for rapid bursts of high-frequency EEG activity (12–14 Hz) that occur intermittently, sometimes just after a K-complex. Although N2 comprises the largest proportion of adult sleep (50–60% of nightly sleep), the very first phase of the night's N2 sleep is usually quite short. Nevertheless, N2 is relatively consolidated sleep, and if you are woken out of N2 (unlike N1) you are much more likely to report that you were asleep and have just been woken up. Interestingly, however, and consistent with the possibility that people with insomnia are more 'aware' of processing information during sleep, when woken from early N2 sleep, people with insomnia were more likely to report still being awake and to describe more mentation than their good-sleep counterparts, in some but not all studies (Borkovec et al., 1981; Mercer et al., 2002; Feige et al., 2021).

We have the deepest part of our sleep, from which it is hardest to waken, during the first couple of hours of the night when there is a rapid transition into NREM stage 3 sleep, now referred to as N3 sleep. It seems like the brain wants to drive us into this *slow wave sleep (SWS)* as quickly as possible, and this reflects a homeostatic function that we will return to shortly. If you look at N3 sleep in Table 1.1 you will see that the EEG trace now reveals higher-amplitude waves occurring at much lower frequency. In other words, the waves have slowed down; they are high rollers! We call these delta (Δ) waves. Their amplitude will be 75 microvolts (μV) or greater, and the wave frequency has now dropped to its lowest at 0.5–2 Hz. SWS is also a form of synchronised sleep because the brain's electrical activity settles to a harmonised rhythm and produces the steady 'beats' that you can see in Table 1.1. It is worth bearing in mind that the use of the term 'deep sleep' as a synonym for SWS is quite common, and it is not necessarily incorrect. However, many people, including our patients, think that deep sleep sounds like exactly the kind of sleep they want and are not getting! In reality SWS only occupies a relatively small proportion of the night even for healthy, good sleepers, and it may be helpful to discuss with your patients that all stages of sleep are needed for a balanced sleep diet.

Rapid Eye Movement (REM) Sleep

So far, then, we can see that the transition from wakefulness through to N3 sleep involves not only a diminution of consciousness and of responding but also a steady change in the EEG wave pattern from fast to slow activity, and that several stages of NREM sleep (N1, N2, N3) can be differentiated. However, in 1953 two researchers in Chicago, Dr Nathaniel Kleitman and his assistant Dr Eugene Aserinsky, made a crucial discovery. They noticed that there was another form of sleep during which the eyeballs moved rapidly, whereas the rest of the body was atonic, that is, effectively paralysed (Aserinsky and Kleitman, 1953). The term 'rapid eye movement sleep (REM sleep)' was coined, and so important was its discovery that the previously described group of stages of sleep actually became known for what it was not – non-rapid eye movement sleep.

Let's move on now to Figure 1.3, which shows a recording of REM sleep from our sleep laboratory. The first thing to do is to examine the three EEG channels. As you can see, the basic PSG is set up to measure EEG from frontal (F4), central (C4), and occipital (O2) electrode sites (see also Figure 1.2). Notice the low-amplitude, mixed-frequency EEG *sawtooth waves*, resembling the blades of a saw, that can be observed. REM sleep is a light form of sleep, and indeed is not too dissimilar to wakefulness (compare it with Table 1.1). Importantly of course, there are characteristic eye movements during REM sleep, and you can see these in the left and right eye EOG traces at the top of Figure 1.3. The eye movements are more similar to those that occur in wakefulness, moving around rapidly in a range of directions, but of course the eyelids are closed, and the movements don't send any visual information to the brain. You can also see the chin EMG trace at the bottom of the figure, illustrating very low muscle tone, or motor atonia, interrupted only by occasional phasic muscle twitches. Because REM sleep is marked by a waking-like EEG pattern coupled with atonia, REM is sometimes aptly referred to as 'active sleep' or 'paradoxical sleep'. Peever and Fuller (2017) provide an excellent overview of the biology and functions of REM sleep.

It seems reasonable to think of REM sleep as an awake brain in a paralysed body, and this makes sense when you think that it is during REM sleep that we do most of our

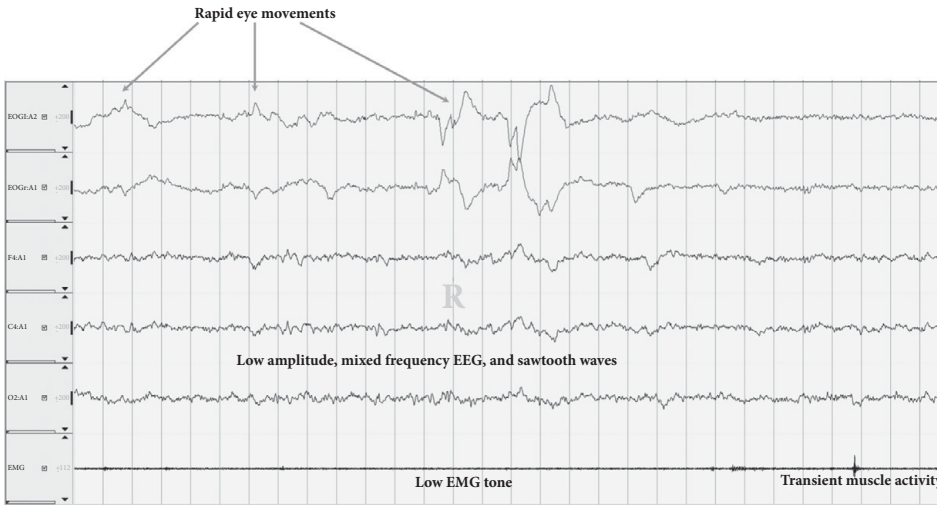


Figure 1.3 PSG output demonstrating REM sleep (annotated here as R) on EEG, EMG, and EOG

dreaming (Schwartz and Maquet, 2002). Indeed, were it not for the fact that our major voluntary muscles are out of commission, we could easily injure ourselves by acting out our dreams, particularly if they are nightmares. In fact, if this does occur it may mean that the person has a problem known as *REM sleep behaviour disorder (RBD)*, which we will discuss later. It is also worth noting that RBD is not the same as sleepwalking, when people clearly do move around, because sleepwalking, as we have seen, occurs during partial arousals out of NREM sleep.

Before we leave this introduction to REM and NREM sleep, it is important to recognise that there are also marked differences between wake, REM, and NREM sleep in levels of global and local brain activity. Much of what we know here has been discovered through functional neuroimaging studies, particularly using PET and functional magnetic resonance imaging (fMRI). Research over the past decades has consistently found a drop in brain activation during NREM sleep compared to wakefulness, particularly in subcortical and cortical regions (e.g., Maquet, 2000), whereas glucose metabolism in REM sleep shows a global level of activity that is not significantly different from wakefulness, with several brain structures actually enhancing their activity compared to waking (Dang-Vu et al., 2010). I told you sleep was fascinating!

The Architecture of Sleep

I sometimes ask patients to draw what they think a night of sleep looks like, starting at the time they first fall asleep until they finally wake up and get out of bed. Most people instinctively want to draw a U-shaped curve, but qualify this by saying they know that there are different bits to sleep, some lighter and some deeper. So, they end up drawing a modified 'U' with a spiky base. It's a useful exercise, and it can prompt you to explain about NREM and REM sleep cycles across the night.

The *hypnogram* in the left panel of Figure 1.4 illustrates the distribution of different stages of sleep in a typical night in both a younger adult (panel A1) and an older adult (panel A2). I use this to explain the architecture of sleep by standing back to see the big

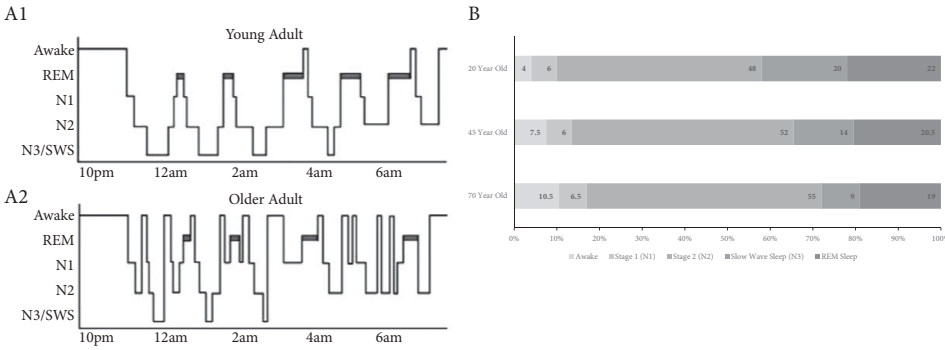


Figure 1.4 Hypnogram of sleep in a young adult (A1) and an older adult (A2); panel B shows the distribution of different stages of sleep in 20-, 45-, and 70-year-olds
(from Scullin and Bilwise, 2015)

picture as it were, and the skyline or landscape of the night emerges. First, on the vertical y-axis we have wakefulness and types of sleep; and second, on the horizontal x-axis we have time. It is important to recognise that the time course is notional, and if you use this illustration with your patient, please bear this in mind. When a person goes to bed, and gets up, the exact amounts of different types of sleep they get and their total sleep time will be quite individual to them. However, there is a recognisable pattern that remains true for most people. Notice the rapid descent into N3 when you first fall asleep. The greatest amount of NREM SWS is obtained in the first third of the night, with the greatest portion in the very first cycle. This helps explain why people often feel they have been in a deep sleep if woken up early in the night, because they have been! It is also why we may feel quite refreshed after a relatively short sleep.

N3 sleep, however, reduces with age as a proportion of the night. In panel B of Figure 1.4 the typical reduction illustrated is from around 20% in a young adult to less than 10% in an older adult. In contrast, REM sleep is relatively preserved with age, occurring in three to five cycles, the first of which is approximately 70–90 minutes after falling asleep (Dement and Kleitman, 1957). REM episodes also tend to get longer and are of increased density towards the morning, consistent with the common experience of drifting in and out of dream states before rising. Comparison of the hypnograms of the younger and older person also show more sleep fragmentation in later life associated with periods of wakefulness. The net effect is that as a developmental norm, the sleep that we do get tends to be lighter (more N2) and more broken as we get older.

Patients usually find these illustrations and learning about sleep parameters helpful. They are often surprised to see that it is normal for lighter sleep stages such as N1, and N2 in particular, which occupies about 50% of the night, to be so prominent (panel B, Figure 1.4). However, they will be quick to point out that they take longer getting to sleep, wake up more often, and struggle to return to sleep, as well as get less sleep than the chart suggests. This is important to acknowledge. However, it is also reassuring to explain that the brain's ability to generate sleep stages and NREM/REM cycles is not fundamentally damaged in insomnia. Although the sleep pattern of the person with insomnia has become troubling, disordered, and unpredictable from night to night, they can still generate sleep. You can also provide some insight about being more wakeful during sleep, and this in my experience chimes with their experience of 'not being

properly asleep' even when they are sleeping. As we shall see the behavioural therapeutic components of CBTx are particularly effective in re-establishing the drivers and boundaries of the underlying sleep need.

There is also limited evidence for sex differences in the broad landscape of REM and NREM sleep or in the distribution of sleep stages across the night. However, women are more likely to experience sleep disruptions and insomnia across their lifespan, and sex differences may underlie the differential risk for and the reporting of sleep disorders (Mong et al., 2011; Mallampalli and Carter, 2014; Won et al., 2020). Sleep in women is particularly impacted by hormonal effects such as during the pre-menstrual week, pregnancy, and the various phases of the menopause. There are also often difficulties in regulating body temperature that result in downstream impacts on sleep. Hormonal and thermoregulatory factors and their wider impacts on physiology and psychology, therefore, often contribute to sleep disturbances or sleep disorders for many women (Shechter and Boivin, 2010; Boivin et al, 2016).

The Regulation of Sleep

We now have a better understanding of what sleep is, its phases and stages, the characteristics of how it presents across the night, and the differences in sleep that we may observe depending on sex, age, and stage of life. Further information on such developmental factors, as related to the presentation of insomnia, may be found in Chapter 3. However, it is important at this point to ask the question, How is sleep regulated?

Biological Rhythms

We have been thinking about cycles and rhythms. The NREM–REM sleep sequence that occurs numerous times during the night is an *ultradian rhythm*, that is, one that has a cycle length of less than 24 hours. Other examples include pulse rate, appetite, and bowel movements. When we consider the regulation of the sleep–wake cycle the term that we use is *circadian rhythm*, because this process recurs naturally on an approximate 24-hour cycle (this has a Latin origin: '*circa diem*', meaning around the day). Often the word circadian is used to refer to the biological clock or the body clock; however, the circadian rhythm is really an effect of the biological clock, because not all biological clocks are circadian. Indeed, biological rhythms that have a period length greater than 24 hours, such as menstruation, hair growth, and seasonality, are known as *infradian rhythms*. I am pleased to recommend books by my Oxford colleague Russell Foster as brief and readable introductions to our fascinating biological clock systems (Foster and Kreitzman, 2009, 2017; Foster, 2022).

Interacting Homeostatic and Circadian Alerting Processes

It is the circadian rhythm that aligns our sleep and wakefulness with night and day. During the day our relatively steady state of alertness is due to the circadian alerting system (known as process 'C') acting in opposition to the mounting pressure of the sleep homeostat (process 'S') that builds up hour by hour from the time we wake up until we sleep again at night. This is often known as the 'two-process model' of sleep regulation (Borbély and Acherman, 1999; Borbély et al., 2016). When the biological clock's alerting signal releases at night it then permits the onset of sleep. The synchronisation of the

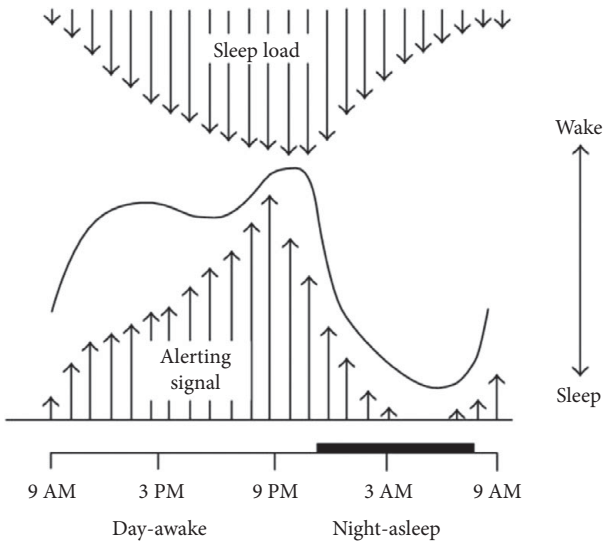


Figure 1.5 The regulation of sleep and wakefulness
(reprinted from Dijk and Edgar, 1999)

sleep–wake schedule and the internal clock is essential to an individual's ability to maintain sleep and wakefulness at predictable and regular times (Dijk and Czeisler, 1995), and this is one reason why sleep and circadian disruption are implicated in a range of disorders (as we shall see in Chapter 2), and not least in psychiatric illnesses (Wulff et al., 2010; Jagannath et al., 2013).

An engaging narrative approach to explain how sleep is regulated is very useful to our patients, while some also benefit from an illustration. I have reproduced in Figure 1.5 what I find to be an accessible representation that you can use. However, it is still a bit complicated, and not all patients want or need this level of detail. Let me talk you through it.

Broadly speaking, the longer we are awake, the sleepier we will become. Normal wakefulness, therefore, increases the homeostatic sleep drive and builds sleep load (process S). In physiology, homeostatic processes are there to restore balance, so sleep reduces the drive for sleep, and wakefulness increases the sleep drive. Much in the same way we become parched if we go without fluid, and homeostatic pressure, in this case thirst, increases as time without fluids increases. Drinking satisfies that thirst and so reduces the drive to drink. The famous sleep researcher, Dr William C. (Bill) Dement, from Stanford University, California, used the helpful analogy of the 'sleep economy' to explain this homeostatic mechanism. With each hour that we spend awake we accumulate an increasing sleep debt. In healthy, good sleepers this debt is then repaid in full by the night's sleep, and they awaken refreshed and back 'in balance' the next morning. Incidentally, it should also be noted that homeostatic drive is evident not just in relation to total sleep but also at the level of sleep stages, particularly REM sleep and NREM N3 sleep (or SWS) (Dijk and Czeisler, 1995). Consequently, a person who is sleep-deprived, or someone that has been on a drug that suppresses a component of sleep (commonly REM), will experience some increased drive for or 'rebound' of that lost sleep during the early hours of recovery nights.

But why do we not feel sleepier and sleepier as the day goes on? This is where an alerting signal to keep us awake during the day comes in. This circadian component

(process C) acts as an opponent to the sleep drive in the daytime, keeping sleep and sleepiness at bay, until at night-time it stops its opposition and opens the sleep gate to allow the accumulated drive for sleep to rush through. During the night the sleep load then reduces because the debt to sleep is paid off, and the alerting signal strengthens again in the morning in readiness to keep us awake once more. Of course, you will see in Figure 1.5 that there is some natural variation in our daytime wakefulness, and how alert we feel, and you will probably be aware of the afternoon circadian dip when we tend to feel temporarily rather more tired. Indeed, in some societies it is normal to have a 'siesta' at this time because it also coincides with the hottest part of the day. Notice also the increased alerting signal that precedes the sleep period. This is sometimes referred to as the *wake maintenance zone* (Lavie, 1986). It is a window of around two hours prior to the onset of *melatonin* secretion when the circadian system must put extra effort into keeping us awake because of the peak in sleep load.

The Neuroscience of the Sleep and Circadian System

We have been learning some neurophysiology, and now it's time to dip into some neuroscience. I did say that sleep is enchanting, but we can't get away from the fact that underlying the apparent simplicity of the sleep–wake cycle that we so much take for granted, there is considerable complexity.

The discovery that mice lacking the retinal rods and cones necessary for sight nevertheless could regulate their circadian rhythms led to the identification of specialised cells in the retina containing melanopsin that were controlled by several genes (Foster et al., 1991). These photosensitive retinal ganglion cells respond to light, particularly the blue part of the spectrum, and communicate with the master circadian pacemaker located in the suprachiasmatic nuclei (SCN) of the anterior hypothalamus (Hankins et al., 2008). It turns out that we need our eyes not just for our sense of sight but for our sense of time.

Different analogies have been used to describe how the SCN operates, for example as the 'ringmaster of the circadian circus' or the 'conductor of the circadian orchestra'. However, either way, the SCN clock turns systems off and on in a way that is integrated with external stimuli, creating the (circus) performance or (orchestral) symphony that we require for our biological and social functioning. But how does our biological clock actually tick? The SCN clock comprises multiple, single-cell circadian oscillators that coordinate outputs to regulate our biological rhythms. Specific clock genes, regulated by their protein products, are involved in numerous complex interactions (called transcriptional-translational feedback loops, or TTFLs) that represent the detailed mechanics inside our molecular clockwork (Reppert and Weaver, 2001). We are able to fall asleep or to snap out of sleep into wakefulness because of a two-way inhibitory relationship between the brain's arousal-promoting and sleep-promoting pathways. This creates the conditions for the so-called flip-flop switch that generates rapid and complete transitions between waking and sleeping states (see Saper et al., 2005 and Saper and Fuller, 2017 for a thorough review). On the 'sleep side' of the switch is the ventrolateral preoptic nucleus (VLPO) in the anterior hypothalamus that sits adjacent to the optic chiasm, and on the 'wake side' are the tuberomammillary nucleus (TMN) neurons in the posterior hypothalamus and brainstem arousal regions. In a similar way, populations of mutually inhibitory neurons in the upper pons form a switch for controlling transitions between REM and NREM sleep.

Like any timepiece the circadian clock needs to be 'set' or 'synchronised'. We call this process *entrainment*. The daily light–dark cycle is the most powerful stimulus for entrainment in our environment. However, other *zeitgebers* (time-givers) such as work demands and planned activities (including mealtimes and exercise) may provide social cues that influence entrainment. Of course, biological and psychosocial influences may also interact and impact entrainment. This is certainly relevant for insomnia (Lack and Wright, 2007; Lack et al., 2008), and we shall return to the role of circadian processes in assessment and treatment in Sections 3 and 4. A well-entrained circadian clock ensures that expressed rhythms in physiology and behaviour are coordinated to the 24-hour day. The hormone melatonin is the best marker of the positioning of our sleep–wake timing (e.g., Pandi-Perumal et al., 2007) because it provides reliable information on what we call the photoperiod (day length) for the organisation of our seasonal physiology (Arendt, 2019). Melatonin is produced in the brain, in the pineal gland. Its production rate is dictated by natural light, so that during hours of darkness (the normal sleep period) melatonin production increases, and as morning approaches and with the coming of daylight, melatonin production is once again shut down. One way to consider melatonin is that it serves as a 'circadian glue' because it acts as a coupling agent to maintain stability in the sleep–wake system.

It is worthy of note that clock influences are important too for the wider field of mental health (see Harvey et al., 2021; Meyer et al., 2022). Consider depression, where withdrawal from social structure and tasks, coupled with reduced light exposure, may compromise timing of the biological clock. For example, disruption of social and biological rhythms might be responsible for triggering the onset of depressive episodes in vulnerable people (Ehlers et al., 1988; Grandin et al., 2006). Asarnow et al. (2014) have suggested that sleep and circadian science may serve as a foundation for behavioural interventions in mood disorders, and in my view *behavioural activation* therapy (Richards et al., 2016) seems the most obvious place to start.

As we can see in Figure 1.6, however, the clock is not just a resident of the central nervous system, with an address in the SCN of the anterior hypothalamus! So-called peripheral clocks that are influenced by the autonomic nervous system are found throughout the tissues and body organs, including the liver, pancreas, kidney and heart (Richards and Gumz, 2012). These molecular clock systems drive the circadian expression of specific genes involved in a variety of physiological functions. As you can see in Figure 1.6 the circadian rhythm of core body temperature is almost the inverse of the melatonin rhythm, with temperature reaching its nadir, or lowest point, soon after melatonin reaches its peak. Many measurable functions such as blood pressure and urine volume reduce during sleep, whereas growth hormone is expressed in the early part of the night, particularly during SWS, and cortisol expression reduces in advance of the sleep period and accelerates during the latter part of the sleep period in preparation for waking.

Of course, Figure 1.6 provides a picture of harmony, but we all know what the cacophony of discordant music sounds like. The conductor of the orchestra can struggle to keep all the players in tune and in time! This is what people feel like when their sleep–wake system is out of sync with local time, or when their various clock systems are out of sync with each other. The common experience of *jet lag* is one example, and this occurs when there is *internal desynchrony*, that is, the various clock systems are out of step with each other. I'm sure we have all had this experience, and it is interesting to recognise, for example, that our hunger drive might recover at a different pace from our temperature

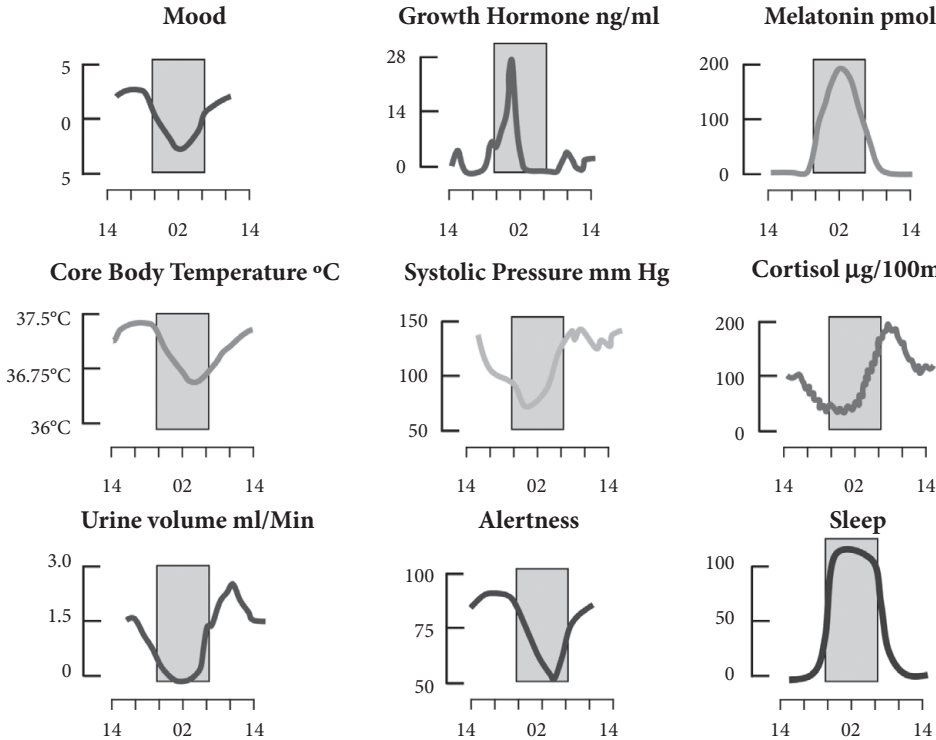


Figure 1.6 Circadian rhythms of a range of physiological and psychological functions (provided by Russell Foster, University of Oxford)

rhythm or sleep–wake schedule, leaving us with a feeling of general malaise that can last for some time depending on the number of time zones we have passed over.

Finally, let me add that the molecular clock is fairly omnipresent at cellular level. Research has demonstrated that if we take a single cell – yes, one solitary cell – and put it in an observation dish with sufficient energy to retain life and function, then that single cell will retain a 24-hour sleep–wake rhythm all on its own (Richards and Gumz, 2012). I find that pretty mind-blowing. The body clock is absolutely not just a figure of speech, but is an integral part of our very being. We need our sleep, and we need it well timed. In the next chapter we will find out more about what sleep does for us.

Narrative Summary

As you have read, sleep is a complex process that is by no means an inactive or static state. Rather, it is a dynamic process that can be measured across behavioural, physiological, and cognitive domains, and the boundaries between sleep and wake are often blurred, not least in the case of insomnia. I hope that you now have a solid understanding of what sleep is, and how it is regulated in association with the circadian system. This will be helpful background knowledge when you assess and treat your patients with insomnia. We will certainly build upon this grounding in subsequent chapters of the book.