A sleep physiologist's view of the drowsy driver

Murray W. Johns [*]

Epworth Hospital, Epworth Sleep Centre, 187 Hoddle Street, Richmond, Melbourne, Vic. 3121, Australia

Abstract

Drowsy driving is dangerous because of the impairment of driving skills that it causes. Unfortunately, the conceptual basis that underlies much of the multi-disciplinary research on this topic is muddled. The same poorly defined terms, such as fatigue and sleepiness, are used differently by different disciplines and researchers. Some new definitions and concepts are proposed here which may be helpful, as least as a stimulus for discussion by others. Drowsiness, sleepiness and fatigue are distinguished. A new conceptual model of sleepiness is outlined, based on a mutually inhibitory interaction between a putative sleep drive and a wake drive. Sleepiness, defined as sleep propensity, is a function of the relative strengths, not the absolute strengths, of the sleep and wake drives. The measurement of sleepiness requires some new variables such as instantaneous sleep propensity, to be distinguished from either the situational or the average sleep propensity. A subject's instantaneous sleep propensity depends on many variables including his average sleep propensity in daily life, the time of day, the duration of prior wakefulness, the subject's posture, physical and mental activity at the time, and individual differences based on psychophysiological traits. The relationship between dozing at the wheel while driving and crashing the vehicle may not be as straightforward as it appears at first. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Driving; Drowsiness; Sleepiness; Vehicle accidents

1. Introduction

Drowsy driving is dangerous for us all, not just for drivers in the transport industry. Our knowledge of the physiology of sleep and wakefulness has increased rapidly in recent years and that knowledge needs to be incorporated into the multi-disciplinary thinking about drowsy driving. This is made difficult by the poorly defined concepts and terms, often used differently by different disciplines and investigators, even when used by the same investigator at different times.

[*] Fax: +61-3-9427-8522.
E-mail address: sleepsda@enternet.com.au (M.W. Johns).

r us to define what we think we are talking about, as I propose to do. I would
hers in this field to do the same.

· is in the state of drowsiness, the intermediate state between wakefulness and
ctro-physiologically by the pattern of brain waves (EEG), eye movements and
t what point drowsiness begins is still uncertain, but it is before the onset of
urrently defined, even before Stage-1 occurs very briefly as "microsleeps"
san, 1995; Tanaka, Hayashi, & Hori, 1996). Drowsiness is accompanied by
ological changes that are anathema to safe driving, such as the impairment of
the inability to maintain visually focussed attention, the impairment of higher
and of volition (Lamond & Dawson, 1999; Thomas et al., 1998). The level of
y fluctuates in the drowsy state. Although an optimum level of arousal of the
tem is necessary, it is not sufficient by itself for normal waking function and the
appropriate focus of attention, which can also be affected by many other
man, 1984).

ess is used here in the sense of sleep propensity, or the probability of falling
lar time, i.e., the chances of making the transition from wakefulness to the
o Stage-1 sleep, whether or not sleep progresses to other stages subsequently
this definition, sleepiness is neither a state nor a process. It is analogous to a
continuum of arousal states from alert wakefulness to sleep. Sleep propensity
hed from what some people call subjective sleepiness which is related to the
ty of a set of feelings and symptoms that accompanies drowsiness, i.e., it is a
an often be introspected. These concepts of sleepiness, sleep propensity and
tively new and, indeed, are still being developed in sleep medicine. They are not
other disciplines.

term fatigue has a much longer history, having arisen from applied psychology
f the twentieth century. Driver fatigue has long been cited as a causal fact in
es (Brown, 1994). However, there is still confusion about the several different
ord, fatigue. In a psychological and behavioural sense, fatigue can be seen as a
subjective experience of tiredness and a disinclination to continue performing
Brown, 1994). Alternatively, fatigue can be seen as a process involving "those
ges in the expression of an activity which can be traced to the continued exercise
artlett, 1953) or, put more simply, the changes in performance of a task over
s cognitive changes that may be caused by mental fatigue, analogous to the
ie of many (but not all) muscles, manifested by a reduction in their strength of
ontinued use. However, the neurophysiological mechanisms that cause mental
ar than they are for physiological fatigue.

quate fatigue with sleepiness, or at least have them as closely overlapping

sleepiness". However, we know that normal subjects can enter the drowsy state at almost any time by changing their posture, lying down and closing their eyes for 15 min. Patients with excessive daytime sleepiness caused by a sleep disorder can doze off while driving, soon after a night's sleep, without any sense of fatigue. Thus, the evidence suggests that sleepiness and fatigue should be distinguished, but there is confusion about both. More careful and broader conceptual analysis is required from each of the disciplines involved.

3. Measuring sleepiness

There is convincing evidence from sleep laboratory experiments that, when tested repeatedly under the same circumstances (e.g., lying quietly in bed), a subject's sleepiness, measured by the time taken to fall asleep, varies slowly with the time of day, usually being maximal at 3–4 a.m. – i.e., there is a circadian rhythm of sleepiness (Borbély, Achermann, Trachsel, & Tobler, 1989). Sleepiness also increases progressively, but not linearly, with the duration of prior wakefulness. There are mathematical models that describe these changes in sleepiness over periods of hours and days (Borbély & Achermann, 1992). However, Johns (1998) has criticised these models for their inadequate conceptual basis. He has pointed out that some of the most important influences on sleepiness are the subject's posture, and physical and mental activity at the time. Environmental stimuli also affect sleepiness – whether hot or cold, noisy or quiet, etc. On a more subtle level, many people find they are more likely to doze off (i.e., they have a higher sleep propensity or sleepiness) "when sitting and reading" than "when sitting inactive in a public place such as a theatre or meeting" (Johns, 1994). Thus, even minor changes in our posture and behaviour can influence our sleepiness. This is evident in EEG recordings made with the subject in different postures (Caldwell & Prazinko, 1999).

To help quantify a subject's sleepiness that is influenced by so many different factors we need to introduce some new concepts:

(a) *Instantaneous sleep propensity*: a subject's sleep propensity at any particular time. This can vary widely over periods of seconds to minutes depending on posture and activity, the time of day, the duration of prior wakefulness, the effects of sedative or stimulant drugs, etc. The instantaneous sleep propensity can be measured to some extent by subjective reports, as in the Karolinska sleepiness scale (KSS) or a visual-analogue scale of alertness/sleepiness (Åkerstedt & Gillberg, 1980). It can be measured objectively by monitoring the subject's EEG or eye and eyelid movement continuously to detect the signs of drowsiness or sleep at a particular time. This has been done successfully for research purposes in train drivers (Torsvall & Åkerstedt, 1987) and truck drivers (Kecklund & Åkerstedt, 1993; Mitler, Miller, Lipsitz, Walsh, & Wylie, 1997). The instantaneous sleep propensity can also be measured in the sleep laboratory from the time taken to fall asleep under a particular set of circumstances, as with each nap in the multiple sleep latency test (MSLT) or the maintenance of wakefulness test (MWT) (Sangal, Thomas, & Mitler, 1992).

(b) *Situational sleep propensity*: the subject's usual sleep propensity when in the same situation repeatedly. This is measured objectively by the mean sleep latency (minutes) in several naps during the day in the MSLT or the MWT. It can also be measured from subjective reports of the subject's usual chances of dozing in a particular situation, such as when driving, when sitting

and reading, or when sitting and talking to someone. The Epworth sleepiness scale (ESS) is based on 8 such situational sleep propensities that refer to a range of different activities and situations, each of a different soporific nature that either enhances or inhibits sleepiness in most people (Johns, 1991). These situational sleep propensities are relatively reliable characteristics of each subject (test–retest $r \approx 0.55$–0.65) (Johns, 1992). However, each involves a subject x situation-specific interaction, depending on how the subject usually perceives and responds to the particular situation. Consequently, different situational sleep propensities in the same subject are usually not highly intercorrelated ($r \approx 0.3$–0.4) even though their correlation coefficient may be statistically significant. That is equally true of objectively measured situational sleep propensities such as the mean latency in the MSLT or the MWT (Sangal et al., 1992), as of subjectively measured situational sleep propensities in the ESS (Johns, 1994). In practical terms, this means that when we attempt to predict what a subject's usual sleepiness would be in a particular situation, based on measurement of his sleepiness in a different situation, such predictions may not be very accurate. This has serious implications for the assessment of a person's general fitness to drive a vehicle based, for example, on measurements of his sleepiness when not driving.

(c) *Average sleep propensity*: a subject's average sleep propensity across a range of specified activities and situations in daily life, measured over periods of days to months. This is influenced by chronic sleep disorders such as obstructive sleep apnea, narcolepsy, or by chronic sleep deprivation, as well as by psychophysiological traits making different subjects more or less "sleepy" or "alert", even within the normal range of sleepiness. The average sleep propensity can be measured reliably by the Epworth score which is the sum of 8 item-scores in the ESS (test–retest $r = 0.7$–0.8) (Johns, 1991, 1992). Despite its reliance on retrospective subjective reports, the Epworth score is more accurate than either the MSLT or MWT in distinguishing the excessive sleepiness of narcoleptics from normals (Johns, 2000). A subject with a high average sleep propensity may not doze at all during the day simply by avoiding sitting down, as many narcoleptics will attest.

4. A conceptual model of sleep and wakefulness

To take account of the postural and behavioural influences on sleepiness, Johns (1998) has described a new model of sleep and wakefulness. This involves a continuous inhibitory interaction between a putative wake drive and a sleep drive, each involving the integrated action of several different neuronal centres in the brain. This is not a new idea (Kleitman, 1963). There is neurophysiological evidence for its basis (Edgar, Dement, & Fuller, 1993), but it has not been invoked by others in discussions of sleepiness. Each drive is postulated to have a primary component due to intrinsic neuronal activity, and a secondary component that is homeostatic. The primary wake drive changes with a circadian rhythm and is reflected in core body temperature, with a peak in the evening and a trough at 3–5 a.m. The secondary wake drive depends on sensory inputs from all parts of the body, for example, reflecting posture and movement, and stimuli from the environment. This secondary wake drive can change rapidly, over periods of seconds to minutes. By contrast, the secondary sleep drive increases progressively during wakefulness and is discharged, over minutes to hours, during non-REM sleep. Whether or not the primary sleep drive varies with time is uncertain, but it can differ significantly between subjects.

At any particular time, our state of sleep or wakefulness depends on the relative strengths of the total wake drive and the total sleep drive, not on their absolute strengths. When we lie down, close our eyes, and relax in bed, this reduces our secondary wake drive, not our sleep drive, which then becomes dominant and causes us to fall asleep. If we try to stay awake all night after being awake all day, our primary wake drive will decrease to a minimum by about 3–4 a.m. Our secondary sleep drive will have increased progressively during our prolonged period of wakefulness. Under those circumstances we can stay awake only by maintaining or increasing our secondary wake drive, by moving or standing up, etc. However, there will be considerable differences between subjects in their ability to stay awake at 4 a.m. because of differences in their primary wake and sleep drives that may be constitutional, in addition to the time of day effect, and to their willingness and ability to maintain their secondary wake drive at the time. The chronically excessive sleepiness of patients with sleep disorders such as narcolepsy or obstructive sleep apnea could be due to a high primary sleep drive or a low primary wake drive or both. Alternatively, their secondary sleep drive could remain high because it is not discharged adequately during sleep.

This model incorporates the useful features of the most widely accepted previous models (Borbély et al., 1989). However, the new model can explain, as no other can, the wide range of phenomena associated with sleepiness in daily life for normal subjects, as well as for those who are acutely deprived of sleep, those taking stimulant or sedative drugs, or with a chronic sleep disorder as they engage in daily activities of all kinds, including driving. It emphasises the critical role of a new variable, the secondary wake drive, in the maintenance of wakefulness. We do not yet have a method for measuring this wake drive directly, but it can be inferred, at least semi-quantitatively, from the results of its interaction with the sleep drive.

5. The drowsy driver

This conceptual background will hopefully enable us to understand more clearly the problem of the drowsy driver. Several characteristics of vehicle crashes that involve the driver being drowsy or falling asleep at the wheel have been described (Pack et al., 1995; Horne & Reyner, 1995a,b). Such crashes are more likely to happen at night between midnight and 7.00 a.m. and, to a lesser extent, in mid-afternoon. The crash is typically a run-off-the-road type at a relatively high speed, where it appears that the driver has not taken any evasive action. The driver is likely to have been alone, often younger than 25 and male. Injuries and property damage are likely to be more severe than in many other crashes.

Much, but not all, of the problem of drowsy driving can be attributed to a combination of effects due to the time-of-day and the duration of prior wakefulness (Horne & Reyner, 1995a). The former can be attributed to the circadian rhythm in the primary wake drive, which is reduced between midnight and about 7.00 a.m., and its shorter, 12-h rhythm causing another small decrease in the wake drive in mid-afternoon. The effect of prior wakefulness is explained by the progressive increase in the secondary sleep drive during wakefulness. Both effects increase the driver's instantaneous sleep propensity, but for different reasons. Not everyone will simply fall asleep when attempting to drive at 4.00 a.m., even after having been awake for more than 20 h.

Yet, we know that such a sleep-deprived person is very likely to fall asleep quickly if he lies down and closes his eyes. The difference can be explained by differences in his secondary wake drive under those circumstances, related in particular to somatosensory inputs to the central nervous system from the contractions of postural muscles, especially of the head, neck and back. Whether a driver falls asleep at the wheel or remains awake at 4.00 a.m. depends largely on his ability to maintain or increase his secondary wake drive at the time. Long-haul truck drivers can tell us about the many ways they achieve that intuitively, e.g., by chewing gum, talking or singing to themselves, or by making frequent changes in their sitting posture, using different muscles in turn, voluntarily. Those muscles may eventually become fatigued and will therefore provide less sensory input to maintain the wake drive. In that way, fatigue may be involved indirectly in sleepiness and drowsiness. Volition is also impaired by drowsiness, so it becomes more difficult for the drowsy driver to make voluntary changes to his posture or anything else. Nevertheless, an understanding of the physiology of drowsiness raises the possibility of developing improved methods for increasing the driver's secondary wake drive under those circumstances.

The driver's posture involves sitting upright with the head unsupported, one foot depressing the accelerator, and the hands holding the steering wheel that may have to be turned only occasionally and slowly on a straight road. The driver's instantaneous sleep propensity would be decreased if he stood up and thereby increased his secondary wake drive, but he cannot do that without stopping the vehicle. If he gets out of the vehicle and walks about for a few minutes, his wake drive will remain increased for only a few minutes after he resumes his driving posture. A dose of caffeine will further increase his wake drive for a few minutes. A nap will decrease his instantaneous sleep propensity by decreasing his secondary sleep drive. However, the sleep onset process also involves active inhibition of the primary wake drive by the sleep drive. If this inhibitory process continues for more than a few minutes, the subject may have an increased rather than a decreased level of drowsiness when he wakes from a nap because of "sleep inertia" until the wake drive has inhibited the sleep drive again. Drowsy drivers may introduce novel environmental stimuli to try and increase their secondary wake drive, such as cool air blowing on their face, or a radio turned on loudly, but these stimuli may have only a mild and short-lived effect (Reyner & Horne, 1998).

6. Who is too drowsy to drive?

The problem of the drowsy driver raises important questions about who is too drowsy to drive at a particular time, or at all. There is recent experimental evidence that a driver who has gone without sleep for 24 h has impaired driving skills, comparable to a driver with an illegally high blood alcohol concentration of 0.1 g/l (Lamond & Dawson, 1999). Despite initial reports that most people with obstructive sleep apnea have an increased frequency of car crashes, recent evidence suggests that this is true only for those with the most severe sleep apnea (George & Smiley, 1999). Those patients are also known to have a high average sleep propensity reflected in their Epworth scores (Johns, 1991). Similarly, narcoleptics, who usually have Epworth scores >15 (Johns, 1991), are also known to have a relatively high crash rate (Aldrich, 1989). Such people with chronically excessive sleepiness because of a sleep disorder are probably unfit to drive at all

until treated. Epworth scores > 15 were present in 2.9% of 5000 truck drivers surveyed recently in Eastern Australia (Swann, 1998).

In those drivers with excessive daytime sleepiness because of a chronic sleep disorder, the frequency of dozing at the wheel while driving is directly related to their Epworth score (Maycock, 1996; Johns & Martyn, 1999). So, too, is their risk of having a drowsy crash. (Stutts, Wilkins, & Vaughn, 1999). Somewhat surprisingly, though, drivers with mildly increased Epworth scores do not necessarily have more car crashes, even those crashes that involve drowsiness at the time, than do drivers with lower Epworth scores (Maycock, 1996; Johns & Patterson, 2000). One explanation for this may be that drivers with chronic excessive sleepiness often modify their driving to cope with their increased tendency to doze at the wheel (Baltzan, 1999; McCartt, Ribner, Pack, & Hammer, 1996). Perhaps an episode of dozing at the wheel is more dangerous in a driver with little experience of such things. McCartt et al. (1996) have described some other differences between dozing episodes that do or do not lead to a crash. Drivers whose dozing caused a crash were more likely to be driving between midnight and 7 a.m., driving alone, having worked night-shift or overtime, having drunk some alcohol or taken some medication, and having been driving for more hours before the crash. We need more research on this subject, including the duration of dozing episodes. We might assume that very brief dozing episodes, lasting only 1 or 2 s, may not be very dangerous. However, they should act as a warning to the driver that worse could follow if no remedial action is taken. The relationship between dozing at the wheel and crashing may not be as straightforward as it appears initially.

Drowsy crashes often involve drivers whose sleepiness is usually normal at other times, but whose instantaneous sleep propensity has temporarily increased because of circumstances at the time of the crash, such as acute sleep deprivation, the time of night, the added effects of alcohol on sleepiness, the boring nature of the driving task, etc. (Horne & Reyner, 1995a,b). The problem of drowsy driving should therefore be helped by making all drivers aware of its danger, as is happening now in Victoria and elsewhere. Perhaps additional strategies for the drowsy driver could be advocated on the basis of a new understanding about the importance of the secondary wake drive in maintaining wakefulness, as described here. Current developments with new methods for monitoring the driver's state of alertness continuously may be useful, at least for some drivers, in the near future (Häkkäanan, Summala, Partinen, Tiihonen, & Silvo, 1999). Such methods will, hopefully, be able to indicate to a driver that he is no longer safe to drive, even before he becomes very drowsy or falls asleep at the wheel. This is the current focus of research at Epworth sleep centre.

References

Åkerstedt, T., & Gillberg, M. (1980). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52*, 29–37.

Aldrich, M. S. (1989). Automobile accidents in patients with sleep disorders. *Sleep, 12*, 487–494.

Baltzan, M. A. (1999). Sleepy patients with suspected sleep apnea syndrome change their driving habits. *Sleep, 22*(Suppl. 1), S80.

Bartlett, F. C. (1953). Psychological criteria of fatigue. In W. F. Floyd, & A. T. Welford (Eds.). *Symposium of fatigue.* London: Lewis.

Borbély, A. A., & Achermann, P. (1992). Concepts and models of sleep regulation: an overview. *Journal of Sleep Research, 1,* 63–79.

Borbély, A. A., Achermann, P., Trachsel, L., & Tobler, I. (1989). Sleep initiation and sleep intensity: interaction of homeostatic and circadian mechanisms. *Journal of Biological Rhythms, 4,* 149–160.

Broughton, R., & Hasan, J. (1995). Quantitative topographic electroencephalographic mapping during drowsiness and sleep onset. *Journal of Clinical Neurophysiology, 12,* 372–386.

Brown, I. D. (1994). Driver fatigue. *Human Factors, 36,* 298–314.

Caldwell, J. A., & Prazinko, B. (1999). Body posture affects slow-wave EEG activity in sleep-deprived volunteers. *Sleep Research Online, 2*(Suppl.1), 518.

Edgar, D. M., Dement, W. C., & Fuller, C. A. (1993). Effect of SCN lesions on sleep in squirrel monkeys: evidence for opponent process in sleep–wake regulation. *Journal of Neuroscience, 13,* 1065–1079.

George, C. F. P., & Smiley, A. (1999). Sleep apnea and automobile crashes. *Sleep, 22,* 790–795.

Häkkäanan, H., Summala, H., Partinen, M., Tiihonen, M., & Silvo, J. (1999). Blink duration as an indicator of driver sleepiness in professional bus drivers. *Sleep, 22,* 798–802.

Horne, J. A., & Reyner, L. A. (1995a). Driver sleepiness. *Journal of Sleep Research, 4*(Suppl. 2), 23–29.

Horne, J. A., & Reyner, L. A. (1995b). Sleep related vehicle accidents. *British Medical Journal, 310,* 565–567.

Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep, 14,* 540–545.

Johns, M. W. (1992). Reliability and factor analysis of the Epworth sleepiness scale. *Sleep, 15,* 376–381.

Johns, M. W. (1994). Sleepiness in different situations measured by the Epworth sleepiness scale. *Sleep, 17,* 703–710.

Johns, M. W. (1998). Rethinking the assessment of sleepiness. *Sleep Medicine Reviews, 2,* 3–15.

Johns, M. W., & Martyn, L. (1999). Dozing at the wheel, daytime sleepiness and sleep disorders in patients at Epworth Sleep Centre. *Sleep Research Online, 2*(Suppl. 1), 382.

Johns, M. W. (2000). Specificy and sensitivity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test (MWT) and the Epworth sleepiness scale; failure of the MSLT as a gold standard. *Journal of Sleep Research, 9,* 5–11.

Johns, M. W., & Patterson, J. (2000). Car crashes in relation to the sleep habits and sleepiness of drivers in the general population (unpublished).

Kecklund, G., & Åkerstedt, T. (1993). Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. *Ergonomics, 36,* 1007–1017.

Kleitman, N. (1963). *Sleep and wakefulness.* Chicago: University of Chicago Press.

Lamond, N., & Dawson, D. (1999). Quantifying the performance impairment associated with fatigue. *Journal of Sleep Research, 8,* 255–262.

Maycock, G. (1996). Sleepiness and driving: the experience of UK car drivers. *Journal of Sleep Research, 5,* 229–237.

McCartt, A. T., Ribner, S. A., Pack, A. I., & Hammer, M. C. (1996). The scope and nature of the drowsy driving problem in New York state. *Accident Analysis and Prevention, 4,* 511–517.

Mitler, M. M., Miller, J. C., Lipsitz, J. J., Walsh, J. K., & Wylie, C. D. (1997). The sleep of long-haul truck drivers. *The New England Journal of Medicine, 337,* 755–761.

Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F., & Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis and Prevention, 6,* 769–775.

Parasuraman, R. (1984). Sustained attention I detection and discrimination. In R. Parasuranam, & D. R. Davies (Eds.). *Varieties of attention.* Orlando: Academic Press.

Reyner, L. A., & Horne, J. A. (1998). Evaluation of "in-car" countermeasures to sleepiness: cold air and radio. *Sleep, 21,* 46–50.

Sangal, R. B., Thomas, L., & Mitler, M. M. (1992). Maintenance of wakefulness test and multiple sleep latency test: measurement of different abilities in patients with sleep disorders. *Chest, 101,* 898–902.

Stutts, J. C., Wilkins, J. W., & Vaughn, B. V. (1999). *Why do people have drowsy driving crashes?* Washington, DC: AAA Foundation for Traffic Safety.

Swann, P. (1998). Personal communication.

Tanaka, H., Hayashi, M., & Hori, T. (1996). Statistical features of hypnagogic EEG measured by a new scoring system. *Sleep, 19,* 731–738.

Thomas, M., Thorne, D., Sing, N., Redmond, T., Balkin, T., Wesensten, N., Russo, M., Welsh, A., Rowland, L., Johnson, D., Aladdin, R., Cephus, R., Hall, S., & Belenky, G. (1998). The relationship between driving accidents and microsleep during cumulative sleep deprivation. *Journal of Sleep Research*, 7(Suppl. 2), 275.

Torsvall, L., & Åkerstedt, T. (1987). Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology, 66*, 502–511.