Sleep Deprivation and Vigilant Attention

JULIAN LIM \textsuperscript{a,b} AND DAVID F. DINGES\textsuperscript{b}

\textsuperscript{a}Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
\textsuperscript{b}Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

Sleep deprivation severely compromises the ability of human beings to respond to stimuli in a timely fashion. These deficits have been attributed in large part to failures of vigilant attention, which many theorists believe forms the bedrock of the other more complex components of cognition. One of the leading paradigms used as an assay of vigilant attention is the psychomotor vigilance test (PVT), a high signal-load reaction-time test that is extremely sensitive to sleep deprivation. Over the last twenty years, four dominant findings have emerged from the use of this paradigm. First, sleep deprivation results in an overall slowing of responses. Second, sleep deprivation increases the propensity of individuals to lapse for lengthy periods (>500 ms), as well as make errors of commission. Third, sleep deprivation enhances the time-on-task effect within each test bout. Finally, PVT results during extended periods of wakefulness reveal the presence of interacting circadian and homeostatic sleep drives. A theme that links these findings is the interplay of “top-down” and “bottom-up” attention in producing the unstable and unpredictable patterns of behavior that are the hallmark of the sleep-deprived state.

Key words: sleep deprivation; psychomotor vigilance; lapses; time-on-task; caffeine; modafinil; amphetamine

Introduction

The complexity of the failures in vigilant attention is underscored by the diversity of compounds that can reverse the deficits caused by sleep deprivation. Caffeine, amphetamines, and modafinil, each of which has different molecular targets, all cause similar improvements when administered to sleep-deprived individuals. Because of its serious real-world consequences, elucidating and understanding these biological substrates is an important and urgent current topic of investigation.

The link between sleep and the capacity to attend to external stimuli is both intimate and inextricable. To the nonexpert, this fact may seem so intuitive as to be almost trivial. Is there any question, for example, that one would prefer to do an important piece of work after a good night of sleep rather than at midnight after 16 hours of continuous wakefulness? Most of us are aware that being sleep deprived for an extended period of time can feel like a physical force acting in the brain, compelling the eyes to shut and mental processing to “switch off,” resulting in intrusive and unwanted lapses in attention.

Early attempts to understand the effects of sleep deprivation (SD) on attention were by necessity restricted to behavioral observation and experimentation. However, with the advent and development of methods such as EEG, PET and fMRI, as well as leaps in our understanding of the molecular biology of the brain, we are beginning to form a clearer picture of the relationship between attention and sleep deprivation at the cortical and cellular level. We are also beginning to understand how these mechanisms are integrated and interact across these observational scales. In the process, researchers are uncovering layers of complexity underlying the simple notion that sleep deprivation compromises our ability to pay attention.

In this chapter, we review the body of research on the effects of sleep deprivation on attention, with special focus on vigilant attention and what is known of its underlying neural substrates. First, we briefly discuss the psychomotor vigilance test, and its importance as an assay of neurocognitive capacity. We then examine four common patterns of behavior that are observed in tests of attention, and particularly the psychomotor vigilance test, following sleep deprivation, and discuss the putative neural basis of these changes. Finally, we turn our attention to the most commonly used wake-promoting drugs, reviewing how our knowledge of their modes of action implicate various molecular systems in the maintenance of attention and its failure after SD.
Why Vigilant Attention?

“Attention” has a wide range of meanings in the psychological literature, and numerous classificatory systems have been suggested to isolate its component parts. William James was among the first psychologists to note that “attention” is not a unitary process, and that its independent components include “voluntary” and “involuntary” attentional processes. These processes correspond approximately to what we now label “top-down” and “bottom-up” attention. Top-down attention is driven by knowledge-based mechanisms that enhance the contrast between signal and noise, and produce biases toward relevant stimulus features. In contrast, bottom-up attention refers to target detection that is driven by alerting features inherent to a stimulus. These are popular and often-discussed constructs, and are of some use in understanding attention following SD. Sarter notes, however, that top-down and bottom-up attention are theoretical constructs that do not necessarily have anatomically distinct correlates. At a neurobiological level, therefore, it is hard to tease apart how or whether these two attentional components are differentially affected.

In a different body of work, Sturm and colleagues have proposed a multicomponent model of attention based on its classification into “intensity” and “selection” aspects. In this model, the intensity (or tonic) aspects of attention, which include alertness and sustained attention, are more fundamental than the selection aspects, which include orienting and executive attention. Robertson and Garavan reiterate this point, arguing that vigilance is functionally distinct from selective aspects of attention, and closely related to the ability to inhibit peripheral or distracting stimuli. The implication of this model is that vigilant attention can wax and wane over the course of seconds, not minutes or hours, as previously hypothesized. Moreover, the ability to carry out the more demanding selective and executive aspects of a task is contingent on an agent being able to sustain attention to the task at hand. Thus, the variance associated with any deficit in selective attention must be at least partially shared with the variance attributable to declines in vigilance.

This theme—the fundamental importance of vigilance to all other aspects of cognition—has been taken up and explored by researchers of sleep deprivation. Indeed, cognitive deficits after SD can be observed in a wide range of domains, including memory, executive function, and the multiple facets of attention mentioned earlier. However, it is likely that much of this poorer performance is attributable to the inability to sustain attention to the task at hand, as this is a prerequisite for all upstream cognitive processing.

A companion hypothesis to this is the controlled-attention model. This theory arose from the observation that complex cognitive tasks tend to be affected far less, if at all, during periods of interference, or in individuals with lower attentional capacity (for example, patients with attention deficit hyperactive disorder [ADHD]). Controlled attention is a synthesis of intensity and selective aspects of attention, and its requirements are greater for tasks that are not intrinsically engaging, because of the greater need in these tests to inhibit non-task-related distractions. Thus, sleep-deprived subjects do not need to exert as much effort to engage controlled attention when performing complex challenges as compared to basic, unengaging tasks, accounting for the much greater declines on tests of vigilant attention.

Although the two preceding hypotheses approach the behavioral data from slightly different perspectives, it is universally agreed that vigilance is the component of cognition that is most consistently and drastically affected by periods without sleep. It is also of great interest to sleep researchers due to the fact that many tasks requiring this facility have good ecological validity; indeed, its failure can result in devastating real-world consequences. For example, characteristics of vigilance tasks very closely resemble the operational demands of those in jobs requiring long periods of sleep deprivation, particularly military personnel and commercial pilots.

For all these reasons, it is critical in both pure and applied science that the behavioral changes in vigilant attention after SD, as well as the biological underpinnings of these changes, be clearly elucidated. We thus spotlight vigilant attention in this chapter to emphasize its fundamental importance in the field, as well as highlight the breadth of research that bears relation to the topic.

The Psychomotor Vigilance Test

Over the last two decades, the instrument that has emerged as the dominant assay of vigilant attention in paradigms of sleep deprivation is the psychomotor vigilance test (PVT). This simple yet extremely informative task has been employed as a marker of attentional deficit in hundreds of studies to date. The test is highly sensitive to SD, and its reliability and validity have been amply demonstrated. Because of its utility and widespread use, we describe the test in some detail here as the archetype of a neurocognitive assay of attention after sleep loss.
The PVT is a test of simple reaction time (RT) to a cue that occurs at random inter-stimulus intervals (ISI). In the standard administration, the ISI varies randomly from 2 to 10 s. Across the duration of the run, the distribution of reaction times is flat in this range. The standard test is 10 min long. During this time, subjects are seated comfortably, and instructed to attend to a small, rectangular area on a dark screen. They are told to respond as rapidly as they are able whenever they perceive the appearance of a bright millisecond counter inside this rectangular area. Stopping the counter allows subjects to view their reaction time, which remains on the screen for a duration of 1 s, and serves as feedback for that particular trial. Button presses when the counter is not displayed on the screen are counted as false starts, or errors of commission, which subjects are instructed to avoid.

Numerous outcome measures can be collected from a single 10-min bout of the PVT. Although the number of lapses, or responses greater than 500 ms, is often used as the primary dependent variable in the test, important information can also be obtained from the median RT, errors of commission, the variability in RTs, and the slope of reciprocal RTs across the run (which is a measure of the time-on-task effect).

Various properties of the PVT make it a particularly suitable test of vigilance for studies of SD. These are summarized briefly later in this chapter; for a more comprehensive discussion of the subject, see Dorrian and Dinges. First, the PVT has a high signal-load; this allows for the collection of a large amount of data in a relatively short period of time. This signal load increases the sensitivity of the test in detecting even relatively small changes in attentional function without making the task so onerous that it depletes motivation. The test is reliable, with intraclass correlations measuring test–retest reliability at 0.826 (P < .0001) for median response times, and 0.888 (P < .0001) for number of PVT lapses. The convergent validity of the test has been demonstrated through its sensitivity to both total and chronic partial sleep deprivation, as well as intervention with psychoactive, wake-promoting drugs. Finally, the test shows very minor learning effects, making it suitable for regular repeated administration over the course of hours or days.

**Effects of Sleep Deprivation on Vigilant Attention**

The widespread use of the PVT as an outcome measure in experiments of SD has yielded a vast amount of data in the field over the past 15 years. Broadly, these results have revealed four large areas of behavioral change, namely that:

1. Sleep deprivation causes a general, overall slowing of reaction times.
2. Sleep deprivation results in increased errors of omission and commission.
3. Sleep deprivation enhances the time-on-task effect.
4. Tests of vigilant attention during periods of SD are sensitive to both circadian and homeostatic drives.

Together, these discoveries have been consolidated into and form the leading theoretical paradigm describing performance after SD: the state-instability theory. According to this theory, numerous competing systems work to exert an influence on behavior during periods of extended SD. Chief among these are an involuntary drive to fall asleep and a counteracting top-down exertion to sustain alertness. The interaction of these drives results in unpredictable behavior, including heightened RT variability, as well as the lapsing and false starts that occur sporadically and randomly throughout each test bout.

The unpredictability of neurobehavioral outcomes displayed by sleep-deprived subjects suggests that extended wakefulness produces a liminal state in which multiple biological modules attempt to gain control of brain and behavior. Indeed, our current state of knowledge indicates that the neurobiology of sleep deprivation is highly complex, and the systems that subserve SD-related changes in behavior are, in all likelihood, not independent. Nevertheless, we attempt to survey and synthesize the literature on the subject here, addressing each of the four changes listed earlier.

**Sleep Deprivation Causes a General, Overall Slowing of Reaction Times**

Although there is substantial interindividual variability in vulnerability to sleep loss, average reaction times on the PVT increase in length overall after a period of sleep deprivation. This generalized response slowing is also reflected through a worsening of the fastest 10% of RTs on both visual and auditory vigilance tasks. The increase is independent of the fact that subjects are also “lapsing” (defined on the PVT as responding more than 500 ms after the stimulus onset), a phenomenon that will be discussed in the following section.

Because virtually all but the fastest RTs on the PVT are affected, it is likely that cognitive slowing is associated with general state-related changes in brain
activity. These changes have been explored using a variety of instruments and methods. By employing electrophysiological measures, steady fluctuations in brain activity have been observed across periods of SD. General drowsiness is associated with frontal increases in theta-band activity, and global increases in resting alpha power.36 In agreement with this, Cajochen et al.37 found progressively increasing power in the 6.25–9-Hz (high theta/low alpha) band of the EEG power spectrum over 34 h of total SD, and drew a link between this and buildup of the homeostatic sleep drive. This result has been replicated in a number of SD paradigms.30,38,39 Cross-correlational analysis has shown that slow-eye movements (SEMs) during periods of eye closure10 and subjective sleepiness ratings40,41 are good predictors of these EEG power changes. Increases in absolute theta power are also moderated by body posture, with attenuation in standing compared to supine subjects.42

A number of EEG markers have been specifically correlated with performance on tests of vigilant attention following SD. Using multivariate EEG, Makeig and Jung43 found that a single principle component of EEG spectral variance was predictive of reaction time on a test of alertness. Of the several frequencies that load onto this component, it has been suggested that decreased beta activity is most strongly associated with vigilance changes.44 Later, Jung43 reported that full-spectrum EEG power was a marginally better predictor of reaction times than the single principle component, or any of its constituent frequencies.

Across the substantial body of work on this topic, it has been generally noted that intersubject EEG outcomes only show strong correlations with performance on vigilance tests under conditions of severe impairment, due in part to the relatively poor signal-to-noise ratio of EEG recording.46 Moreover, scalp electrophysiology is limited in its ability to localize changes in function to specific neural regions. To achieve this, researchers have turned to neuroimaging methods such as PET and fMRI, both of which have spatial resolution adequate to the task. Brain regions affected by SD are to a large extent task-dependent; however, certain areas—in particular the thalamus, anterior cingulate cortex, middle prefrontal gyrus, and inferior parietal lobes—do appear to show hypoactivation after SD across a large number of paradigms. Although a metaanalysis of the literature would provide stronger evidence for the claim, we tentatively suggest that this brain network is responsible for the main, state-related changes in vigilance that drive global slowing of response time.

Early PET researchers were among the first to report these changes. Using a continuous-performance test, Wu and colleagues37 revealed strikingly different patterns of metabolic activity in sleep-deprived compared to rested subjects. Although global mean cerebral metabolic rate did not change, decreases were seen in the thalamus, basal ganglia, and cerebellum. The frontal and temporal lobes also showed significant decreases in absolute metabolic rate after SD. Greater decreases in vigilant attention after SD (measured by reaction times) were associated with greater decreases in absolute metabolic rates.

This shift in functional hemodynamics after SD has been replicated in a number of other PET studies. In an experiment with a shared but expanded subject pool to their original work,37 Wu et al.46 found the same thalamocortical decreases in activation after 24 h of SD, and also observed that these decreases were only partially reversed by one night of recovery sleep. Thomas et al.49,50 found both global and regional decreases in cerebral metabolic rate across an 85-h SD period for subjects performing a serial addition/subtraction task. Relative decreases were observed in bilateral prefrontal cortex, dorsal and ventral anterior cingulate, dorsal and ventral thalamus, middle and inferior temporal gyrus, and in medial temporal cortex. On a group level, decreases in activation in the thalamus, parietal, and prefrontal cortices were correlated with both alertness and cognitive performance over time.

fMRI paradigms have found decreased activation in a very similar network of areas. In sustained attention tasks administered to well-rested subjects, good performance is most closely linked to activation in a right fronto–parietal network of regions.51,52 The thalamus and reticular activating pathway has also been implicated with rapid responding.53 Finally, the importance in task-related deactivation of certain brain regions has been noted by several authors to be critical to maintaining a high level of vigilance.51,53–55 These areas constitute the “default-mode network,” a set of brain areas that show higher levels of cerebral blood flow at rest than during cognitive task engagement.56 Default areas show anticorrelated activity to attention-related areas, and both networks are vital to optimal task performance.57

Portas et al.58 studied the effect of arousal as a moderator of brain activity to an attentional task, using manipulations of caffeine (high arousal), and sleep deprivation (low arousal). The task had a short duration in order to equalize performance across conditions and ensure that differences were attributable solely to arousal state. The authors found that the thalamus was the area primarily affected by arousal-state changes,
with greater activity in this region after SD. Areas associated with attention but not modulated by arousal were superior and posterior parietal cortices, anterior cingulate cortex, and dorsolateral prefrontal cortex.

A series of experiments of performance and brain activation on a Sternberg-like memory task have led to the suggestion that the inferior parietal sulcus (IPS) plays an important role in modulating attention following SD. On this task, the closest associations between brain activation and behavior were between declines in surrogate measures of attention (as opposed to working memory) and reductions in activation in bilateral inferior parietal cortex after 24 h of total SD. Using a task that measures short-term memory capacity, Chee and Chuah also found task-related reductions in IPS activity that were disproportionately lower than performance declines after SD, lending credence to the suggestion that the area is implicated in attentional dysfunction in sleep-deprived individuals.

**Sleep Deprivation Results in Increased Errors of Omission and Commission**

Lapsing, or failing to respond in a timely fashion to a presented stimulus, is a hallmark of the sleep-deprived state. On the PVT, this is defined as any reaction exceeding 500 ms in length. Sleep-deprived individuals experience so-called “microsleeps” and slow eyelid closures, and these are typically the intervals during which prolonged lapses occur. Lapses also grow more frequent as cumulative wakefulness increases, making this variable a useful outcome measure of the PVT.

An important but unanswered question in the field is whether “lapses” as sleep researchers traditionally understand them are simply slower responses, or have qualitatively different neural signatures. Certainly, microsleeps are discrete phenomena, detectable by EEG and ocular tracking, where a subject is momentarily but definitively in stage I or II sleep. However, not all suboptimal responses on the PVT can be explained in this fashion, particularly since subjects occasionally lapse even in a well-rested state. As a result, researchers of lapses have worked from the assumption that many slow responses simply fall on the extreme end of a continuum, and are a result of perceptual, processing, or executive failures in the central nervous system. SD amplifies the tendency of this system to fail, both more often and for longer periods of time.

In support of this formulation, a strong correlation has been found between the number of lapses made by a subject and the duration of those lapses, with a correlation of .75 between the two variables after controlling for intersubject variance and repeated measures.
FIGURE 1. The average number of lapses on a PVT bout is highly correlated with the average lapse duration. The top left panel contains data from an 88-h total sleep deprivation paradigm, and the remaining panels depict performance over a 2-week chronic sleep-restriction protocol (4-, 6-, and 8-h time-in-bed). After controlling for subject and repeated measures effects, the correlation between the variables was .75 (P < .0001).

group did not find differences on the P300 component), and Corsi-Cabrera et al.74 on a simple reaction-time test. These ERP changes occur over a diverse area of cortical regions, with the P300 thought to originate from the anterior cingulate cortex, prefrontal areas, as well as the temporal–parietal junction, and the N1 found over sensory cortex. This suggests there is no specific area responsible for lapsing, but rather a network of regions that contributes to failures of attention.

The neural correlates of slow responding have been explored in a couple of recent fMRI studies. Drummond et al.54 imaged PVT performance in 20 healthy adults at rested baseline and after a period of 24 hours of SD. Using an event-related analysis of individual reactions, the authors found that fast reactions were associated with greater responses in the cortical sustained-attention network, as well as subcortical arousal and motor systems, while, particularly after SD, relatively slower responses were associated with a failure to disengage default areas.

Weissman and colleagues55 used a more sophisticated fMRI analysis method to investigate lapsing, albeit only in the rested state. By employing individualized regressors based on the reaction time for each trial on a global–local selective attention task, the authors were able to uncover regions of the brain that showed patterns of hemodynamic response that varied according to response speed. Slow responses were predicted by the failure of early engagement of anterior cingulate cortex, middle, and inferior frontal gyrus, as well as increased target-related activity in the default-mode network. This failure in preparedness presumably led to degraded perceptual input, which resulted in greater, compensatory fMRI response in frontal and parietal cortices. Finally, the authors found that activity in the right temporal–parietal junction and right inferior frontal cortex was related to reaction time on the subsequent trial, and suggested that these areas are implicated in re-orienting mechanisms that facilitate future performance. This network of regions corresponds well to the origins of evoked potentials in the studies discussed earlier.

When considered in tandem, the results of Drummond et al.54 and Weissman et al.55 support the conclusion that many, though not all, lapses after SD are
FIGURE 2. Number of sleep attacks, or 30-s lapses on a 20-min PVT (every 2 hours) over 42 h of total sleep deprivation. Bars show mean (SEM) number of failures to respond for n = 14 healthy adult subjects. Numbers above bars are percentage of subjects who had at least one failure to respond at each test bout time. Failures began at midnight after 16 h awake and peaked at 0700 (7 A.M.) at 26 h awake.

simply an enhancement of a tendency already nascent in well-rested individuals, although this remains to be empirically tested. Although further research is needed to identify the specific elements of failure in this complex system, current neurobiological evidence supports the notion that sleep deprivation does not fundamentally change the nature of regular responding, but rather slows it down due to a number of bottlenecks in neural processing.

What then, of the more catastrophic lapses that are not simply slow, but a result of progression into true sleep? Microsleeps, or lapses into true sleep lasting several seconds, have been attributed to a weakening of the inhibitory mechanisms in the ventrolateral preoptic nucleus (VLPO). The VLPO has been identified as a “flip-flop switch” that is stable during rested wakefulness, but highly sensitive to small neurochemical perturbations after SD. Because of this instability, the sleep switch is prone to being involuntarily tripped after SD, most often without prior warning. This is only one putative mechanism among numerous plausible candidates; other brain regions and neurotransmitter systems include the ascending reticular activating system, the basal forebrain, the orexin-hypocretins, as well as specific monoaminergic and brainstem nuclei (e.g., the locus coeruleus).

The features of catastrophic lapse occurrence emphasize their qualitative distinctness. On the PVT, catastrophic lapses are those in which there is no response for 30 s. These extreme lapses—which we believe are functional sleep attacks—start appearing after 5–6 days of chronic sleep restriction (4–6 hours TIB), and, unlike shorter lapses, are completely absent in individuals with a full quota of sleep (FIGS. 2 and 3). Even though time-outs are relatively rare, they are nevertheless modulated by circadian and homeostatic drives, first appearing after 16-hours time awake with peak occurrences at the circadian nadir. However, unlike lapses of shorter duration, 30-s sleep attacks are not completely unwarned. Plotted retrospectively, average RTs 5 min prior to a sleep attack increase in a linear fashion, suggesting that greater instability in the waking state puts subjects at greater risk for falling into involuntary true sleep.

It is worthwhile noting that errors of commission, or false alarms, also increase in number during SD, and show the same pattern of circadian modulation as lapses. The appearance of false alarms may reflect a compensatory response to drowsiness; however, little work to date has been done to investigate this behavior.

The two features described so far—generalized response slowing and lapsing—explain the change in the
distribution of reaction times that occurs following SD (Fig. 4). As cumulative wakefulness increases, the right tail of the RT distribution shifts, increasing the skewness of the left portion of the curve, as well as its second modal peak; this is one of the defining characteristics of the impact of sleep deprivation, as opposed to other challenges, on tests of vigilant attention.

More recently, it has been suggested that gamma curves may be useful in synthesizing information across multiple test bouts of the PVT. This method involves plotting a cumulative distribution function (CDF) of reaction times across multiple tests during a baseline period, as well as the relevant periods of sleep deprivation (Fig. 5), and then computing a difference function by subtracting one curve from the other. The maximum difference between the two curves (baseline versus SD) serves as a single coefficient reflecting the level of impairment of an individual. Preliminary data suggest that this is a valid way of classifying individuals as vulnerable or resistant to the deleterious effects of SD, and that these phenotyping coefficients are stable over periods of time. Difference values also have utility in investigating the sensitivity and specificity of the PVT, for instance, computing the threshold reaction time that maximally discriminates rested individuals from individuals first experiencing sleep-deprivation-related impairment.

**Sleep Deprivation Enhances the Time-on-Task Effect**

The time-on-task (TOT) effect describes the phenomenon whereby performance worsens across the course of a cognitive task owing to fatigue or other factors (e.g., boredom or diminishing motivation). This decline is usually measured using either the change in reciprocal response speed or number of lapses over time. Originally thought to be present only in tasks of considerable duration (30 min or greater), it has since been found that TOT decrements are measurable within the first several minutes of performance in sleep-deprived individuals.

Sleep deprivation greatly enhances the TOT effect, especially in operations with high cognitive demand. However, as hours of wakefulness increase, these TOT decrements do not change in a straightforward fashion. Initially, SD changes the slope of responses across a task. However, as time awake increases further, the intercept, representing the average reaction time of the first few PVT trials, shifts downwards, causing the slope to level off. This change arises when subjects are no longer to compensate for their attentional deficits, even for very short periods of time. Finally, the downward movement of the intercept stops and the slope of responses once again increases.
Various moderators have been found that influence the degree of impairment caused by TOT. In an experiment using both a clock vigilance test (monotonous) and a driving simulator (engaging), Richter and colleagues found relatively greater TOT impairment in the monotonous task. Steyvers and Gaillard reported that TOT-related declines can be reversed by incentives or rewards. Both these results suggest that there is at least a certain degree of top-down compensation that subjects can exert to bolster performance; the reality, however, is that declining motivation is in and of itself an integral part of the TOT effect, and should permissibly be treated as such in future experimental work.

The biological basis of individual subject and task differences in TOT after SD has not been explored, but it is likely that many of the same systems that subserve general response speed slowing play a role in this deficit. For example, caffeine and modafinil administration during the period of SD partially attenuates TOT effects, while a study of insomniacs versus controls has shown that it is enhanced in the clinical population.

**The PVT Is Sensitive to Both Circadian Effects and Homeostatic Sleep Drive**

Human sleep–wake behavior is most commonly modeled using the two-process model of sleep regulation. This model consists of two interacting components: a circadian process, which is a sinusoidal oscillator with a 24-h period, and a homeostatic process, which increases exponentially with time awake, and dissipates in a similar exponential fashion. Initially, this model was applied to predict sleep propensity; however, it soon became clear that aspects of cognitive function could be forecast using the model in a similar fashion.

Performance on the PVT is affected by both circadian and homeostatic drives; FIGURE 6 shows PVT data from an 88-h total SD paradigm, in which both a steadily increasing linear trend and an oscillating circadian rhythm can clearly be seen. The number of lapses and the slowest (10th percentile) reaction times are particularly sensitive in tracking this pattern. Critically, the pattern in this figure informs us that decline in performance over time during SD is not unidirectional, for example, “circadian rescue” can account for better performance in the morning after 48 hours of continuous wakefulness compared to the preceding hours of the night. Additionally, as time awake increases, the homeostatic drive interacts with, and exerts a multiplier effect on the circadian cycle, thus amplifying performance deficits at each circadian nadir.

Edgar et al. elaborated on the neurobiological basis of the circadian process in their “opponent process” model of sleep–wake regulation. This updated model highlights the probable role of the suprachiasmatic nucleus (SCN) of the anterior hypothalamus as an endogenous circadian pacemaker. The SCN has direct axonal projections to the posterior hypothalamus,
and regulates arousal through the action of melatonin and hypocretins. Although there are substantial interspecies differences, the general effect of SCN lesions is to disrupt the consolidation of sleep (i.e., during periods when the circadian drive to sleep is typically high), without decreasing total sleep time.

The neural substrate of the homeostatic process remains in dispute. One putative mechanism is thought to be adenosine, although this contention is still under debate. With increasing time awake, brain glycogen and ATP levels in the animal brain are steadily depleted due to metabolic demand, with adenosine as

FIGURE 5. Gamma distributions as a way of representing PVT data. All individual reaction times across a period of interest are plotted as a cumulative distribution function. The left panel shows data from an 88-h total sleep-deprivation (SD) protocol; the right panel shows data from a 14-day partial sleep-restriction protocol (4-h time-in-bed). One potential use of these curves is to calculate a cutoff point for lapses that produces maximum discriminability between different groups of interest. For example, using a 500-ms threshold, there is a 19.2% difference between subjects at baseline and after 88-h SD, and a 16.7% difference between baseline and performance after 14 days of chronic sleep restriction.
FIGURE 6. Errors of omission and commission are modulated by circadian and homeostatic drives. White circles represent subjects undergoing 88 h of total SD, and black squares represent control subjects (8-h time-in-bed).

the final product of the breakdown of ATP. In rodents, sleep deprivation leads to an increase in the levels of endogenous intra- and extracellular adenosine in the basal forebrain.\(^90\)–\(^92\) An \textit{in vivo} PET study of the adenosine \(A(1)\) receptor in humans after prolonged wakefulness has also demonstrated its up-regulation in widespread cortical and subcortical regions.\(^93\) Adenosine is thought to have an inhibitory effect on wake-promoting neurons in the basal forebrain,\(^94\) as well as increase sleep pressure by acting on the VLPO.\(^95\) This parsimonious hypothesis has been recently challenged by Blanco-Centurion \textit{et al}.\(^96\), however, who showed that direct lesions in the basal forebrain of rats disrupted the build-up of extracellular adenosine, but did not affect homeostatic sleep drive after 12 h of wakefulness. Moreover, the administration of an adenosine agonist to the basal forebrain induced sleep in these lesioned rats. Thus, adenosine’s involvement in sleep promotion is not as a direct mediator, and further work needs to be done to address this question.

The consolidation of circadian and homeostatic information from the SCN and the basal forebrain is thought to be carried out in the midbrain structures of the medial preoptic area and the anterior paraventricular thalamic nucleus.\(^97\) Ascending projections from the diencephalon subsequently feed this information forward to higher cortical areas through a number of neurotransmitter systems,\(^98\) and it is likely that these pathways are responsible for the modulation of gross aspects of attention. In support of this theory, \textit{in vivo} neuroimaging studies have shown that thalamocortical activation does vary based on both the sleep homeostat\(^49\),\(^50\) as well as the time of day during which the scan was conducted.\(^60\)

### Attention-Modulating Compounds and Their Molecular Targets

Although the cellular and molecular changes that occur during periods of sleep deprivation are becoming increasingly well-understood, little research has been conducted that directly links these neurochemical changes to the modulation of arousal and alertness. Nevertheless, strong inferential conclusions can be drawn from results of studies of psychoactive compounds that temporarily attenuate or reverse the effects of SD. The most commonly used compounds at present are caffeine, amphetamines, and the relatively new drug modafinil; these stimulant and wake-promoting pharmacological agents have been extensively studied for their ability to enhance cognitive functioning following periods of SD.\(^99\) An extensive discussion of each of these compounds is beyond the scope of this review, so here we present a brief survey of the literature.
**Caffeine**

Caffeine is the most commonly used legal psychoactive stimulant in the world, and can be readily self-administered in coffee, tea, and many carbonated colas. Its arousing effects can be observed within an hour of ingestion, both through decreased subjective sleepiness and improvement in objective measures of vigilance. Caffeine routinely improves performance on the PVT,\textsuperscript{100–102} longer visual vigilance tasks,\textsuperscript{103,104} more complex executive attention tasks,\textsuperscript{103} and reduces errors made on a driving simulator.\textsuperscript{105} Subjects also report lower levels of subjective sleepiness when ingesting caffeine versus placebo.\textsuperscript{105,106} The pharmacological effects of caffeine have been reasonably well studied. Caffeine is an alkaloid compound that acts as an antagonist on adenosine receptors, causing downstream increases in dopaminergic and glutamatergic activity.\textsuperscript{107,108} Its stimulant effects have been specifically attributed to binding at adenosine A(2A), and not A(1) subtype receptor sites.\textsuperscript{109} As discussed, the accumulation of adenosine in the basal forebrain has been implicated in escalating homeostatic sleep drive—the fact that caffeine both counteracts the neurochemical changes and reverses the behavioral consequences of SD provides buttressing evidence for this conjecture. Adenosine A(2A) receptor mRNA is found exclusively in portions of the ascending arousal pathway, including the ventral and dorsal striatum,\textsuperscript{110,111} making it a plausible candidate molecule responsible for the declines in sustained attention following SD. Indeed, preliminary evidence in a rat model has directly linked this adenosinergic build-up with performance impairments on a rodent version of the PVT.\textsuperscript{112}

More recent studies have focused on the association between individual differences in caffeine sensitivity, which are mediated by a polymorphism (c.1083T > C) of the adenosine A(2A) receptor subtype gene (ADORA2A).\textsuperscript{113,114} Differences in ADORA2A predict the effects of caffeine on both the subjective quality of sleep and the differences in the EEG correlates invoked by caffeine consumption during subsequent nREM sleep.\textsuperscript{114} Retey et al.\textsuperscript{115} found that caffeine-sensitive individuals showed greater impairment than caffeine-insensitive individuals on PVT performance after 24 hours of SD, but that caffeine administration reversed this pattern of impairment. The anterior–posterior distribution of EEG theta power was also different between groups, with caffeine-sensitive individuals showing greater increases in a frontal derivation. The authors suggest that these adenosine receptor polymorphisms form the neurobiological basis of the stable interindividual vulnerabilities to SD observed by Van Dongen et al.\textsuperscript{27} and Leproult et al.\textsuperscript{116}

**Amphetamines**

The two main classes of amphetamines are l- and dextro (d-)amphetamine, which differ in their isomer composition. Of these, dextroamphetamine is the more abundant subtype, and is also almost twice as potent as l-amphetamine in its arousing effects.\textsuperscript{117} Relative to placebo, amphetamines improve vigor and reaction-time performance during sleep deprivation,\textsuperscript{118–121} and are considered among the most effective compounds for doing so.

Various studies have specifically tested the effects of d-amphetamine on vigilance. Cochran et al.\textsuperscript{122} found that 20 mg of d-amphetamine administered after 63 h of SD improved performance on an auditory RT task relative to placebo. Visual vigilance was also boosted by 20 mg d-amphetamine administration after 40.5 h of SD.\textsuperscript{118} Hartmann and colleagues\textsuperscript{123} compared the vigilance-improving effects of d- and l-amphetamine, and concluded that the former is more effective in its enhancement of this cognitive facility.

Outside of the laboratory, amphetamine use by military pilots has been a focus of study over the last decade. Dextroamphetamine has been reported to have been used by pilots flying combat missions requiring extended wakefulness, and its users overwhelmingly show objectively improved performance, and report subjective benefits of taking the stimulant.\textsuperscript{124–126} The outcome measures in these field experiments are typically identical to predictors of costly errors in actual combat situations.

Endogenously, amphetamines bind to monoamine transporters, thus blocking the reuptake of, and subsequentially increasing postsynaptic levels of dopamine, serotonin, and noradrenaline.\textsuperscript{127,128} Of these neurotransmitters, it is thought that dopamine is the molecule primarily responsible for mediating the arousing effects of amphetamine.\textsuperscript{129,130} For example, dopamine knockout mice do not experience the arousing effects of consuming classic stimulants,\textsuperscript{131} and Drosophila with mutations in the dopamine transporter gene display high levels of activity and reduced sleep need.\textsuperscript{132}

**Modafinil**

Modafinil is a wake-promoting atypical stimulant drug prescribed for the treatment of narcolepsy and excessive daytime sleepiness. When taken after a period without sleep, modafinil reverses deficits in vigilant attention,\textsuperscript{23,118,133–135} speed of processing,\textsuperscript{136} executive attention,\textsuperscript{102,118,137} and performance on operational tasks.\textsuperscript{138,139} Modafinil attenuated the TOT effect on the PVT in subjects who had been sleep deprived for 54.5 h.\textsuperscript{23} In many of the experiments cited earlier, it has
been noted that the beneficial effects of modafinil are dose-dependent, with 50–100 mg every 24 h producing no significantly greater effect than placebo, and 200–400 mg every 24 h maintaining vigilance near baseline levels.23,102

The clinical benefits of modafinil are evident in the treatment of shift-work sleep disorder. In a study of 32 subjects undergoing simulated night-shift work, modafinil administration resulted in significant improvements in the number of lapses and slowest 10% of reaction times on the PVT.140 In a group of patients with shift-work sleep disorder, being on the drug improved objective performance (measured by PVT performance) and reduced the number of accidents and near-accidents experienced during commutes to and from work, although these patients continued to show impaired performance compared to controls.141 In military settings, modafinil is being advocated as a superior drug to amphetamine; although the two substances produce comparable behavioral effects, modafinil has fewer harmful side effects and greater acceptability.142

The receptor targets of modafinil are not entirely clear, although several studies have implicated dopamine, serotonin, and in particular norepinephrine transporters143–145 as potential mediators. The drug is hypothesized to increase the inhibition of sleep-promoting neurons in the VLPO via the action of norepinephrine; evidence for this comes from increases in plasma and urine norepinephrine levels following modafinil administration,146 as well as in vivo studies in rhesus monkeys.144

The effect of modafinil vis-à-vis the orexin-hypocretins has also been considered. Orexins are thought to be responsible for stabilizing the waking state, and their levels are reduced in narcoleptic individuals. There is disagreement, however, on the precise effect of modafinil on this system. On the one hand, modafinil increases FoS expression in wake-promoting, orexin-rich areas of the brain (the perifornical area).147 However, in a study of orexin (−/−) versus wild-type mice, modafinil surprisingly increased wake time in the mutant more than the null strain.148 This suggests that orexin may contribute to the arousing effects of modafinil, but does not on its own fully mediate the relationship between drug and behavior.

Although caffeine, amphetamines, and modafinil all produce improvements in vigilance in the sleep-deprived state, it is apparent that their receptor targets are somewhat dissimilar. Thus, even at the molecular level, it is probable that multiple systems subserve the changes in attention already discussed. Future research should produce better descriptions of the interactions of these neurotransmitter systems, as well as more closely integrate our knowledge of attention from the level of the neuron to the level of human behavior.

Summary and Conclusions

The impact of SD on attention is far from straightforward. The layers of complexity in this story arise from the fact that extended wakefulness affects a number of neurobehavioral systems that then influence one another in synergistic or antagonistic ways. Attention is also affected by the interaction of drives from multiple cortical and subcortical networks, including sleep pressure conceivably from diencephalic and basal forebrain areas, compensatory, top-down effort to remain alert from the prefrontal cortex, and motivation and engagement associated with dopaminergic pathways. Given this, it is perhaps unsurprising that performance following SD is both unpredictable and open to extremely high levels of interindividual variation.

Identifying the biological underpinnings of vigilant attention may aid us in understanding the neurocognitive effects of SD. The PVT, a simple, yet highly sensitive and valid task, has already afforded us significant insight into the behavioral features that characterize sleep-deprived individuals. Future research will doubtless focus further on the biological basis of these behavioral features, and allow us to implement improvements upon our present pharmacological solutions to counteract the deleterious effects of SD.

Indeed, the need for these solutions is urgent. It has been estimated that as many as 1 in 3 healthy adults obtain insufficient lengths of sleep,149 at considerable cost to society. Sleep deprivation leads to an increase in the risk of motor-vehicle accidents and near-accidents,150,151 as well as increases in on-the-job errors in a wide range of occupations, from truck drivers and train operators to medical professionals who are required to work long shifts with little-to-no intervening rest opportunity.152–154 It is arguable that many of these errors are attributable, at least in part, to failures of vigilant attention. By coming to a fuller understanding of SD and its deleterious effects, scientists put themselves in a better position to both create effective interventions and educate the public on its harmful and sometimes devastating consequences.

Competing Interest

The authors declare no competing interest.
References

36. STRIJKSTRA, A.M. et al. 2003. Subjective sleepiness correlates negatively with global alpha (8–12 Hz) and positively with central frontal theta (4–8 Hz) frequencies in...
64. DINGES, D.F. et al. 1999. Chronic sleep restriction: neurobehavioral effects of 4 hr, 6 hr, and 8 hr TIB. Sleep 22(Suppl. 1): 115–116.
72. GOSSELIN, A., J. DE KONINCK & K.R. CAMPBELL. 2005. Total sleep deprivation and novelty processing:


149. BONNET, M.H. & D.L. ARAND. 1995. We are chronically sleep deprived. Sleep 18: 908–911.

Acknowledgments

The work reported in this article was funded by National Institutes of Health Grant NR04821, by National Space Biomedical Research Institute through NASA Grant NCC 9-58, and by Air Force Office of Scientific Research Grant FA9550-05-1-0293.