Managing neurobehavioral capability when social expediency trumps biological imperatives

Andrea M. Spaeth†, Namni Goel‡ and David F. Dinges‡,*

† Department of Psychology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA
‡ Division of Sleep and Chronobiology, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract: Sleep, which is evolutionarily conserved across species, is a biological imperative that cannot be ignored or replaced. However, the percentage of habitually sleep-restricted adults has increased in recent decades. Extended work hours and commutes, shift work schedules, and television viewing are particularly potent social factors that influence sleep duration. Chronic partial sleep restriction, a product of these social expediencies, leads to the accumulation of sleep debt over time and consequently increases sleep propensity, decreases alertness, and impairs critical aspects of cognitive functioning. Significant interindividual variability in the neurobehavioral responses to sleep restriction exists—this variability is stable and phenotypic—suggesting a genetic basis. Identifying vulnerability to sleep loss is essential as many adults cannot accurately judge their level of impairment in response to sleep restriction. Indeed, the consequences of impaired performance and the lack of insight due to sleep loss can be catastrophic. In order to cope with the effects of social expediencies on biological imperatives, identification of biological (including genetic) and behavioral markers of sleep loss vulnerability as well as development of technological approaches for fatigue management are critical.

Keywords: sleep deprivation; sleep duration; neurobehavioral functions; fatigue management; individual differences; genetics; biomarkers.

Introduction

“Sleep persists in predators and prey, carnivores and vegetarians, on the land and in the water, in most mammals as they lie down relaxed, in ruminants while they stand, in birds while they perch,
and in dolphins which constantly swim ... in the smartest and in the dumbest of all mammalian species” (Rechtschaffen, 1998). Although an organism cannot procreate, gather food, or protect itself or offspring during sleep, sleep has been evolutionarily conserved from the sloth to the human. Indeed, some evidence suggests that sleep is vital to life. A number of studies from 1894 to 1962 reported lethal outcomes from total sleep deprivation in dogs, rabbits, and rats (Kleitman, 1963). However, these studies lacked proper controls for the procedures implemented to keep the animals awake. This methodological failure was corrected in 1983 when Rechtschaffen and colleagues developed a novel disk-over-water methodology with yoked control procedures (Rechtschaffen et al., 1983). Using this paradigm, mortality was more likely in the sleep-deprived rats (death occurred after 5–33 days) compared with yoked control rats that experienced the same forced ambulation but were allowed to rest. Research using Drosophila also showed that prolonged sleep deprivation could be deadly (Shaw et al., 2002). The results obtained from these more recent studies as well as earlier studies strongly suggest prolonged sleep deprivation can be lethal and can cause death even more rapidly than food deprivation (Everson et al., 1989).

Sleeping, like eating, is a biological imperative that consumes approximately one-third of human life. Just as missing a day’s worth of meals produces strong feelings of hunger, being deprived of a night’s sleep leads to overwhelming feelings of fatigue. Sleep resists being deprived; if exhausted enough, humans can fall asleep under even dangerous circumstances (Nansen, 1999). When sleep is denied, there is an increase in the frequency and duration of sleep episodes and an elevation in the intensity of sleep (Banks et al., 2010; Borbély and Tobler, 1996; Brunner et al., 1993; Goel et al., 2009a, 2010; Van Dongen et al., 2003). These properties of sleep and its ubiquitous manifestation across species indicate that sleep must serve an adaptive function. Despite our understanding that sleep plays a necessary and adaptive evolutionary purpose, the universal function of sleep, if one exists, remains elusive (Rechtschaffen, 1998; Siegel, 2005; Tobler, 1995). Our lack of knowledge regarding sleep’s fundamental function has led a select few to assume that humans need only a small “core” amount of sleep to function normally (Horne, 1985, 1988); however, a large body of research indicates that most humans who undergo chronic partial sleep deprivation exhibit profound neurobiological and physiological changes that can impede performance and negatively impact health (Banks et al., 2010; Dement and Greenberg, 1966; Dinges, 2004; Goel et al., 2009a; Leproult and Van Cauter, 2010).

Healthy human sleep

Normal human sleep comprises two states—rapid eye movement (REM) and non-rapid eye movement (NREM)—that alternate cyclically during a sleep episode. Using electroencephalography (EEG), characteristics of these sleep states have been well defined (Rechtschaffen and Kales, 1968; Silber et al., 2007). NREM sleep shows synchronous cortical EEG (sleep spindles, K-complexes, and slow waves) and is associated with low muscle tone and minimal psychological activity whereas REM sleep shows desynchronized EEG and is associated with muscle atonia and dreaming (Kryger et al., 2011). In healthy humans, the timing, intensity, and duration of sleep are primarily regulated by two processes: homeostatic regulation and circadian timing (Borbély, 1982, 1998).

Sleep homeostasis

Sleep homeostasis describes the drive for sleep that increases progressively during wakefulness and decreases progressively during NREM sleep (Borbély, 1994). Organisms strive to maintain sleep at a constant level; therefore, the propensity for sleep is increased after prolonged wakefulness and decreased after prolonged sleep. Similarly, sleep restriction leads to an increase in the
intensity and duration of subsequent sleep, whereas excess sleep leads to decreased sleep intensity and duration (Borbély, 1982). Researchers have found that slow waves, low-frequency EEG, likely represent a measure of sleep intensity—extended waking leads to increases in slow wave energy (SWE) during the subsequent recovery night and the extent of this SWE increase is a function of prior wake duration (Åkerstedt et al., 2009; Brunner et al., 1990). In addition to SWE, other biological markers of sleep homeostasis have been identified including extracellular adenosine, central nitrous oxide levels, and salivary amylase; levels of these markers increase with prolonged sleep–wakefulness and thus may reflect an increased sleep drive (Kalinchuk et al., 2006; Porkka-Heiskanen and Kalinchuk, 2011; Scharf et al., 2008; Seugnet et al., 2006). The homeostatic process of sleep–wake regulation interacts with but is independent from circadian control (Dijk et al., 1989).

Sleep as a circadian process

The circadian process that controls sleep is described as the 24-h oscillatory variation in the propensity for sleep. This 24-h period is the time it takes the Earth to rotate about its own axis which generates daily environmental cycles of ambient temperature and illumination. The alternation of light and darkness directly entrains an organism’s circadian rhythms and thus influences its life patterns, creating species that are active primarily during the light (diurnal), the dark (nocturnal), twilight periods (crepuscular), or during the light and the dark (cathemeral) (Pittendrigh, 1981). Environmental light is transmitted from the retina to the suprachiasmatic nuclei in the anterior hypothalamus which then transmit this information to, among others, the pineal gland via a multisynaptic pathway; the pineal gland secretes melatonin, a hormone that regulates various biological functions (Czeisler, 1995; Lucas et al., 1999; Ralph et al., 1990; Sadun et al., 1984; Watts, 1991). In humans, melatonin increases during the dark cycle which coincides with a period of inactivity and sleepiness and decreases during the light cycle which coincides with a period of activity and wakefulness (Aschoff and Wever, 1981; Shanahan and Czeisler, 1991). This circadian process interacts with but is independent from sleep homeostasis (Åkerstedt and Froberg, 1978).

Scientific evidence shows that human sleep is naturally regulated by these two processes, with sunrise and sunset providing the photic signals necessary to entrain the sleep–wake cycle (Wehr et al., 2001). However, with the introduction of artificial light and other technologies, television and alarm clocks have replaced these natural signals and thereby may not allow optimal sleep duration. Additionally, industries that operate 24 h a day create light and noise that can interfere with sleep (Basner et al., 2011) and require employees to work during hours usually devoted to sleep. Finally, with the increase in international business and air travel, many adults are frequently traveling across time zones which affects the circadian timing system. Thus, many humans are challenging basic biological pressures in order to accommodate social norms and obligations.

Human sleep duration

For most healthy adults, physiological sleep duration appears to range between 7.0 and 8.5 h; however, habitual sleep duration among adults is determined by a variety of factors and shows considerable variance within and between individuals (Van Dongen et al., 2005). Data from the 2005–2008 National Health and Nutrition Examination Survey suggest that self-reported sleep duration is distributed approximately normally and 2004–2007 National Health Interview Survey-Sample Adult data reveal that 7.8% of adults report sleeping less than 5 h per night, 20.5% report sleeping 6 h per night, 30.8% report sleeping 7 h per night, 32.5% report sleeping 8 h per night, and 8.5% of adults report sleeping more than 9 h per night (Krueger and Friedman,
One limitation of large epidemiological studies is that sleep duration is examined using self-report which can be subject to inherent biases and thereby inaccurate.

Indeed, when comparing self-reported sleep duration to objectively measured sleep duration using actigraphy (an unobtrusive measure of gross motor activity which is analyzed to objectively identify sleep periods), subjectively reported sleep averaged 0.80 h longer than objectively measured sleep (Bradshaw et al., 2007; Lauderdale et al., 2008). Although wrist actigraphy provides an objective measure of inactivity, it still overestimates polysomnography (PSG)-measured sleep duration, which is the gold standard for determining physiological sleep. One study showed this overestimation was approximately 18 min on average (Blackwell et al., 2008). Therefore, although epidemiological studies are valuable in that they include data based on a large number of subjects, the actual amount of physiological sleep that adults receive is typically less than what is self-reported in these studies.

There is considerable debate as to whether or not sleep duration has been decreasing among adults in recent decades and if so, whether this reduction is resulting in higher rates of chronic sleep restriction or sleep debt (Dinges, 2004; Horne, 2004). However, according to the Center for Disease Control Morbidity and Mortality Weekly Report (2008), the percentage of adults who reported an average of less than or equal to 6 h of sleep within a 24-h period significantly increased from 1985 to 2004 (in both females and males and among all age groups 18–75). Among employed American adults, two epidemiological studies found that the prevalence of being a short sleeper (either ≤6 h/day (Knutson et al., 2010) or <6 h/day (Luckhaupt et al., 2010)) has increased significantly in the past few decades. Studies have suggested that habitual short sleepers do not require less sleep than other adults; rather, these individuals gradually accrue sleep debt over time (Aeschbach et al., 1996; Bradshaw et al., 2007; Klerman and Dijk, 2005). Evidence from the American Time Use Survey indicates that adult sleep duration is significantly shorter during weekdays compared to weekends, suggesting that adults attempt to recover sleep debt by extending sleep when it is presumably more convenient and when schedules are likely more flexible (see Fig. 1a; Basner et al., 2007).

Causes of chronic partial sleep restriction

Paid work

Compensated work time may be the most potent determinant of sleep duration (see Fig. 1b and c; Basner et al., 2007). There is a higher prevalence of short sleep duration among full-time employed adults when compared to part-time workers, students, retired individuals, homemakers, or unemployed adults (Knutson et al., 2010; Luckhaupt et al., 2010), and longer work hours are associated with shorter sleep duration (Hale, 2005; Nakashima et al., 2010; Virtanen et al., 2009). Adults working more than 8 h/day have the same bedtime compared to adults who do not work more than 8 h/day, but they wake up much earlier (Basner and Dinges, 2009). Using a prospective study design, Virtanen et al. (2009) found that working more than 55 h/week is a risk factor for the development of shortened sleep and for difficulty falling asleep. Although the average number of hours spent working has remained relatively stable during the past few decades, the prevalence of individuals who work greater than 48 h/week has increased (Rones et al., 1997). The duration of required working hours varies by occupation, with managerial, professional, manufacturing, and transportation industries typically involving longer work hours; interestingly, the prevalence of short sleep duration is also highest in these occupations (Luckhaupt et al., 2010). By contrast, the lowest prevalence of short sleep duration is observed in the “agriculture, forestry, fishing, and hunting industry” (Luckhaupt et al., 2010). Although workers within this category (such as farmers) often work long hours and begin work
early in the morning, their work schedules are more in sync with the natural light–dark cycle compared to other occupations, and thus their work hours presumably interfere less with sleep quality and duration.

In contrast to the schedules of agriculture, forestry, fishing, or hunting industries, shift work schedules often require work to occur during the night when the circadian system is promoting sleep and require sleep to occur during the day when the circadian system is promoting wakefulness. This work schedule creates a conflict between the worker’s internal circadian time and his/her required sleep–wake schedule, leading to impaired wakefulness and disturbed sleep (Åkerstedt, 2005; Kolla and Auger, 2011). According to the Bureau of Labor Statistics (2005), approximately 15% of full-time wage and salary earners work shifts that are not during the daytime. Night workers sleep 2–4 h less per day than day workers (Åkerstedt, 2003) and workers either permanently on a night shift or on a rotation including the night shift are significantly more likely to sleep less than 6 h and experience excessive sleepiness during situations requiring a high degree of attention (Ohayon et al., 2010). Although some countermeasures, such as strategic light exposure (Burgess et al., 2002), or the use of melatonin, stimulants, and other

Fig. 1. (a) Average sleep duration with respect to age and day of the week. Average sleep time across age groups was longest on Sunday followed by Saturday and was markedly shorter during the week, gradually decreasing from Monday to Friday for most age groups. (b and c) Average change in weekday (b) (Mon–Fri) and weekend (c) (Sat+Sun) waking activity time depending on sleep time category (N = 23,325) based on multiple linear regression models adjusting for age, gender, ethnicity, educational attainment, income, presence of partner, and presence of children. The 7.5 h to < 8.5 h sleep time category served as a reference. For work time, separate models were run for the whole group (Work) and those only who worked on the interview day (Work > 0 min). The largest reciprocal relationship to sleep was found for work time, followed by commute/travel time. Short sleepers spent more time socializing, relaxing, and engaging in leisure activities, while both short and long sleepers watched more television than the average sleeper. (Reprinted with permission from Basner et al., 2007.)
pharmacological agents related to sleep–wake regulation can promote wakefulness or improve sleep among shift workers (Kolla and Auger, 2011), currently there is no treatment to completely normalize a shift worker’s sleep duration that can be widely used in real-world situations (Åkerstedt and Wright, 2009).

**Commute/travel**

According to the Census American Community Survey Report (2011), American workers spend nearly an hour commuting to and from work each day and 33.5% of workers have commutes that are greater than 30 min each way. Travel time (composed of traveling to work [commute], stores, schools, and social events) is negatively associated with sleep time (see Fig. 1b and c; Basner et al., 2007). A majority of Americans (55.4%) leave their homes between 06:00 and 08:30 in order to arrive at work on time (Census American Community Survey Report, 2011), and they do not arrive home until late in the evening in part due to traffic; the result is an increase in the length of the workday and a decrease in time for other activities such as socializing, relaxing, and sleeping.

**Leisure**

Socializing, relaxing, and engaging in other leisure activities are also negatively related to sleep duration (see Fig. 1b and c; Basner et al., 2007). For many adults, this time usually involves sitting in front of some type of screen, such as a television, computer, or smartphone. For example, watching television was the most common activity during the 2-h period before bedtime, suggesting that nighttime television viewing and sleep onset are tightly correlated (Basner and Dinges, 2009; Basner et al., 2007). Adults are now able to access e-mail, engage in social networking and online gaming, and use the web 24 h a day; devices have become so portable that they can easily be brought into bed and used right before going to sleep. Among adults, computer and mobile phone use in the bedroom is associated with greater variability in sleep–wake schedules and poorer sleep habits (Brunborg et al., 2011). In addition to the time used engaging in these activities, exposure to light emitted from these screens may also impede sleep. Blue (short wavelength) light represents the most potent portion of the light spectrum for suppressing melatonin and thus promoting wakefulness (Brainard et al., 2001; Lockley et al., 2003); moreover, blue light from light-emitting diodes elicits a dose-dependent suppression of melatonin (West et al., 2011). Exposure to blue light immediately prior to bedtime (via flat screen televisions, smartphones, and tablet computers) may cause circadian phase delays and disrupt sleep (Cajochen et al., 2006, 2011).

**Chronotype**

Although humans are diurnal, some individuals prefer activity in the morning (larks) whereas others prefer activity in the evening (owls)—this preference influences the timing of sleep–wake cycles. Morning-type and evening-type individuals differ endogenously in the circadian phase of their biological clocks (Baehr et al., 2000; Kerkhof and VanDongen, 1996), which is partially determined by genetic polymorphisms (Goel, 2011). In addition to genetic factors, age, and gender also influence morningness–eveningness (Roenneberg et al., 2007). The Earth’s light/dark cycle and the work schedules of industrial societies complement individuals who function best in the morning rather than in the evening. Owls experience heightened alertness in the late evening and typically stay awake longer and have a delayed bedtime compared to larks; however, due to typical work schedules owls often use an alarm clock to wake up early in the morning, which produces chronic sleep restriction, extended wakefulness, and accumulated sleep debt during the work week.
(Korczak et al., 2008; Roenneberg et al., 2003; Roepke and Duffy, 2010; Taillard et al., 2003).

Social jetlag, a term coined by Roenneberg and colleagues (Wittmann et al., 2006), describes the misalignment of biological and social time caused by social activities such as work, commuting, and television viewing. For many adults, social jetlag and/or a preference for eveningness can lead to the accumulation of sleep debt during the work week due to repeated episodes of shortened sleep. This repeated and chronic partial sleep deprivation has been shown to have serious neurobehavioral and physiological consequences unlikely to be reversed by increasing sleep time on weekends and days off (Banks et al., 2010; Goel et al., 2009a; Van Dongen et al., 2003).

**Consequences of chronic partial sleep restriction**

According to the National Transportation Safety Board (1989, 1995), 30–40% of all U.S. truck accidents are fatigue related and several catastrophes including the Exxon Valdez accident, Chernobyl and Three Mile Island nuclear plant meltdowns, and Space Shuttle Challenger tragedy were partially due to human error resulting from sleepiness and fatigue (Mitler et al., 1988). Sleepiness due to long work hours has also been implicated in medical errors: interns, surgeons, physicians, and residents make significantly more errors, including those that are injurious or fatal, after working for an extended period of time and/or after chronic and acute sleep loss (Czeisler, 2009; Lockley et al., 2004). As such, gaining a greater understanding of the relationship between sleep and waking performance is an important focus of ongoing research.

**Regulation of human performance**

Similar to the regulation of sleep duration and intensity, distinct sleep/wake-related physiological processes regulate alertness and neurobehavioral performance. The homeostatic sleep-dependent process (process S) balances sleep propensity by keeping track of recent sleep history (Daan et al., 1984). As hours of wakefulness increase, homeostatic drive increases the propensity for sleep. The endogenous circadian process (process C) tracks changes in light exposure (as well as other zeitgebers or synchronizers) and entrains sleep propensity to the light–dark cycle; when sleep propensity is increased (during the night) waking performance is often degraded (Czeisler et al., 1999). Sleep inertia, a transient process that occurs immediately after awakening, causes deficits in waking performance; its effects depend on prior sleep history, circadian phase, and the depth of sleep at the time of awakening (Dinges, 1990). Finally, a second homeostatic process (process U) influences waking performance by monitoring sleep–wake on a longer time scale (nightly sleep duration) and interacting with the other processes (McCauley et al., 2009). Process U builds up over several days of prolonged wakefulness, when sleep debt is accruing and the need for sleep is increasing, and dissipates during sleep. The status of process U, similar to a sleep reservoir being full or empty, codetermines the rate by which the original homeostatic process (process S) increases sleep propensity and thus decreases waking performance (McCauley et al., 2009).

All of these processes interact to form a dynamic relationship between sleep propensity and waking performance (Raslear et al., 2011; Van Dongen and Dinges, 2005). During a 24-h period, the interaction of these processes leads to a period in the early morning when waking performance is particularly vulnerable to sleep restriction; neurobehavioral deficits are largest at 08:00 and become progressively smaller during the day, especially between 16:00 and 20:00 (Mollicone et al., 2010). The incidence of single-vehicle automobile accidents, human error, and accidents related to work performance and several industrial/engineering disasters also reveal early morning hours as particularly vulnerable to sleep restriction (Mitler et al., 1988). However, when an individual lacks a sufficient amount of
sleep, either as a result of acute total sleep deprivation or chronic partial sleep restriction, homeostatic pressure can increase to the point whereby waking cognitive functions will be degraded even during periods of time when the circadian drive for wakefulness peaks (Doran et al., 2001).

**Sleep propensity**

Individuals who report sleeping less than 7.5 h/day on average are more likely to exhibit a stronger sleep propensity (the tendency to fall asleep)—they report unintentionally falling asleep during the day and nodding off or falling asleep while driving (McKnight-Eily et al., 2011; Punjabi et al., 2003). In the laboratory, sleep propensity—operationalized as the speed of falling asleep in both sleep-conducive and non-conducive conditions—is among the most well-validated measures of sleepiness (Roehrs et al., 2005). The effects of chronic sleep restriction on daytime physiological sleep propensity have been evaluated using the Multiple Sleep Latency Test (MSLT) (Carskadon and Dement, 1981) and the Maintenance of Wakefulness Test (MWT) (Mitler et al., 1982). During the MSLT, the subject is instructed to close his/her eyes and try to fall asleep while lying in a supine position, whereas during the MWT, the subject is instructed to look straight ahead and try to stay awake while seated upright. In both tests, PSG recordings are made (including EEG, electrooculogram and electromyogram) and the time taken to fall asleep is a measure of sleep propensity.

Whether attempting to fall asleep or resist sleep, the latency from waking to sleeping is reduced by chronic partial sleep restriction (see Fig. 2a; Banks et al., 2010; Belenky et al., 2003; Carskadon and Dement, 1981; Dinges et al., 1997; Guilleminault et al., 2003; Rupp et al., 2009). As the number of days of sleep restriction increases or the amount of sleep available during each night of sleep restriction decreases, the latency to fall asleep decreases (Belenky et al., 2003; Carskadon and Dement, 1981; Devoto et al., 1999; Guilleminault et al., 2003; Rupp et al., 2009). Spending 10 h in bed daily for 1 week prior to experiencing a week of sleep restriction may attenuate the effects of sleep restriction on sleep propensity as measured by MWT (Rupp et al., 2009). By contrast, after experiencing partial sleep restriction, one recovery night of up to 10 h time in bed (TIB) is insufficient to return sleep propensity back to presleep restriction levels (see Fig. 2b; Banks et al., 2010). MWT sleep latencies appear normalized after two 8 h TIB nights of recovery sleep (Rupp et al., 2009); however, due to the small sample size used in this study, replication of this finding is needed.

**Behavioral alertness: Microsleeps and wake-state instability**

The increased propensity for sleep due to restricted sleep time can lead to the occurrence of “microsleeps,” very brief sleep episodes that intrude into wakefulness despite an individual’s best effort to stay awake (Åkerstedt et al., 1987; Bjerner, 1949; Torsvall and Åkerstedt, 1987). Wake-state instability refers to the moment-to-moment shifts in the neurobiological systems mediating the motivated desire to sustain waking alertness and those mediating the involuntary homeostatic drive to fall asleep (Doran et al., 2001; Saper et al., 2005). This type of interaction between drives results in unpredictable behavior, including increased variability in cognitive performance and lapsing (i.e., brief periods of half a second to many seconds of no response) (Doran et al., 2001). Although individuals may not realize they are experiencing microsleeps and performance decrements, over time this instability can progress into full blown sleep attacks—when individuals will not spontaneously wake without additional stimulation (Dinges and Kribbs, 1991; Harrison and Horne, 1996; Kleitman, 1963). Sleep restriction can increase the incidence of microsleeps and produce decreased behavioral
alertness, even during goal-directed behaviors (e.g., motor vehicle operation). The rapid intrusion of sleep into wakefulness and consequent wake-state instability can have profound adverse consequences, especially if experienced by those in safety-sensitive occupations, such as police officers, firefighters, health care providers, and motor vehicle operators (Barger et al., 2009; Lombardi et al., 2010; Trew et al., 2011). Unfortunately, workers in these professions are often subjected to chronic sleep restriction due to shift work schedules and long work hours and are thus particularly vulnerable to microsleeps and wake-state instability.

Behavioral alertness, measured using sustained attention tasks in the laboratory, has been shown to be sensitive to sleep restriction. According to
the state instability hypothesis, sleep drive escalates instability in attention which creates increasingly variable neurobehavioral performance (periods of accurate responding are interrupted by errors of omission [lapses] comingled with errors of commission) (Doran et al., 2001). Studies support this hypothesis; cognitive performance variability is influenced by prior wakefulness, circadian phase, and the amount of time spent on the task (Doran et al., 2001; Graw et al., 2004; Mollicone et al., 2010; Zhou et al., 2011b). When sleep is chronically restricted, sleep debt accumulates across each night leading to greater impairment over time (Banks et al., 2010; Belenky et al., 2003; Dingess et al., 1997; Goel et al., 2009a; Rupp et al., 2009; Van Dongen et al., 2003; Wu et al., 2010). Thus, as the need for sleep increases, the brain’s ability to maintain alertness becomes progressively more overwhelmed by the activation of sleep processes which leads to microsleeps and consequent neurobehavioral instability (Doran et al., 2001). Even when highly motivated, an individual’s attempt to compensate for excessive sleepiness by engaging in various behaviors ultimately fails to prevent intrusions of microsleeps and impaired neurobehavioral performance (Horne and Pettitt, 1985; Hsieh et al., 2010).

Driving performance

Driving a vehicle requires sustained attention and is sensitive to the reduced alertness associated with sleepiness. Every year, thousands of automobile crashes, injuries, and fatalities are due to drivers falling asleep (Strohl et al., 1997). Adults who are employed, work more than 60 h/week, work irregular hours, or work at night, as well as those who are sleep deprived or chronically sleep restricted are more likely to have a car accident (Abe et al., 2010; Åkerstedt et al., 2005; Brown, 1994; Gander et al., 2005; Horne and Reyner, 1999; Philip et al., 2003, 2005; Scott et al., 2007; Stutts et al., 2003). In the laboratory, studies have primarily focused on short-term sleep restriction and have found that driving performance on simulators decreases (resulting in more crashes) after short sleep duration (4–6 h TIB) (De Valck and Cluydts, 2003; Lenne et al., 2004; Macchi et al., 2002; Otmani et al., 2005; Vakulin et al., 2007). Similarly, the less sleep obtained during chronic partial sleep restriction, the more driving performance (increase in number of driving simulator accidents) is impaired (Russo et al., 2003).

Cognitive performance

In addition to its negative effects on alertness, sleep deprivation degrades aspects of cognitive performance (Harrison and Horne, 2000; Kleitman, 1963; McCoy and Strecker, 2011). There are hundreds of published studies on the effects of total sleep deprivation on cognitive performance (Alhola and Polo-Kantola, 2007; Goel et al., 2009b; Lim and Dingess, 2010), but far fewer studies on the cognitive-impairing effects of chronic partial sleep restriction.

Sleep deprivation induces a wide range of effects on cognitive functions (Killgore, 2010; Poe et al., 2010); however, cognitive tasks vary considerably in their sensitivity to sleep loss. In general, regardless of the task, cognitive performance becomes progressively worse when time on task is extended; this is the classic “fatigue” effect that is exacerbated by sleep loss (Bjerner, 1949; Wilkinson, 1969). However, performance on brief cognitive tasks that measure speed of cognitive “throughput,” working memory, and other aspects of attention are also sensitive to sleep deprivation (Dinges, 1992). Two confounding factors that can obscure the effects of sleep loss on many cognitive tasks are intersubject variability and intrasubject variability (Dorrian et al., 2005). The intersubject confound describes differences in aptitude: one individual’s poorest performance during sleep deprivation may be superior to the best performance of a non-sleep-deprived individual (Goel et al., 2009b). The intrasubject confound
describes a learning effect: a person may be cognitively diminished by sleep loss, but continue to improve on a repeated task due to the effects of learning (Goel et al., 2009b). The nature of the dependent variables selected for examining the cognitive effects of sleep deprivation can also be problematic. Sleep deprivation increases variability within subjects (i.e., state instability) and between subjects (i.e., differential vulnerability to the effects of sleep deprivation); therefore, the effects of sleep loss on cognitive measures may be missed due to insensitive metrics or data analyses (Olofsen et al., 2004; Van Dongen et al., 2004a; Whitney and Hinson, 2010). To provide an accurate and useful measure of performance during sleep loss as well as the dynamic expression of waking neurobehavioral integrity as it changes over time, cognitive assessments must be valid and reliable reflections of fundamental waking functions altered by sleep deprivation (Goel et al., 2009b). As such, measures of attention, vigilance, and declarative memory are often used, with reaction time as the dependent variable.

The Psychomotor Vigilance Test (PVT), a measure of sustained attention, is free of aptitude and learning effects and is widely used in sleep studies due to its sensitivity to sleep loss, sleep pathology, and functioning at an adverse circadian phase (Balkin et al., 2004; Dorrian et al., 2005; Lim and Dinges, 2008). The primary outcome measures of the PVT are reaction time and errors, including omissions (lapses) and commissions (false responses). In 2003, two large-scale experimental studies found that when sleep is reduced to 3, 4, 5, or 6 h TIB for several nights, sustained attention performance (PVT) as well as working memory and cognitive throughput (Serial Addition/Subtraction Task, Digit Symbol Substitution Task) decreases in a dose-dependent manner (see Fig. 2c; Belenky et al., 2003; Van Dongen et al., 2003). As sleep debt accumulates across days, performance becomes progressively worse over time; 14 days of chronic sleep restriction (4 or 6 h TIB each night) produced comparable cognitive deficits to those produced by 24–48 h of total sleep deprivation (Van Dongen et al., 2003). This finding has been replicated several times (Axelsson et al., 2008; Banks et al., 2010; Bliese et al., 2006; Cote et al., 2008; Dingess et al., 1997; Fafrowicz et al., 2010; Goel et al., 2009a; Rupp et al., 2009; Wu et al., 2010).

Research has also shown that the recovery rate from chronic sleep restriction may be slower than from acute total sleep loss (Banks et al., 2010; Rupp et al., 2009). In a large-scale experimental study examining dose–response effects of one night of recovery sleep after five nights of sleep restriction on PVT performance, recovery to either a subject’s own baseline values or values recorded for the sleep-satiated control group was not achieved at any of the doses examined, including the highest dose of 10 h TIB (see Fig. 2d; Banks et al., 2010). Due to circadian limitations on sleep duration, it is unlikely that recovery from 5 days of chronic sleep restriction can occur in one night of more than 10 h TIB; therefore, residual attentional deficits still present after one night of recovery may potentiate the effects of a subsequent sleep restriction period (Banks et al., 2010). Thus, attempting to recover lost sleep from a work week by extending sleep on a weekend night is likely insufficient for recuperating impaired alertness and sustained attention. Other studies examining recovery after chronic sleep restriction have found that PVT performance remains substandard even after 3, 5, or 7 recovery nights of 8 h TIB (Axelsson et al., 2008; Belenky et al., 2003; Rupp et al., 2009). However, similar to results for sleep propensity, prophylactic napping prior to sleep deprivation significantly improved reaction time performance and spending 10 h TIB daily for 1 week prior to experiencing chronic partial sleep restriction attenuated the decrement in PVT performance during sleep loss and facilitated improvement of PVT performance during recovery (Dingess et al., 1987; Rupp et al., 2009). Thus, chronic sleep restriction induces slow changes in neural processes mediating alertness and attention that cause performance to become progressively worse over time, producing the accumulation of severe performance
decrements. Similarly, the slow recovery rate from chronic sleep restriction suggests that this type of sleep loss may induce long-term neuromodulatory changes in brain physiology (Banks et al., 2010; Rupp et al., 2009).

Researchers have proposed that attention and working memory, considered basic cognitive actions, are essential to virtually all other cognitive processes; therefore, sufficient attention and memory performance are necessary for optimal functioning of neural circuits that mediate higher cognitive functions, such as executive function (Balkin et al., 2008). Interestingly, there is a positive relationship between the level of sustained attention required to perform a task and the degree to which task performance is impaired by sleep loss (Jennings et al., 2003). Although there is evidence that sleep deprivation adversely affects prefrontal cortex-related executive attention and working memory abilities, these cognitive effects are often not as prominent or easy to measure as those involving basic processes such as cognitive and psychomotor speed (Goel et al., 2009b). Therefore, more complex cognitive tasks involving higher cognitive functions have been regarded as insensitive to sleep deprivation by some researchers (Harrison and Horne, 2000).

Sleep restriction affects mental flexibility, attention shifting, and the inhibition of automatic responses, all reflective of executive function (McCoy and Strecker, 2011). Executive function can be defined as “the ability to plan and coordinate a willful action in the face of alternatives, to monitor and update action as necessary and suppress distracting material by focusing attention on the task at hand” (Jones and Harrison, 2001). Many tasks believed to engage different aspects of executive function have been used in sleep deprivation studies. Examples of such tasks include the Brixton Spatial Anticipation Test, Controlled Oral Word Association Test (COWAT), Hayling Sentence Completion Task, Torrance Tests of Creative Thinking, Tower of London, Thurstone’s Verbal Learning Task, and Wisconsin Card Sorting Task (Harrison and Horne, 2000; Jones and Harrison, 2001). In experimental studies, individuals who are sleep restricted exhibit deficits on the Wisconsin Card Sorting Task (Herscovitch et al., 1980) and Stroop test (Stenuit and Kerkhofs, 2008) but do not exhibit deficits on the Hayling, Brixton, COWAT, or Tower of London Tasks (Goel et al., 2009a). Although subjects are able to accurately perform many cognitive assessments, compensatory mechanisms become engaged in response to chronic sleep restriction which impedes the ability to think flexibly and increases the use of automatic processing, impulsivity, and rigid rule adherence to complete tasks (Stenuit and Kerkhofs, 2008; Swann et al., 2006).

Plessow and colleagues studied new parents, a population of healthy adults who are often subjected to chronic sleep restriction for several months due to nightly feedings and duties related to caring for a new infant. Parents were administered an explicit-cueing version of the task-switching paradigm to assess their ability to flexibly adapt to changing environmental demands. Sleep-deprived new parents exhibited significantly slower reaction times during task switches compared to new parents reporting adequate sleep; correlational analyses showed that greater self-reported sleep debt was associated with greater impairment on the task (Plessow et al., 2010). These results indicate that chronic sleep restriction can impair an individual’s ability to implement task goals in order to switch tasks in a fast-changing environment—factors important for doctors, pilots, and individuals in the armed forces, who constantly rely on fast and flexible goal shifting and are likely to experience chronic sleep restriction.

Interindividual differences in response to sleep restriction

As stated previously, average sleep duration is normally distributed and widely differs depending on the individual (Van Dongen et al., 2005). In addition, although humans are diurnal mammals, there are individual differences in the
Individuals are highly vulnerable, somewhat vulnerable, or highly resistant to the neurobehavioral effects of sleep restriction (Goel and Dinges, 2011). It remains unknown whether the same individuals are vulnerable to the adverse effects of both acute total sleep deprivation and chronic sleep restriction.

A few studies have found that these interindividual differences may be task dependent; one individual’s sustained attention may be most affected by sleep loss whereas another individual’s working memory may be most affected (Frey et al., 2004; Van Dongen et al., 2004b). In addition, interindividual sensitivity to sleep loss is trait like; when neurobehavioral performance was assessed repeatedly (during separate laboratory visits), the same individuals remained particularly vulnerable or resistant to the effects of sleep loss even when previously sleep satiated or sleep restricted (Van Dongen et al., 2004b). Finally, these stable interindividual differences account for a substantial proportion of variance in cognitive performance decrements induced by sleep loss. Demographic factors (age, sex, IQ), baseline functioning, various
aspects of habitual sleep timing, and circadian chronotype have not accounted for the robust and stable differences in neurobehavioral responses to sleep restriction (Doran et al., 2001; Goel and Dingels, 2011; Van Dongen and Dingel, 2003; Van Dongen et al., 2004b).

Due to the trait like, or phenotypic, nature of these interindividual differences, research has focused on understanding which genetic polymorphisms may underlie vulnerability to sleep loss. Several genetic mechanisms have been examined in relation to the effects of chronic sleep restriction including the PERIOD3 (PER3) variable number tandem repeat (VNTR) polymorphism and the DQB1*0602 allele. Although the VNTR polymorphism of the circadian gene PER3 (PER34/4, PER34/5, and PER35/5) has been shown to influence executive function deficits in response to total sleep loss (Groeger et al., 2008; Vandewalle et al., 2009; Viola et al., 2007), it does not appear to influence neurobehavioral effects of chronic sleep restriction. A large-scale experimental study examining individuals with different PER3 genotypes found that PER34/4, PER34/5, and PER35/5 genotypes exhibited similar baseline performance on neurobehavioral assessments and demonstrated large but equivalent cumulative increases in sleepiness and sleep propensity and cumulative decreases in cognitive performance and physiologic alertness across five nights of sleep restricted to 4 h per night (see Fig. 3; Goel et al., 2009a). Daily intersubject variability also increased across sleep restriction days similarly in all genotypes (Goel et al., 2009a). Although PER3 genotypes did not differ at baseline in habitual sleep, physiological sleep structure, or sleepiness, during sleep restriction, PER35/5 subjects had elevated sleep homeostatic pressure, measured physiologically by EEG SWE during NREM compared with PER34/4 subjects (Goel et al., 2009a). Goel and colleagues also examined the human leukocyte antigen DQB1*0602 allele which is closely associated with narcolepsy. Subjects who were either positive or negative for the DQB1*0602 allele differed in levels of sleep homeostatic pressure, baseline sleepiness, and sleep physiology but showed comparable cumulative decreases in cognitive performance and increases in sleepiness in response to chronic sleep restriction (Goel et al., 2010). Additional investigations are underway to identify other genetic polymorphisms that may mediate an individual’s neurobehavioral vulnerability to sleep restriction and acute total sleep deprivation (Goel, 2011).

Identifying correlates of interindividual differences in neurobehavioral vulnerability to chronic sleep restriction, such as biomarkers (including genetic polymorphisms as well as adenosine and salivary amylase levels) and behavioral predictors, will provide a viable means to determine those individuals in the general population who need longer habitual sleep durations and/or who require effective interventions and countermeasures for unavoidable sleep loss (Goel and Dingel, 2011). This is a particularly important area of research since adults are not adept at accurately judging how affected they are by chronic sleep restriction (see Figs. 3 and 4; Banks and Dingel, 2007; Van Dongen et al., 2003; Zhou et al., 2011b). Subjective sleepiness is measured using two widely used and well-validated questionnaires: the Stanford Sleepiness Scale (SSS) which requires subjects to rate how sleepy they feel on a scale of 1–7 (Hoddes et al., 1973) and the Karolinska Sleepiness Scale (KSS) which requires subjects to rate how sleepy they feel on a scale of 1–9 (Åkerstedt and Gillberg, 1990). Although subjects are able to detect the rapid and severe changes in levels of alertness and sleepiness in response to total sleep deprivation, subjects are much less sensitive in detecting the smaller changes in levels of alertness and sleepiness that accumulate during each day with sleep restriction (see Figs. 3 and 4; Banks and Dingel, 2007; Goel et al., 2009a; Van Dongen et al., 2003; Zhou et al., 2011a). After a week (or two) of sleep restriction, subjects show marked cognitive impairment and severe decreases in alertness but rate themselves as only moderately sleepy (Belenky et al., 2003; Van Dongen et al., 2003). Consequently, a subject’s rating of subjective sleepiness does not accurately parallel the continuing accumulation of cognitive performance deficits associated with sleep loss. This finding
suggests that people who are chronically sleep restricted underestimate the impact of sleep restriction and may overestimate their levels of alertness and ability to perform various cognitive tasks (Banks and Dinges, 2007; Czeisler, 2009; Van Dongen et al., 2003; Zhou et al., 2011a).

Management of sleep restriction-induced neurobehavioral impairment

Due to the high prevalence of chronic partial sleep restriction in industrialized societies, the severe consequences of sleep loss on alertness and cognitive function, and our inability to accurately monitor our own sensitivity to chronic sleep loss, objective actions need to be employed to manage sleep debt. Currently, fatigue management within transportation, health professions, and other occupational settings are primarily based on policy; a regulatory organization imposes limits to daily/weekly shift duration as well as the number of consecutive shift or rest hours (Czeisler, 2009; Gander et al., 2011). Regulatory organizations may also screen potential employers for sleep disorders when occupational duties pose safety risks (Czeisler, 2009; Gander et al., 2011). Although these policies can improve workplace safety and productivity, many regulations are not strictly enforced or adhered to and it is nearly impossible for an employer to ensure that workers are actually obtaining an appropriate amount of sleep each night (Czeisler, 2009; Gander et al., 2011).

Balkin et al. (2011) summarized various technological approaches to fatigue management that accurately predict and/or objectively monitor sleepiness and fatigue effects in real time and that allow for optimally timed administration of interventions or countermeasures (such as the use of properly dosed pharmaceuticals or napping).

Fig. 4. Restriction of nocturnal sleep in healthy adults resulted in near-linear increases in PVT lapses (>500 ms reaction times) of attention across 14 days (coefficients of change near 1.0), but subjective ratings of sleepiness, measured using the Stanford Sleepiness Scale (SSS), showed a nonlinear coefficient below 0.5 for change over days. As objective performance continued to decline near-linearly, there were only minor further increases in the subjective ratings of sleepiness. By the end of the 14 days of sleep restriction, when PVT performance was at its worst levels, subjects in the 4 and 6-h sleep period conditions reported feeling only slightly sleepy. Therefore, unlike performance measures, sleepiness ratings appeared to show adaptation to sleep restriction (N=35 subjects; 8 h condition n=9, 6 h condition n=13, and 4 h condition n=13). (Reprinted with permission from Van Dongen et al., 2003.)
When discussing the ideal monitoring system, the authors state that technological approaches should be valid, reliable, sensitive, specific, and generalizable. In order to adhere to these criteria, the approach would have (a) the ability to predict fatigue, based on the factors that produce it (sleep history, hours of wakefulness, circadian rhythms, time of day, and biomarkers or behavioral correlates of sensitivity to sleep loss); (b) the ability to measure and monitor fatigue/performance online in the operational environment in order to detect downward trends in alertness or performance before such trends reach the threshold at which operational performance is actually impacted; and (c) the ability to effectively intervene when potential deficits are identified or anticipated, with interventions calibrated so as to restore and sustain alertness/performance as long as needed (e.g., until the operator returns home after the work shift and can safely obtain adequate, recuperative sleep) (Balkin et al., 2011). Examples of instruments currently being examined for such use include EEG measurement of brain wave activity, ocular measures, video-based monitoring systems and portable assessments of attention, and reaction time. Fatigue detection and prediction instruments, implemented in addition to fatigue-related regulations, will help ensure workplace and road safety and adequate levels of alertness in workers while on duty (Balkin et al., 2011).

Conclusion

Despite being treated as a commodity that can be traded for other activities, sleep is essential for optimal neurobehavioral functioning. The 24/7 schedule of many occupations in current industrialized societies as well as other social- and work-related expediencies has produced a high percentage of habitually sleep-restricted adults, a trend that will to continue in the decades ahead. Although chronic partial sleep restriction leads to increased sleep propensity and neurobehavioral deficits, adults are not adept at accurately judging how affected they are by this sleep loss. These effects can have serious adverse consequences including fatal medical errors, catastrophic oil spills, and vehicle crashes. Finding biological and behavioral markers of sleep loss vulnerability and developing technological approaches to fatigue management are two domains of research essential for offsetting the negative effects of societal factors on the biological need for sleep.

Acknowledgments

Supported by NIH NR004281; the National Space Biomedical Research Institute through NASA NCC 9-58; and the Department of the Navy, Office of Naval Research (Award No. N00014-11-1-0361).

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>COWAT</td>
<td>Controlled Oral Word Association Test</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>KSS</td>
<td>Karolinska Sleepiness Scale</td>
</tr>
<tr>
<td>MSLT</td>
<td>Multiple Sleep Latency Test</td>
</tr>
<tr>
<td>MWT</td>
<td>Maintenance of Wakefulness Test</td>
</tr>
<tr>
<td>NREM</td>
<td>Non-rapid eye movement</td>
</tr>
<tr>
<td>PSG</td>
<td>Polysomnography</td>
</tr>
<tr>
<td>PVT</td>
<td>Psychomotor Vigilance Test</td>
</tr>
<tr>
<td>REM</td>
<td>Rapid eye movement</td>
</tr>
<tr>
<td>SSS</td>
<td>Stanford Sleepiness Scale</td>
</tr>
<tr>
<td>SWE</td>
<td>Slow wave energy</td>
</tr>
<tr>
<td>TIB</td>
<td>Time in bed</td>
</tr>
</tbody>
</table>

References


regulation is not mediated by the three cone photopic visual system. *Journal of Clinical Endocrinology and Metabolism*, 86, 433–436.


