San Jose State University

SJSU ScholarWorks

Fall 2020

Influence of Night Work on Performance during Lunar Telerobotic Operations

Zachary Luke Glaros
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Glaros, Zachary Luke, "Influence of Night Work on Performance during Lunar Telerobotic Operations" (2020). *Master's Theses*. 5143.
DOI: https://doi.org/10.31979/etd.58xt-r8s3
https://scholarworks.sjsu.edu/etd_theses/5143

INFLUENCE OF NIGHT WORK ON PERFORMANCE DURING LUNAR
TELEROBOTIC OPERATIONS

A Thesis

Presented to

The Faculty of the Department of Psychology

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Zachary L. Glaros

December 2020

The Designated Thesis Committee Approves the Thesis Titled

INFLUENCE OF NIGHT WORK ON PERFORMANCE DURING LUNAR
TELEROBOTIC OPERATIONS

by

Zachary L. Glaros

APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

SAN JOSÉ STATE UNIVERSITY

December 2020

Sean Laraway, Ph.D.   Department of Psychology

Susan Snycerski, Ph.D.   Department of Psychology

Erin E. Flynn-Evans, Ph.D.   NASA Ames Research Center

ABSTRACT

INFLUENCE OF NIGHT WORK ON PERFORMANCE DURING LUNAR TELEROBOTIC OPERATIONS

by Zachary L. Glaros

Real-time, reactive telerobotic mission control operations require personnel to actively operate remotely controlled vehicles or robots in real time. Due to the physical separation of the vehicle from the operator, such operations present additional factors that can influence fatigue (degraded mental performance) and workload (mental and physical cost of task requirements), making it difficult to assess how long an individual can conduct operations safely. The upcoming Volatiles Investigating Polar Exploration Rover will involve remotely controlling a lunar vehicle from an Earth-based mission control station. In order to determine how long personnel could successfully maintain alertness and performance while operating a rover, we studied seven trained operators in a simulated mission control environment. Operators completed two five-hour simulations in a randomized order, beginning at noon and at midnight. Performance was evaluated every 30 minutes using the Psychomotor Vigilance Task (PVT), Karolinska Sleepiness Scale (KSS), and NASA Task Load Index (NASA-TLX). On average, participants rated themselves as sleepier on the midnight drives compared to the day drives. Workload was rated higher during the noon drives compared to midnight. Lastly, participants had no change in average reaction time between the two drives. From the analysis, performance showed degradation after approximately three hours of driving. Our findings suggest that rotating drivers at least every three hours would be prudent to allow for breaks, and to minimize performance degradation, particularly during midnight shifts.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Telerobotic operations have been increasingly utilized across a variety of occupational situations, ranging from extraplanetary exploration (e.g., NASA rovers), military operations (e.g., hazard-seeking or surveillance drones), search and rescue operations, and robotic surgery (Chen et al., 2007). Telerobotic operations often involve a user operating a robot in a different location from where the vehicle is operating from, such as when an operator controls a Mars rover from an Earth-based mission control station. As a result, the user is provided with limited sensory feedback and instead must rely on feedback provided on visual displays. Such operations require the operator to maintain vigilant monitoring of displays for extended durations of time in order to preserve the robot and mission. However, it is unclear how long telerobotic operators can maintain such duties before experiencing elevated fatigue or workload which may increase the potential of making an error or having an accident.

Fatigue can arise from acute or chronic sleep loss and circadian misalignment (i.e., being awake during the biological night). Although every working professional experiences fatigue to some extent, it is especially problematic for shift workers who are required to work long and irregular hours around the clock such as physicians and nurses, pilots, police officers, control center operators, and many more (Åkerstedt, 1995; Fitzpatrick et al., 1999; Scheiman et al., 2018). Extending wakefulness through the night or for an extended period of time (i.e., over 16 hours), as often happens in shift work, can cause deficits in cognitive performance similar to being legally intoxicated, with cumulative performance deficits as wakefulness continues (Dawson & Reid, 1997;

Lamond & Dawson, 1999; Williamson & Feyer, 2000). Similarly, working consecutive night shifts results in circadian misalignment, leading to elevated risk of accidents among shift workers (Åkerstedt, 1995; Dinges, 1995).

Excessive workload is an ongoing threat to shift workers' fatigue as their mental and physical processes are continuously being utilized. With the addition of late night shifts, there are added stressors that are not present during the day, such as being active and physical during times when one is normally at rest. Continuous operations require personnel to work at safe levels of performance to maintain the integrity of the mission. The combination of mental, physical, and temporal stressors adds to one's overall workload, potentially causing faster burnout with the addition of fatigue. Many NASA missions are telerobotic in nature due to the physical separation from the device, and thus require attentive personnel to monitor safety.

The NASA Volatiles Investigating Polar Exploration Rover (VIPER) mission, anticipated to launch late 2023, aims to locate and sample water throughout the moon. The upcoming VIPER mission is one of many that will help prepare for the future Artemis program—justly named after the twin sister of Apollo in Greek mythology— which plans to return humans to the moon by 2024. Due to the relative lack of prior research on sustained (i.e., 24-hour) real-time, reactive mission control operations, we aimed to evaluate fatigue and workload during a simulated VIPER operation to identify how sleep loss, circadian misalignment, and workload might interact to inform scheduling and staffing requirements for the mission.

Fatigue

"Fatigue" is a broad term that can be defined in many ways. In the context of this report, the term fatigue can be defined as arising from acute (i.e., one or a few nights) or chronic (i.e., accumulation of weeks, months, or years with fragmented sleep) sleep loss, as well as sleep inertia (i.e., grogginess experienced during the time after waking), and circadian misalignment (e.g., switching time zones or being awake at night) (Åkerstedt & Wright, 2009; Boivin & Boudreau, 2014). The drive for sleep has been described by a two-process model involving homeostatic and circadian processes (Dijk & Schantz, 2005). The homeostatic process corresponds to the build-up of sleep pressure as wake continues and is reduced during sleep. The circadian process corresponds to the circadian pacemaker located in the suprachiasmatic nucleus in the hypothalamus, which responds to the day-night cycle through regulating many aspects of biological function, including core body temperature and melatonin (Deboer 2018; Dijk & Schantz, 2005).

The circadian rhythm promotes sleep at night and wakefulness during the day. Most individuals have a circadian clock that is slightly longer than 24 hours on average, estimated to be around 24.18 hours (Czeisler et al., 1999; Wright et al., 2001). In order to stay aligned with the 24-hour solar day, an individual must be exposed to light each day (Lockley et al., 2007). Under normal circumstances this occurs passively, with an individual awake during the day and being exposed to sunlight. When an individual works during the night and sleeps during the day, a state of circadian misalignment occurs between the drive for sleep generated by the internal clock and the solar light-dark cycle, making it difficult to fall asleep and stay asleep during the day and hard to stay

3

awake and perform well at night. Circadian misalignment also occurs during "jet lag," when the social drive for waking is out of sync with one's internal clock (Boivin & Boudreau, 2014). Although it is possible to shift the circadian rhythm, the process is slow, and without strong, appropriately timed photoperiodic cues, individuals do not synchronize with their imposed work schedule (Smith & Eastman, 2012). In addition, when an individual is in a circadian misaligned state, performance during the biological night will be impaired even when a long sleep episode has occurred during the day.

Continuous (i.e., 24-hour) operations are typically impacted by both circadian misalignment and sleep loss due to the need to schedule staff around the clock, resulting in decrements in alertness in cognitive performance (Pilcher & Huffcutt, 1996). Every hour of wakefulness increases the homeostatic drive to sleep, which affects a variety of neurobehavioral functions, such as increased lapses of attention, decreased reaction time, and reduced memory capability (Goel et al., 2013). Even moderate amounts of sleep deprivation (e.g., loss of 2-3 hours per night) have been shown to cause significant decrements in neurobehavioral functioning that remain present for up to 48 hours even during rest (Carskadon & Dement, 1981; Dinges et al., 1997). These performance impairments often appear quickly (e.g., one or a couple days) and do not subside until adequate rest is achieved (Van Dongen et al., 2003). This becomes especially problematic among shift workers for whom sleep is consistently fragmented, not allowing for enough time to catch up with their "sleep debt," which is the accumulation of lost sleep through either acute or chronic sleep loss.

Effects of Shift Work

Shift Duration

Fatigue accumulates over the course of a work shift with increasing number of hours awake, such that longer shifts are typically associated with higher levels of fatigue. The most common durations of shifts for 24-hour operations are 8- and 12-hour shifts, although shifts of other durations may be optimal depending on the work environment. There are many considerations that must be accounted for related to the total duration of the work shift and the duration of time on task that may be allowable given the nature of the work.

Compared to 12-hour shifts, 8-hour shifts are generally considered safer (Mitchell & Williamson, 2000), and result in better performance (Fitzpatrick et al., 1999). Nurses on 12-hour shifts, for example, are reported to have impaired judgement and clinical decision-making, as well as having decreased reaction time relative to 8-hour shifts (Bowers-Hutto & Davis, 1989). More recently, nurses working 10- to 12-hour night shifts experienced increased fatigue and burnout relative to 8-hour shifts, and the 12-hour night shift resulted in nearly half of all nurses expressing concern about making mistakes (Estryn-Béhar & Van der Heijden, 2012). A field study examining the differences between 8- and 12-hour shifts among electrical power station operators found that there were both improvements and deficits when adopting the 12-hour schedule (Mitchel & Williamson, 2000). Individuals rated their quality, duration, and habits of sleep to be improved during the 12-hour shifts; however, there were higher instances of errors occurring at the end of the 12-hour shifts for both the day and night (i.e., reaction time

and grammatical ability). Among the Australian coal mining industry, no differences in performance were found when comparing 8- and 12-hour schedules; however, there was a marked increase in absenteeism for 12-hour shifts when overtime was required for the operation (Baker et al., 2001).

On the contrary, in some situations 12-hour shifts improve work communication by minimizing "handovers," which is the process of informing the upcoming shift of the previous work that preceded that shift. A comprehensive study of nurses working 8- and 12-hour shifts found that most nurses preferred 12-hour shifts over 8-hour shifts, reporting that 12-hour shifts provided them with enough time to organize their day and to ensure the overlapping shift would be set up accordingly (e.g., in terms of handing over relevant information; Baillie & Thomas, 2019). Nurses working early morning, 8-hour shifts reported feeling rushed to complete work and guilty if they were forced to handover uncompleted work to the next shift. Researchers investigating 8- and 12-hour shift schedules among chemical workers found no significant differences between alertness and performance, and once again found a general increase in well-being, sleep quality, and better social and domestic quality of life (Tucker et al., 1996). However, the authors note that these results may be attributable to their fast, forward-rotating schedules.

Effects of Shift Start and End Time

The time at which a shift begins impacts fatigue and sleep opportunity, such that an earlier start time on a day shift can reduce overall sleep length and increase subjective sleepiness during the subsequent shift (Kecklund et al., 1997). This is largely due to the

influence of the wake maintenance zone, which promotes wakefulness in the hours prior to one's habitual bedtime, making sleep difficult at that time (Jewett & Kronauer, 1999). The start time of day and evening shifts can also influence worker performance and their ability to achieve recovery sleep.

There is a wide range of shift start times for work that is considered a "day shift." In a study of experienced train drivers with work start times at 05:49, 07:49, and 09:49, sleep duration was shortest, and sleepiness was highest among those with the earliest work start time (Ingre et al., 2004). Similarly, airline pilots working a series of early starts beginning around 06:00 obtained significantly less sleep, reported increased sleepiness, and had reduced performance compared to shifts starting later in the day (Flynn-Evans et al., 2018). Together, these findings suggest that daytime shifts should begin later in the morning in order to allow for adequate rest prior to an individual beginning work.

The shift start time for evening and night shifts can also determine how well an individual will sleep and perform. Among airline pilots, evening shifts ending around midnight were associated with reduced alertness and performance compared to shifts starting around 09:00 (Flynn-Evans et al., 2018), but did not reduce sleep duration, likely due to the ability of most individuals to "sleep in" in the morning. Such late finishes following evening work have the potential to increase the risk of motor vehicle crashes upon commuting home (Crummy et al., 2008), making it advisable to end the evening shift earlier than one's typical sleep time.

Working at night is associated with increased sleepiness and reduced performance in general, but strategic shift design can allow some individuals to achieve better sleep and

performance outcomes. Takeyama et al., (2005) compared the differing start times for night shifts among firefighters (i.e., 22:00 – 07:00, split between five, two-hour schedules) and found that those who worked between 01:30 – 05:15 achieved shorter sleep with increased complaints of fatigue due to the earlier start times. This is likely because shift start times that are only slightly later than one's habitual bedtime do not allow for one to go to bed at a normal time to achieve a nap before the shift, resulting in an extended duration of time awake (Boivin & Boudreau, 2014). In contrast, when night shifts are scheduled to start somewhat earlier, individuals may be more likely to engage in a prophylactic nap (i.e., a nap in preparation for extended wakefulness) in the afternoon, allowing them to begin work at a lower sleep pressure.

The end of a night shift should be scheduled so that it does not coincide with the circadian low. The circadian rhythm strongly promotes sleep between 01:00 and 07:00 for most daytime entrained individuals (Wright et al., 2001). Single-vehicle accidents are increased during this time, especially when commuting home following an extended shift, suggesting that individuals are more likely to have unintentional microsleeps and attentional lapses during this time (Barger et al., 2005). As a result, work shifts should not be scheduled to end until after this time, when the circadian system is typically promoting waking.

Number of Consecutive Shifts

Working too many consecutive day or night shifts can adversely affect performance. Due to the cumulative effect of sleep loss and circadian misalignment, working consecutive night shifts is more difficult than working consecutive day shifts. Tucker et

al. (2010) reported that working seven consecutive 8-hour night shifts resulted in significantly higher levels of fatigue, accident risk, and work-life disruption when compared to working only three or four consecutive nights. Similarly, Geiger-Brown et al. (2012) found that nurses on their third consecutive night shift experienced greater sleepiness compared to the first two nights. Muller et al., (2008) also found similar results among a mineral processing plant in Australia where operators were on shift for eight consecutive 12-hour night shifts. The authors found that at the end of the first two nights, operators showed significantly elevated levels of fatigue and slower response times (i.e., percent change in reaction time through the Mackworth Clock Vigilance task) when compared to the beginning of the shift. The authors also analyzed performance comparing it to high levels of blood alcohol concentration; after the end of one 12-hour night shift, operators had performance equivalent to someone with a 0.05% blood-alcohol concentration.

The number of consecutive day shifts that an individual can work without experiencing performance degradation depends on the characteristics of the work. For example, short-haul airline pilots flying an "easy" schedule were able to sustain high levels of alertness and performance for five days in a row when their first flight began in the mid-morning and ended in the early afternoon. However, when the same pilots were scheduled to fly five consecutive days with an early (e.g., 06:00 start) or late start (e.g., 17:00 start) or with elevated workload (e.g., multiple short flights as opposed to fewer long flights), their performance degraded each consecutive day (Flynn-Evans et al., 2018).

Time off for recovery between shifts impacts the number of consecutive days that one can work. For example, day workers reported that fatigue was higher when returning to work after three or four consecutive day shifts but with only one rest day when compared to those who received two (Tucker et al., 2010). In medicine, where workload is high and attentional requirements are similar to mission control operations, recommendations for junior doctors limit consecutive nights to two or three, with at least two days off for every two night shifts worked (Cappuccio et al., 2009; Horrocks et al., 2006). Similarly, researchers have suggested that more than one day off is required following two consecutive night shifts (Horrocks et al., 2006).

Work Shift Designs

As described above, the duration and number of consecutive shifts that one is scheduled to work can greatly influence levels of fatigue. There are many potential combinations of shifts that can be designed to cover work requirements. The optimal design of work schedules depends on the nature of the work and the number of hours that need to be covered and must account for handing over work duties. For example, some shift designs require individuals to work five 8-hour shifts, while others use 10- or 12-hour shifts or combinations of these shift lengths. These shifts are then organized so that individuals maintain a stable shift schedule or rotate between day, evening, and night.

One consideration when scheduling is whether to schedule rotating work shifts or to have workers maintain a stable day, evening or night shift schedule. Rotating schedules for continuous operations are typically comprised of either three 8-hour shifts, or two 12-hour shifts in each 24-hour period. It has been reported that rapidly rotating shifts do not

allow for circadian adaptation to night shifts, resulting in poorer performance relative to daytime work, particularly on the second night shift (Baulk et al., 2009). Pilcher et al. (2000) performed a meta-analysis on rotating night shift nurses working a mixture of day and night shifts and found that they had similar performance decrements to nurses that solely worked night shifts, suggesting that both types of schedules lead to performance impairment. However, individuals working permanent night shifts can take up to three weeks to adapt (Fossum et al., 2013), making scheduling individuals for permanent night shifts during short duration missions impractical. Furthermore, researchers examining circadian adaptation to night shifts suggests that individuals are typically unable to adapt completely to a full-time night schedule because they typically revert to a diurnal schedule during days off (Smith & Eastman, 2012). Therefore, the persistent circadian misalignment that occurs when individuals are scheduled for permanent night duty favors scheduling individuals for rotating shifts instead of stable shifts.

Although rapidly rotating shifts are preferable to permanent night shifts, the sequence of the shift rotation and individual differences—for example, whether or not individuals are more "morning" or "evening types" based on their individual preference and performance—all determine how well workers will manage on night shifts. Shifting schedules to a later, rather than earlier, rotation has been shown to increase alertness and productivity in factory operations (Czeisler et al., 1980). This is likely because most individuals have an endogenous circadian period that is slightly longer than 24 hours (Czeisler et al., 1999). This means that it is easier for most of the population to stay awake later, rather than go to bed earlier, making later rotating shifts easier to adapt to for

most individuals (Brooks & Swailes, 2002). The preference of time of day is associated with circadian phase, and those who indicate that they are moderate (i.e., neutral) or evening types will typically have an easier time with a later rotating schedule (Duffy et al., 2001). As a result, shifting rotating schedules from day, to evening, to night, followed by days off is preferable to other shift designs.

Shift schedules that do not follow sleep and circadian principles provide valuable lessons about how to avoid common pitfalls in scheduling. For example, in 2016, the United States Air Force School of Aerospace Medicine conducted a fatigue and workload study on crewmembers working on the MQ-9 Reaper unit—an unmanned aerial vehicle capable of autonomous flight operations—in which shifts were increased from 8- to 12-hours and individuals rotated between five day and five night shifts with two days off between (Scheiman et al., 2018). The transition to longer shifts with more than two consecutive night shifts resulted in half of the participants reporting extreme, persistent fatigue, even on days off. In addition, the new shift schedule resulted in increased headaches with eye strain, decreased appetite, decreased reaction time and processing speed, and increased frustration and irritability (i.e., towards co-workers and/or family members). Participants also reported difficulty staying awake while commuting.

Handover

Although there is no evidence that handover of information is not in itself a contributor to fatigue, schedules that exclude sufficient time for handover can lead to increased error. Proper handover procedures are essential in mission control operations to maintain the safety and success of the mission. In the medical industry, handovers are

recommended to be received as written documents alongside verbal communication (Randell et al., 2011). There are many threats to shift handover, such as time constraints, absence of written documents, and lack of interdisciplinary communication (Fealy et al., 2018). Handover procedures are performed best when they are documented and are received during overlap shifts. Another barrier to handovers is the high potential for being interrupted due to working in a noisy environment, which is highly probable during mission control operations.

Workload

NASA Task Load Index

The cost of accomplishing a particular mission requirement for the human operator is referred to as "workload" (Hart, 1988). Workload is difficult to measure, because it is largely a subjective construct and it can be difficult to ask individuals about their workload without interfering with their work activities. There are several workload measures that have been designed to attain individuals' perceived workload during or after an activity. Due to the increasing prevalence of flight and military operations, the NASA Task Load Index (NASA-TLX) was created to evaluate the multiple attributes surrounding workload: Mental Demands, Physical Demands, Temporal Demands, Frustration, Effort, and Own Performance (Hart & Staveland, 1988). Today, the NASA-TLX is used across many areas involving automated technology, including by not limited to aviation, automobiles, combat, and medicine (Hart, 2006). The NASA-TLX has also been tested for reliability and validity and has been shown to have good internal consistency (Xiao et al., 2005).

The three demand characteristics refer to the mental and perceptual activity required for the task (i.e., Mental Demands), physical activity required (i.e., Physical Demands), and the time pressure or sensitivity of the task (i.e., Temporal Demands). The other three subscales refer to the individual, such as how successful one felt in accomplishing the task (i.e., Own Performance), the amount of mental and physical work needed for the task (i.e., Effort), and how stressful or irritating the task was (i.e., Frustration) (Bunce & Sisa, 2002; Young et al., 2008). Each workload domain can be analyzed individually but is mainly used to observe overall workload through one aggregate workload score.

Although constructed for flight operations, the NASA-TLX is now utilized across many areas of field research. Early on, Weinger & Englung (1990) demonstrated that excessive workload and fatigue, in conjunction with poorly scheduled shifts, resulted in cognitive impairment (i.e., vigilance and thought processing) and irritability. In addition, poorly designed systems, such as those that do not consider human factors control compatibility, cluttered displays, or illegibility, can further increase workload during time on task (Hooey et al., 2018). In 2011, Dorian et al. assessed the workload and fatigue of Australian rail industry employees in the field, and found that regardless of their role (e.g., driver, controller, guard, signaler, etc.), major decrements were found in either Mental Demands or Effort. Specifically, the signalers and controllers rated Mental Demand highest, while employees in other roles rated Effort highest. Although differences were reported in work type, sleep length, and workload, there were no differences in subjective fatigue levels. As reported in this study, the type of work also plays a key role in fatigue and workload, especially in mission control operations.

Time on Task

Beyond the type of work performed and the workload required, the time in which one spends on each task and switching between task types (e.g., physical versus mental workload), plays a crucial role in how successful one will be in completing a given task. In 2002, researchers explored the relationship between task difficulty, time on task, and workload among graduate and undergraduate students performing simple pursuit and memory tasks, requiring the participants to keep a target moving on a circumference between two lines (i.e., through turning a wheel) without intersecting them (Haga et al., 2002). Utilizing the NASA-TLX and physiological measures (i.e., a combination of electroencephalography to measure electrical brain activity, and electrooculography to measure eye movements), the authors found that difficult tasks that required a higher degree of concentration resulted in Mental Demand building up more rapidly under more temporally demanding conditions. In other words, difficult and cognitively demanding tasks tended to burn out the operators when they were under greater time pressures, as is common in mission control operations. The converse was also found, such that easier tasks with lower degrees of concentration resulted in Mental Demand increasing due to less temporally demanding conditions, such as when the time pressures were not as severe, or when the tasks have become monotonous (e.g., less direct input from operators due to automation).

When working extended shifts (i.e., 12-hours), workers typically report higher levels of fatigue throughout the day which interacts with increased workload, leading to reports of fatigue accumulating at an even higher rate (Rosa, 1995). MacDonald and Bendak

15

(2000) investigated the effects of varying workload (i.e., mental and physical) during both standard and extended work hours (i.e., 8- and 12-hour shifts, respectively) and its effect on performance and fatigue in both a field and laboratory study. In the laboratory study, results indicated that high instances of physical workload led to increased overall workload during both 12- and 8-hour shifts despite identical task demands. Cognitive processes such as alertness and grammatical ability were subjectively lower during the 12-hour shifts as well. Similar results were found across the field study, suggesting that the combination of high workload and 12-hour workdays resulted in greater fatigue through decreased alertness and overall cognitive abilities and increased bodily discomfort. Similarly, nurses working a variety of differing intensive care units (e.g., surgical, trauma, pediatric, etc.) showed increased overall workload during the 12-hour shifts, especially in Physical Demands and Frustration (Hoonakker et al., 2011).

Mission Control Operations

There are three major subcategories of control center operations: (1) pre-planned sequence evaluation and generation operations, where controllers intermittently receive and upload new sequences of information, with no active control of the remote vehicle (e.g., Phoenix Mars Lander); (2) real-time, reactive human operations, where a controller is directing or overseeing the actions of a remote human operator (e.g., ISS mission control); and (3) real-time, reactive telerobotic operations, where a controller actively operates a remotely controlled vehicle or robot in real time (e.g., VIPER). These operating models present their own unique challenges in managing fatigue and workload due to the variation across the goals of the mission. Previous researchers have

successfully implemented a fatigue management program for personnel working in a pre-planned sequence evaluation and generation control center (Barger et al., 2012), but have yet to systematically evaluate real-time, reactive telerobotic operations. The upcoming VIPER mission—formerly the Resource Prospector mission—is aimed towards expanding lunar exploration through sending a remotely operated rover to the moon to both map out and excavate the moon for concentrations of ice. This will be the first resource mapping mission on another celestial body and will aid in the future Artemis 2024 mission. Due to the nature of this upcoming VIPER mission (i.e., being a continuous operation with human-operated, remotely controlled technology) there is a need to systematically evaluate missions such as these from a fatigue risk-management perspective.

Real-time, Reactive Telerobotic Operations

Real-time, reactive telerobotic operations require a significant fraction of time for monitoring, managing, and maintaining the health and safety of the onboard systems (McCann & Spirkovska, 2005). This differs from the other types of missions in two key ways: (1) it will be a continuous operation meaning it needs to be operated at all times around the clock; and (2) there will be a direct lack of immediate feedback with the technology due to the remote and distant location, as well as not having direct human interaction and feedback. In space-related missions, long-term exposure to environmental stressors are constant (e.g., circadian disruption or misalignment), which can be problematic when information processing and decision-making capabilities need to be at their maximum potential. On top of that, the sustained nature of these real-time, reactive

operations (i.e., performing operations 24-hours a day until completion) require personnel to rotate shifts consistently throughout the day and night. Previous missions have had various complications arise due to mission personnel experiencing decreased sleep quality and circadian misalignment during prolonged, continuous mission operations (e.g., Barger et al., 2012). Despite these findings, more research is required to ascertain the extent of this fatigue in relation to other factors, such as how it might affect workload, initial shift timing, staff scheduling, and use of fatigue countermeasures.

Volatiles Investigating Polar Exploration Rover

The upcoming VIPER mission aims to locate and extract water (in the form of ice) on the moon. Projected to launch late 2023, mission control operators will continuously (i.e., rotating, 24-hour shifts) operate the rover and evaluate the flow of data coming in for 100 days. Roughly the size of a golf cart, the VIPER will land in the south pole of the moon to locate sources of water using four distinct science measures: (1) the Neutron Spectrometer System (NSS) will detect "wet" areas of the moon; (2) The Regolith and Ice Drill for Exploring New Terrain (TRIDENT) will drill into the moon's surface up to 1 meter below; (3) Lastly, the Mass Spectrometer Observing Lunar Operations (MSolo) and (4) Near InfraRed Volatiles Spectrometer System (NIRVSS) will analyze the volatiles in terms of composition and concentration of the water. For the purposes of this experiment, a simulation of the VIPER mission was utilized to assess the major roles of the future mission (i.e., rover driver and real-time scientist, with guidance from an acting flight director).

Purpose of Study

The purpose of this study was to evaluate fatigue and workload during real-time, reactive control center operations, using the computer-based simulated control center environment at NASA Ames Research Center (ARC). Using participants trained for the Resource Prospector mission (a precursor to VIPER), we assessed the duration of drives based on operators' fatigue and workload, as well as the differences in performance related to the time of day (i.e., one beginning at 12:00, and one beginning at 24:00). We hypothesized that the performance of the drives (i.e., in terms of vigilance, workload, and sleepiness) would worsen over time during the two five-hour drives. We also hypothesized that the midnight drives would result in worse performance relative to the noon drives, and that midnight drives would result in quicker burnout than the noon drives (i.e., indicative of increased workload, subjective fatigue, and reaction time).

Method

Participants

This study was approved by the NASA IRB (HRI-359). A total of 16 individuals were trained to operate the VIPER simulation and were invited to participate. Of the eligible participants, seven (n = 1 female) agreed to participate (i.e., five drivers; two real-time science operators). Each trial consisted of one driver paired with one real-time scientist (RTSci) operating the simulation; the duties performed by each role are detailed below. Due to the uneven distribution of drivers and RTSci, two researchers from the Fatigue Countermeasures Laboratory at NASA ARC acted as RTSci (i.e., the last six trial runs were with confederates). The driver was always paired with the same RTSci (or confederate) during the study. For the purposes of this study, there were no exclusion criteria beyond requiring the participants to be trained operators for the VIPER mission.

Driver Activities

The role of the driver was to operate and manage the safety of the rover. During the simulation, the driver viewed a 3D projected world space of the moon. There were two levels of the world space: (1) the drivable terrain, and (2) the projected image of the surface of the moon. A high-resolution visual environment of the moon's surface was displayed (i.e., 1 km x 1 km), though the drivers only had a portion of that (i.e., 320 m x 320 m) available for driving. If the rover was driven past the physical terrain space, the simulation would error; therefore, drivers were instructed to remain within the mapped simulation. The driver issued a series of commands to the rover, such as precise directional input, waypoints to travel to, simulated images to capture, and projected

hazards (e.g., rock, divot, or crater in immediate vicinity). Capable of turning in any

direction, the rover would navigate to a waypoint submitted by a driver. To simulate the

slight delay in communications during a real lunar operation, there was a six second

buffer for uplink of commands to the rover and a 10 second buffer for the downlink of

telemetry, plus an additional bandwidth calculation—the larger the telemetry item, the

longer it took to download—resulting in pictures taking roughly 30 seconds to download

after the command was received. The monitor displayed each command that was recently

uploaded in the top left corner of the screen, as well as a timer for subsequent commands.

The rover traveled at a speed of 10 cm per second while in motion, but on average, it

traveled roughly 1 cm per second when buffer time, operator decision time, and other

factors are considered, equaling roughly 36 m an hour. The driver also utilized other

commands for navigational purposes to determine the safest route, such as the navigation

and hazard cameras.

The navigation camera, or NavCam, took photo-realistic images of the projected

world space, allowing the driver to see if there were obstacles in their path. There was a

total of two NavCams, both located at the top of the rover situated roughly two meters

high to simulate the angle that an average sized human would observe. These cameras

had an effective range of eight meters. The NavCam pair took a 70-degree wide image of

the terrain around the rover and displayed the projected hazards immediately in that field

of view with a color-coded visual. Flat surfaces were highlighted green to indicate that

they were safe to traverse, while difficult terrain (e.g., rocks, divots, or craters) was

highlighted in shades of red; the greater the threat, the darker the saturation of red. In

addition to the NavCams, there were an additional four cameras located at each wheel, with one in the back of the rover as well. The hazard cameras, or HazCam, displayed hazards to the immediate terrain around the wheels with a range equal to two meters. The drivers continuously scanned their environment in order to avoid crashing the rover. The pitch—the rover pointing up or down a slope—or roll—the rover moving across a slope—was not supposed to exceed a 15-degree tilt, as that was the agreed upon limit for safe travel. It was estimated that a tilt of 25-degrees or more would result in a crash, either through flipping the rover and or general loss of control.

The main task for the drivers was to traverse the path that was programmed for the simulation. Each traverse plan was the same for every trial run, so that the drivers would experience the same obstacles and challenges during the day compared to the night. In order to minimize learning effects, the traverse plan was reversed for each drive (i.e., the midnight traverse plan was the reverse of the noon traverse plan). While driving the traverse plan, the drivers needed to be mindful of the location of the sun; solar panels were located on all sides of the rover except the front, meaning that drivers were advised to avoid driving in the direct direction of the sun to avoid loss of power. However, this was often difficult when maneuvering around the lunar surface, often forcing drivers to drive backwards to ensure the sun remained on the panels. Each traverse plan contained two Area of Interest Maps (AIM) which showed a colored square around the projected world space, indicating areas that would be of interest to the science operator.

Real-Time Scientist Activities

The role of the real-time scientist (RTSci) was to observe the flow of data collected from the hydrogen sensors on the rover. As the goal of the future mission is to excavate the moon for volatiles, specifically water, the RTSci were consistently monitoring a two-channel count rate, comparing water-equivalent hydrogen abundance, and burial depth. This informed the RTSci about the concentration of water as well as how deep below it would be located. As the driver traversed the path and AIM, the RTSci may provide input on where specifically the rover should travel based on the flow of data. For example, when traveling across the traverse plan, the RTSci may observe a spike in the data, and would suggest going back over that location to see if truly was indicative of a high water concentration, or just an anomaly (e.g., statistical error). The areas of high hydrogen concentration were color coded, such that areas of low concentration were colored yellow and progressively transitioned to green as concentration increased, with blue indicating the highest degree of hydrogen concentration. Areas with too little data were marked black.

Measures

Pre-screening

Prior to the start of the experiment, participants were debriefed on the study procedures. Participants were given a background questionnaire that included general demographic information (e.g., age, BMI, and typical sleep and wake schedule; see Appendix A) and questionnaires to assess whether they are more morning or evening types, as well as their quality of sleep. The Morningness-Eveningness Questionnaire

(MEQ) distinguishes individuals who identify as morning types from individuals who identify as evening types in terms of how well they feel subjectively (e.g., overall mood) and how well they perform in general, and has been demonstrated to correlate with the circadian phase of the participant (Duffy et al., 2001; Horne & Östberg, 1976). In the present study, the MEQ was assessed for reliability through Cronbach's alpha and was rated internally consistent with a score of 0.83.

The Pittsburgh Sleep Quality Index (PSQI) and Epworth Sleepiness Scale (ESS) are questionnaires that determine the likelihood of a person having a sleep disorder (Buysse et al., 1989; Johns, 1991). In the present study, the ESS and PSQI were assessed for reliability through Cronbach's alpha and were rated internally consistent with a score of 0.57, and 0.88, respectively. These were gathered early to provide a well-rounded picture of their prior sleep history.

Sleep Assessment

During the initial meeting, each participant was given an activity monitor (Actiwatch Spectrum, Respironics Inc®, Bend, OR, USA) to wear on their non-dominant wrist. The Actiwatch was used to evaluate participants' sleep timing, duration, and quality leading up to and during the stimulation. Participants were also asked to complete a daily sleep diary consisting of periods of sleep, including naps, to supplement the actigraphy analysis (see Appendix B).

Subjective Fatigue Assessment

Participants' self-rated fatigue levels were assessed through the Karolinska Sleepiness Scale (KSS), which is a highly sensitive subjective measurement for sleepiness

(Åkerstedt & Gillberg, 1990). The KSS is a one-item scale that asked participants to rate how sleepy they felt in that instance; on a nine-point Likert scale (i.e., 1 = extremely alert to 9 = fighting sleep), participants were asked to rate how fatigued they felt throughout the duration of the drive (i.e., one prior to the drive, and repeated testing every subsequent 30 min). In the current study, the KSS was assessed for reliability through Cronbach's alpha and was rated internally consistent with a score of 0.30 for the noon drives, and 0.46 for the midnight drives.

Workload

The NASA-TLX is a multi-dimensional scale designed to obtain estimates of workload from operators and was originally developed at NASA ARC (Hart & Staveland, 1988). Since the original pen-and-paper version, the NASA-TLX was recently adapted to a fifth generation, 32-GB Apple iPod with the same characteristics as the original. Participants completed a baseline assessment of workload prior to both drive sessions in order to create the individual weighted ratings for each participant. This consisted of a 15-item comparison in which the participants were asked to compare which workload characteristics were more demanding. Each of the six workload domains—Mental Demands, Physical Demands, Temporal Demands, Frustration, Effort, and Own Performance—were weighted against individuals' perceived workload sensitivities. In other words, the weight for each demand was multiplied by the raw score provided, and then divided by 15 (i.e., the total number of comparisons during the baseline assessment). Raw scores were obtained through a series of visual analog scales for each domain, allowing participants to select their perceived workload for each domain

along a continuum (see Appendix C). In the present study, the NASA-TLX was assessed

for reliability through Cronbach's alpha and was rated internally consistent with a score

of 0.74 for the noon drives, and 0.87 for the midnight drives.

Vigilance

A handheld, 5-minute visual Psychomotor Vigilance Task (NASA-PVT) was utilized

to assess participants' reaction times and lapses in performance (Arsintescu et al., 2019;

Dinges et al., 1997; Dinges & Powell, 1985). As with the NASA-TLX, the NASA-PVT

was developed at NASA ARC on a fifth generation, 32-GB Apple iPod with the same

characteristics as the original PVT-192. This version of the PVT required participants to

hold a horizontal facing iPod that displayed a black screen with a red rectangle in the

middle (see Appendix D). Once the red rectangle began counting in milliseconds, the

participants were required to tap the screen with their dominant thumb as quickly as they

could. The intervals at which the stimulus (i.e., the millisecond counter) appeared varied

between two to 10 seconds, requiring the participants to remain vigilant throughout the

five-minute duration. Participants were instructed how to use the NASA-PVT prior to

testing and performed a baseline test prior to the start of the simulation. The main

parameters used in the analysis of reaction time data were mean reaction time (RT),

response speed (1/RT), and number of lapses (i.e., responses > 500 ms).

Study Protocol

A within-subjects, randomized experimental design was used for the study utilizing

the simulated control center environment at NASA ARC. The experimental conditions

consisted of the two times at which the drives commenced, 12:00 – 17:00 and 00:00 –

05:00. In order to simulate the full experience of a real-time, reactive telerobotic operation, the major roles of the VIPER mission were examined (i.e., rover driver and real-time science operator). Four participants had their noon drive first and three participants had their midnight drive first, and both RTSci had their noon drive first. Each participant completed two drives each, ranging from a minimum of two to a maximum of 10 days between.

The day drive began at noon, while the night drive began at midnight. Each participant drove for five hours total, with the option to end the simulation if they felt too fatigued. Prior to the start of the simulation, the acting flight director provided a brief rundown on the drive procedures (e.g., purpose of drives, traverse plan, and refresher on controls). The driver and RTSci (or confederate) were also asked about their most recent caffeine consumption prior to testing, although caffeine was not accounted for in the final analysis. After this, participants completed a baseline cognitive test battery that took nearly 10 minutes completing an initial KSS, NASA-TLX, and NASA-PVT. Once the baseline assessment was complete, participants began their session. Roughly every 25 minutes, participants were asked to pause operating the simulator to complete their cognitive battery. In total, participants were scheduled to complete nine test batteries (not including the baseline assessment), until they completed five hours of driving. During each cognitive battery, participants were moved to a nearby, isolated meeting space to minimize distractions. Participants were monitored by a study staff member for compliance during all tests.

Statistical Analysis

Analyses were calculated using R statistical software (version 3.6.1) on the Ubuntu

Linux Operating System (version 18.04). Sleep data were calculated from actigraphy

using a standard algorithm (Phillips Respironics). Sleep outcomes included sleep duration

in the week prior to the simulation, sleep duration in the 12 and 24 hours prior to the

simulation, and wake after sleep onset (WASO), which is a measure of sleep disruption.

As is in the name, the WASO refers to how long an individual was awake after they

started their sleep period. For example, individuals with sleep apnea or other sleep

disorders might have a high WASO simply due to waking at multiple periods through the

night.

We calculated the following metrics from the NASA-PVT: mean reaction time (RT),

number of lapses (i.e., RT > 500 ms), optimum response times (i.e., the fastest 10% of

response times for all trials), and cognitive slowing (i.e., the slowest 10% of response

times for all trials). We also assessed the correlation between average duration of sleep

achieved 24 hours prior to the drive with NASA-PVT performance (i.e., mean RT) using

a Pearson correlation.

Workload was evaluated on the NASA-TLX by calculating the weighted average for

each participants' trial. At baseline, participants completed a 15-item comparison in

which they selected which workload domain was most taxing to them for their given task.

These weights were used to calculate the weighted average by multiplying individual

weights to individual raw scores and dividing the sum of those numbers by 15 (i.e., the

number of comparisons). Individual workload metrics were also analyzed to measure differences between the noon and midnight drives.

To assess operational performance during each drive, we determined how often drivers exceeded a 15-degree tilt during their drives. We also evaluated the magnitude of the tilt as a measure of severity.

We aimed to evaluate performance difference in three ways. First, we used paired t-tests to evaluate the overall difference between the noon and midnight drives. In order to determine the time when performance deterioration reached an unacceptable level, we identified the NASA-PVT tests when RT exceeded two standard deviations above the baseline for each individual. This was calculated through averaging each trial and getting the difference in RT from the baseline. The standard deviations of those differences were calculated to determine the two-standard deviation limit. This method was also used for the NASA-TLX as well. For all analyses, confidence intervals will surround the mean differences observed. Lastly, we utilized Hedge's *g* to measure effect size for a more accurate representation of the population (i.e., due to utilizing the sample-based standard deviation) as well as the low sample size (see Eq. 1 below). Because our sample was within-subjects, we calculated Hedge's *g* utilizing the pooled standard deviation equation from Cohen (1988).

$$\text{Hedge's } g \ = \ \frac{M_1 - M_2}{SD * pooled} \tag{1}$$

Results

Demographics

Baseline demographic characteristics are shown in Table 1. Participants differed on many characteristics. One participant scored above the clinical threshold on the ESS and PSQI, while two others scored above the threshold for just the PSQI which are indicative of the participants potentially having undiagnosed sleep disorders (see Table 1). Sleep duration was also variable, ranging from pre-study sleep duration of four to almost 10 hours (see Table 2). According to the MEQ, most participants were either neutral (i.e., neither morning nor evening types), or had scores close to neutral; two participants were scored to be moderate morning types and one participant was scored to be a moderate evening type. Sleep durations prior to the drives are shown in Figure 1 below, as well as the amount of WASO (see Figure 2).

Table 1

Demographic Information

Variable	*M*(*SD*)	Range
Age	47.14 (12.32)	28 – 65
BMI	26.30 (5.52)	21.6 – 27
Weekday Bedtime	23:17 (1:02)	22:00 – 01:00
Weekday Waketime	06:34 (0:43)	06:00 – 08:00
Weekend Bedtime	23:42 (0:59)	22:00 – 01:00
Weekend Waketime	07:34 (0:56)	07:00 – 09:30
MEQ	53.36 (8.09)	41 – 64.5
ESS	6.57 (3.26)	2 – 10
PSQI	6.57 (3.39)	2 – 12

Note. Age = years; BMI = Body Mass Index.

Table 2

Sleep Information

Variable	*M*(*SD*)	Range
Sleep Duration for prior week (h)	6.50 (1.75)	4.10 – 9.58
WASO for prior week's sleep	50.58 (40.52)	0.00 – 3.40
Sleep in prior 8 hours (noon)	1.89 (0.69)	1.83 – 2.90
Sleep in prior 12 hours (noon)	5.28 (0.29)	4.97 – 5.73
Sleep in prior 24 hours (noon)	6.02 (0.83)	4.97 – 7.19
Sleep in prior 8 hours (midnight)	1.19 (1.21)	0.00 – 3.50
Sleep in prior 12 hours (midnight)	1.19 (1.21)	0.00 – 3.50
Sleep in prior 24 hours (midnight)	7.44 (2.54)	4.12 – 11.00

Note. Sleep in prior # hours = Sleep duration (h) prior to a drive session.

Figure 1

Sleep Durations
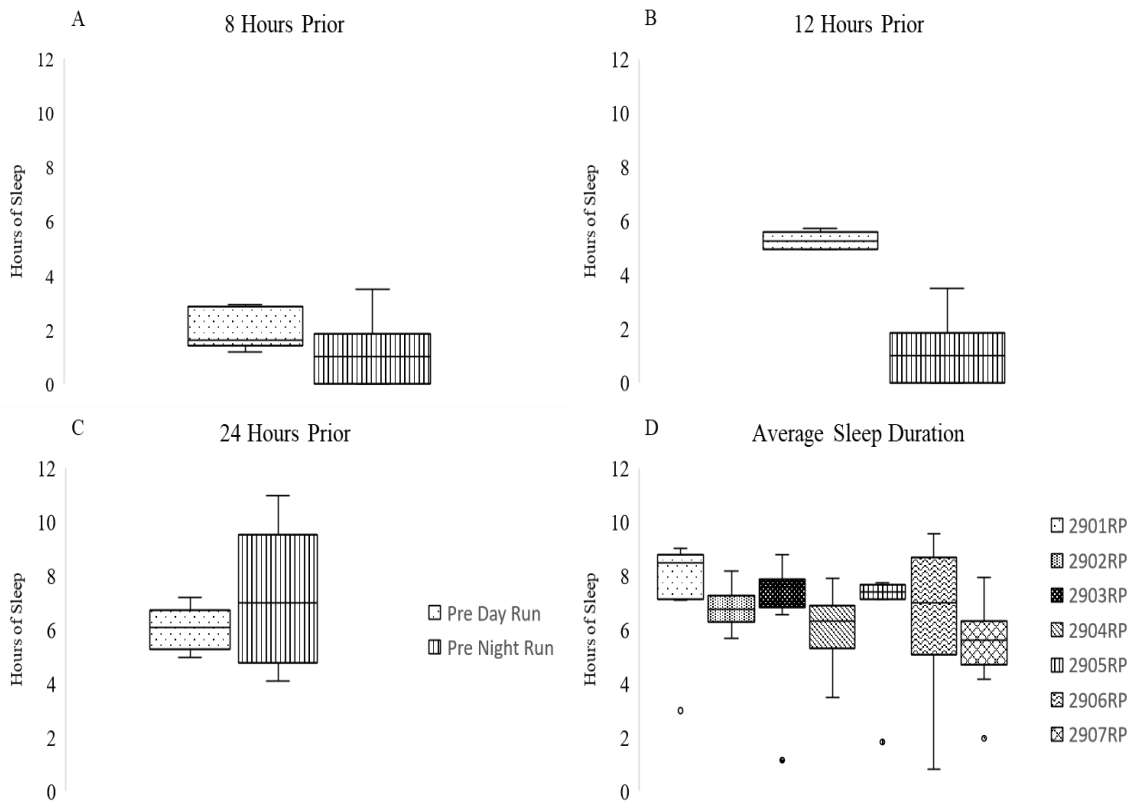
Note. Median duration of sleep in hours (y-axis) prior to their drive dates. Panels A, B, and C represent the median duration of sleep received in the eight, twelve, and twenty-four hours before their noon and midnight drives (respectively); and as represented by the legend, the lightly-dotted box represents the time prior to a noon drive (12:00), while the striped line represents the time prior to a midnight drive (00:00). The upper and lower hinges represent the inner and outer quartiles for the range of scores. Panel D represents all participants' median sleep duration throughout the study, which each participant in numbered order from left to right. The four dots below each represent minimum sleep durations.

Figure 2
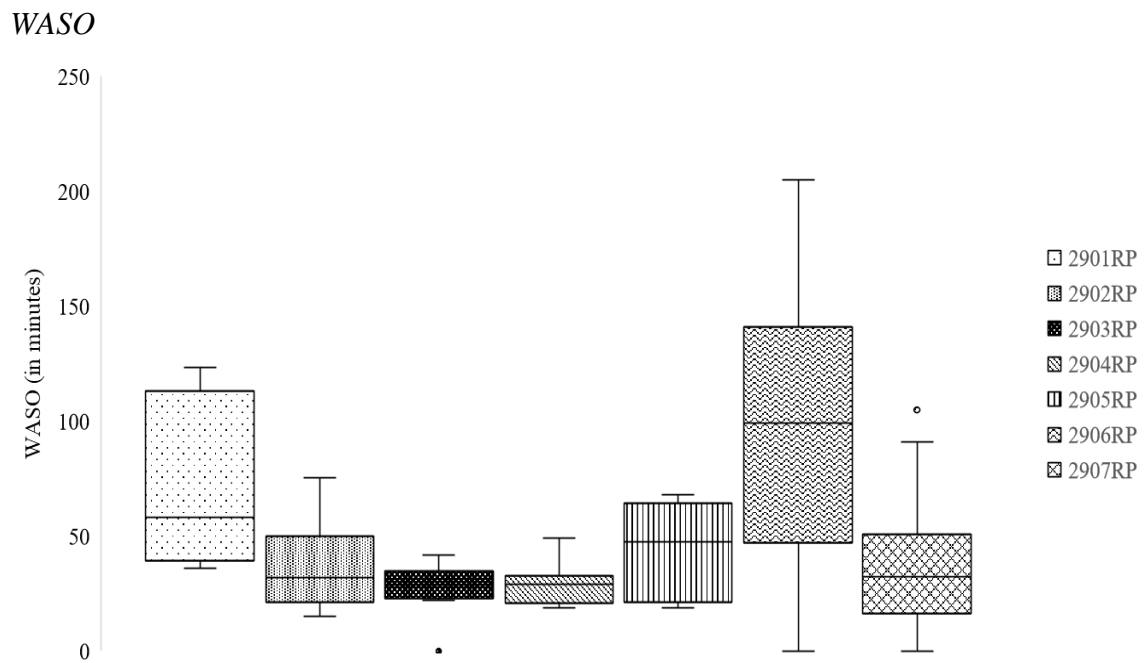
WASO

Note. Median wake after sleep onset (WASO) for each participant (in minutes; y-axis) prior to their drives.

Noon versus Midnight

KSS

There was a significant increase in mean KSS scores from the noon drive ($M = 3.12$, $SD = 1.44$), to the midnight drive ($M = 5.06$, $SD = 2.28$), $t(65) = -9.13$, $p < .001$, 99% CI [-2.37, -1.30], $g = 0.85$, meaning the participants felt sleepier on average during the midnight drive (see Figure 3).

Figure 3

KSS Mean Score
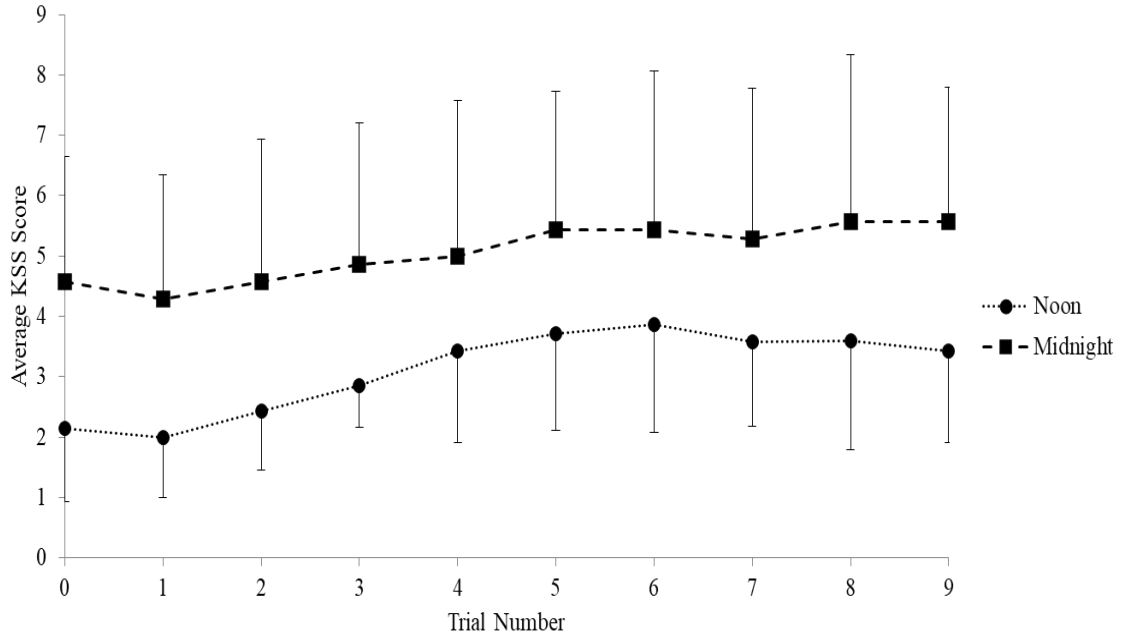
Note. Mean KSS score (y-axis; larger number indicate greater sleepiness) plotted by trial number (x-axis); The noon scores are represented by the dotted line with circles, with midnight represented by the hashed lines with squares. Error bars represent standard deviation.

NASA-PVT

There were no significant differences across any mean measures on the NASA-PVT. RT during the noon drive (M = 237.73, SD = 60.99) was not significantly different from the midnight drive (M = 238.73, SD = 57.85), $t(65)$ = 0.12, p = .91, 99% CI [-9.13, 9.95], g = 0.04 (see Figure 4). There were no significant differences in RT for the fastest 10% of responses comparing the noon drives (M = 170.83, SD = 12.23) to the midnight drives (M = 175.11, SD = 18.41), $t(65)$ = -1.73, p = .088, 99% CI [-8.96, 1.89], g = 0.27. There were no significant differences in RT for the slowest 10% of responses for the noon (M = 363.0, SD = 79.91) and midnight (M = 364.4, *SD*

= 99.25) drives, $t(65) = -0.19$, $p = .85$, 99% CI [-40.87, 35.51], $g = 0.02$. Lastly, there were no significant differences across lapses in performance between the noon ($M = 0.35$, $SD = 0.54$) and midnight drives ($M = 0.46$, $SD = 0.88$), $t(65) = 0.61$, $p = .55$, 99% CI [-0.25, 0.41], $g = 0.27$.

Figure 4

NASA-PVT Mean Reaction Time
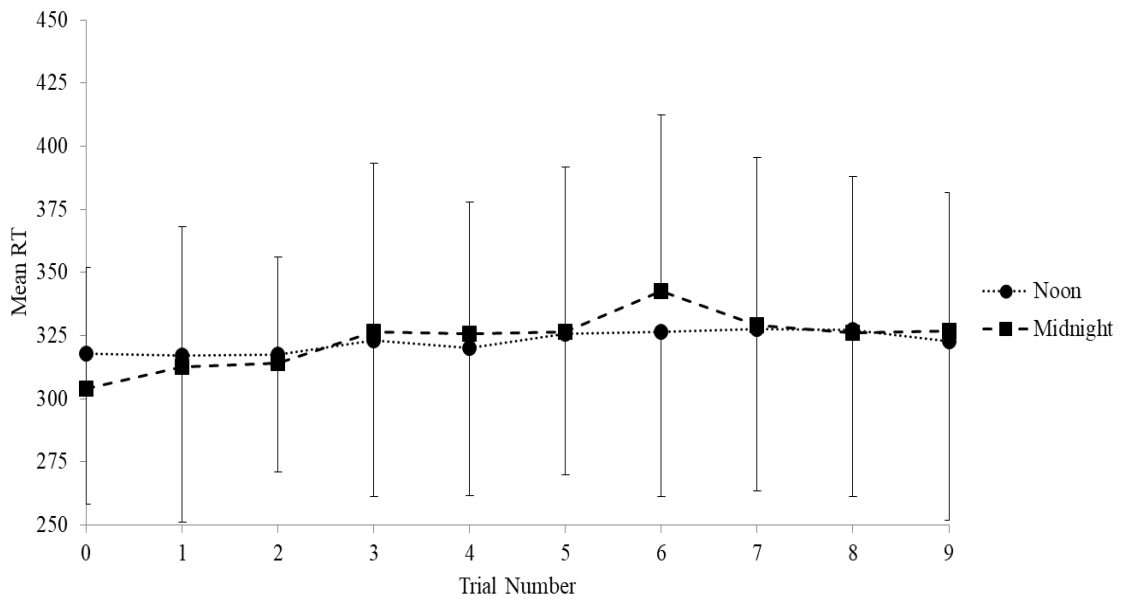
Note. NASA-PVT performance (RT; y-axis) plotted by trial number (x-axis); Noon performance is represented by the dotted line with circles, with midnight represented by the hashed lines with squares. Error bars represent standard deviation.

Relationship between Sleep and Performance

Sleep duration was significantly positively correlated with NASA-PVT performance for the noon drives, $r(64) = .30$, $p = .013$, 99% CI [0.06, 0.51], accounting for 9% of the total variance. This would imply that performance got worse as the hours of sleep attained prior increased, which is counterintuitive based on the literature. Sleep duration was also found to be significantly negatively correlated with

NASA-PVT performance for the midnight drives, $r(68) = -0.34$, $p = .004$, 99% CI [-0.53, -0.11], accounting for 11% of the total variance. This would suggest that performance was best when sleep duration prior was increased.

NASA-TLX

There was a significant increase in TLX scores from the noon drive ($M = 37.93$, $SD = 20.09$), compared to the midnight drive ($M = 32.09$, $SD = 21.74$), $t(65) = 2.81$, $p = .007$, 99% CI [0.29, 10.91], $g = 0.26$, meaning the participants felt their overall workload was higher during the noon drive (see Figure 5).

Figure 5

NASA-TLX Weighted Workload
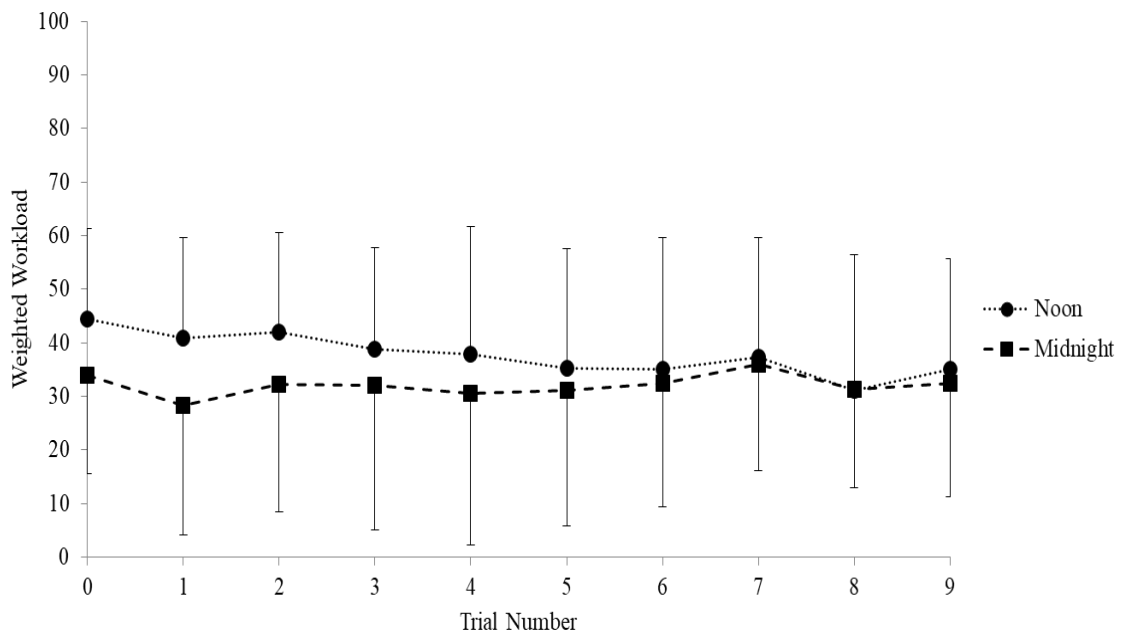
Note. NASA-TLX performance (weighted workload; y-axis) plotted by trial number (x-axis); Noon performance is represented by the dotted line with circles, with midnight represented by the hashed lines with squares. Error bars represent standard deviation.

For individual workload dimensions, Temporal Demand was significantly higher in the noon drive ($M = 40.15$, $SD = 28.74$) compared to the midnight drive ($M = 32.50$, $SD = 27.04$), $t(65) = 2.62$, $p = .011$, 99% CI [-0.46, -0.06], $g = 0.26$. Frustration was also significantly higher in the noon drive ($M = 31.36$, $SD = 30.44$) compared to the midnight drive ($M = 22.57$, $SD = 28.09$), $t(65) = 3.34$, $p = .0014$, 99% CI [16.45, 1.89], $g = 0.31$.

Temporal and Mental Demand, closely followed by Effort, were the highest rated workload demands during the noon drives (see Figure 6). Conversely, Mental Demand, closely followed by Own Performance and Effort, were the highest rated workload demands during the midnight drives. Notably, physical demands were rating slightly higher in the midnight drives ($M = 11.36$, $SD = 20.78$) compared to the noon drives ($M = 8.94$, $SD = 14.26$).

Figure 6

NASA-TLX Individual Domains
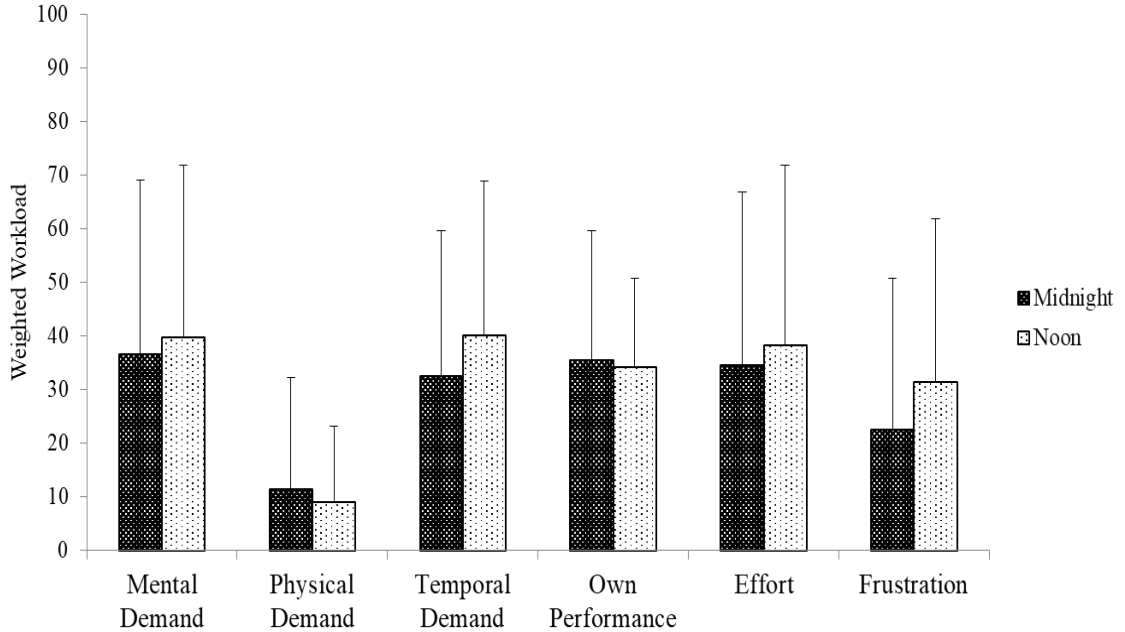
Note. NASA-TLX performance (mean weighted workload; y-axis) plotted against individual workload domains (x-axis). Noon performance is represented by the lightly dotted bars, and midnight with the heavily dotted bars. Error bars represent standard deviation.

Rover Health

Four out of the five drivers exceeded a pitch or roll tilt of 15-degrees during the noon drives, while three of the five exceeded during the midnight drives (Figure 7). In the noon drives, one participant exceeded only once, two exceeded twice, and one exceeded four times; the simulation needed to be reset twice due to loss of control. Of the midnight drives, one participant exceeded four times, while the other two exceeded five times; the simulation needed to be reset once due to loss of control.

Figure 7

Rover Health
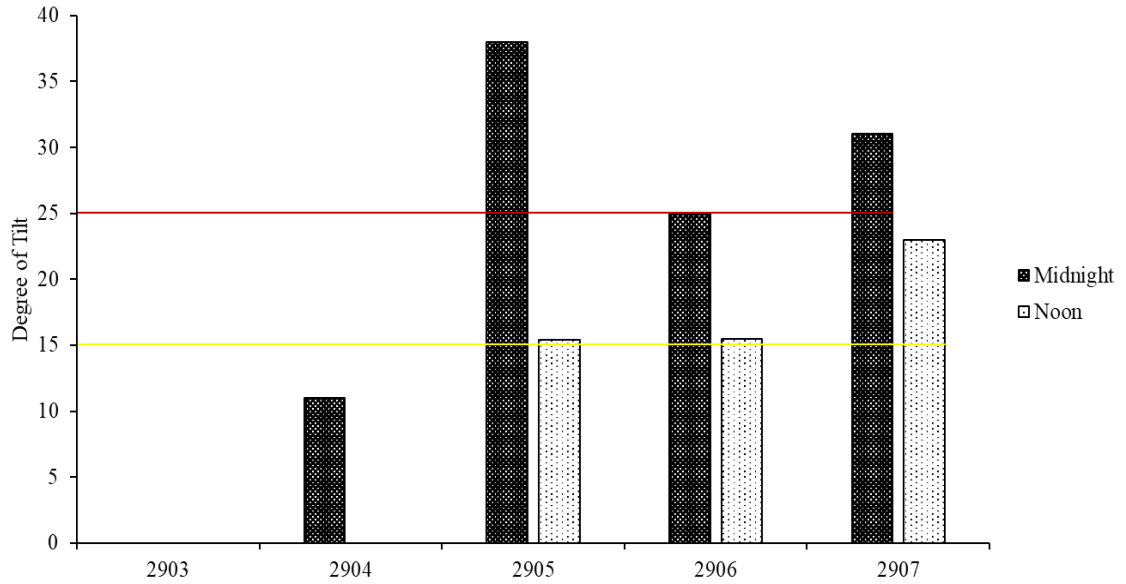
Note. Rover health indicated by max degree of tilt (accounting both pitch and roll) achieved by the driver during both the noon (lightly dotted) and midnight (heavily dotted) drives. A tilt of 15° was agreed as the limit for safe travel (yellow line), where a tilt exceeding 25° would be indicative of a crash (red line). It should be noted that tilt was only recorded for drivers that approached 15°, hence the empty cells for 2903, and 2904's Noon drive.

Discussion

We evaluated how long an individual could operate a lunar rover before the effects of fatigue and workload began to impair performance. We also evaluated the changes in performance over time during work shifts starting at noon and work shifts starting at midnight. We found that on average, participants rated themselves as subjectively sleepier with a greater range of scores during the midnight drives compared to the noon drives. Objective performance did not differ between the noon and midnight drives on average but was more variable at night. Our evaluation of performance over time suggests that drivers should rotate into a different role after three hours of driving during the day and night. In the following sections, we highlight our results compared to current literature findings that have assessed similar populations in terms of operators' subjective sleepiness, psychomotor vigilance, workload, and operational performance.

Subjective Sleepiness

We found that participants rated themselves to be much sleepier during the midnight drives, with a mean difference of almost two points. Previous researchers have found subjective sleepiness to increase when work shifts are either extended (Barger et al., 2014), or scheduled during the biological night (Kazemi et al., 2016). In 2014, Amann et al. assessed pilots' subjective sleepiness during extended operations and found it to be rated the highest (i.e., a KSS score greater than 7) for missions that exceeded 20 hours. Similarly, when operators perform during the biological night, sleepiness tends to increase as time continues, before tapering off as the morning approaches (Kazemi et al., 2016). This has been consistently shown across a variety of operational environments,

where sleepiness tends to reach well above average during night shifts (Garbarino et al., 2002; Harma et al., 2002). However, researchers assessing underground miners found that years of experience seemed to mitigate this subjective fatigue, in that the only reported difference was at the end of their night shift (Legault et al., 2017). In the present study, we found that almost all participants, regardless of years of experience, rated their subjective sleepiness to be higher during the midnight drives. Only one driver reported lower levels of sleepiness throughout both drives; however, this participant's rating spiked on the very last trial, which coincides with the literature suggesting sleepiness to be highest at the end of the night shift (Dinges, 1995). Often coupled with this increase in subjective sleepiness are other cognitive deficits that have been shown to worsen throughout the shift (i.e., in terms of reaction time and lapses).

Psychomotor Vigilance

Our findings support that an increase of time on task in driving the lunar rover leads to overall poorer performance. We found that performance on the NASA-PVT exceeded two standard deviations from baseline after approximately three hours of driving during both the noon and midnight drives. At night, performance was more variable between participants with worse performance towards the end of the night shift. This coincides with much of the literature on shift work; In 2011, Ferguson et al. assessed miners' psychomotor performance through a hand-held NASA-PVT while working a variety of shift schedules (e.g., comprising a combination of consecutive day and night shifts, with few days off in between) and found a significant main effect of test timing, in that reaction time was significantly lower during the end of the night shifts compared to any

other time reported. The authors also compared performance based on the hours of sleep

attained in the prior 24-hours and found those who exceeded seven hours of rest resulted

in significantly better performance (i.e., in terms of mean differences of reaction time).

We observed a wide spread of average sleep durations across the participants; we found

that those who slept less leading up to their shift performed much worse compared to

those who slept more and was much more apparent during the midnight drive.

In 2019, Naerri et al. (2019) conducted a simulated flight study utilizing pilots from

the Department of Aviation at the University of Oklahoma found that as the number of

manual tasks required of the pilot increased, performance worsened. The authors also

correlated psychomotor performance with an eye tracking paradigm and found a decrease

in the number of average eye fixations, an increase in the average eye fixation duration,

as well as an increase in random eye movements as psychomotor performance declined.

Taken together, perhaps a combination of a psychomotor vigilance and an eye tracking

paradigm should be investigated to provide some context of a threshold of operator

performance when operating in real-time.

Workload

Contrary to our hypotheses, we found that overall weighted workload was slightly

higher during the noon drives compared to the midnight drives. When coupled with

subjective sleepiness levels, it seems that moderate levels of workload results in lower

subjective fatigue, with lower levels of workload resulting in higher levels of fatigue.

This exact phenomenon was observed among Naval crewmembers operating in a 5-day,

24-hour setting in that moderate levels of workload resulted in lower levels of subjective

sleepiness; the authors also reported that high levels of workload (exceeding the moderate range) resulted in an increase in fatigue once again (Grech et al., 2009). This coupling of subjective sleepiness and workload has been observed across a variety of occupational contexts as well. In 2019, Berg et al. assessed cabin crew pilots on ultra long-haul flights (over 12 hours) and found that higher levels of workload resulted in significantly higher levels of fatigue with a moderate effect. Arsintescu et al. (2020) found similar results with short-haul flights (under 3 hours) as well, but with a weaker effect size. Similarly, short-haul pilots were reported to have the highest workload during morning shifts, potentially a result of compensation for the negative effects of fatigue on performance (Bourgeois-Bougrine et al., 2018). Additionally, over 300 short-haul transport drivers reported that the four strongest contributors to their fatigue and workload were: (1) long driving hours, (2) insufficient night/pre-work sleep, (3) insufficient rest breaks, (4) and monotonous driving routes (Friswell & Williamson, 2008). Therefore, it seems that moderate levels of workload are required to maintain appropriate levels of fatigue, so long as adequate rest is achieved prior to the shift.

Another factor that seems to impact overall workload is the degree of autonomy in the simulation. Although not directly measured in this study, other researchers have found that tasks that involved manual operations resulted in a significant increase in workload among train operators when compared to autonomous tasks (Huang et al., 2019). This would also help explain the Naeeri et al. (2019) study mentioned previously, where pilots' performance became significantly more decremented as the number of manual tasks increased. Researchers investigating workload in various unmanned vehicle systems

also found that increasing the levels of automation may directly reduce workload; for example, in high-demanding missions with various visual displays, additional modalities of the display (e.g., auditory and or haptic cueing) can be used to further reduce workload and aid navigation (Hooey et al., 2017). Perhaps if some of these tasks used in the simulation were converted to be more autonomous, it would help preserve some of the operators' cognitive performance.

Operational Performance

We measured operational performance in terms of the overall "health" of the rover. Primarily, this consisted of monitoring the degree of tilt in which the simulated rover was placed; anything over 15-degrees was deemed as "unsafe," while anything exceeding 25-degrees was estimated to result in a definite loss of control of the rover. Our results, although not directly comparable to most operational contexts, does coincide with the available literature. Truck drivers operating 12-hour shifts around the clock were reported having a greater percent of failures in operational performance (i.e., through the Truck Operator Proficiency system which is a dual-axis tracking task that monitors speed and steering in controlled environments with its own divided attention task) during the biological night, as well as during extended (i.e., over 8 hours) shifts (Charlton & Baas, 2001).

In the present study, we found that three of the five drivers during the noon drives exceeded the agreed upon safe limit for travel (i.e., over 15-degrees), but did not exceed 25-degrees. During the midnight drive, the same three drivers all exceeded 25-degrees, with two reaching upwards of 30- to 35-degrees. For the midnight drives, two of the three

participants achieved less than 5 hours of sleep in the prior 24-hours which may have

resulted in their poor performance. However, one of the three received upwards of 9

hours and yet still resulted in a crash. Looking solely at previous sleep, it would be

contradictory to claim that one performed poorly due to lack of sleep, while others

performed safely with similar minimal hours of rest. Regardless, the worst operational

performance was observed during the end of the midnight drives, once again coinciding

with the literature that performance degrades much quicker towards the tail end of a night

shift. Taken together, it may be beneficial to preserve the last hour on shift for something

less cognitively demanding (e.g., documenting information for subsequent shifts). It may

also provide useful to rate the task complexity for each task in a given role before

scheduling operators. Two researchers compared the task complexity measure known as

TACOM with the NASA-TLX among 18 nuclear powerplant crewmembers and found

them to be correlated (Park & Jung, 2006). Therefore, it may be beneficial to assess

certain tasks with the TACOM prior to creating a shift roster to properly allocate

crewmembers to tasks (with a larger emphasis towards the tail end of night shifts) to

prevent accidents and overall burnout.

Implications

One discovery that was not predicted in our initial hypothesis was the higher overall

workload in the noon drives compared to the midnight. This could suggest that the noon

drive was more temporally and mentally demanding, requiring greater overall effort. On

the other hand, this could have been reflective of the participants' stresses of the day

considering they took five hours away from their normal work schedule to participate in

this study. This would also help explain why Mental, as well as Physical Demands were higher in the midnight drive as they were not accustomed to working against their body's biological clock and were therefore required to exert more effort.

Our findings have practical implications for individuals working in similar environments. For example, drone operators typically control an unmanned aerial system from a ground control station and is typically performed around the clock. Due to the anticipated increase in prevalence of such operations (i.e., where an individual is controlling a vehicle from a ground control station), our findings may provide benefit for future operations.

In 2016, the United States Air Force School of Aerospace Medicine conducted a fatigue and workload study on crewmembers working on the MQ-9 Reaper unit (Scheiman et al., 2018). The concern was stemmed from the increase of daily shift work from 8 to 12 hours, where crews were required to work five consecutive 12-hour days with only two days off, including rapidly alternating between shifts adding up to 60+ hours a week. A total of 72 operators participated, assessing both pilots and sensory operators. During the end of the shifts, around 18% of pilots reported a substantial increase in degraded alertness with roughly 23% of sensory operators. This split between the pilots and sensory operators should be considered when creating schedules, such that the tasks performed by both do not reflect the same workload. In the present study, drivers were paired with RTSci operators; and although we had an uneven split between the two groups, we found that drivers tended to have more variable performance with greater levels of subjective fatigue.

As mentioned earlier, Scheiman et al. (2018) reported that 30% of pilots and around 16% of sensory operators expressed fatigue prior to the mission altogether, suggesting that they may live in a state of persistent fatigue. This was also observed in the present study as many of the drivers came into their drives with very little average sleep beforehand. The authors reported symptoms that stemmed from their fatigue, including headaches with eye strain, decreased appetite while on shift, and 30% expressing increased frustration and irritability (i.e., towards co-workers and or family members). Participants also reported in interviews that it was not uncommon to fight fatigue when commuting home due to the long hours. This increase in frustration can negatively impact the work environment by disrupting crucial communication between the operators, not to mention the dangers in commuting home while fatigued. Future operations should take these into consideration when creating rosters, and perhaps provide lodging to those who need to commute far distances to and from work.

Due to the similarity between nurses, miners, drone and mission control operators (i.e., in terms of the 24-hour operations, staffing requirements, and similar levels of cognitive output required), many of the findings reported in the literature are generalizable to the context of this study. Researchers have found that the temporal sensitivity of an operation while performing situational awareness tasks led to the largest increases in workload across all other tasks (Hendy, 1995). This may help explain why workload was rated higher during the noon drives, simply due to the time pressures of participating in this study while still having work of their own to do. Regardless, it is imperative to assess the effects of the shift start time during these missions. Identifying

how much time a driver can reasonably operate during both the day and night can provide better insight regarding staffing requirements, and guidance to future mission operations. Although fatigued workers may continue to function, the potential for error will only increase as their quality of work diminishes, especially when operating a mentally and or physically straining task.

These results may have various research and practical implications for shift workers who need to perform rigorous duties around the clock. Although our study was simulation based, it remains one of the few studies to evaluate a specific type of mission control operation (i.e., real-time, reactive) from a fatigue-risk management perspective. This study should also provide some insight in preparing for the future VIPER mission through proper shift scheduling, and insight into the effects of fatigue on operators' performance. As of now, many such operations have multiple tasks that each individual is assigned to. However, the issue occurs when one individual becomes stuck performing one role and is forced to sometimes work over 12 hours in one sitting. Based on our results, operators need some form of break or relief of duty after three hours of continuous work in order to be at their maximum performance levels. Proper scheduling should also ensure that no one individual is forced to work one role for a prolonged period.

Limitations

While our study was one of the first to evaluate a specific real-time, reactive telerobotic simulation from a fatigue perspective, it was not without limitation. Limitations include dose dependency for either melatonin and or caffeine use as these

were not controlled for during the study. It was quite common for most if not all the

participants to be have consumed caffeine either hours or minutes prior to the start of the

drives. We did not assess participants' use of melatonin either. Also, there was the

potential for sleep inertia to have been a factor from waking from recuperative naps. For

example, we observed a wide range of sleep durations prior to the midnight drive, which

suggests that some attempted to nap prior while others did not. There may not be a "be-

all-end-all" solution for fatigue as it encompasses many environmental and physiological

factors; therefore, some form of sleep-monitoring technology should be present during

these shifts to prevent accidents and errors due to fatigue. As mentioned previously,

perhaps some form of psychomotor vigilance and eye tracking paradigm can be utilized

as studies have found correlations between the two. This system could alert personnel if

their performance drops below "optimum performance levels;" however, there is

currently no gold standard as to what that may be.

Due to the small number of participants and representation of only one female, our

study findings may not be generalizable to all individuals completing future mission

operations. Although our study population represents nearly half of those currently

trained for such duties, it is not enough to be able to generalize to other individuals that

may be trained to conduct these operations in the future. In addition, we also utilized two

members from the Fatigue Countermeasures Laboratory to act in the remaining three

RTSci roles due to uneven distribution of drivers to RTSci operators. Although their data

was not utilized in the analyses, the unfamiliarity of the lab members acting in this role

may have affected the drivers in some unpredicted way, such as through providing a

sense of laxness of the simulation due to lack of technical knowledge, for example. Although the lab members were instructed how to perform in their role, it still may have biased the dynamic between the driver and our confederate RTSci operator. Similarly, all participants were instructed to avoid carrying any side conversation unrelated to the simulation, though this was not always the case throughout both the noon and midnight drives as each pair of drivers and RTSci operators held side chatter at some point throughout. However, we were told early on from key members of the future VIPER mission that side chatter will be very common in and around the control room during the real operation.

As with any laboratory simulation, our study lacked consequences for errors made during the simulation. Although the participants were instructed by both the researchers and flight director, there were no repercussions for "crashing" the rover. During a real mission operation, there is likely to be a much greater component of stress than was simulated here due to the vast amount of time, money, and resources that goes into these missions. In addition, there were many more individuals in and around the control room during the noon drives and no other individuals present during the midnight drives, which may explain why some of the workload measures were unexpectedly higher during the day. As mentioned in the previous paragraph, we were told that during a real mission environment the control room will always be occupied due to the nature of the continuous operation. Therefore, our lack of a busy control room during the midnight drives may have confounded the data. Although some side conversation occurred during these

midnight drives that could have potentially kept them more alert, we still saw a decline in performance and an increase in overall sleepiness during the midnight drive.

Lastly, other limitations that may have affected the study results that were not anticipated were technical difficulties with the simulation itself. There was one occurrence during a noon drive where the acting flight director could not be present. Taking their place, another highly trained individual in the future VIPER mission acted in this role. However, there were some nuances of the simulation that were not fully translated (e.g., how certain displays were set up and configured), therefore taking roughly 30 extra minutes to begin. This may have confounded the results in some unseen way, despite accounting for the extra 30 minutes towards the end.

Future Directions

Motivation and Consequence

As the VIPER mission nears, a great deal of time will be spent developing further simulation tests and trainings to prepare as much as possible. It would be beneficial to continue analyzing operators during this time to gain a deeper understanding of their fatigue and workload. Although the participants were trained to operate the simulation and were instructed to take it as seriously as a real mission, we still lacked consequences/motivation for failure/completion. Future researchers should attempt to implore some motivational and consequential component, such that participants treat it as closely to a real mission as possible with a simulation. Future researchers should also study handover procedures. Although not heavily discussed in the literature, some researchers have found issues with handover due to time constraints, absence of written

documents, and lack of interdisciplinary communication (Fealy et al., 2018). Therefore, it would be beneficial to study this transfer of information from shift to shift, to ensure there is no loss of vital information. Although most operations likely have handover's scheduled, it would be prudent to allocate at least the last hour to documenting vital information.

Schedule Types

In the present study, we evaluated only two schedule types with the same duration: noon versus midnight, each lasting a total of 5 hours each. As discussed previously, the literature on shift work has shown repeatedly that poorly scheduled shifts can result in fatigue and cognitive impairment (Åkerstedt, 1995; Dinges, 1995; Weinger & Englung, 1990). Therefore, future researchers should also explore the different schedule types, comprised mainly of forward rotating schedule types as the literature suggests those to be most beneficial to one's fatigue and workload (Czeisler et al., 1980; Tucker et al., 1996). Some researchers have found fast, forward rotating shifts to be detrimental to performance, though this was demonstrated for solely night shift work (Baulk et al., 2009). Other operational professions requiring shift work (e.g., doctors) limit the consecutive night shifts to two or three and require at least two days off for every two nights worked (Cappuccio et al., 2009; Horrocks et al., 2006). Taking the MEQ into account should help construct these schedules, so that morning types can take a larger portion of morning shifts before rotating to evening, while those who prefer evenings can be scheduled with only a few mornings before rotating to the evening (Duffy et al., 2001).

Shift Lengths

Lastly, and in conjunction with evaluating schedule types, it would be beneficial to determine proper shift lengths for certain roles (e.g., driver vs RTSci operator). In our analyses, only one of the RTSci operators had worse performance during their midnight drive (i.e., indicative of mean RT), while the second operator remained resilient throughout. Perhaps this was indicative of the hours of sleep attained beforehand or their preference indicted by the MEQ; however, both RTSci operators had comparable hours of rest attained for both the noon and midnight drives, and both responded neutrally in the MEQ prescreen questionnaire. There are many other factors that could have accounted for this, such as motivation, personal stress, susceptibility to fatigue, or years of experience, but were not investigated in this study. Future researchers should therefore gather a larger sample—one that is gender balanced and consisting of different operator roles—and study them from a fatigue-risk management perspective to ensure both roles are operating safely and without decrement to their performance.

Projected Terrain

One of the ways to potentially reduce cognitive workload produced from strenuous tasks or frustration to time delays would be to externalize information that would normally reside in the operator's short-term memory (Chen et al., 2006). For example, one or a few operators during their initial drives mistakenly clicked off the hazard mapping display which ultimately erased all previous scans. One feature that should be implemented is some form of permanent/autonomous hazard mapping system that will continuously display each hazard that the rover traveled to prevent loss of important data.

This feature was not implemented in the simulation because each NavCam image would eventually accumulate error due to the drivable surface being a projected terrain, creating a major discrepancy towards where the rover thought it went versus where it actually went. For example, one of the drivers mentioned that they drove through rocks not projected on either the Haz- or NavCam, likely the result of this accumulated error. This increased difficulty in estimating vehicle position, speed, and the presence of environmental hazards can all significantly increase workload over time (Hooey et al., 2017). Perhaps a permanent mapping system that reset itself after a certain period, while informing the driver it is doing so, may help prevent loss of information and accumulation of statistical error. For example, considering this was a simulation, each scan performed by the operator created an aggregate estimate of what the program "thought" the moon's surface was. As the number of scans increased, it increased the overall accumulation of error which created a less accurate picture of the surface, making it increasingly difficult for the operator to both traverse and interpret the terrain.

Image Adjustment

Another feature that should be implemented in the future VIPER mission that was present in the simulation is the ability to adjust the brightness and contrast of the NavCam images taken. Some of the operators had expertise in specific functions of the simulation, such as the user interface and how it displayed the real-time data and elements of the NavCam, as well as the ability to adjust the brightness and contrast on the NavCam images. This made it easier to differentiate between rough and smooth terrain. Although not expressed by everyone, every driver spent time early in their drives

adjusting the contrast and continued to do so throughout the drives when faced with ambiguous images. For example, some images had simulated glare from the sun, making it much harder to differentiate between the terrain. This feature should remain in future missions, as previous telerobotic missions have named perception as being one of the many detriments to operator performance (Chen et al., 2007).

Image Labeling

Lastly, each image taken by the rover should have an appropriate labeling system. On a few instances, some of the operators were faced with ambiguous images from the NavCam displays. For example, one driver towards the end of their night shift viewed what looked to be old rover tracks and was confused if it was an old image or a newer one. They concluded that it could have been taken behind the rover as opposed to in front but were ultimately unsure. Considering fatigue is likely to interfere with short-term memory of image retention, timestamps as well as the location of the camera (or simply displaying which NavCam is taking the picture) should be displayed on each image. This would free up required cognitive resources from the operator, allowing them to easily understand the situation they are in based on the image.

Conclusion

In summary, we evaluated how long a remote operator could drive a lunar rover before the effects of fatigue and workload began to impair performance. Participants, on average, rated themselves as sleepier during the midnight drives with less overall workload demand compared to the noon drives. No significant differences were found in vigilance, though most participants during the midnight drives reached poor performance

thresholds after roughly three hours. Nearly half of the participants performed well throughout the noon drive, with only one of those same participants remaining resilient during the midnight drive. Our findings support limiting the duration of time that a telerobotic operator engages in the driving task.

Despite our limitations, our findings nonetheless have practical implications for individuals working in similar environments. For example, drone operators typically must control an unmanned aerial system from a ground control station through a series of scheduled shifts around all hours of the day of night, often causing operators to experience fatigue and performance degradation (Arrabito et al., 2010). These operations are becoming more prevalent today through the increasing use of technology and should therefore be further investigated. Our findings may provide useful for such operations, especially as NASA prepares for the VIPER launch planned for late 2023. Our findings also support the implementation of thoughtful schedule design that accounts for shift duration, timing, rotation, and time off to maintain control center staff alertness and performance. Therefore, future researchers should consider our findings and further investigate differing schedule types (i.e., comprised of fast, forward rotating shifts) and duration lengths of main role activities to prevent fatigue and cognitive degradation.

References

Åkerstedt, T. (1995). Work hours, sleepiness and the underlying mechanisms. *Journal of Sleep Research*, *4*, 15–22. https://doi.org/10.1111/j.1365-2869.1995.tb00221.x

Åkerstedt, T. (2003). Shift work and disturbed sleep/wakefulness. *Occupational Medicine*, *53*(2), 89–94. https://doi.org/10.1093/occmed/kqg046

Åkerstedt, T., Wright, K. P. (2009) Sleep loss and fatigue in shift work and shift work disorder. *Sleep Medicine Clinics, 4*(2), 257–271.

Amann, U., Holmes, A., Caldwell, J., & Hilditch, C. (2014). Sleep and sleepiness of pilots operating long-range airplane emergency medical missions. *Aviation, Space, and Environmental Medicine*. *85*, 954 – 9. https://doi.org/10.3357/ASEM.3986.2014

Arrabito, G. R., Ho, G., Lambert, A., Rutley, M., Keillor, J., Chiu, A., Au, H., & Hou, M. (2010). *Human factors issues for controlling uninhabited aerial vehicles: preliminary findings in support of the Canadian Forces Joint Unmanned Aerial Vehicle Surveillance Target Acquisition System Project* (Report No. DRDC Toronto TR 2009-043). Defense Technical Information Center. https://apps.dtic.mil/sti/citations/ADA543186

Arsintescu, L., Chachad, R., Gregory, K. B., Mulligan, J. B., & Flynn-Evans, E. E. (2020). The relationship between workload, performance and fatigue in a short-haul airline. *Chronobiology International,* 1–3. https://doi.org/10.1080/07420528.2020.1804924

Arsintescu, L., Kato, K. H., Cravalho, P. F., Feick, N. H., Stone, L. S., & Flynn-Evans, E. E. (2019). Validation of a touchscreen psychomotor vigilance task. *Accident Analysis & Prevention*, *126*, 173–176. https://doi.org/10.1016/j.aap.2017.11.041

Baillie, L., & Thomas, N. (2018). Changing from 12-hr to 8-hr day shifts: A qualitative exploration of effects on organising nursing care and staffing. *Journal of Clinical Nursing*, *28*(1-2), 148–158. https://doi.org/10.1111/jocn.14674

Baker, A., Heiler, K., & Ferguson, S. A. (2001). The effects of a roster schedule change from 8- to 12-hour shifts on health and safety in a mining operation. *Journal of Human Ergology*, *30*, 65–70.

Barger, L. K., Sullivan, J. P., Vincent, A. S., Fiedler, E. R., McKenna, L. M., Flynn-Evans, E. E., Gilliland, K., Sipes, W. E., Smith, P. H., Brainard, G. C., & Lockley, S. W. (2012). Learning to live on a Mars day: Fatigue countermeasures during the Phoenix Mars Lander mission. *Sleep: Journal of Sleep and Sleep Disorders Research*, *35*(10), 1423–1435.

Barger, L. K., Wright, K. P., Burke, T. M., Chinoy, E. D., Ronda, J. M., Lockley, S. W., & Czeisler, C. A. (2014). Sleep and cognitive function of crewmembers and mission controllers working 24-h shifts during a simulated 105-day spaceflight mission. *Acta Astronautica, 93*, 230–242. https://doi.org/10.1016/j.actaastro.2013.07.002

Berg, M. J., Signal, T. L., & Gander, P. H. (2019). Perceived workload is associated with cabin crew fatigue on ultra-long range flights. *The International Journal of Aerospace Psychology, 29*(3–4), 74–85. https://doi.org/10.1080/24721840.2019.1621177

Boivin, D., & Boudreau, P. (2014). Impacts of shift work on sleep and circadian rhythms. *Pathologie Biologie*, *62*(5), 292–301. https://doi.org/10.1016/j.patbio.2014.08.001

Bourgeois-Bougrine, S., Gabon, P., Mollard, R., Coblentz, A., & Speyer, J. (2018). Fatigue in aircrew from shorthaul flights in civil aviation: The effects of work schedules. *Human Factors and Aerospace Safety,* 177–187. https://doi.org/10.4324/9781315194035-5

Bowers-Hutto, C., & Davis, L. L. (1989). 12-hour shifts: Panacea or problem? *Nursing Management*, *20*(6). https://doi.org/10.1097/00006247-198908000-00019

Brooks, I., & Swailes, S. (2002). Analysis of the relationship between nurse influences over flexible working and commitment to nursing. *Journal of Advanced Nursing*, *38*(2), 117–126. https://doi.org/10.1046/j.1365-2648.2002.02155.x

Bunce, D., & Sisa, L. (2002). Age differences in perceived workload across a short vigil. *Ergonomics*, *45*(13), 949–960. https://doi.org/10.1080/00140130210166483

Carskadon, M. A., & Dement, W. C. (1981). Cumulative effects of sleep restriction on daytime sleepiness. *Psychophysiology, 18*(2), 107–113. https://doi.org/10.1111/j.1469-8986.1981.tb02921.x

Charlton, S. & Baas, P. (2001). Fatigue, work-rest cycles, and psychomotor performance on New Zealand truck drivers. *New Zealand Journal or Psychology*, *30*(1), 32–39.

Chen, J. Y. C., Haas, E. C., & Barnes, M. J. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, *37*(6), 1231–1245. https://doi.org/10.1109/tsmcc.2007.905819

Chen, J. Y. C., Haas, E. C., Pillalamarri, K., & Jacobson, C. N. (2006). *Human robot interface: Issues in operator performance, interface design, and technologies* (ARL-TR-3834). U.S. Army Research Laboratory. https://doi.org/10.21236/ada451379

Crummy, F., Cameron, P. A., Swann, P., Kossmann, T., & Naughton, M. T. (2008). Prevalence of sleepiness in surviving drivers of motor vehicle collisions. *Internal Medicine Journal*, *38*(10), 769–775. https://doi.org/10.1111/j.1445-5994.2008.01629.x

Czeisler, C. A., Duffy, J. F., Shanahan, T. L., Brown, E. N., Mitchell, J. F., Rimmer D. W., Ronda, J. M., Silva E. J., Allan, J. S., Emens, J. S., Dijk, D. J., & Kronauer, R. E. (1999). Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science*, *284*(5423), 2177–2181. https://doi.org/10.1126/science.284.5423.2177

Czeisler, C., Weitzman, E., Moore-Ede, M., Zimmerman, J., & Knauer, R. (1980). Human sleep: Its duration and organization depend on its circadian phase. *Science*, *210*(4475), 1264–1267. https://doi.org/10.1126/science.7434029

Dawson, D. & Reid, K. (1997). Fatigue, alcohol and performance impairment. *Nature*, *88*, 235.

Deboer, T. (2018). Sleep homeostasis and the circadian clock: Do the circadian pacemaker and the sleep homeostat influence each other's functioning? *Neurobiology of Sleep and Circadian Rhythms*, *5*, 68–77. https://doi.org/10.1016/j.nbscr.2018.02.003

Dinges, D. F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*, *4*, 4–14. https://doi.org/10.1111/j.1365-2869.1995.tb00220.x

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., & Pack, A. I. (1997). Cumulative sleepiness, mood disturbance and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep, 20*(4), 267–277.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments & Computers*, 17(6), 652–655. https://doi.org/10.3758/BF03200977

Dijk, D. J., & Schantz, M. V. (2005). Timing and consolidation of human sleep, wakefulness, and performance by a symphony of oscillators. *Journal of Biological Rhythms*, *20*(4), 279–290. https://doi.org/10.1177/0748730405278292

Duffy, J. F., Rimmer, D. W., & Czeisler, C. A. (2001). Association of intrinsic circadian period with morningness-eveningness, usual wake time, and circadian phase. *Behavioral Neuroscience*, *115*(4), 895–899. https://doi.org/10.1037//0735-7044.115.4.895

Estryn-Béhar, M., & Van der Heijden, B. I. J. M. (2012). Effects of extended work shifts on employee fatigue, health, satisfaction, work/family balance, and patient safety. *Work*, *41*, 4283–4290. https://doi.org/10.3233/wor-2012-0724-4283

Ferguson, S. A., Paech, G. M., Dorrian, J., Roach, G. D., & Jay, S. M. (2011). Performance on a simple response time task: Is sleep or work more important for miners? *Applied Ergonomics, 42*(2), 210–213. https://doi.org/10.1016/j.apergo.2010.06.010

Fitzpatrick, J. M., While, A. E., & Roberts, J. D. (1999). Shift work and its impact upon nurse performance: current knowledge and research issues. *Journal of Advanced Nursing*, *29*(1), 18–27. https://doi.org/10.1046/j.1365-2648.1999.00861.x

Flynn-Evans, E. E., Arsintescu, L., Gregory, K., Mulligan, J., Nowinski, J., & Feary, M. (2018). Sleep and neurobehavioral performance vary by work start time during non-traditional day shifts. *Sleep Health*, *4*(5), 476–484. https://doi.org/10.1016/j.sleh.2018.08.002

Folkard, S. (2008). Do permanent night workers show circadian adjustment? A review based on the endogenous melatonin rhythm. *Chronobiology International*, *25*(2-3), 215–224. https://doi.org/10.1080/07420520802106835

Fossum, I. N., Bjorvatn, B., Waage, S., & Pallesen, S. (2013). Effects of shift and night work in the offshore petroleum industry: A systematic review. *Industrial Health*, *51*(5), 530–544. https://doi.org/10.2486/indhealth.2013-0054

Friswell, R., & Williamson, A. (2008). Exploratory study of fatigue in light and short haul transport drivers in NSW, Australia. *Accident Analysis & Prevention, 40*(1), 410–417. https://doi.org/10.1016/j.aap.2007.07.009

Garbarino, S., Nobili, L., Beelke, M., Balestra, V., Cordelli, A., & Ferrillo, F. (2002). Sleep disorders and daytime sleepiness in state police shiftworkers. *Archives of Environmental Health: An International Journal, 57*(2), 167–173. https://doi.org/10.1080/00039890209602932

Geiger-Brown, J., Rogers, V. E., Trinkoff, A. M., Kane, R. L., Bausell, R. B., & Scharf, S. M. (2012). Sleep, sleepiness, fatigue, and performance of 12-hour-shift nurses. *Chronobiology International, 29*(2), 211–219.

Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Circadian rhythms, sleep deprivation, and human performance. *Progress in Molecular Biology and Translational Science Chronobiology: Biological Timing in Health and Disease*, 155–190. https://doi.org/10.1016/b978-0-12-396971-2.00007-5

Haga, S., Shinoda, H., & Kokubun, M. (2002). Effects of task difficulty and time-on-task on mental workload. *Japanese Psychological Research*, *44*(3), 134–143. https://doi.org/10.1111/1468-5884.00016

Harma, M., Sallinen, M., Ranta, R., Mutanen, P., & Muller, K. (2002). The effect of an irregular shift system on sleepiness at work in train drivers and railway traffic controllers. *Journal of Sleep Research, 11*(2), 141–151. https://doi.org/10.1046/j.1365-2869.2002.00294.x

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(11), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology Human Mental Workload,* 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9.

Hendy, K. C. (1995, April 24–28) Situation awareness and workload: Birds of a feather? *Situation Awareness: Limitations and Enhancement in the Aviation Environment* [Symposium]. AGARD AMP Symposium, Brussels, BD, EUR.

Hooey, B., Kaber, D., Adams, J., Fong, T., & Gore, B. 2018. The underpinnings of workload in unmanned vehicle systems. *IEEE Transactions on Human-Machine Systems. 48*(5).

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4,* 97–110.

Horrocks, N., Pounder, R., & RCP Working Group (2006). Working the night shift: Preparation, survival and recovery--a guide for junior doctors. *Clinical medicine (London, England)*, *6*(1), 61–67. https://doi.org/10.7861/clinmedicine.6-1

Hoonakker, P., Carayon, P., Gurses, A. P., Brown, R., Khunlertkit, A., Mcguire, K., & Walker, J. M. (2011). Measuring workload of ICU nurses with a questionnaire survey: The NASA Task Load Index (TLX). *IIE Transactions on Healthcare Systems Engineering*, *1*(2), 131–143. https://doi.org/10.1080/19488300.2011.609524

Huang, Y., Li, L., Liu, Z., Zhu, H., & Zhu, L. (2019). Assessment of urban railway transit driver workload and fatigue under real working conditions. *Transportation Research Record: Journal of the Transportation Research Board, 2673*(11), 891–900. https://doi.org/10.1177/0361198119826071

Ingre, M., Kecklund, G., Åkerstedt, T., & Kecklund, L. (2004). Variation in sleepiness during early morning shifts: A mixed model approach to an experimental field study of train drivers. *Chronobiology International*, *21*(6), 973–990. https://doi.org/10.1081/cbi-200036895

Jewett, M. E., & Kronauer, R. E. (1999). Interactive mathematical models of subjective alertness and cognitive throughput in humans. *Journal of biological rhythms*, *14*(6), 588–597. https://doi.org/10.1177/074873099129000920

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth Sleepiness Scale. *Sleep*, *14*(6), 540–545. https://doi.org/10.1093/sleep/14.6.540

Kazemi, R., Haidarimoghadam, R., Motamedzadeh, M., Golmohamadi, R., Soltanian, A., & Zoghipaydar, M. R. (2016). Effects of shift work on cognitive performance, sleep quality, and sleepiness among petrochemical control room operators. *Journal of Circadian Rhythms*, *14*(1), 1–8. https://doi.org/10.5334/jcr.134

Kecklund, G., Åkerstedt, T., & Lowden, A. (1997). Morning work: Effects of early rising on seep and alertness. *Sleep*, *20*(3), 215–223. https://doi.org/10.1093/sleep/20.3.215

Lamond, N. & Dawson, D. (1999). Quantifying the performance impairment associated with fatigue. *Journal of Sleep Research, 8*(4), 255–262, https://doi.org/10.1046/j.1365-2869.1999.00167.x

Legault, G., Clement, A., Kenny, G. P., Hardcastle, S., & Keller, N. (2017). Cognitive consequences of sleep deprivation, shiftwork, and heat exposure for underground miners. *Applied Ergonomics, 58*, 144–150. https://doi.org/10.1016/j.apergo.2016.06.007

Lockley, S. W., Arendt, J., & Skene, D. J. (2007). Visual impairment and circadian rhythm disorders. *Dialogues in Clinical Neuroscience*, *9*(3), 301–314.

McCann, R. S., & Spirkovska, L. (2005). *Human factors of integrated systems health management on next-generation spacecraft*. First International Forum on Integrated System Health Engineering and Management in Aerospace, Napa, CA.

Mitchell, R. J., & Williamson, A. M. (2000). Evaluation of an 8 hour versus a 12 hour shift roster on employees at a power station. *Applied Ergonomics*, *31*(1), 83–93. https://doi.org/10.1016/s0003-6870(99)00025-3

Muller, R., Carter, A., & Williamson, A. (2008). Epidemiological diagnosis of occupational fatigue in a fly-in–fly-out operation of the mineral industry. *The Annals of Occupational Hygiene*, *52*(1), 63–72. https://doi.org/10.1093/annhyg/mem058

Park, J., & Jung, W. (2006). A study on the validity of a task complexity measure for emergency operating procedures of nuclear power plants—Comparing with a subjective workload. *IEEE Transactions on Nuclear Science, 53*(5), 2962–2970. https://doi.org/10.1109/tns.2006.882149

Pilcher, J. J., & Huffcutt, A. I. (1996). Effects of sleep deprivation on performance: A meta-analysis. *Sleep, 19*(4), 318–326. https://doi.org/10.1093/sleep/19.4.318.

Pilcher J. J., Lambert B. J., & Huffcutt A. I. (2000). Differential effects of permanent and rotating shifts on self-report sleep length: A meta-analytic review. *Sleep, 23*(2), 155–163. https://doi.org/10.1093/sleep/23.2.1b

Randell, R., Wilson, S., & Woodward, P. (2011). The importance of the verbal shift handover report: A multi-site case study. *International Journal of Medical Informatics, 80*(11), 803–812. https://doi.org/10.1016/j.ijmedinf.2011.08.006

Rosa, R. R. (1995). Extended workshifts and excessive fatigue. *Journal of Sleep Research*, *4*, 51–56. https://doi.org/10.1111/j.1365-2869.1995.tb00227.x

Scheiman, T., Chappelle, W., & Sanford, E. (2018). *U.S. Air Force Special Operations Command remotely piloted aircraft: Operator fatigue levels and compensatory strategies* (AFRL-SA-WP-TR-2018-0013). Air Force Research Laboratory: https://apps.dtic.mil/docs/citations/AD1057349

Smith, M. R. & Eastman, C. I. (2012). Shift work: Health, performance and safety problems, traditional countermeasures, and innovative management strategies to reduce circadian misalignment. *Nature and Science of Sleep*, *4*, 111–132. https://doi.org/10.2147/nss.s10372

Takeyama, H., Itani, T., Tachi, N., Sakamura, O., Murata, K., Inoue, T., Takanishi, T., Suzumura, H., & Niwa, S. (2005). Effects of shift schedules on fatigue and physiological functions among firefighters during night duty. *Ergonomics*, *48*(1), 1–11. https://doi.org/10.1080/00140130412331303920

Tucker, P., Barton, J., & Folkard, S. (1996). Comparison of eight and 12 hour shifts: Impacts on health, wellbeing, and alertness during the shift. *Occupational and Environmental Medicine*, *53*(11), 767–772. https://doi.org/10.1136/oem.53.11.767

Tucker, P., Brown, M., Dahlgren, A., Davies, G., Ebden, P., Folkard, S., Hutchings, H., & Åkerstedt, T. (2010). The impact of junior doctors' worktime arrangements on their fatigue and well-being. *Scandinavian Journal of Work, Environment & Health*, *36*(6), 458–465. https://doi.org/10.5271/sjweh.2985

Van Dongen, H. P., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep, 26*(2), 117–126. https://doi.org/10.1093/sleep/26.2.117.

Weinger, W. B., & Englung, C.E. (1990). Ergonomic and human factors affecting anesthetic vigilance and monitoring performance in the operating room environment. *Anesthesiology*, *73*, 995–1021.

Williamson, A. M., & Feyer, A. M. (2000). Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and Environmental Medicine*, *57*(10), 649–55.

Wright, K. P., Hughes, R. J., Kronauer, R. E., Dijk, D. J., & Czeisler, C. A. (2001). Intrinsic near-24-h pacemaker period determines limits of circadian entrainment to a weak synchronizer in humans. *Proceedings of the National Academy of Sciences*, *98*(24), 14027–14032. https://doi.org/10.1073/pnas.201530198

Xiao, Y. M., Wang, Z. M., Wang, M. Z., & Lan, Y. J. (2005). The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index (English translation). *Chinese Journal of Industrial Hygiene and Occupational Diseases, 23*(3), 178–181.

Young, G., Zavelina, L., & Hooper, V. (2008). Assessment of workload using NASA Task Load Index in perianesthesia nursing. *Journal of PeriAnesthesia Nursing, 23*(2), 102-110. https://doi.org/10.1016/j.jopan.2008.01.008

Appendix

A. Demographics Questionnaire

The information contained in the form will be kept confidential. All analyses conducted on the information contained in this form will report aggregate statistical findings only, and will not identify individuals. Your cooperation is appreciated.

Date of Birth:_____

Height: _____
Weight: _____

What is your typical bedtime on WEEKDAYS?   _____
What is your typical wake-time on WEEKDAYS?   _____

What is your typical bedtime on WEEKENDS?   _____
What is your typical wake-time on WEEKENDS?   _____

B. Sleep Wake Log

Name:_____ Phone:_____ Date:_____

INSTRUCTIONS: Please make your form entry every night before going to bed and every morning upon awakening.

Date at bedtime	Day of week	Time into bed	Est. time to fall asleep	Day at wake time	Final wake time	Time out of bed

C. NASA Task Load Index

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

D. NASA Psychomotor Vigilance Task