Neuroscience and Biobehavioral Reviews 34 (2010) 108–118

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Sleepiness as a need for sleep: When is enough, enough?

Jim Horne *

Sleep Research Centre, Human Sciences, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK

ARTICLE INFO

Article history:
Received 19 July 2009
Accepted 20 July 2009

Keywords:
Sleep
Sleepiness
Multiple Sleep Latency Test
Psychomotor Vigilance Test
Sleep need
Sleep debt
Statistical significance

ABSTRACT

This review considers the relationship between sleep need and sleepiness. In healthy adults, objective measures of sleepiness (e.g. Multiple Sleep Latency Test; Psychomotor Vigilance Test) and subjective indices (e.g. Stanford Sleepiness Scale) often poorly inter-correlate and have been seen as orthogonal dimensions. This is perhaps not surprising as the methodology of these tests is quite different in, for example, their duration, testing environment, whether they are experimenter versus participant-paced, and the understanding and expectancy of participants. It is argued, here, that 'sleepiness', the 'propensity to fall asleep' and the 'need for sleep' are not synonymous, but qualitatively different. They may represent different positions on a dimension ranging from essential to non-essential sleep/sleepiness, and the position on this dimension is detected to varying extents by the different measures. As these tests can detect – and perhaps induce – levels of sleepiness which would be undetectable by, and of little concern to people under everyday situations, they can reveal a sleepiness having the potential to be misinterpreted as sleep debt.

Contents

1. Three dimensions? . . . 108
2. Measures . . . 109
3. Time and space . . . 109
4. Scaling of measures and individual differences . . . 111
5. The afternoon 'dip'—normal sleepiness? . . . 113
6. Recovery sleep . . . 113
7. Adaptation to less sleep. . . . 114
8. The underlying dimension: from essential to superfluous sleep and sleepiness? . . . 115
9. Semantics of sleepiness . . . 116
10. Conclusions—uncoupling sleep and sleepiness from sleep need . . . 117
Acknowledgements . . . 117
References . . . 117

> The subject of sleeplessness is once more under public discussion. The hurry and excitement of modern life is held to be responsible for much of the insomnia of which we hear; and most of the articles and letters are full of good advice to live more quietly and of platitudes concerning the harmfulness of rush and worry. The pity of it is that so many people are unable to follow this good advice and are obliged to lead a life of anxiety and high tension. Editorial. Br Med J, 1894, Sept 29 p. 279

1. Three dimensions?

Ostensibly, 'sleepiness', the propensity to fall asleep and the need for sleep, seem to be synonymous terms, with the most popular instruments for measuring sleepiness expected to be highly positively correlated with each other. But this is not always the case, as has again been demonstrated, most recently, by Franzen et al. (2008), following a one-night sleep loss study in healthy young adults. The investigators used physiological, psychological and subjective measures of sleepiness, being respectively, the Multiple Sleep Latency Test (MSLT, Carskadon et al., 1986), the Psychomotor Vigilance Test (PVT, Dinges and Kribbs, 1991; Lim and Dinges, 2008) and subjective scales. These

* Tel.: +44 01509 223004; fax: +44 01509 228480.
E-mail address: j.a.horne@lboro.ac.uk.

0149-7634/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neubiorev.2009.07.009

indices will be described more fully in the next section. The salient outcome was that the authors found these three groups of measurements (physiological, psychological performance and subjective) to reflect separate independent and orthogonal dimensions of sleepiness that were not equivalent, even though all show increasing sleepiness, to various extents. Similar disparities between these measures have also been reported by Van Dongen et al. (2003, 2004) and Frey et al. (2004), although there are clear intra- and inter-individual differences in the extent of this 'orthogonality' (Frey et al., 2004).

This review will examine how these frequently and well-cited methods for assessing sleepiness (i.e. MSLT, PVT and subjective sleepiness), are used and will argue that:

(i) Their apparent independence is largely an artefact of the method of presentation.
(ii) Instead, 'sleepiness', the 'propensity to fall asleep' and the 'need for sleep' are not synonymous, but qualitatively different, and represent different positions on a dimension which ranges from essential to non-essential sleep/sleepiness, which is detected to varying extents by these different measures.
(iii) Minor levels of sleepiness or small, albeit statistically significant changes to sleepiness, only evident under exacting laboratory settings, may be relatively inconsequential for everyday performance, especially as any apparent sleep loss underlying this particular sleepiness is acceptable to most people leading ordinary lives, who are not prepared to change waking habits to facilitate more daily sleep for what appears to be easily masked sleepiness and will offer only a small return in terms of waking alertness.
(iv) 'Recovery sleep', which is also used to gauge sleep need, also reflects a position on this 'essential' versus 'non-essential' dimension.

Sleepiness in the general population is a matter of current debate as, increasingly, the sleep literature points to an apparently widespread 'sleep debt' in western society. Seemingly, we are becoming sleepless, with many healthy adults having chronically insufficient sleep (e.g. Spiegel et al., 1999; Dinges, 2004; Dement, 2005), and who largely seem unaware of daytime sleepiness (Van Dongen et al., 2003). This review's perspectives on sleepiness also encompass issues relevant to sleep debt.

2. Measures

The MSLT and PVT are the most commonly used objective measures of sleepiness. Although there are other physiological indices, such as pupil dilation and blink rate (cf. Franzen et al., 2008), these can be difficult to measure and are seldom utilised. The MSLT consists of four, sometimes five, sessions over the day, commencing at 10:00 h and given at two-hourly intervals. For each session, participants, who have sleep EEG electrodes attached, retire to a quiet and dimly lit bedroom, and are told 'to lie down, relax, close your eyes and try and go to sleep'. Here, they remain for up to 20 min, or until the appearance of three consecutive epochs of stage 1 sleep or deeper (cf. Rechtschaffen and Kales, 1968) containing no more than 49% wakefulness per 30 s epoch, whichever event is the sooner. The session is ended, and latency to this sleep onset, from the commencement of the test until its termination is logged. If the participant fails to sleep then a score of 20 min is assigned. The four or five sleep onset values are averaged to give the overall MSLT score—the lower the score the greater the sleepiness, of course.

For the PVT, participants usually sit facing a computer screen, in a sound attenuated cubicle, devoid of distractions, and with their thumb or index finger of the dominant hand, push a button in response to a digital millisecond clock that appears with a random inter-stimulus interval of between 5 and 12 s. The clock stops and remains in view for 1–2 s, to provide the participants with feedback. Typically, the task lasts 10 min. The reaction time distribution changes as a result of decreases in vigilance. These changes are reflected in an increase in the number of responses with a latency greater than 500 ms (these are classified as 'lapses') and an increase in the mean reaction time of responses less than 500 ms. There is also a decrease in the fastest 10% of responses also, but this value (indicative of decreasing kutosis in the reaction time distribution) is seldom used because it is statistically less reliable unless the sleepiness becomes profound, following one or two nights of total sleep loss.

Subjectively, there are various sleepiness scales, with amongst the most common being the Stanford Sleepiness Scale (SSS, Hoddes et al., 1973) and the Karolinska Sleepiness Scale (KSS, Åkerstedt and Gillberg, 1990). The SSS is rather problematic, and this issue will be covered later. The KSS in its later modified form (cf. Horne et al., 2008) uses the following 9 point scale:

1 = extremely alert, 2 = very alert, 3 = alert, 4 = rather alert, 5 = neither alert nor sleepy, 6 = some signs of sleepiness, 7 = sleepy, no effort to stay awake, 8 = sleepy, some effort to stay awake, 9 = very sleepy, great effort to keep awake, fighting sleep.

There are also analogue sleepiness scales, having a 10 cm line anchored at either end by the terms 'very alert' or 'very sleepy', requiring a cross to be placed at the appropriate point.

As will be seen, all these measurement devices are, in their own ways, sensitive to low levels of sleepiness, whereas it usually takes at least 30 h of prolonged wakefulness in order to produce more subtle impairments to 'executive' function (affecting: decision making, working memory, perseveration, distractibility, word fluency, risk taking, euphoria, etc.); functions that are heavily reliant on the prefrontal cortex (cf. Harrison and Horne, 2000). As measures of executive function are insensitive to sleepiness per se, it is an area that will not be covered here. Also largely excluded, will be retrospective, subjective measures of chronic sleepiness, for example the Epworth Sleepiness Scale (ESS, Johns, 1991), that do not assess concurrent sleepiness, but requires respondents to think back over several weeks about instances of inadvertently falling asleep.

3. Time and space

Other than circadian factors, there are at least four methodological confounds affecting the sensitivity of each of these apparently orthogonal aspects of sleepiness, which might explain or at least contribute to the orthogonality: (i) duration of testing ('time on task'), (ii) whether the task is experimenter or participant-paced, (iii) the physical environment of the test, and (iv) instructions to the participant and their understanding of, and attitude towards the measurement tool. 'Individual differences' may well affect all these confounds.

Subjective scales are usually completed within a minute of the participant sitting down, unlike the 10 min for the PVT, that often follows a brief set of practice trials. The MSLT can last up to 20 min. So, for example, neither the PVT nor MSLT are useful in detecting sleepiness if only administered for a minute or two (i.e. a timescale which would be comparable to that required to administer a subjective scale). To some extent, detecting sleepiness with either method depends not just on time per se, but also on the element of tedium which develops during the course of measurement and when self-motivated 'alerting' wanes. Moreover, these objective tests have the further advantage that they are presented in quiet, non-distracting and relaxing settings, whereas this is often not the case with subjective measures. Unfortunately, few studies report