Simulated manual spacecraft docking as a measure of operational performance under the influence of space flight-related stressors

DISSERTATION

submitted in partial fulfilment of the requirements for the degree of
„Doktor der Naturwissenschaften (Dr. rer. nat.)“
to the Faculty of Psychology at Ruhr University Bochum

by
Sarah Piechowski

First Examiner: Prof. Dr. Oliver T. Wolf
Second Examiner: Prof. Dr. Jörn Rittweger

April 2023
Tag der mündlichen Prüfung: 23.08.2023

Abstract

Astronauts live and work in an isolated, confined, and extreme environment. Mission success and crew safety relies on their ability to maintain high levels of performance. However, if and to what extent the stressors associated with space flight have a detrimental impact on cognitive performance is still controversial. Research has been lacking operational measures to evaluate performance in realistic and relevant tasks. The manual spacecraft docking simulation *6df* addresses this gap as a research tool for the evaluation of skill acquisition, maintenance, and performance reliability in a safety-critical operational task. Based on the *6df* tool, the aims of this thesis were the advancement of learning efficiency (study I), the identification of objective predictors of performance (study II), and the evaluation of the impact of sleep loss – a common stressor in space – on performance (study III).

Manual spacecraft docking is based on the skill to control an object with six degrees of freedom. The first study addressed the special challenge of three-dimensional (3D) control on the basis of a two-dimensional (2D) screen. To facilitate the mental representation of position and orientation in relation to the docking point, a stereoscopic presentation of the *6df* docking tasks was developed. Participants absolved the *6df* learning program during the course of 20 sessions in a standardized environment with simulated microgravity, i.e. during a six-degree head-down tilt bed rest study. Twelve participants were allocated to the standard 2D program and the other twelve to the 3D visualization that supplemented the first part of the acquisition process. Participants in the stereoscopic condition initially learned faster, but this advantage was not persistent over time. Likewise, there was no consistent effect of 3D presentation on learning success as measured by performance in a higher-fidelity docking simulation. Until further evidence is available, the more realistic 2D standard visualization can be favored.

In the second study, eye tracking was applied as a measure of visual attention allocation during the docking simulation. Manual docking is a visually demanding task that requires high levels of situation awareness. In the aviation sector, many examples support the notion that considering pilots' gaze behavior can enhance training and flight safety. In the 2D group from study I, number and total duration of dwells to predefined regions of interest on the simulation screen and their relationship with docking accuracy were analyzed. Participants concentrated most of their visual attention on the vizor that was used to orientate the spacecraft in the direction of the docking point. Frequency and duration of instrument checks were significantly associated with docking performance. These results pose an interesting starting point for the

development of tailored performance feedback and training interventions based on gaze behavior.

The disturbance of sleep quality and quantity is one stressor that is highly prevalent in space. Although sleep deprivation has been shown to have a detrimental impact on various cognitive domains, it is still unclear whether this transfers into more complex operational performance. Therefore, the third study evaluated the impact of 24 hours of total sleep deprivation on manual docking performance in a counterbalanced repeated measures cross-over design. In addition to *6df* performance, the *Psychomotor Vigilance Task* (PVT) was considered as a measure of sustained attention that is highly sensitive to sleep loss and a gold standard in sleep deprivation research. Our results showed that docking accuracy in difficult *6df* task levels decreased significantly after sleep deprivation in comparison with performance after eight hours of sleep. These impairments were partly explained by decrements in sustained attention. Participants with larger decrements in PVT response speed following sleep deprivation also exhibited larger decrements in docking accuracy. As a result, the PVT in combination with operational measures like *6df* seems promising to assess readiness for duty under sleep loss in demanding work environments.

Future long-duration missions, e.g. to Mars, will pose unprecedented challenges for human cognitive functioning. As a consequence, autonomous on-board systems are needed that are able to support efficient training and maintenance of critical operational skills as well as feedback on the operator’s state to facilitate self-monitoring. The results of this dissertation highlight the need to protect operational performance against sleep deprivation, a common stressor in space. Eye tracking and short cognitive tests like the PVT will represent valuable additions to *6df* to evaluate operators’ fitness for duty and to timely detect performance risks, e.g. due to insufficient instrument scanning and sleep loss. Promising approaches to advance learning efficiency arise in the further development of training systems based on virtual reality and gaze behavior feedback. In conclusion, the presented results add to the further development of *6df* as a motivating operator support tool that fosters the crew’s safety and autonomy during long-duration missions.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D	two-dimensional
3D	three-dimensional
AGBRESA	Artificial Gravity Bed Rest Study with European Space Agency
CC	control condition
DoF	degrees of freedom
HDT	head-down tilt
ISS	International Space Station
KSS	Karolinska Sleepiness Scale
LME	linear mixed effects model
PVT	Psychomotor Vigilance Task
ROI	region(s) of interest
SDC	sleep deprivation condition
TORU	Telerobotically Operated Rendezvous Unit
VR	virtual reality

1 General Introduction

"Merely gazing upon the dark sky lights up the imagination, presenting an abundance of possible futures and potentialities to the mind" (Rovetto, 2013).

Humanity's curiosity and urge to explore new worlds is not confined to Earth. After Yuri Gagarin's first space flight in 1961, more than 500 astronauts have followed him to space (Smith et al., 2020). Whereas in the beginning, scientists were afraid that prolonged microgravity might be fatal, humans turned out to adapt even to this most extreme environment (De la Torre, 2014). Human space flight aspires to promote knowledge, global cooperation, and technological advancement (Rovetto, 2013). For the astronauts involved, certainly the unique change in perspective on our planet might be one of the most remarkable effects of space flight (Yaden et al., 2016). But even for those who are confined to the Earth's surface, space is an infinite source of inspiration and creativity.

1.1 Human space flight as a psychological challenge – from low Earth orbit to Moon and Mars

Space is a particularly extreme environment that poses numerous challenges to the humans living and working in its hostile and unfamiliar setting. For space missions, Kanas and Manzey (2008) established four categories of stressors, i.e. features of the environment that affect the individual: physical, habitability, psychological, and interpersonal stressors. Physical stressors comprise the characteristics of the space environment, e.g. acceleration and radiation. Of course, the most obvious difference to Earth is microgravity, which greatly impacts physiological functioning (Demontis et al., 2017). Furthermore, the day-night cycle is reduced from 24 hours on Earth to 90 minutes on the International Space Station (ISS), which may disrupt the circadian rhythm (Basner & Dinges, 2014). These circumstances deviate from the boundary conditions of human evolution on Earth and, therefore, require extensive adaptation of the individual (Kanas & Manzey, 2008). Habitability stressors encompass the living environment in the station, e.g. noise levels, temperature, lighting schedules, vibration, and air quality. Psychological factors include isolation and confinement, as well as the threats and dangers astronauts are exposed to during their missions. Depending on crew schedules and mission phase, both monotony and high workload can characterize the working environment. Lastly, interpersonal stressors evolve around crew composition and size. Conflicts may arise due to leadership issues or differences in personality and cultural background. Space habitats provide only limited volume and a lack of privacy (Palinkas, 2007). Contact with family and friends is restricted, whereas contact with other crew members is enforced. In addition to these four categories of stressors, Morphew (2001) highlights the importance of human factors, such as limited equipment, risk of equipment failure, and technology-interface challenges in microgravity. All the stressors described and their interactions can result in various stress reactions, including disturbances of sleep, mood, and crew cohesion. Although many stressors are obligatory conditions during space flight, some can be attenuated by careful crew selection and training or in-flight monitoring and support (Kanas & Manzey, 2008).

At the beginning of human space flight, psychological issues were oftentimes disregarded (Harrison & Fiedler, 2011). Astronauts with "the right stuff" and military background were deemed to be insusceptible to the stressors of space flight. This invulnerability assumption proved itself wrong when reports of depressive symptoms, human errors, work overload, and interpersonal conflicts started to emerge in space (Palinkas, 2007). Increases in mission duration, crew size, and astronaut diversity raised the awareness of psychosocial factors for mission success (Harrison & Fiedler, 2011). In previous decades, space flight had developed from a competition between political systems to an international collaboration, leading to culturally diverse teams. The proportion of female astronauts has increased from 2.1% in the 1960's to 20% in the 2010's (Smith et al., 2020). Recently, women and men are selected in equal parts for astronaut classes. Whereas astronauts were mostly military test pilots in the early phases of space flight, they now include engineers, scientists, and physicians as well. Furthermore, the operation of orbital space stations like Salyut, Skylab, Mir, and ISS has led to an increase in mission duration from hours and days to months (Smith et al., 2020). Currently, most astronauts visit the ISS for about six months in a crew of six (Dempsey & Barshi, 2020). The record for the longest stay in space is 437.75 days, held by Valery Polyakov since 1995 (Smith et al., 2020). As a result of these developments in human space flight, the "right stuff" was redefined to include resilience and coping abilities for living in an isolated, confined, and extreme environment. Individuals that are able to regard stressful events as comprehensible, manageable, and meaningful, are not only able to cope, but might even gain strength as a buffer against future disruptions (Suedfeld, 2005). Today, space psychology is integrating countermeasures and prevention with positive psychology that fosters not only the absence of problems, but well-being and optimal performance of the individuals sent to space.

In the future, the impact of stressors associated with space missions will gain even more importance. Whereas today human space flight is focused on the ISS located in low Earth orbit, future missions will target Moon and eventually Mars (Hufenbach et al., 2014; National Aeronautics and Space Administration, 2018). As an intermediate step to Mars, Gateway is planned as a platform orbiting the Moon for long-term missions to the lunar surface and starting point for deep space exploration (Creech et al., 2022). Developments and experiences from low Earth orbit and Moon as well as robotic missions to Mars are supposed to be the foundation for the first human landing on Mars (National Aeronautics and Space Administration, 2018). With these ambitions, unprecedented risks and challenges arise that found a growing need for psychological countermeasures to preserve crew cohesion, performance, and well-being over extended time in even more extreme conditions (Sandal & Leon, 2011). The cruise to Mars alone is expected to take about six months, a whole mission about two and a half years (Stuster et al., 2018). This is more than twice as long as any human has ever been in space. Mission duration will likely influence performance and psychological functioning, as astronauts might be able to tolerate stressors and compensate for performance decrements only for limited durations (Kanas & Manzey, 2008). We lack knowledge about the influence of time in mission for Mars due to the unprecedented duration and time profile, including long transfer phases and activities on Martian surface. A Mars mission may include severe fluctuations in task load, from overload during critical mission phases (e.g. landing on Mars) to monotony and boredom (Whiteley & Bogatyreva, 2018). Therefore, motivation and mood will be harder to maintain – especially during the return flight, when all the important work has been done and there is not much left to prepare. Whereas astronauts during shorter missions oftentimes experience overload and hyperarousal, longer missions will likely also include phases of sensory deprivation. The psychological consequences may include irritability as well as impairments in sleep, concentration, and performance (Summers et al., 2005). In comparison with current missions in low Earth orbit, communication with family and friends as well as other psychological support (e.g. care packages, private conferences) will be very limited (Kanas et al., 2013; Patel et al., 2020). In case of an emergency, crew rescue or evacuation will not be possible, which may lead to chronic feelings of threat and anxiety (Kanas, 2011). In opposition to the awe that results from seeing the Earth from above (Yaden et al., 2016), we might evidence an Earth-out-of-view phenomenon when Earth disappears out of sight entirely. Feelings of isolation and loneliness could aggravate, possibly leading to depressive symptoms (Kanas & Manzey, 2008). Few individuals have been in space for more than a year and there is only one high-fidelity isolation study that comprised the duration that is expected for a Mars mission

(Basner et al., 2014). Only two out of six crew members showed no signs of distress or behavioral disturbances throughout the study. Disrupted sleep-wake periodicity, partial sleep deprivation associated with vigilance deficits, and impaired sleep quality were problems that occurred during the study (Basner et al., 2013). These findings hint at the increased risks for psychological functioning and performance that can be expected for extensive mission durations.

A crucial feature of Mars exploration will be the unprecedented autonomy of the crew (Dempsey & Barshi, 2020; Love & Reagan, 2013). Whereas communication delay to Moon is barely disruptive (about 1.3 seconds), communication from Mars to Earth will take up to 22 minutes one-way (Frank et al., 2013), making real time engagement from ground impossible. Today, crew autonomy is strictly limited. Mission control is responsible for the daily schedule and most of the decision-making; astronauts are monitored and supported throughout the day (Goemaere et al., 2016; Krikalev et al., 2010). Higher autonomy is usually welcomed by crew members (Slack et al., 2016) and can even produce a positive impact on performance (Roma et al., 2011). However, this comes with substantial risks, especially in the case of unexpected events and emergencies.

Another challenge, related to autonomy and long mission duration, is the maintenance of skills until the arrival on Mars, e.g. the operation of landers, rovers, or robotic arms. Without in-flight training of such critical skills, the duration between acquisition and application would exceed any safety limits. On ISS, novice astronauts additionally have the possibility to learn from their experienced colleagues during handover. But for distant exploratory missions, there will be no exchange of crew and even no experienced personnel (Barshi & Dempsey, 2016). Therefore, skill maintenance and likely also part of the acquisition training will have to take place in-flight with autonomous training systems, e.g. during the six-month cruise to Mars (Stuster et al., 2018). These considerations are especially important for sensorimotor skills that are acquired in 1g, maintained during 0g, and applied under varying gravitational forces, including 1/3g on Mars or 1/6g on Moon. Whereas the ISS crew receives ten days of preflight training for each day of their mission, this ratio will be unrealistic for prolonged mission durations to Mars (Barshi & Dempsey, 2016). Therefore, some skills will have to be acquired during the flight and it remains questionable if learning under such extreme conditions will be as efficient as on Earth (Kanas & Manzey, 2008). As threat and remoteness increase, our knowledge derived from prior missions will most likely be insufficient and specific research is needed to address the new circumstances of long-term missions (Hockey et al., 2011). In summary, exploratory missions to Mars will involve challenges to astronauts that largely exceed those on ISS today (Patel et al., 2020). The influence of these challenges on performance requires deeper understanding to optimally prepare and support astronauts for future long-duration missions.

1.2 Living and working in an adverse environment – a risk to cognitive and operational performance?

"Perhaps the most dangerous symptom is impairment to cognitive function – we have to be able to perform tasks that require a high degree of concentration and attention to detail at a moment's notice, and in an emergency, which can happen anytime, we need to be able to do those tasks right at the first time. Losing just a fraction of our ability to focus, make calculations, or solve problems could cost our lives" (Kelly, 2017).

This quote from Scott Kelly, who spent a whole year in space, gives a clear impression of the worries that are related to cognitive performance during long-term space flight. Space missions involve a complex work environment that poses high demands on cognitive and psychomotor functioning. Astronauts operate and maintain diverse technical systems and conduct complex experiments; errors are easily expensive or perilous. Still, the impact of microgravity on the body is better understood than its impact on cognition (Arshad & Ferré, 2022). Some newer approaches on this topic used neuroimaging to evaluate the impact of space flight on neurocognitive functioning. After astronauts return to Earth, changes in brain structure and functional connectivity can be observed (Mhatre et al., 2021). Especially cerebellum, cortical motor areas, and pathways related to vestibular functioning were affected, resulting in sensorimotor and vestibular deficiencies (Van Ombergen et al., 2017). Koppelmans et al. (2016) found decreases in gray matter volume in temporal and frontal areas, but increases in gray matter volume regarding the somatosensory and motor cortex – the latter probably associated with neuroplasticity during adaptation to microgravity. In general, it is expected that not all brain alterations are dysfunctional processes that will negatively impact health and performance. Some changes are rather seen as adaptive neural compensation, e.g. to adjust motor processes to microgravity conditions (Demertzi et al., 2016; Jillings et al., 2020). However, there is first evidence to suggest that brain alterations become more pronounced with increasing mission duration (Roy-O'Reilly et al., 2021) and that these also correlate with postflight changes in cognitive and motor performance (Roberts et al., 2019). All in all, neuroimaging studies in astronauts are still too scarce for a conclusive evaluation (Hupfeld et al., 2021). Also, they mostly focused on the effects of 6-month ISS missions, whereas the impact of longer mission durations on the brain is still unclear. Radiation and chronic stress for example – risks that will aggravate during missions that exceed low Earth orbit – have the potential to negatively affect critical brain structures for memory and learning, e.g. the hippocampus and basal ganglia (Steinberg, 2019; Strangman et al., 2014). Such potential brain alterations during space flight raise some concerns over cognitive performance, but astronauts themselves have also repeatedly uttered anecdotal reports of cognitive deficits, which are referred to as “space fog” (Kanas et al., 2001; Welch et al., 2009). Especially during future long-term missions, the maintenance of optimal performance is critical for mission safety and success, because ground control has only limited options to intervene. To achieve this goal, it is necessary to understand how the stressors associated with space flight affect performance (Kanas & Manzey, 2008).

The long-term exposure to various stressors such as high workload, sleep deprivation, or the isolated, confined, and extreme environment can put optimal cognitive functioning at risk (Kanas & Manzey, 2008; Morphew, 2001). A plethora of “Earth-based” studies has shown the degradation of mood and performance in response to extreme environmental stressors (Lieberman et al., 2002). Stress can impair cognitive flexibility (Shields et al., 2016), working memory (Schoofs et al., 2008; Schoofs et al., 2009; Shields et al., 2016), and long-term memory retrieval (Meir Drexler & Wolf, 2017; Wolf, 2017). Other undesirable consequences of stress include attentional selectivity, decrements in alertness (Hockey & Robert, 1997), and disadvantageous and risky decision-making (Starcke & Brand, 2016). Additionally, the tradeoff between speed and accuracy can be influenced, which means that stressed individuals tend to perform faster but less accurate (Staal, 2004). Also visuomotor performance tends to degrade under stress (Staal, 2004), especially if the situation is evaluated as a threat rather than a challenge (Vine et al., 2016). As astronauts live and work in an extreme and demanding environment, performance decrements due to stress are expected to occur during a mission (Patel et al., 2020). Stress levels might increase for future long-term missions, because of the unprecedented risk and isolation, as well as fluctuations between monotony and very high workload.

Next to the effects of chronic exposure to a variety of stressors, also microgravity itself can cause changes in brain functioning and performance. The absence of gravity leads to various physiological changes and adaptations. Cardiovascular (e.g. fluid shifts into the upper part of the body) and vestibular or sensory-motor effects evolve fast and will impair astronauts during the first days or weeks of a mission. In contrast, musculo-skeletal effects of microgravity will show later, but become more severe with increasing mission duration (Kanas & Manzey, 2008). In microgravity, the otolith organs of the vestibular system are not able to provide information on the vertical orientation of the body. The lack of congruence between visual, vestibular, and proprioceptive signals induces sensory conflicts which then may elicit spatial disorientation and space motion sickness (Davis et al., 1988; De la Torre, 2014), including symptoms such as fatigue, lack of initiative, vomiting, and performance decrements (Bloomberg et al., 2015). Usually, gravity is used as a frame of reference to determine one's own position relative to the external space (Glasauer & Mittelstaedt, 1998). Because the gravitational vertical is missing in space, the perceived spatial position of oneself and objects in the environment depends more on an egocentric reference frame, e.g. "up" is perceived where the head is located (Kanas & Manzey, 2008). Even spatial illusions are common, such as the feeling of hanging upside down or falling (Kornilova, 1997; Kornilova et al., 1996). There is evidence that distances are underestimated in space and also the estimation of height and depth of objects is distorted (Clément et al., 2013). Such errors in the perception of distances and sizes might lead to serious operational consequences when controlling a spacecraft or robotic arm. In addition to vestibular disturbances, another challenge is that motor control is learned and programmed in Earth's gravity conditions (Manzey et al., 1993). This results in a mismatch between the anticipated consequences of a movement and the actual sensory experience in space (Carriot et al., 2021). Adaptation of motor programs to microgravity requires information-processing resources and cognitive effort to deal with this sensorimotor discordance – resources which are scarce regarding the usually high workload (Bock, 1998; Kanas & Manzey, 2008). During adaptation, the underestimation of masses in microgravity leads to disturbances in psychomotor functioning in space, e.g. regarding pointing and tracking movements (Berger et al., 1997; Bock et al., 2001; Heuer et al., 2003; Watt, 1997). Fortunately, these perceptual-motor problems usually ameliorate after the first days of exposure to the new gravity condition (Carriot et al., 2021; Kanas & Manzey, 2008). However, impairments in motor speed and higher reliability can continue for more than a month (Strangman et al., 2021). In summary, microgravity alters spatial orientation, mental rotation, and other perceptual functions (De la Torre, 2014). Sensory conflicts and the lack of familiar vestibular and proprioceptive feedback are expected to impede the execution of sensorimotor tasks at least during adaptation after a gravitational change (Carriot et al., 2021).

Although fine motor and cognitive processes are interdependent and disturbed by similar stressors (Beard, 2019), there is no conclusive evidence of other substantial decrements to basic cognitive functioning in space (Kanas & Manzey, 2008). In their review, Strangman et al. (2014) reported decrements in dual-task and divided attention paradigms (Bock et al., 2010; Heuer et al., 2003; Manzey, 2000; Manzey et al., 1998), whereas other facets of attention seemed to be unaffected. There were no conclusions possible regarding memory; especially the effects on long-term memory have not been sufficiently investigated. Experience shows that effective learning is possible during space flight, but it is still unclear if differences to learning on Earth are present (Strangman et al., 2014). Some evidence from extreme analog environments, e.g. Antarctic stations, indicates that learning of new strategies could be impaired in such stressful circumstances (Sauer et al., 1999a, 1999b). There was also no sufficient research available regarding executive and higher-order cognitive functioning during space flight (Strangman et al., 2014). Interestingly, in parabolic flight studies even positive effects of microgravity on performance in mental arithmetic tasks have been reported (Wollseiffen et al., 2019; Wollseiffen et al., 2016). In general, it is likely that astronauts can stabilize their performance on a high level at least after an adaptation period that is characterized by disturbances of mood and well-being as well as an increase in perceived effort to complete tasks (Kanas & Manzey, 2008). However, some deficits have been reported to persist for several months, e.g. in dual-task performance (Strangman et al., 2021). To summarize, in many domains the evidence on cognitive functioning in space is still scarce and inconclusive.

The question whether reported cognitive impairments in astronauts are originated in microgravity itself or stress effects not specific to space is still a matter of debate (Schneider et al., 2008; Schneider et al., 2007). Decrements in perceptual-motor performance during initial exposure to microgravity are likely direct effects of microgravity, whereas other cognitive decrements, e.g. in dual-task performance, are more probably originated in unspecific stressors such as workload, sleep deficiency, or the isolated, confined, and extreme environment (Kanas & Manzey, 2008; Kanas et al., 2013). For example, impairments of sleep will most likely affect vigilance and the astronauts' ability to focus (Morphew, 2001). An integrative theoretical framework on how microgravity could affect not only perceptual-motor, but even other cognitive functions was suggested by Arshad and Ferré (2022). They observed that effect sizes in studies involving altered gravity were usually much higher in the sensorimotor domain than in the cognitive and socio-affective domains. Based on this observation, they proposed a stepwise cascade framework: vestibular alterations lead to impairments in sensorimotor functioning, which in turn lead to alterations in cognitive and socio-affective domains. For example, vestibular disturbances in microgravity cause impairments of manual dexterity, which can result in slower and less accurate key responses and, therefore, decrements in cognitive test performance. Taken together, space missions constitute a complex working environment including various stressors – specific and unspecific to space – that both have the potential to affect human performance capability in sensorimotor and cognitive domains as well. However, severity and origin of cognitive effects are still debated.

1.2.1 Human performance in space – from basic cognitive functioning to complex skills

Many studies on the influence of space flight concentrated on basic cognitive functions by using standard laboratory tasks (Kanas et al., 2013). As summarized in the previous paragraph, many cognitive domains have not yet been conclusively studied, but oftentimes no or only small effects on performance were found. However, this does not necessarily imply that concerns about cognitive functioning are unwarranted. It is uncertain to what extent performance in basic cognitive tests actually generalizes to performance in operational tasks. The generalizability of experimental results relies on the capability of the task to mirror the demands of the actual work environment (Hockey et al., 2011). However, many laboratory tasks fail to be representative of the complex demands of the mental and psychomotor work that is typical for astronauts in space. Steinberg et al. (2015) addressed this issue in a parabolic flight study using a realistic power plant instrument-control task that was supposed to resemble astronauts' work demands more closely. They found decreased control efficiency and hand velocity during microgravity phases, accompanied by higher ratings of physical strain. Increments in psychological strain or task load were not evident, suggesting that stress did not mediate the change in control efficiency. It was concluded that in microgravity, more resources have to be allocated to the motor system and, therefore, deficits in attention, concentration, or multitasking might be responsible for the performance decrement. These results are in line with earlier assumptions that perceptual-motor performance is likely the domain most sensitive to microgravity and that the additional cognitive resources needed for adaptation may compromise attention on the actual task.

Due to the scarcity of research, it is not clear if operational performance is more or less vulnerable to space flight than more basic cognitive performance. From a theoretical perspective, stress does not necessarily lead to overt performance decrements in complex operational tasks. To a certain degree, individuals can compensate for stress and stabilize their performance, e.g. by applying extra effort (Kanas & Manzey, 2008). According to the compensatory control model by Hockey and Robert (1997), the effects of stress on performance can be masked if individuals are able to apply compensation strategies. Astronauts are highly trained and motivated, so they are usually capable of protecting mission critical task goals even under extreme stress or workload (Hockey et al., 2011). Complex tasks offer ample opportunities for task modifications, such as working faster under time pressure or concentrating only on the core task features. However, this comes at the risk of more subtle performance decrements, because increased effort to stabilize performance is a limited resource. If a compensatory control process is maintained over long time periods, this will have physiological or behavioral costs. Because primary task goals necessary for mission success are protected, detrimental effects on complex performance will rather be latent and not easily detected by standard laboratory tasks. Examples for such latent decrements are the lack of attention to peripheral goals (narrowing of attention), impairments in subsidiary tasks, increased strain, anxiety, and fatigue, or the use of short cuts and risky decision-making (Hockey & Robert, 1997). In spatially complex tasks, peripheral elements are sometimes disregarded under high-stress conditions (Hockey et al., 2011). This can be dangerous for operational tasks like the manual docking of a spacecraft, if critical instrument information is not sufficiently processed. To summarize, complex human performance in space should be regarded as the result of an adaptive response to manage a task under given environmental conditions (Hockey et al., 2011). It can be assumed that performance is moderated by constraints due to the current individual state and the costs of maintaining task goals over prolonged time periods, possibly at the expense of other goals. As a consequence, performance decrements can be present although primary mission tasks are completed successfully (Kanas & Manzey, 2008).

1.2.2 Obstacles to the evaluation of performance in space

The inconclusive evidence on the impact of space flight on cognitive and operational performance is not only caused by a scarcity of research, but also by typical methodological difficulties. For once, crew time is a scarce resource during space missions, but also in complex analog studies. Therefore, studies oftentimes suffer from small sample sizes and short observation periods that restrict generalization to long-duration missions (Strangman et al., 2014). Due to the limited number of astronauts with longer mission durations, observation periods of more than six months in space are usually case studies. For example, a case study on one astronaut evidenced impairments in tracking performance in the first three weeks in-flight, but no other decrements in basic cognitive functioning over a mission duration of more than 13 months (Manzey et al., 1998). The National Aeronautics and Space Administration Twins Study (Garrett-Bakelman et al., 2019), another case study spanning a 340-day mission duration, found only minor in-flight effects on few cognitive domains as well. However, there was a post-flight decline in speed and accuracy measures that lasted for up to six months after the mission. Such decrements could possibly affect post-landing activities in future Mars missions. It should be noted that novel environments lead to a high variance of cognitive effects between as well as within individuals. Therefore, generalizing from case studies or small sample sizes can be deceptive (Bloomberg et al., 2015; Strangman et al., 2014). The high between-subject variability also suggests that some individuals are more resilient to the space flight environment than others (Strangman et al., 2021).

Additional problems in the literature are the lack of matched control groups and normative baseline data for astronaut populations. Replication and validation of measures is rare (Arshad & Ferré, 2022). Studies are often difficult to compare, because they tap on a variety of different cognitive tasks and domains, measured at different points in time. This makes systematic meta-analyses difficult (Basner et al., 2015). According to Strangman et al. (2021), 85% of studies conducted in space suffer from confounding effects of time (e.g. task fatigue), task repetition (e.g. learning effect), or space flight-related stressors. Space is not a controlled laboratory environment, which makes it hard to identify the cause of an effect (Shelhamer, 2017). Mission characteristics and demands as well as unexpected events will influence the results, and the effects of microgravity cannot be separated from other operational stressors (Gushchin et al., 2019). Additionally, common cognitive tests often suffer from ceiling effects when applied to astronauts, continued learning effects during testing, or do not engage work motivation (Hockey et al., 2011). Many tests are less sensitive in highly trained astronauts because they were designed for clinical or lower aptitude populations. Such tests can identify severe impairments, e.g. following brain trauma, but may not detect subclinical deficits that interfere with optimal performance (Basner et al., 2015). Additionally, many studies only investigated single cognitive tasks, therefore, no comprehensive picture of cognitive functioning is available (Basner et al., 2015). To standardize research on cognitive performance in space, specialized test batteries were designed to assess relevant cognitive domains and timely identify performance decrements (Kane et al., 2005). The most recent approach has been the *Cognition* test battery that was developed to increase data comparability across studies and has been validated in a high-aptitude population (Basner et al., 2015; Moore et al., 2017).

1.2.3 Cognition in space – a mismatch between anecdotal and empirical evidence

To summarize the literature on performance in space, there seems to be a mismatch between anecdotal reports of cognitive decrements in space and the lack of empirical evidence thereof (Hockey et al., 2011; Strangman et al., 2014). Although astronauts are concerned about their level of cognitive functioning in space, previous studies mostly failed to observe substantial impairments in cognitive domains such as memory, logical reasoning, or mental arithmetic (Kanas et al., 2013). If performance decrements were observed, they have been mostly linked to difficulties in psychomotor adaptation during the first weeks in space and again on Earth (Morphew, 2001). During adaptation to a change in gravity, detrimental effects can be expected in sensorimotor functioning, dual-task management, and attention (Li & Qu, 2021). A possible explanation for the mismatch between anecdotal reports and empirical research could be that astronauts possess a high awareness of changes in their cognitive reserve (Strangman et al., 2014). Although they are able to perform at high levels in space, they might still report increased effort and cautiously perceive this as a potential risk. Other reasons for the mismatch are likely the scarcity of research in general and the methodological weaknesses discussed in the last paragraph. For many cognitive domains, we are not yet able to estimate the impact of space flight conclusively. However, the success of many space missions until today is proof of the astronauts' general ability to perform in complex operations in space. For durations of about six months like on ISS, there are likely no alarming impairments in cognitive performance (Strangman et al., 2014) – but there is still a dearth of evidence on extended time periods in extreme isolation and on actual operational performance.

1.3 Investigating operational performance in the context of space flight

The scarcity of knowledge regarding operational performance in space is problematic, because we are not able to reliably gauge the impact of space stressors on the safety of mission-critical tasks. Many operational tasks combine cognitive and motor demands, particularly in robotic and vehicle control. There is some evidence that this combination might be especially vulnerable to performance decrements in space (Seidler & Mulavara, 2021). Sensory deficits, as observed during the first days to weeks of space flight, could have detrimental impact on operations like extravehicular activities, docking and landing, or robotic control (Jones, 2010). Such sensorimotor operational tasks might be impacted strongly during adaptation following gravity transitions, which will be even more relevant for missions to Moon and Mars. Gravity transitions affect manual coordination and spatial orientation (Carriot et al., 2021; De la Torre, 2014), and incorrect perceptions of acceleration or orientation could easily lead to errors in manual control (Steinberg, 2019). Additionally, gravity transitions often coincide with critical mission phases that are based on operational tasks with a sensorimotor component, such as landing or docking (Milstead, 2022). Relevant skills are initially learned on Earth, but have to be applicable in microgravity or partial gravity of Moon and Mars as well (Hockey et al., 2011). Although it is very likely that gravity influences manual control abilities, there is not much evidence on the impact on operational performance in vehicle control. Experience with Space Shuttle landings on Earth and landings on Moon indicates that deviations from predefined performance specifications occurred, presumably due to the change in gravity forces and spatial disorientation after being exposed to microgravity (Bloomberg et al., 2015). Spatial disorientation, i.e. the failure to estimate the own position and orientation in space, is one of the leading causes for aviation accidents and responsible for many fatalities (Gibb et al., 2011; Newman & Rupert, 2020). The risk of losing spatial orientation is expected to increase for Mars landings due to the long transit in microgravity. But even in astronauts shortly after return from ISS, performance decrements in a car driving task have been observed (Moore et al., 2019). In the same study, impairments in manual dexterity and dual-task performance were apparent that point to post-flight limitations in motor functioning and available central processing resources. The authors emphasized that performance impairment was subtle in cognitive and sensorimotor test batteries, but more pronounced in the operational driving task. They attributed the effect to an accumulation of small physiological changes. Self-assessments to estimate fitness for duty prior to critical tasks and in-flight refresher trainings were proposed as countermeasures. Although the general cognitive capacity is preserved in space, sensorimotor operational tasks are likely more vulnerable to performance decrements – especially following changes in gravitational force.

A major characteristic of operational tasks is the small margin of error. Even subtle performance impairments can potentially lead to catastrophic outcomes and jeopardize mission success and safety. Performance failures during space flight are fortunately rare, although there is evidence on the detrimental effects of stress on performance from laboratory and field studies (Staal, 2004). However, critical incidents may be underreported and their underlying cause is not easily detected, because environmental conditions cannot be controlled (Hockey et al., 2011). It is known that human error plays a critical role in the majority of aviation accidents (Weigmann & Shappell, 1997), often due to insufficient training, cockpit design or factors related to stress and fatigue (Morphew, 2001). This is most likely also true for vehicle control in space. One frequently mentioned and almost fatal incident happened in 1997 during a manual docking maneuver. Due to problems with the availability of the automated docking system, a Progress cargo spacecraft was supposed to be remotely controlled and docked to the space station Mir by using the manual backup system. Because the radar system was turned off to avoid interferences, there was no range data available and the operator had to rely on the size and position of Mir from the video image provided by Progress. Thus, distance and velocity were not adequately estimated and the spacecraft finally collided with the station, damaging its hull and solar arrays. The crew was able to lock the ruptured module and thereby barely avoided the evacuation of Mir. The cause of this incident has been seen in an unfavorable accumulation of stressors and circumstances (Ellis, 2000). Due to the missing range data and poor resolution of the video monitor, the operating cosmonaut lost situational awareness and spatial orientation (Bloomberg et al., 2015). Additionally, operator training was not sufficient for the attempted maneuver and dated back several months without any refresher training, because there were no simulation systems on the station (Oberg, 1998). When the accident happened, the crew had already been exposed to a variety of system failures on Mir that resulted in long hours, considerable stress, and constant tension. Finally, the operator had very little rest time and reported low sleep quality during the weeks before the accident (Ellis, 2000). Taken together, the accumulation of stressors in the adverse space environment can compromise performance drastically in operational tasks with typically small error margins. Portable simulations of critical tasks are crucial training tools to reduce this risk.

Many safety-critical operational tasks, such as spacecraft docking, are usually (and will be even more so in the future) carried out automated to reduce the associated risks. Paradoxically, this approach can bring about its own risks if automation fails or manual control is necessary for other reasons. The operator’s ability to manually control such systems is far from obsolete.

Failures of automation have occurred in the past and will likely occur in future missions (Moore et al., 2019). Moreover, automation cannot be provided for every possible situation, whereas manual control offers more flexibility and the ability to react to unexpected incidents (Brody, 1988; Ellis, 2000; Moore et al., 2019). Therefore, the high reliance on automation comes with drawbacks. Manual control skills are applied only infrequently, leading to skill degradation and insufficient experience. During the manual takeover, increased workload and lack of situation awareness about the system's status are risk factors (Hainley Jr et al., 2013). Additionally, operators have to be highly skilled in automated contexts, because if they take over control, oftentimes a system failure or anomaly has taken place. To resolve such situations, operators have to be more proficient and need more free cognitive capacity than in standard situations (Bainbridge, 1983; Hancke, 2020). In conclusion, tasks that are usually automated require rigorous maintenance training to preserve manual control skills on a highly proficient level. Otherwise, mission safety is easily compromised if a critical incident occurs.

We have seen that continuous skill maintenance in space is of utmost importance to protect performance against environmental stressors and preserve fitness for duty in an operational task. Performance in manual docking for example has been shown to deteriorate below a safe level after three months in space without training (Salnitski et al., 2001). This critical time period was replicated by Bosch Bruguera et al. (2021) in a Soyuz simulator, whereas performance could be stabilized when participants received monthly training during winterover in Antarctica. The duration of the mission itself can also pose a serious problem, especially during future long-duration exploration missions. Whereas the influence of mission duration on mood and simple cognitive performance has been studied, there is still a lack of studies on performance in operationally relevant tasks. Stankovic et al. (2022) investigated performance in a simulated lunar landing task during a 45-day isolation study. They found peak performance in the second or third quarter of the experiment, whereas performance decreased again afterwards. Isolation seemed to interrupt the learning effect that was present in previous studies (Hainley Jr et al., 2013). The decrease in performance was partly similar to the often-cited third-quarter effect of decreased mood during this time in mission (Bechtel & Berning, 1991; Stuster et al., 2000); though there was no recovery towards the end of the mission. In contrast, Bosch Bruguera et al. (2021) found no evidence of a comparable third-quarter effect regarding manual docking performance during an Antarctic winterover. The reliability of performance though was decreased in Antarctica compared with a control group not exposed to isolation and hypoxia. In conclusion, we have some reason to believe that long mission duration and isolation can impede normal learning processes in complex operational tasks. Again, simulations of operational tasks are key to prevent performance decrements with effective training schedules.

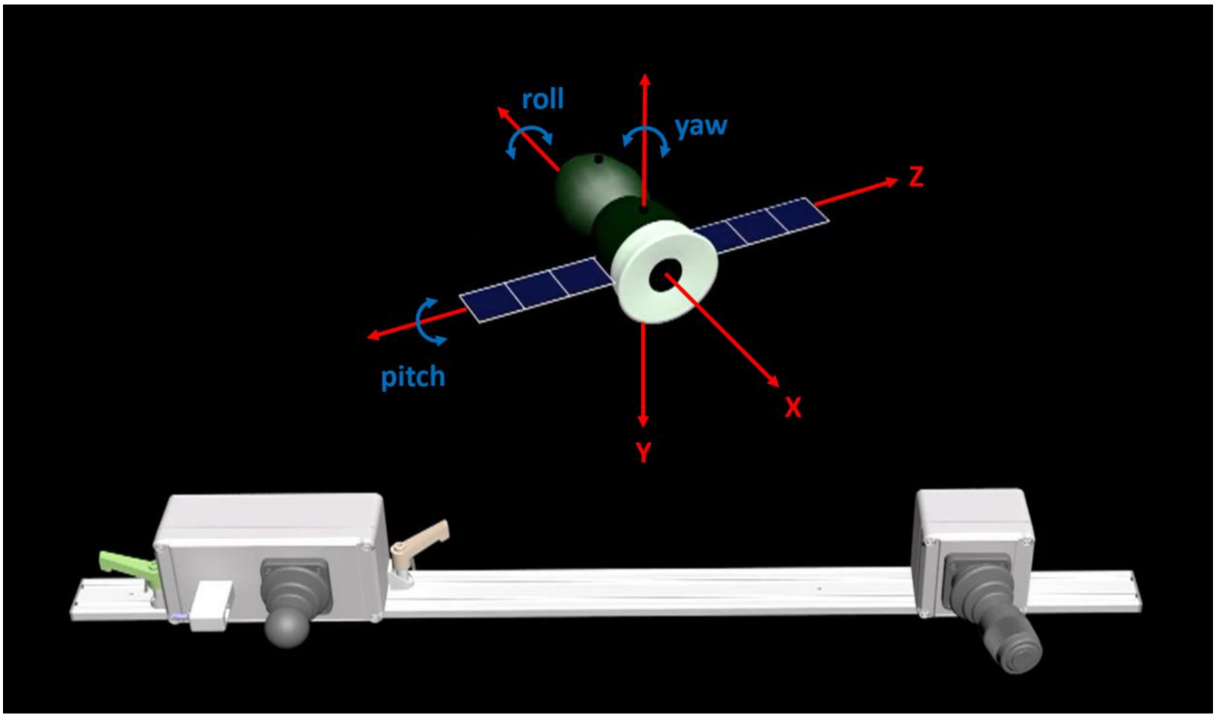
In conclusion, future missions will heavily rely on the interaction of humans with various robotic systems such as rovers and robotic arms (Hambuchen et al., 2021). Human performance is a limiting factor in man-machine-interactions, especially if sensorimotor skills are needed that are vulnerable to gravity transitions (Fong et al., 2013) and have small error margins. Manual control skills may be applied only infrequently due to the high level of automation, but remain mission critical. Therefore, high levels of manual control skills must be preserved throughout a mission – considering the adverse environmental conditions including stress, high risk, and isolation. For the success of a space mission, it is important to investigate more deeply if the concerns regarding cognitive functioning in space actually translate to performance decrements in the real work environment. To achieve this goal, objective, real-time performance metrics of operationally relevant performance are important (Duda et al., 2015). Operational monitoring allows for the assessment of operator's state and skill level and enables targeted feedback. Thereby, performance decrements can be mitigated timely, e.g. via additional tailored training (Kanas et al., 2013). In-flight maintenance of critical skills can also serve as a meaningful activity to support astronauts' motivation and well-being, especially during phases of minor workload (Holland, 2000). Questionnaires and psychological tests can be perceived as exhausting, particularly if they are applied regularly for long mission durations. If participants lack motivation, acquired data will be less reliable (Gabriel et al., 2012). In contrast to other test procedures, operational tasks are usually well accepted and perceived as highly motivating because of their relevance.

1.3.1 The *6df* tool – a manual docking simulation for research purposes

The *6df* tool is a manual docking simulation that was designed at the German Aerospace Center DLR for research on operational performance in a space flight-relevant task (Johannes et al., 2017). Manual control and docking of a spacecraft are based on the control of an object in six degrees of freedom (DoF). On Earth, most individuals are used to control vehicles along two axes, with only two translational DoF: forwards-backwards and left-right. A spacecraft can additionally move up and down along a third axis and also rotate around each of these three axes (roll, pitch, and yaw). Two hand controls are used to navigate the simulated spacecraft: the left one controls translation and the right one rotation (Figure 1). The control of robotic manipulators on ISS, used for maintenance and payload handling, follows similar principles (Currie & Peacock, 2002).

Figure 1

Six degrees of freedom in spacecraft control and 6df hand controls

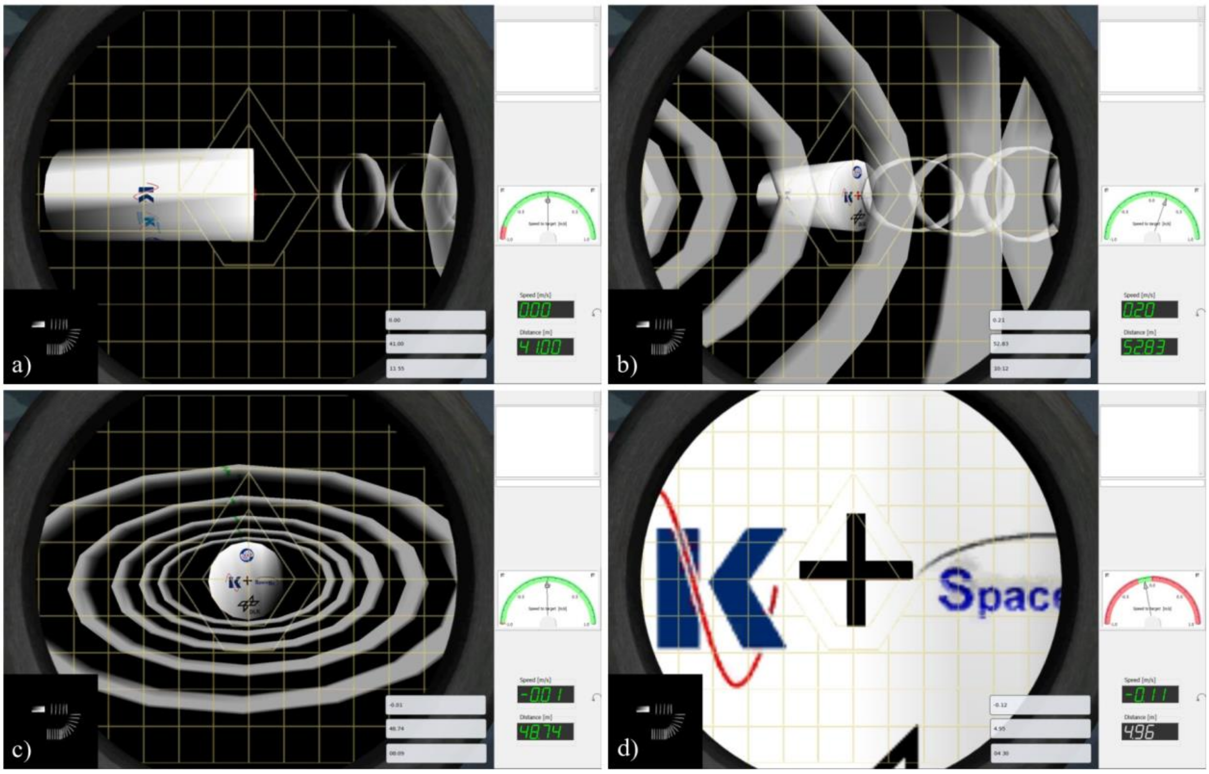


Note. The left hand control is used to navigate along all three axes (translation, red), the right hand control to rotate around these axes (orientation, blue).

6df has been developed as a self-sufficient learning program that helps to investigate the individual learning process, improve training efficiency, and diagnose decrements in manual control performance (Johannes et al., 2017). Flight mechanics and hand controls were designed to resemble the Soyuz spacecraft. However, the abstract design and variety of featured tasks is supposed to facilitate the generalizability of the acquired skill to control six DoF, e.g. to other space vehicles or robotic arms. The operator is looking out of the simulated spacecraft they are moving, resembling the video footage usually available for docking. The simulation is desktop-based and easily portable, suitable for the application with a single commercial laptop in settings with very limited space. *6df* as a learning program was designed following training principles for high performance skills (Johannes et al., 2011; Schneider, 1985). To avoid overload, it starts with part-task training of single manual control components and ends with a standard docking maneuver, as it is usually demanded in space. Such a standard maneuver includes a curved flight around the station, stabilization at safety distance in front of the docking point, linear approach on the center line that leads straight to the docking point, and finally docking contact with the black target cross (Figure 2).

Figure 2

Standard manual docking maneuver in 6df

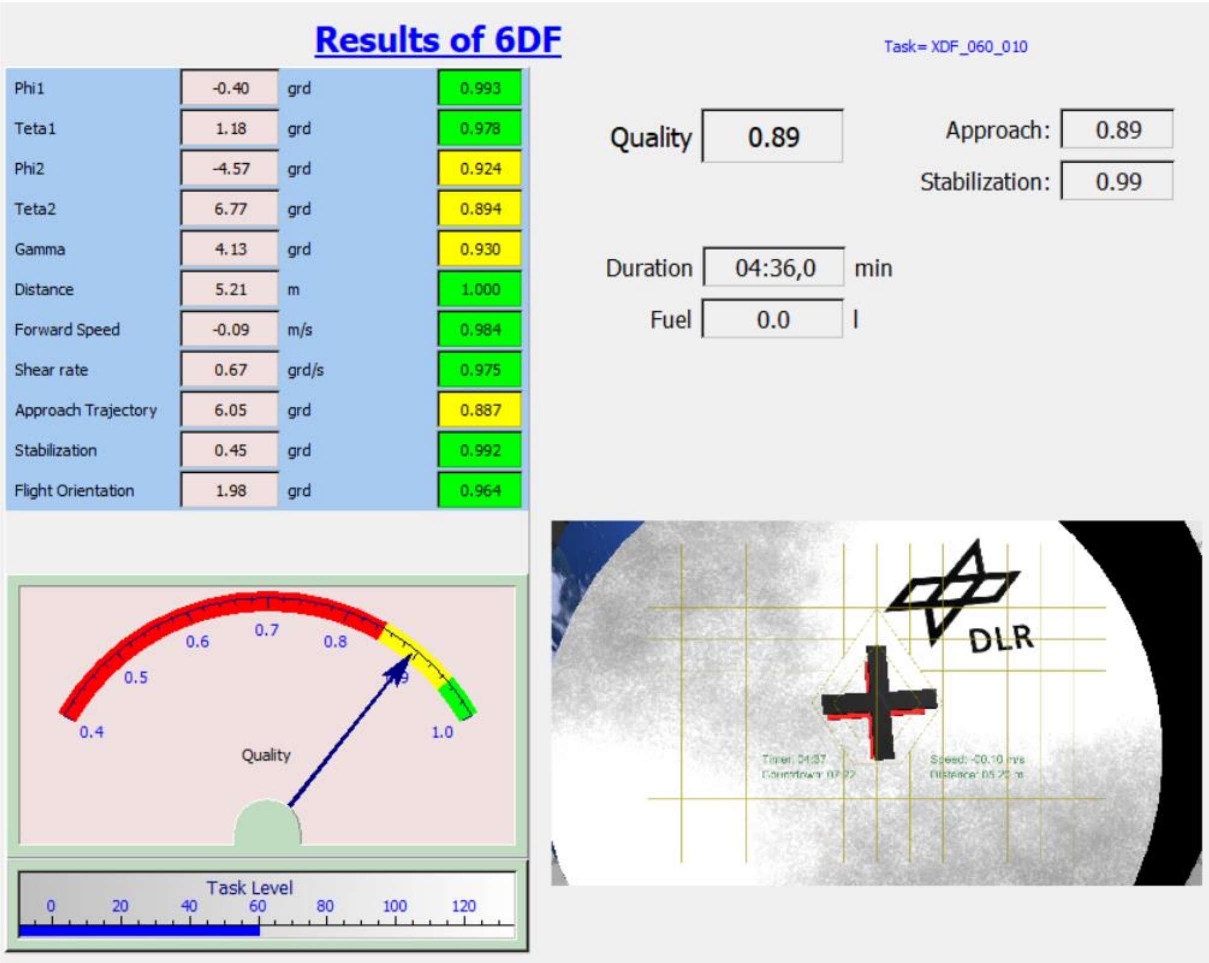


Note. Screenshots of a level 50 *6df* task, depicting start position (a), curved flight in safety distance (b), stabilization on the center line (c), and docking contact (d). If docking is precise, the red cross disappears behind the black target cross.

The adaptive program provides the necessary amount of repetitions in twelve difficulty levels depending on the operator's current skill level. Guiding rings indicate the flight path when a new maneuver is introduced. Video instructions are provided to introduce six DoF, hand control functions, important new maneuvers, and performance parameters. Additionally, illustrated written instructions are provided ahead of every single docking task. Detailed performance feedback is given after every task to make the source of errors evident (Figure 3). A description of all difficulty levels (Table 2) and performance parameters (Table 3) used in the *6df* tool is provided in the appendix. Performance assessment has been derived from the manual control system called *Telerobotically Operated Rendezvous Unit* (TORU) used for Soyuz and Progress spacecrafts and the Russian simulation of the same name that is used for cosmonaut training (Johannes et al., 2016). The overall performance score (docking accuracy) ranges from 0 to 1 and is categorized as follows: failed docking attempt (red, accuracy score < .85), sufficient performance (yellow, accuracy .85-.95), and desirably good performance (green, accuracy ≥ .95). When green accuracy scores are achieved, a more difficult task is presented. Yellow scores lead to the repetition of the task and red scores to the presentation of an easier task. The non-aggregated performance scores are displayed as absolute values and as standardized color-coded scores ranging from 0 to 1.

Figure 3

Screenshot of performance feedback after a 6df docking task



Note. Reported are single performance scores (accuracy of stabilization, continuous orientation to the station, speed and angle errors during docking contact) as well as an overall docking accuracy score (quality) ranging from 0 to 1. Scores are color-coded to indicate insufficient (red), sufficient (yellow), and good (green) performance.

The *6df* software was developed by *Spacebit GmbH* (Eberswalde, Germany). Hardware is produced by *KORA Industrie-Elektronik GmbH* (Hambühren, Germany), including hand controls and portable devices for the synchronous collection of various physiological data (e.g. electrocardiogram and skin conductance response). The simulation can also be supplemented with secondary tasks, eye tracking, and questionnaires. Up to now, the tool has been used to characterize the learning process during 6 DoF skill acquisition (Johannes et al., 2019) and to determine free cognitive capacity during docking via electroencephalography (Johannes, Bubeev, et al., 2021). In space, research with the *6df* tool comprised two decades on Mir and ISS, revealing increases in cosmonauts' docking reliability as a result of advanced training schedules (Johannes, Bronnikov, et al., 2021).

Manual docking is a prime example of a complex task that relies on cognitive as well as motor skills and has high implications for the success and safety of a mission. The control of six DoF is a high-performance skill, which means that more than 100 hours of training are necessary, yet some individuals will fail to reach proficiency. Additionally, expert performance differs qualitatively from novice performance (Johannes et al., 2011). The task cannot be trained in the real environment; therefore, astronauts are lacking experience outside of the simulator. Microgravity, additional stressors, and the pressure to succeed during the real maneuver might facilitate human error. Tasks incorporating the control of six DoF tap on various cognitive domains, such as spatial orientation, mental rotation, working memory, executive functions, and decision-making, as well as fine motor control (Ivkovic et al., 2019; Marshburn et al., 2003). Spatial orientation was found to be especially important for manual docking performance (Menchaca-Brandan et al., 2007; Wang et al., 2014). This includes perspective taking and mental rotation: Perspective taking is defined as the ability to imagine the look of an object or scene from another perspective than the observer's and requires a change of the egocentric reference frame. Mental rotation means the ability to mentally manipulate an array of objects within a fixed egocentric reference frame (Menchaca-Brandan et al., 2007). Basner et al. (2020) investigated the relationship between *6df* docking performance and subtests of *Cognition*, a test battery specifically developed to monitor cognitive performance in the high-aptitude astronaut population (Basner et al., 2015). The strongest predictor of docking performance was response time in the *Digit Symbol Substitution Test*, a measure of processing speed that includes working memory and visual scanning. High spatial orientation efficiency and sustained attention as well as low impulsivity were also found to be relevant for *6df* performance.

Realistic operational task designs that are also suitable for research purposes are still scarce. For manual control tasks in space, Ivkovic et al. (2019) developed a research version of the *Robotics On-Board Trainer* that is currently used to train the control of a robotic arm on ISS (Canadarm2) to track and capture spacecrafts. Many other simulators have either lower fidelity and operational relevance (Petit et al., 2019) or rely on a large setup that is not suitable for the use in space or confined analog environments. *6df* requires only a set of hand controls and a custom laptop, but provides realistic flight behavior. In comparison to other six DoF simulators (Bosch Bruguera et al., 2022; Duda et al., 2015), *6df* was designed with abstract graphics instead of platform specific high-resolution images. Such a medium fidelity simulation is suitable for research as well as application in space, because it mirrors the essential demands of the operational task, but is still portable and easy to apply in isolation and confinement (Hockey et al., 2011). *6df* also features autonomous and adaptive training as well as a variety of task designs that can be tailored to specific experimental conditions or objectives.

1.4 Aims of the empirical studies

Microgravity itself and other stressors related to the work environment in space are risk factors for human performance and mission success. The goals of this thesis were to further the understanding of how stressors inherent to space flight relate not only to basic cognitive, but operational performance and how the skill to control six DoF can be acquired and maintained more efficiently. Along these objectives, three empirical studies were conducted. The *6df* tool was chosen as the central methodology because it represents a safety-critical task and is associated with cognitive performance measures that are relevant for space flight. The sensorimotor and spatial orientation components of manual control are vulnerable to performance decrements in space, particularly if the relevance of gravity transitions increases during future long-term missions.

1.4.1 Study I: Stereoscopic learning aid for manual docking

As we have seen, sensorimotor performance and spatial orientation are the cognitive aspects most threatened during space missions, at least during adaptation to microgravity (Kanas & Manzey, 2008). Manual docking and other tasks involving the control of six DoF highly depend on intact spatial orientation (Menchaca-Brandan et al., 2007; Wang et al., 2014). The own orientation and position in relation to the target has to be monitored continuously. Additionally, free movement in three dimensions (3D) is unfamiliar and challenging, particularly because there is a lack of reference points in the space environment. Therefore, the acquisition process is lengthy and often accompanied by difficulties with spatial orientation. Extended training is needed to achieve sufficient proficiency (Dempsey & Barshi, 2020) and the absence of frequent refresher training may lead to disastrous outcomes, as seen in the discussion of the Progress crash with Mir (Ellis, 2000). In contrast, skills that are highly practiced require less cognitive resources and are more resistant against the influence of stressors (Dismukes et al., 2015; Raaijmakers, 1990). Training and maintenance of complex operational skills during long-duration missions will be an enormous challenge in the future (Kanas et al., 2013; Rector et al., 2021), as the required information density will be exceptionally high (Gabriel et al., 2012). Consequently, there is a need to develop effective training tools that are suitable for the autonomous application by astronauts to ensure fitness for duty throughout a mission (Barshi & Dempsey, 2016).

A special challenge of teleoperation is that 3D movement has to be realized via a two-dimensional (2D) screen. Additionally, the provided video image is oftentimes of low quality. This deprives the operator from relevant cues for the estimation of distances and speed. Virtual reality (VR) has been one approach to render remote control more intuitive and natural, while also being efficient regarding time and costs (Pirker, 2022). In comparison with a 2D interface, VR adds an immersive view to robotic control that can increase situation awareness, safety, and time efficiency (Goecks et al., 2017; Wonsick & Padır, 2021). VR is a collective term that comprises a plethora of technological implementations and use cases. Broadly speaking, it can be defined as “the use of computer technology to create the effect of an interactive three-dimensional world in which the objects have a sense of spatial presence” (Bryson, 2013). Varying levels of immersion are incorporated, from desktop-based applications to the complete perception of being physically present in the virtual world (Freina & Canessa, 2015). VR has been applied to support psychomotor performance, spatial abilities, and navigation skills (Abich et al., 2021) in various training contexts, e.g. laparoscopic surgery and transportation (Seymour, 2008; Xie et al., 2021). In the aviation domain, VR and augmented reality have been used to provide pilots with additional information to facilitate their situation awareness and spatial orientation, for example when the visual flight environment is degraded (Brown et al., 2021). Human-machine interaction and teleoperating performance in robotics has been improved in a similar manner, for example by overlaid cues that facilitate orientation when display and controls are misaligned (Chintamani et al., 2010; Chintamani et al., 2011; Maida et al., 2007).

VR approaches also have been used frequently to support astronaut training (Gabriel et al., 2012; Lonchakov et al., 2017; Olbrich et al., 2018). They are suitable for ground-based as well as on-board training, particularly when other training forms would be hazardous or demand large and costly facilities (Ennis et al., 2021; Olbrich et al., 2018). VR has been applied to familiarize astronauts on ground with the layout of the space station (Liu et al., 2016), simulate emergency procedures (Aoki et al., 2007; Finseth et al., 2020), or facilitate remote control of a lunar rover while considering communication delay between Earth and Moon (Cheng et al., 2015).

The *6df* tool was developed with the objective to reduce the number of costly instructor-supported training hours in a high-fidelity simulator (Johannes et al., 2017). The addition of a VR training aid that facilitates spatial orientation could further accelerate training, which is important for long-term missions with very limited time for pre-mission training. Moreover, VR could facilitate the acquisition of six DoF skills in-flight, when spatial orientation is additionally impaired by microgravity. Especially at the beginning of skill acquisition, interindividual differences in complex tasks are pronounced (Schneider, 1985), which has also been observed for the *6df* learning program (Johannes et al., 2019). Accordingly, there is still potential to support especially the slower learners in the early stages of skill acquisition. To circumvent the difficulty of 3D navigation using a 2D screen, a stereoscopic version of the *6df* learning program was developed that features the 3D view of the controlled spacecraft's vizor and the target space station during the first training levels. Simulation is of course an essential part of manual spacecraft docking training and there are other VR solutions available. Bosch Bruguera et al. (2019) for example supplemented their manual docking simulation using a VR headset and hand tracking. Whereas these authors focused on high graphical realism and immersion into the virtual world, our objective was to preserve the simplicity of the *6df* tool by exclusively adding the stereoscopic view. Thereby, we can support the training with additional spatial cues during initial skill acquisition. After that, the operator can easily switch to the standard 2D visualization that is more representative of the actual operational conditions. Only 3D goggles (Nvidia 3D Vision 2 wireless glasses) had to be added to the usual *6df* setup.

This new stereoscopic version of *6df* was tested during the six-degree head-down tilt (HDT) bed rest study AGBRESA ("Artificial Gravity Bed Rest Study with European Space Agency"). In space, operational needs have priority and there is no opportunity to control sleep, nutrition, or medication (Shelhamer, 2017). In contrast, bed rest studies provide a standardized environment that allows for strict experimental control (e.g. scheduled sleep, standardized nutrition, no caffeine) and larger sample sizes. Because the design simulates a fluid shift similar to that occurring in microgravity (Hargens & Vico, 2016), we were able to test training efficiency in a context that at least partly resembles the working environment of astronauts. The 89-day bed rest study also provided the possibility for the participants to absolve the complete training program at a constant and comparable training interval.

In summary, the goal of this first study was to introduce a desktop-based VR addition to the *6df* tool and test its effects on training efficiency in comparison with the standard 2D simulation display. The stereoscopic view was supposed to facilitate the mental representation of one's own position, orientation, and motion in a 3D environment. We hypothesized that the additional spatial information provided by the stereoscopic view would thereby lead to faster learning progress in comparison with the 2D visualization. Additionally, we investigated whether the 3D group would profit from the VR training even after completion of the *6df* training program. At the end of the study, participants completed five manual docking tasks with the high-fidelity simulation TORU, that is used for the actual docking training of cosmonauts, to evaluate learning success and transferability of the skill learned with *6df*.

1.4.2 Study II: Eye tracking as a performance indicator for manual docking

The second study of this thesis took place in the context of the same bed rest study as study I, but involved only the participants that were allocated to the standard 2D visualization of *6df*. The objective of this study was to obtain additional information on the operator's attentional processing that could support the maintenance of high docking reliability and the assessment of readiness for duty. Eye movements are tightly linked to cognitive processes and can be used as objective indicators of visual attention (Duchowski, 2007; Holmqvist et al., 2011; Just & Carpenter, 1980). Importantly, eye tracking is applicable in operational contexts, because it is unobtrusive and does not interfere with the task investigated. As a first step, we wanted to establish if and how eye tracking measures were related to docking performance at all. Knowledge on successful gaze behavior then might be implemented into future training routines; thereby representing a second opportunity to increase training efficiency – next to the stereoscopic visualization in study I. Objective measures of operator proficiency and training effectiveness are still scarce (Barshi & Dempsey, 2016). They are particularly important for future long-duration missions, because astronauts will have to reach a higher level of proficiency when ground support is no longer available. However, the ratio of preflight training duration to mission duration will be shorter than today to be feasible (Dempsey & Barshi, 2020). As a result, self-monitoring of astronauts regarding their performance and functional state will be essential (Manzey et al., 1995). Eye tracking could be a valuable supplement to an operational task if it provides additional information on performance, workload, or fatigue. Greater variety of information on one's own performance provides possibilities for precise improvement and facilitates skill development (Hockey et al., 2011).

In aviation, eye tracking has been used frequently to increase the safety and reliability of piloting (Peißl et al., 2018; Ziv, 2016). Efficient instrument monitoring constitutes a critical factor of piloting performance. Lefrancois et al. (2016) for example identified suboptimal scanning patterns in pilots that failed to stabilize their manual approach. Information on visual attention allocation is also valuable to quantify situation awareness (van de Merwe et al., 2012). The ability to timely detect relevant changes in the cockpit is fundamental to avoid piloting errors (Wickens et al., 2008). Once gaze behavior is linked to piloting performance, actual scanning behavior can be compared to the recommendations (Colvin et al., 2005; Haslbeck et al., 2012) and specific training can take place to individually improve performance (Chuang et al., 2013).

Although state-of-the-art eye tracking devices are lightweight, unobtrusive, and easy to apply, they have rarely been used in space. First studies have investigated eye tracking in simulations of spacecraft control, for example to assess operator fatigue and workload during the maneuver (Tian et al., 2018). In aircraft cockpits, experts use shorter dwells, but check relevant instruments more frequently (Bellenkes et al., 1997; Glaholt, 2014; Kasarskis et al., 2001). In spacecraft control as well, gaze behavior differs between experts and novices (Huemer et al., 2005; Matessa & Remington, 2005). The association between eye tracking metrics and performance has also been studied in a simulation of the Canadarm2 robotic arm (Guo et al., 2021). Increased visual attention, measured as mean fixation duration, was associated with better control performance. These results illustrate the potential eye tracking has to assess fitness for duty in a space flight context.

In summary, this second study aimed at complementing the scarce knowledge on gaze behavior in spacecraft cockpits and their association with performance. Prediction of performance based on eye tracking could support efficient training and mission safety, as already employed in the aviation domain. Manual docking poses high demands on visual attention, because controller input must be updated continuously according to the current speed and position of the spacecraft. We expected visual attention allocation to be associated with docking performance throughout the *6df* learning program. To identify eye tracking-based indicators of performance, we focused on relevant regions of the display and assessed the amount of attention as the total dwell time devoted to the target station, the spacecraft's vizor, and the instruments. We additionally investigated how often information was retrieved by counting the number of dwells on each relevant region.

1.4.3 Study III: Manual docking performance after sleep deprivation

As delineated earlier, there is no clear-cut evidence on serious detrimental effects of space flight on human cognitive performance. Nevertheless, some stressors prevalent in space are tightly associated to cognitive performance on Earth. One prime example for such a stressor is sleep loss. Insufficient sleep is a common problem faced by many individuals on Earth, especially concerning night and shift workers (Åkerstedt, 2003). The working environment in space involves many risk factors regarding sleep as well. During space missions, astronauts frequently reported difficulties sleeping and expressed related concerns about their performance (Stuster, 2010). Subjective sleep quality is often lower in space than on Earth (Dijk et al., 2001) and the use of sleep-promoting drugs highly prevalent (Barger et al., 2014). Objective measurements detect a shorter, more fragmented, and less efficient sleep as well (Dijk et al., 2001; Hockey et al., 2011). Generally, at least seven hours of sleep per night are recommended to preserve health and cognitive functioning in adults (Hirshkowitz et al., 2015; Watson et al., 2015). However, astronauts in orbit sleep on average only six hours per night, which accumulates to a significant sleep debt (Milstead, 2022). Comparable sleep restriction has been shown to significantly impair cognitive performance in laboratory studies (Belenky et al., 2003; Van Dongen et al., 2003) and increase accident risk in aviation (Bendak & Rashid, 2020).

Reasons for the pronounced impairments of astronaut sleep can be found in the stressors of the space environment. Microgravity itself, background noise, lighting conditions in the station, and the unusual sleeping position appear as disruptive factors (Basner & Dinges, 2014). Next to the discomfort of the living environment, high workload and stress due to tight schedules can disturb sleep (Flynn, 2005). Depending on mission demands, the work schedule sometimes requires working or sleeping at an inappropriate circadian phase (Hockey et al., 2011). Due to operational requirements, e.g. launch windows, sleep shifts of several hours ('slam shifts') can be necessary for the astronauts to be awake at time. For this reason, circadian misalignment is a frequent condition in space, associated with decreased sleep duration and increased use of sleep medication (Flynn-Evans, Barger, et al., 2016). Missions to Mars will likely exacerbate sleep problems, because the circadian system must be entrained to the 24.65 h Martian day (Scheer et al., 2007).

The extent of sleep loss, circadian misalignment, and work overload that is experienced by astronauts has been associated with performance decrements in ground-based studies (Flynn-Evans, Gregory, et al., 2016). Sleep deprivation has detrimental effects on various cognitive domains (Goel et al., 2009), such as sustained attention, working memory, and decision-making. Other consequences of sleep deprivation include instability in attention-intensive performance, impairment of learning, and loss of situational awareness. The speed-accuracy tradeoff is affected: cognitive slowing will occur during self-paced tasks to preserve accurate performance, whereas error frequency will increase under time pressure. Compensatory effort is needed to stabilize performance and attention may be reduced to only the most essential task features (Durmer & Dinges, 2005). Additionally, decrements in visuomotor performance have been reported after sleep deprivation, possibly due to impaired spatial attention in connection with impaired oculomotor functioning (Alhola & Polo-Kantola, 2007). The effects of sleep deprivation on cognition are far from negligible – decrements in psychomotor performance after 24 hours of wakefulness can even be equivalent to the effects of 0.10% blood alcohol concentration (Dawson & Reid, 1997). Although sleep deprivation clearly impairs cognitive performance, studies linking sleep and performance of astronauts in space are still scarce. As early as during Space Shuttle missions, Dijk et al. (2001) observed circadian rhythm abnormalities, lower subjective sleep quality, and sleep loss in space – as well as performance deterioration in sustained attention and a probed recall memory task. In a more recent study, a sleep duration of less than six hours in astronauts on ISS was associated with impairments in sustained attention and mood on the following day (Jones et al., 2022). However, both studies were not suitable to establish a causal link between sleep disturbances and performance.

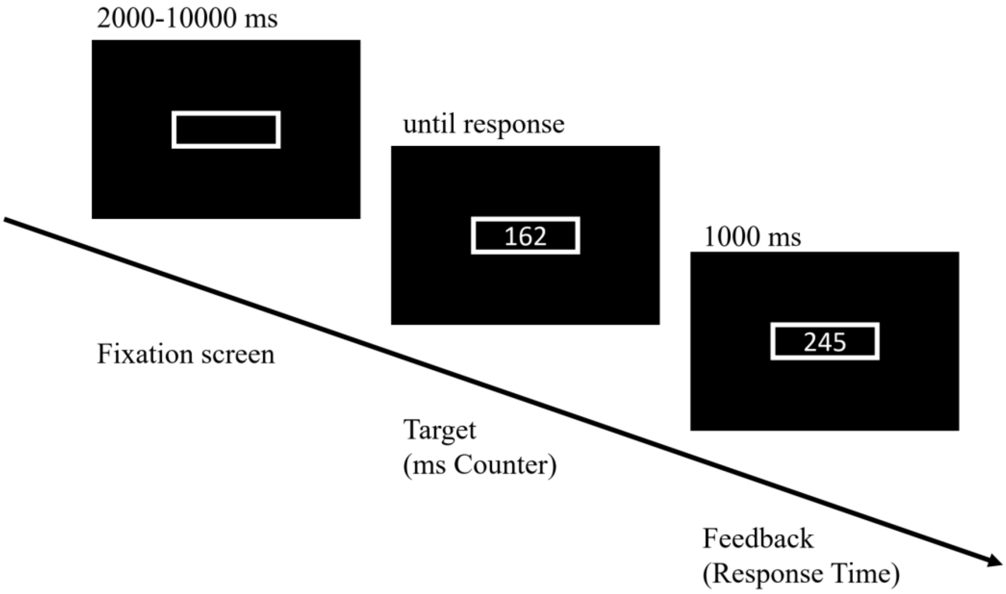
Regarding the effects of both sleep and space flight on human performance, there is a similar scarcity of studies on operational performance. Although the influence of sleep deprivation on a variety of cognitive tasks is established, it remains challenging to predict its impact on tasks in the real work environment (Flynn-Evans, Gregory, et al., 2016; Jackson et al., 2013). Indeed, accident analyses are concerning: sleepiness is a risk factor involved in many motor vehicle (Czeisler et al., 2016; Gottlieb et al., 2018) and aviation accidents (Caldwell et al., 2004). Sleep-deprived pilots display involuntary lapses into sleep as well as difficulties in psychomotor performance and attention to flight instruments (Caldwell, 2012). Total sleep deprivation has been shown to deteriorate flight simulator performance even in experienced pilots, accompanied by decrements in situation awareness and sustained attention (Caldwell et al., 2004). Previc et al. (2009) quantified a performance decrease of 15% due to sleep deprivation compared with baseline performance in a flight simulation.

Only very few studies have investigated the impact of sleep deprivation on simulator performance in tasks relevant for space flight. Wong et al. (2020) observed continued learning in a grappling and docking task during the course of 28 hours of sleep deprivation. In another docking task performed during magnetic resonance imaging, Strangman et al. (2005) found no performance decrements after sleep deprivation, but compensatory cerebral responses that might be related to the stabilization of performance. Oftentimes increased motivation to compensate sleepiness is suspected to be the cause of missing or small effects on complex tasks (Harrison & Horne, 2000). Although both manual docking studies were not able to identify performance decrements after total sleep deprivation, methodological problems persist, most notably small sample sizes or the lack of an appropriate control group. In an electroencephalography study conducted on ISS, slower reaction times in a docking simulation were associated with more global local sleep-like events, a marker of sleep pressure (Petit et al., 2019). These results provided first evidence on increased sleep pressure during space flight and potential negative consequences for visuomotor performance. Again, firm conclusions are not possible – the sample included only five astronauts and the docking task was too simplified to mirror the demands of a complex operational task. Another hint at the deleterious effect of sleep loss in space is found in the crash of Progress and Mir in 1997. Sleep deprivation due to long hours and stress as well as impaired sleep quality were made out as contributors to the manual docking accident (Ellis, 2000). In conclusion, sleepiness increases error rate and accident risk (Dinges, 1995), which is critical in high-risk operational contexts that require reliable performance to prevent catastrophic outcomes (Porcu et al., 1998). More experimental evidence is needed to gauge the consequences of sleep loss in space, particularly with respect to performance in operationally relevant tasks (Flynn-Evans, Gregory, et al., 2016). Therefore, the objective of this third study was to investigate the effect of 24 h total sleep deprivation on manual docking performance in the *6df* task. Knowledge about the influence of stressors like sleep loss on operational performance can be a valuable starting point to identify training needs and improve the prediction of performance.

6df performance has been associated with performance in the *Digit-Symbol Substitution Task*, a measure of processing speed, visual tracking, and working memory, and the *Psychomotor Vigilance Task* (PVT), a measure of sustained attention (Basner et al., 2020). These cognitive domains are not only relevant for docking performance, but also sensitive to sleep loss (Goel et al., 2009; Lim & Dinges, 2010). Decreased sustained attention is even the most reliable effect of sleep deprivation and a likely explanation for accidents and errors attributed to fatigue (Goel et al., 2013; Lim & Dinges, 2010). Because the PVT is a gold standard for investigating the cognitive effects of sleep deprivation and the ability to sustain attention is a basic process essential for cognitive functioning in general, we additionally aimed at assessing the importance of sustained attention for *6df* performance during sleep loss. Sustained attention is defined as the ability to maintain a high level of attention directed at a specific focus over a length of time (van Schie et al., 2021). We used this term instead of vigilance, which has been defined as the ability to be aware of relevant, unpredictable changes in the environment (but without a specific direction or focus) and is a prerequisite of attention in general (van Schie et al., 2021). However, both terms are often used interchangeably in the literature. The PVT as a measure of sustained attention is characterized by a minimal learning curve, briefness, and easy administration (Dorrian et al., 2005). A schematic sequence of the simple task is provided in Figure 4. Participants respond to a millisecond counter appearing at pseudo-random intervals by pressing a button. The complete test continues for 10 minutes. Sleep deprivation consistently leads to longer reaction times as well as errors of omission and commission in response to the randomly occurring stimuli (Basner & Dinges, 2011).

Figure 4

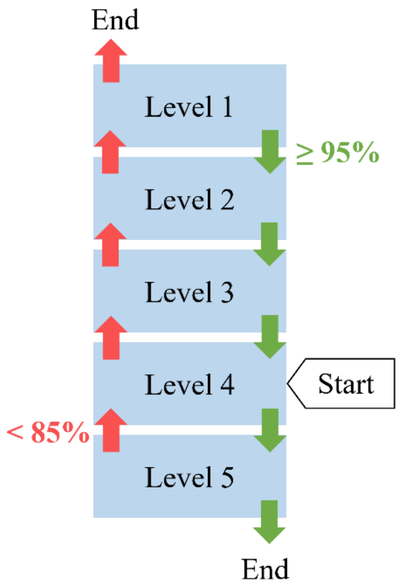
Schematic procedure for the PVT



To investigate the influence of sleep deprivation on docking performance, the *6df* tool had to be adapted. Because the participants were only available for two sessions during their stay in the laboratory, extensive training was not feasible. Instead of the full training program used in studies I and II, participants absolved a short version of *6df* that could be repeated in both sessions and did not require previous training. Five difficulty levels from the original program were chosen; spanning the easiest task and a full standard docking maneuver (Table 2). Every session started with the same task on level 4. Thereafter, the program adapted the difficulty level according to the participant's docking accuracy (Figure 5). This task design allowed for the application of an operational task that usually requires lengthy training in a study of shorter duration and with novice participants. Thereby, a control condition and a larger sample size than in previous studies on sleep deprivation and the control of six DoF could be realized.

Figure 5

Sequence of difficulty levels in the short version of 6df used in study III



Note. The session always started with the same task on level 4; the difficulty of subsequent tasks was adapted with respect to the docking accuracy achieved. A docking accuracy below 85% led to the selection of an easier task and scores of 95% or higher to a more difficult task. For performance scores between 85% and 95%, the same task was repeated. Every level included two or three related tasks that had to be successfully completed to ascend to the next level. The *6df* session ended after successful completion of level 5, insufficient accuracy in level 1, or after 35 minutes.

We hypothesized that sleep deprivation would not only impair sustained attention in the PVT, as is established in the literature (Lim & Dinges, 2008), but also operational performance in the *6df* task. High levels of sustained attention are expected to be important for safety during manual docking, because relevant changes of instruments or the environment demand fast detection and reaction. Therefore, we wanted to assess if sustained attention could explain docking performance under sleep deprivation. If this was the case, administration of the PVT prior to safety-critical operational tasks might be valuable as an early warning for performance decrements due to sleepiness.

1.4.4 Summary of study objectives

In this introduction, space was illustrated as an adverse environment that involves numerous risks and challenges for human performance. At the same time, mission success heavily depends on reliable levels of performance. Performance in complex tasks that are relevant in the actual work environment has been rarely investigated systematically, partly due to the methodological difficulties of research in space. Simulations of operational tasks like the *6df* tool enable both Earth-based research as well as the subsequent application in space. The goal of this dissertation was to add to the understanding of successful operational performance in space, using the example of manual spacecraft docking as a complex psychomotor skill. The following empirical studies (Table 1) investigated how docking training can be facilitated by 3D visualization (study I), how performance can be predicted via specific gaze features (study II), and how performance is affected by sleep deprivation as a typical stressor of space flight (study III).

Table 1

Overview of studies

Study	*N*	Design	*6df* version	Performance measures
I	24	30-day 6° HDT bed rest; between-subjects: 2D vs. 3D task visualization	learning program (12 levels)	learning speed
II	10	30-day 6° HDT bed rest; 2D group of study I: eye tracking	learning program (12 levels)	docking accuracy
III	62	24 h sleep deprivation; within-subjects: sleep deprived vs. control condition performance	shortened adaptive paradigm (5 levels)	docking accuracy; highest difficulty level achieved

2 Study I: Stereoscopic learning aid for manual docking

Sarah Piechowski, Willi Pustowalow, Michael Arz, Jörn Rittweger, Edwin Mulder,
Oliver T. Wolf, Bernd Johannes, & Jens Jordan:

Virtual reality as training aid
for manual spacecraft docking

Published 2020 in *Acta Astronautica*, *177*, 731-736.

Acta Astronautica 177 (2020) 731–736

Contents lists available at ScienceDirect

Acta Astronautica

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

Virtual reality as training aid for manual spacecraft docking

Sarah Piechowski [a,*], Willi Pustowalow [a], Michael Arz [a], Jörn Rittweger [a,c], Edwin Mulder [a], Oliver Tobias Wolf [d], Bernd Johannes [a], Jens Jordan [a,b]

[a] Institute of Aerospace Medicine, German Aerospace Center (DLR), 51147, Cologne, Germany
[b] Chair of Aerospace Medicine, University of Cologne, 50923, Cologne, Germany
[c] Department of Pediatrics and Adolescent Medicine, University of Cologne, 50923, Cologne, Germany
[d] Department of Cognitive Psychology, Ruhr University Bochum, 44801, Bochum, Germany

ARTICLE INFO

Keywords:
Manual docking
Skill acquisition
Virtual reality
Stereoscopy

ABSTRACT

The ability to manually dock a spacecraft to a space station can be crucial for astronauts during space missions. The computer-based self-learning program *6df* is an abstract docking simulation for acquisition and maintenance of the underlying skill to control six degrees of freedom. One of the difficulties of this complex task is to construct a mental representation of the own position and orientation in space, based only on two-dimensional information. To facilitate this and possibly further improve the learning process, a new three-dimensional (3D) stereoscopic presentation of the program is tested. This study investigates whether there is faster learning progress with 3D presentation compared to standard 2D presentation. 24 participants of the Artificial Gravity Bed Rest Study with ESA (AGBRESA) participated in the *6df docking* experiment. Each of them completed 20 training sessions which lasted approximately 45 min and were conducted twice a week. The learning program is self-sufficient and adapts itself to individual learning speed. Half of the participants were presented with an UNITY-based stereoscopic visualization of docking, whereas the other half used the standard 2D version of the learning program *6df*. Learning progress was measured as the number of tasks needed to reach a target task. Results overall indicate a slightly faster learning progress when using 3D technology, but no long-term performance advantages. The small benefit might not justify the usage of costlier and operationally limiting 3D systems.

1. Introduction

Manually controlled docking of a spacecraft to a space station can be crucial for space mission safety, as automatic docking may fail or more flexibility may be needed [1,2]. The complex task requires the ability to proficiently control objects in six degrees of freedom (DoF), which is almost unique to space. In space, objects can be moved along three axes (translation) and rotated around each axis (orientation). During docking, the left hand control operates three DoF of translation (movement along x-, y- and z-axis) and the right hand control three DoF of rotation (controlling yaw, pitch and bank). In contrast, when driving a car, only two DoF have to be controlled. The ability to control six DoF has to be trained intensely on simulators and with experienced instructors. The task is challenging, as internal frames of reference have to be constructed, i.e. a representation of one's own position, orientation, and motion within the physical environment. New cognitive, perceptual, and motor skills have to be acquired. The two hand controls have distinct functionalities: the translation control resembles a set of on-off switches and each impulse must be compensated with an equally strong impulse in the exact opposite direction to stop the movement. Thus, stabilizing the spacecraft in all axes is demanding. By contrast, the orientation hand control is an analogous one. The difficulty here is that handling is counterintuitive for most people, as the hand control must be moved to the right if one wants to turn left. All these challenges occur in addition to the adverse conditions of space flight, which may impair performance in astronauts and cosmonauts with potentially fatal consequences. Indeed, according to Ellis [2], workplace stress, sleep deprivation, and insufficient training for skill maintenance predisposed to an accident during manual docking in 1997.

We developed the *6df* training tool to facilitate acquisition and maintenance of the complex manual ability of controlling six DoF [3]. The learning program acquaints participants without prior knowledge to the handling of six DoF and features individually paced self-learning without an instructor. Moreover, the tool is designed for continuous

* Corresponding author.
E-mail address: sarah.piechowski@dlr.de (S. Piechowski).

https://doi.org/10.1016/j.actaastro.2020.08.017
Received 14 April 2020; Received in revised form 22 June 2020; Accepted 11 August 2020
Available online 20 August 2020

training to maintain docking skills on a safe level, for example during long term space flights. Furthermore, the learning process and operator's skill can be investigated simultaneously, as in a previous study by Johannes et al. [4]. In this study, all participants were able to perform a standard docking maneuver task following the *6df* course. However, some needed considerably more training time and repetitions. Therefore, we were interested in methodologies to further enhance learning and training efficiency. Constructing an appropriate frame of reference is critical for learning success and has to be newly learned in space, as there is no fixed plane for orientation. Additionally there is another difficulty: manual docking is based on a two-dimensional screen that impedes perception of one's own position and spatial relations. We reasoned that perception of relations in space and training efficiency could be improved using a three-dimensional (3D) stereoscopic version of the *6df* learning program as a desktop-based virtual reality (VR) approach.

A plethora of 3D and virtual reality applications have been designed in the past years, not only for entertainment purposes. Fuchs, Moreau, and Guitton [5] provide a definition of VR as a computer-based simulation of the behavior of 3D entities in a virtual world, which interact with the user in real time via sensorimotor channels. According to Freina and Canessa [6], different levels of immersion are possible, which can create a feeling of actual presence in the virtual world. As Freina and Ott [7] review, VR methodology has been discovered for educational purposes in many different fields. In medical training for example, Seymour summarizes VR to provide effective skill transfer into the operating room [8]. In space flight contexts, "real-life training" is often expensive, demanding for large facilities or even impossible, therefore VR is used to efficiently extend astronaut training possibilities [9–13]. For example, Aoki and colleagues tested a VR navigation training for facilitating orientation within International Space Station in the case of an emergency egress [14,15]. Olbrich et al. [16] followed a similar idea with the development of a VR environment that allows astronauts to train for a possible case of fire emergency in a simulated lunar base. Another VR training application, examined by Stroud, Harm and Klaus [17], has been the prevention of motion sickness and spatial disorientation in space. Bosch Bruguera, Ilk, Ruber and Ewald [18] lately developed their Soyuz spaceflight simulator by adapting it for future missions with the Russian Spacecraft "Federatsiya" to Lunar Orbital Platform – Gateway and by adding immersion using a VR headset and hand tracking. Their approach is focused on achieving high graphical and physical realism.

We tested the hypothesis that stereoscopic presentation of the learning program will enhance participants' ability to understand spatial relations, and thereby the construction of an appropriate frame of reference. More precisely, we anticipated that additional spatial information should facilitate mental representation of spatial relations, and eventually lead to faster learning progress compared to the standard 2D view. We had the unique opportunity to test our approach in the setting of a head-down tilt bed rest study, which is an established terrestrial model for microgravity [19,20]. Our primary goal is the development of a tool that is applicable in space. After *6df* has been tested for general suitability as a learning tool [4], we wanted to show that it would be likewise applicable under space analog conditions.

2. Material and methods

2.1. Participants

24 healthy individuals (8 women and 16 men, 24–55 years old) participated in our experiment, which was part of the "Artificial Gravity Bed Rest study with European Space Agency" (AGBRESA) at the :envihab facility of the Institute of Aerospace Medicine at the German Aerospace Center (DLR) in Cologne, Germany. AGBRESA was a large joint project of ESA, NASA, and DLR, designed to accumulate knowledge about the effects of microgravity in an experimental ground-based analog environment for long-term human spaceflight. The study was prospectively registered with the German register for clinical studies (www.drks.de) with the identifier DRKS00015677 and comprised two campaigns (March–June and September–December) in 2019 with twelve participants each. After 15 days of familiarization and baseline measurements, participants spent 60 days in 6° head-down tilt bed rest to simulate the effects of microgravity and to explore the effectiveness of short-arm centrifuge training as a countermeasure against degradation processes in weightlessness. After re-ambulation, participants stayed in the facility for another 14 days of regeneration and post measurements. Our sub-study consisted of a training course on how to maneuver an object with six DoF using the *6df* tool. One of the participants first used the Russian version of the learning program, but switched to the default German version later during the course. Another participant chose the English version of the instructions. Participants were granted monetary compensation for the whole bed rest study. The study has been approved by the ethics committee of the medical association North-Rhine in Dusseldorf, Germany and participants provided written informed consent.

2.2. Docking task

The training tool named *6df* used in this experiment has already been described in detail and tested for applicability before by Johannes et al. [3,4]. In short, *6df* is a computer-based and self-sufficient learning program that simulates the control of an object with six DoF, in this case the manual docking of a spacecraft to an abstract space station. Flight dynamics and controller responsiveness are based on the Russian docking training system TORU (Teleoperatiya Ruchnogo Upravleniya – teleoperated manual control) and the actual Soyuz spacecraft. However, the tool is not designed to be a realistic Soyuz simulation, but to teach the principles of the control of any object in space abstractly. Participants are first familiarized with the controller handling and are then gradually instructed to control up to six DoF. Each task starts with an illustrated instruction text, sometimes including example videos. After each task, feedback about various specific parameters such as forward speed, pitch, bank, and yaw is given, as well as an aggregated general performance measure, with zero being the worst and 1.0 the best possible accuracy, following TORU methodology [21]. The program adapts to individual learning speed, so that tasks are repeated when errors occur. If a task is mastered with a general performance score of at least 0.95, the next (and more difficult) task is presented. The program is structured in twelve ascending levels labeled between 1 and 60, most of them containing a small number of different tasks that are similar in difficulty. At the end of the learning program participants should be able to dock a virtual object to the docking point in a standard docking maneuver including flight-around, stabilization on the center line and final approach. The *6df* software was developed by SpaceBit GmbH (Eberswalde, Germany) and hand controls were produced by Koralewski Industrie-Elektronik oHG (Hambühren, Germany).

2.3. Setup and procedure

During experimental sessions, participants remained in a 6° head-down tilt position without a pillow. A computer screen was attached on a rack above the participants' heads at a distance of approximately 60 cm to present the *6df* training program. Hand controls for docking were also mounted to the rack so that they could be used conveniently in a lying position. Complete laboratory setup and equipment is illustrated in Fig. 1. Each participant completed at least twenty *6df* training sessions, each of which took approximately 45 min. In an earlier study using *6df*, an average of 20 sessions sufficed to pass the course and reach the standard docking level [4]. Sessions were scheduled on average twice a week during the study course, three sessions before bed rest and the remaining sessions during the 60-day bed rest period. Sessions were minimally one day and maximally seven days apart, but the usual interval was every three to four days. Each single docking task comprised

Fig. 1. Laboratory setup for the stereoscopic *6df* version in 6° head-down tilt.

up to 12 min without instructions and feedback, depending on level and participant's speed. The number of tasks in each session also varied depending on these factors. Participants in each campaign were randomly assigned to two groups: one group was presented with the conventional two-dimensional learning program, and the other group used a newly designed stereoscopic *6df* version. Therefore, in each campaign six participants were assigned to the stereoscopic version and six participants to the standard program. The 3D program was equivalent to the standard version, but displayed a three-dimensional view of visor and station based on *Unity* (Unity Technologies, San Francisco, CA, USA). The visor resembles a cross to target the docking point and adjust the orientation of the spacecraft to the station. Participants wore *Nvidia 3D Vison 2* wireless glasses (Nvidia Corporation, Santa Clara, CA, USA). Because a docking maneuver in space has to be performed based on a two-dimensional screen, the stereoscopic view only supported the first learning steps. Three-dimensional viewing was only used until a participant reached the task in the middle of level 15. After achieving this landmark, the program automatically switched to the standard two-dimensional view, so conditions were similar for both groups thereafter and the 3D group would be able to adjust to 2D view during the rest of level 15. Level 15 was chosen because the stabilization and correct orientation of the spacecraft previous to the final docking approach are trained. To stand still in open space is of high difficulty, an important milestone in the learning process and necessary to solve all following tasks. We assumed that the new technology might be most helpful up to that point, but should then be omitted to familiarize participants with the standard two-dimensional presentation, as in reality docking is also based on a 2D screen.

On the day after being released from bed rest (five days after the last *6df* training), we verified learning success in an additional session. This session contained a fixed series of five docking tasks of the Russian training system TORU that was provided by S.P. Korolev Rocket and Space Corporation Energia, Korolyov, Russia. TORU tasks applied the same hand controls and require identical skills to control six DoF based on the Soyuz spacecraft. Nevertheless, these tasks are more demanding, as they additionally take into account orbital mechanics and spacecraft inertia. The same procedure was applied for regular cosmonaut training onboard the International Space Station 2008–2011 [21].

2.4. Data analysis

For data processing and statistical analysis, we applied *SPSS Statistics 21* (IBM, Armonk, NY, USA). Some levels include a predefined flight path, marked with rings the participants had to move through without touching – otherwise the task is terminated. Excluded from the analysis were 256 attempted tasks that ended in such a ring collision. We operationalized learning speed using the number of tasks "flown" by the subjects. As *6df* is adaptive, fewer tasks up to a criterion task or level mean fewer errors, faster progress through the program and therefore faster learning. The dependent variable of interest was, therefore, how fast, which means after how many tasks, participants reached the critical task on level 15. Since the number of tasks was significantly non-normally distributed according to Kolmogorov-Smirnov test ($D = 3.32$, $p < .001$), we compared learning speed between 2D and 3D group using the nonparametric Mann-Whitney test. In the same way we also tested if there was a difference in the number of tasks needed from the beginning of the program up to level 60, which resembles a standard docking maneuver in space and is therefore the end of the learning program. This was done in order to test whether the initial 3D-training had any longer lasting effects on the learning process. Cohen's d is reported as measure of effect size. Additionally we also applied a multilevel linear mixed effect model (LME) to the data to test the effects on the number of tasks (learning speed) throughout all training sessions. For this purpose raw data were approximated to normal distribution as far as possible using Box-Cox transformation ($D = 2.84$, $p < .001$). Another eight tasks were excluded as extreme outliers (task number values more than three standard deviations above mean). Thereby, we achieved normal distribution of residuals for LME modelling. The model included level, group (2D or 3D), campaign (spring or autumn) gender and age (median split: ≤ 33 and >33 years old) as fixed effects, as well as the interactions of level with group and level with gender. Participants were included as a random effect using *variance components* as covariance structure. The model was applied to the whole learning data as well as separately to both training halves before and after the switch from 3D to 2D. Finally TORU data were analyzed to test whether 3D visualization would influence not only learning speed, but final docking performance after the course. The TORU docking performance score (ranging from 0 to 1.0) was the dependent variable of this LME. Group and the interaction of group with TORU task (1–5) were included as fixed effects, participants as random effect.

3. Results

Overall, participants completed 3395 valid training tasks. On average, participants of the 2D group "flew" $M = 89.67$ tasks (Median = 81.50, $SD = 32.33$) before reaching the shift task from 3D to 2D on level 15, whereas participants of the 3D group required on average $M = 76.17$ tasks (Median = 78.50, $SD = 11.65$) (see Fig. 2). Although 3D participants did learn faster descriptively, the difference between the two groups was not statistically significant ($U = 59.00$, $z = -0.75$, $p = .45$, $d = 0.55$). Regarding learning speed up to the final standard docking maneuver, 2D participants completed on average $M = 147.75$ (Median = 151, $SD = 18.40$) tasks and 3D participants $M = 140.58$ tasks (Median = 136.50, $SD = 25.65$) (see Fig. 3). Likewise, the Mann-Whitney test did not result in significant group differences ($U = 56.50$, $z = -0.90$, $p = .37$, $d = 0.32$).

A descriptive view on the learning process based on all single levels (Fig. 4) reveals that 3D participants on average needed fewer tasks to reach every level in comparison with the 2D group; especially during the 3D phase up to level 15. Yet, after switching to 2D, differences between groups were only marginal. Therefore, we assessed the whole learning process using LME to further clarify the results. Not surprisingly, level predicted the number of tasks significantly for the whole learning program ($F(11, 3338.57) = 4215.41$, $p < .001$) as well as for the first 2D vs 3D half ($F(4, 1867.69) = 2910.58$, $p < .001$) and the second all 2D portion of the training program ($F(7, 1447.39) = 1202.34$, $p < .001$); as participants did ascend to higher levels with increasing task number. There was no significant main effect of campaign in any model (whole course: $F(1, 18.98) = 0.10$, $p = .75$; first half: $F(1, 19.14) = 0.07$, $p = .79$; second half: $F(1, 18.89) = 0.20$, $p = .66$), and, therefore, no differences in learning speed between study cohorts. 2D/3D group did not predict learning speed, neither for the whole program ($F(1, 19.09) = 0.17$, $p = .68$), nor for the first 3D half ($F(1, 19.14) = 0.52$, $p = .48$) or the second section after switching to 2D ($F(1, 18.91) = 0.04$, $p = .85$). Although there was no main effect of the 3D presentation, the interaction of level with group did predict the number of tasks for the whole program ($F(11, 3338.54) = 3.55$, $p < .001$) as well as for the first portion of the program ($F(4, 1867.91) = 9.51$, $p < .001$). However, there was no significant interaction for the second half of the training alone ($F(7, 1447.28) = 1.03$, $p = .41$). The efficacy of 3D presentation in augmenting learning speed is, therefore, dependent on task difficulty level – and is not carried over into the all 2D training phase.

The model also included age and gender as possible predictors. Age had a significant main effect in all models (whole course: $F(1, 18.98) = 7.80$, $p = .01$; first half: $F(1, 19.12) = 5.34$, $p = .03$; second half: $F(1, 18.88) = 8.67$, $p = .01$). As shown in Fig. 5, younger participants did learn faster in comparison to older participants. Gender predicted the number of tasks significantly for the whole program ($F(1, 19.11) = 5.39$, $p = .03$) as well as for the second portion ($F(1, 18.97) = 5.80$, $p = .03$); but not for the first portion of the learning program ($F(1, 19.08) = 2.83$, $p = .11$). The interaction of level with gender, however, significantly predicted number of tasks in all models (whole course: $F(11, 3338.62) = 34.00$, $p < .001$; first half: $F(4, 1867.82) = 65.42$, $p < .001$; second half: $F(7, 1447.43) = 5.75$, $p < .001$). Whereas there was barely a gender difference during the very first levels, men did learn faster than women in the middle and higher difficulty ranges (see Fig. 6).

Fig. 7 illustrates performance in the five Russian TORU docking tasks. For the first three tasks, the stereoscopic group's average

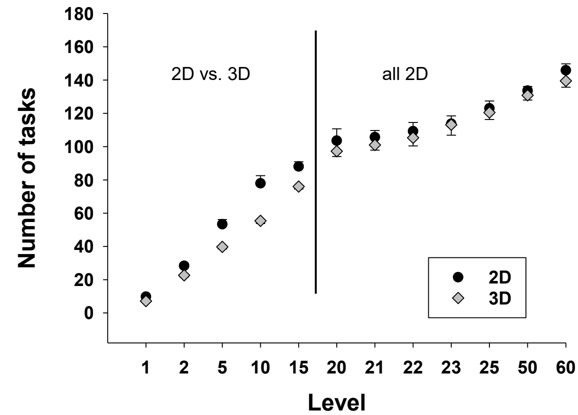


Fig. 4. Average number of tasks needed to reach each *6df* level by experimental group. Whiskers indicate 95% confidence interval.

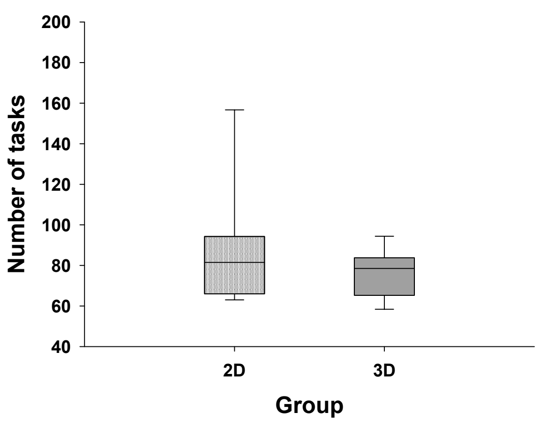


Fig. 2. Learning speed until the criterion task on level 15 by presentation group. Depicted are the median number of tasks, the interquartile range (box) as well as minima and maxima (whiskers).

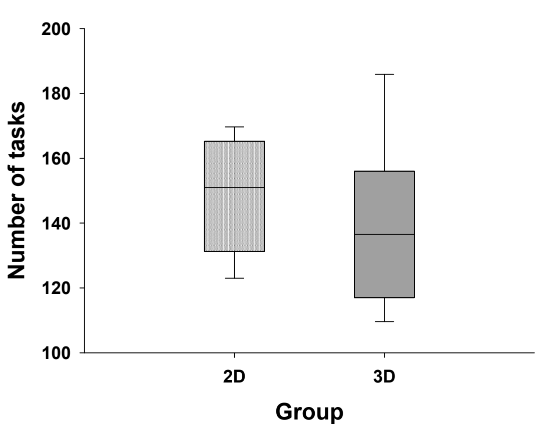


Fig. 3. Learning speed until the end of the learning program by presentation group. Depicted are the median number of tasks, the interquartile range (box) as well as minima and maxima (whiskers).

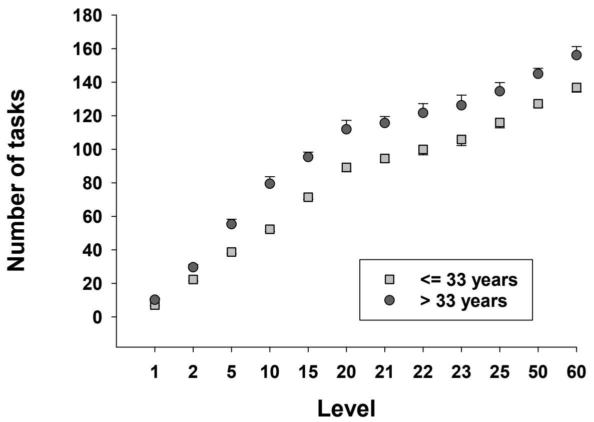


Fig. 5. Average number of tasks for each *6df* level by age group (33 years or younger; older than 33 years). Whiskers indicate 95% confidence interval.

performance scores were higher than those of the 2D group; for the last two tasks there was no difference. Averaged over all TORU tasks 3D participants achieved a performance score of $M = .86$ and 2D participants of $M = .81$. The LME resulted in no significant main effect of 2D/3D group ($F(1, 21.78) = 0.58, p = .46$). However, there was a significant interaction of TORU task and group, which predicted the performance outcome ($F(8, 85.18) = 8.96, p < .001$). Stereoscopic presentation, therefore, did not only have an influence on learning speed in the 6df program, but also on learning success regarding performance during the first Russian docking tasks. Nevertheless, at the end of the TORU tasks, 2D participants performed as good as the 3D group – so once again there was no persisting effect of stereoscopic presentation.

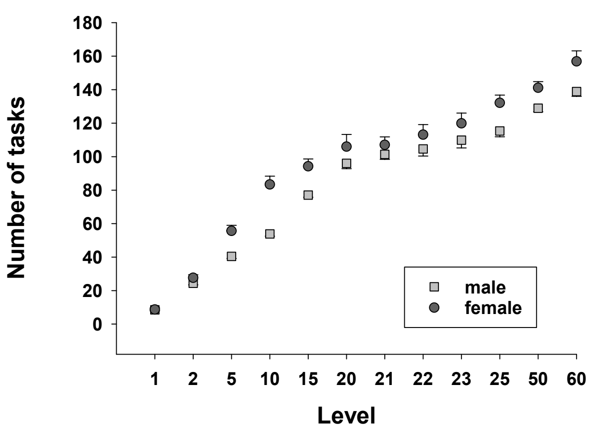


Fig. 6. Average number of tasks for each *6df* level by gender. Whiskers indicate 95% confidence interval.

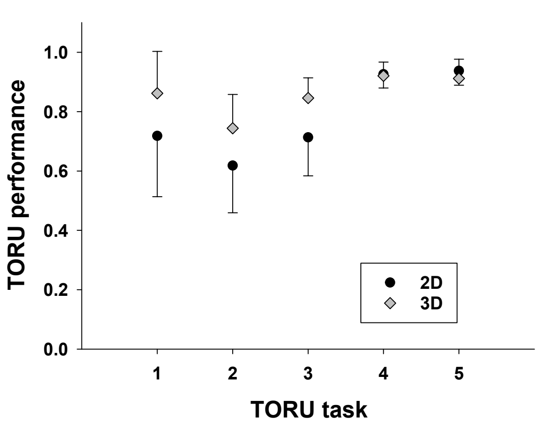


Fig. 7. Average docking performance score in the five TORU tasks for 2D and 3D group. Whiskers indicate 95% confidence interval.

4. Discussion

We did not observe a consistent benefit of 3D presentation compared with 2D. Although 3D participants descriptively required fewer tasks on average than the 2D participants to reach each individual training program level, statistical analysis of these differences yielded equivocal results. Whereas group comparisons by Mann-Whitney tests at measuring points level 15 and 60 failed to reach significance, having a closer look at the whole learning process via LME modelling could significantly confirm a positive effect of the 3D presentation that was dependent on the difficulty level. While we observed no general advantage for participants in the 3D group, they did learn faster at least during the first portion of the course according to the mixed model. The hypothesis that stereoscopic presentation of the *6df* learning tool does facilitate the learning process compared to standard 2D can be confirmed only partly and with restrictions: 3D seems to facilitate learning; however, the effect was attenuated after switching to the standard 2D course. Gender and age both affected learning speed, with steeper learning curves in younger individuals and in men. In a previous study of Johannes et al., a similar effect of age was also present, whereas there was no significant influence of gender [4]. Experience with computer games or simulations might be a possible explanation for this (in this study, all participants who reported being passionate video game players were male and in the younger age group). Video gaming could have an impact as it has been associated with improvements in spatial abilities, as reported in a review by Spence and Feng [22]. Gaming experience occasionally also relates to performance in operational tasks, for example in a robot navigation study of Gomer and Pagano [23].

In addition to learning speed, learning success was measured as performance in a series of docking training tasks that have been used by cosmonauts in space. Although there was no general effect of the stereoscopic presentation, 3D participants outperformed 2D participants during the first three tasks. Participants in the 3D group seem to have adapted faster to the new circumstances, as TORU tasks follow different mechanics (e.g. high inertia of the spacecraft) and, therefore, require generalization of the acquired skill. We speculate that the stereoscopic group may have built up a more robust sense for their position and orientation in space. Taken together, stereoscopic presentation seems to have a positive, but rather small impact on learning to control six DoF.

Our ambiguous findings may be explained in part by the large interindividual variance in learning speed. Indeed, the number of completed tasks until the level 15 breakpoint ranged from 56 to 160. While the sample size was relatively large for a complex bed rest study, it was not sufficiently large to reduce the impact of exceptionally slow or fast learners or for detailed subgroup analysis. The latter could serve to detect predictors for individuals who are more likely to benefit from 3D presentation as training aid. Former experience or familiarization with 3D glasses or virtual reality environments for example might have an impact, as well as general computer affinity. Future studies might also account for spatial orientation ability, which could conceivably contribute to variability in performance. Wang et al. [24] discovered that perspective taking and mental rotation ability are associated with manual docking performance, which might be particularly relevant for novices, according to Du et al. [25]. 3D presentation could be more beneficial for those with limited spatial orientation beforehand, whereas skilled persons might benefit less. Despite the small sample size, bed rest provided an exceptional opportunity to investigate performance under extreme conditions that at least partly resemble the adverse conditions astronauts are facing.

Stereoscopic viewing has been associated with symptoms ranging from discomfort to motion sickness due to visual conflicts [26]. None of the participants reported adverse events when wearing 3D glasses. Still, some experienced the stereoscopic presentation as more strenuous and tiring for the eyes compared to 2D. A few participants reported difficulties integrating stereoscopic double images into one consistent three-dimensional view. These issues might have reduced the benefit of the 3D view or even hindered some participants in the learning process. Often full immersion is mentioned as an important component of VR, which can be achieved for example through head-mounted displays and creates the perception of being physically present in the virtual environment [6]. Nevertheless, there have been non-immersive approaches like in our study, using virtual 3D environments that are presented on a conventional monitor. This "desktop VR" is capable of creating at least mental or emotional immersion, as suggested by Robertson, Card and Mackinlay [27]. We decided for 3D glasses instead of a head-mounted display to reduce the risk of cybersickness, but also because participants should be able to look freely at the hand controls – especially during familiarization with their functioning. The potential advantage of immersive VR is the opportunity to blind out reality and get fully absorbed in the simulation. Whether immersive VR further improves the learning process compared to 3D deserves to be studied. Nevertheless, Aoki, Oman, Buckland and Natapoff observed that non-immersive VR is not necessarily inferior to immersive VR in navigation training contexts [14]. In reality, docking is also done in front of a 2D screen while seeing the hand controls. Therefore, immersion would require virtual imaging of hands and controls, e.g. by using wired gloves. Our main interest was to visually clarify spatial relations, which the simpler 3D glasses are sufficient for.

In conclusion, stereoscopic presentation during the acquisition of the ability to control objects with six DoF had only small effects on learning speed and success. Whilst facilitating learning during the first few sessions, this benefit did not persist throughout the course and seemed to fade away as soon as there was a switch to standard 2D. 3D training may have slightly improved the ability to accustom to varying mechanics like during TORU. In the end, however, performance in this task did not differ between groups. Nevertheless, there are some interesting starting points for further research on possible learning aids as well as on factors influencing the learning process. As manual docking in space relies on 2D screens and not all participants seem to benefit from 3D, we favor the simpler, less costly, but more realistic 2D learning program.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the German Aerospace Center (DLR) acting as space agency [DARA-Grants 50WB9128, 50WB93401, 50WB93401-ZA; DLR-Grants 50WB 96220, 50WP0306, 50WP0501, 50WP0602, 50WP0603, 50WP1104, 50WP1304, 50WP1609].

References

[1] A.R. Brody, Spacecraft Flight Simulation: A Human Factors Investigation into the Man-Machine Interface between an Astronaut and a Spacecraft Performing Docking Maneuvers and Other Proximity Operations, NASA-CR-177502, NASA/Ames Research Center, Moffett Field, CA, USA, 1988.

[2] S.R. Ellis, Collision in space, Ergon. Des. 8 (1) (2000) 4–9.

[3] B. Johannes, S. Bronnikov, Y. Bubeev, A. Dudukin, H.-J. Hoermann, T. Frett, J. Rittweger, A. Gaillard, A tool to facilitate learning in a complex manual control task, Int. J. Appl. Psychol. 7 (4) (2017) 79–85, https://doi.org/10.5923/j.ijap.20170704.01.

[4] B. Johannes, S. Bronnikov, Y. Bubeev, M. Arz, S. Piechowski, J. Rittweger, Individual learning curves in manual control of six degrees of freedom, Int. J. Appl. Psychol. 9 (2) (2019) 47–51, https://doi.org/10.5923/j.ijap.20190902.01.

[5] P. Fuchs, G. Moreau, P. Guitton, Virtual Reality: Concepts and Technologies, CRC Press, Boca Raton, 2011.

[6] L. Freina, A. Canessa, Immersive vs desktop virtual reality in game based learning, in: R. Munkvold, L. Kolas (Eds.), 9th European Conference on Games Based Learning, Academic Conferences and Publishing International Limited, Steinkjer, Norway, 2015, p. 195.

[7] L. Freina, M. Ott, A Literature Review on Immersive Virtual Reality in Education: State of the Art and Perspectives, The International Scientific Conference eLearning and Software for Education, Bucharest, 2015.

[8] N.E. Seymour, VR to OR: a review of the evidence that virtual reality simulation improves operating room performance, World J. Surg. 32 (2) (2008) 182–188, https://doi.org/10.1007/s00268-007-9307-9.

[9] A. Twombly, J. Smith, K. Montgomery, R.D. Boyle, The virtual glovebox (vgx): a semi-immersive virtual environment for training astronauts in life science experiments, J. Syst. Cybern. Inf. 2 (3) (2006) 30–34.

[10] J. Gancet, O. Lamborelle, K. Chintamani, P. Letier, M. Ilzkovitz, D. Churkin, L. Ferra, S. Ghiste, J. McIntyre, A. Preumont, R. Aked, E. Motard, F. Didot, Force Feedback and Immersive Technologies Suit (FITS): an Advanced Concept for Facility-Less Astronaut Training, International Symposium on Artificial Intelligence, Robotics and Automation in Space, Turin, Italy, 2012.

[11] R.B. Loftin, Virtual environments for aerospace training, in: Proceedings of WESCON'94, IEEE, 1994, pp. 384–387. https://doi.org/10.1109/WESCON.1994.403567.

[12] R.J. Stone, P.B. Panfilov, V.E. Shukshunov, Evolution of aerospace simulation: from immersive Virtual Reality to serious games, in: Proceedings of 5th International Conference on Recent Advances in Space Technologies-RAST2011, IEEE, 2011, pp. 655–662. https://doi.org/10.1109/RAST.2011.5966921.

[13] A.D. Garcia, J. Schlueter, E. Paddock, Training astronauts using hardware-in-the-loop simulations and virtual reality, in: AIAA Scitech 2020 Forum, 2020. https://doi.org/10.2514/6.2020-0167.

[14] H. Aoki, C.M. Oman, D.A. Buckland, A. Natapoff, Desktop-VR system for preflight 3D navigation training, Acta Astronaut. 63 (7–10) (2008) 841–847, https://doi.org/10.1016/j.actaastro.2007.11.001.

[15] H. Aoki, C.M. Oman, A. Natapoff, Virtual-reality-based 3D navigation training for emergency egress from spacecraft, Aviat. Space Environ. Med. 78 (8) (2007) 774–783.

[16] M. Olbrich, H. Graf, J. Keil, R. Gad, S. Bamfaste, F. Nicolini, Virtual reality based space operations – A study of ESA's potential for VR based training and simulation, in: J. Chen, G. Fragomeni (Eds.), Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation. VAMR 2018. Lecture Notes in Computer Science, 10909, Springer, Cham, 2018.

[17] K.J. Stroud, D.L. Harm, D.M. Klaus, Preflight virtual reality training as a countermeasure for space motion sickness and disorientation, Aviat. Space Environ. Med. 76 (4) (2005) 352–356.

[18] M. Bruguera Bosch, V. Ilk, S. Ruber, R. Ewald, Use of virtual reality for astronaut training in future space missions-spacecraft piloting for the Lunar Orbital Platform-Gateway (LOP-G), in: 70th International Astronautical Congress, 2019. Washington D.C.

[19] A.R. Hargens, L. Vico, Long-duration bed rest as an analog to microgravity, J. Appl. Physiol. 120 (8) (2016) 891–903, https://doi.org/10.1152/japplphysiol.00935.2015.

[20] A. Pavy-Le Traon, M. Heer, M.V. Narici, J. Rittweger, J. Vernikos, From space to Earth: advances in human physiology from 20 years of bed rest studies (1986–2006), Eur. J. Appl. Physiol. 101 (2) (2007) 143–194, https://doi.org/10.1007/s00421-007-0474-z.

[21] B. Johannes, V. Salnitski, A. Dudukin, L. Shevchenko, S. Bronnikov, Performance assessment in the PILOT experiment on board space stations mir and ISS, Aero. Med. Human Perform. 87 (6) (2016) 534–544, https://doi.org/10.3357/AMHP.4433.2016.

[22] I. Spence, J. Feng, Video games and spatial cognition, Rev. Gen. Psychol. 14 (2) (2010) 92–104, https://doi.org/10.1037/a0019491.

[23] J.A. Gomer, C.C. Pagano, Spatial perception and robot operation: selection, training, gaming, and design, Nav. Eng. J. 124 (1) (2012) 131–140.

[24] C. Wang, Y. Tian, S. Chen, Z. Tian, T. Jiang, F. Du, Predicting performance in manually controlled rendezvous and docking through spatial abilities, Adv. Space Res. 53 (2) (2014) 362–369, https://doi.org/10.1016/j.asr.2013.10.031.

[25] X. Du, Y. Zhang, Y. Tian, W. Huang, B. Wu, J. Zhang, The influence of spatial ability and experience on performance during spaceship rendezvous and docking, Front. Psychol. 6 (2015) 955, https://doi.org/10.3389/fpsyg.2015.00955.

[26] A.D. Hwang, E. Peli, Instability of the perceived world while watching 3D stereoscopic imagery: a likely source of motion sickness symptoms, i-Perception 5 (6) (2014) 515–535, https://doi.org/10.1068/i0647.

[27] G.G. Robertson, S.K. Card, J.D. Mackinlay, Three views of virtual reality: nonimmersive virtual reality, Computer 26 (2) (1993) 81.

3 Study II: Eye tracking as a performance indicator for manual docking

Sarah Piechowski, Bernd Johannes, Willi Pustowalow, Michael Arz, Edwin Mulder,
Jens Jordan, Oliver T. Wolf, & Jörn Rittweger:

Visual Attention Relates to Operator Performance in Spacecraft Docking Training

Published 2022 in *Aerospace Medicine and Human Performance, 93*(6), 480-486.

Visual Attention Relates to Operator Performance in Spacecraft Docking Training

Sarah Piechowski; Bernd Johannes; Willi Pustowalow; Michael Arz; Edwin Mulder; Jens Jordan; Oliver T. Wolf; Jörn Rittweger

BACKGROUND: Manually controlled docking of a spacecraft to a space station is an operational task that poses high demands on cognitive and perceptual functioning. Effective processing of visual information is crucial for success. Eye tracking can reveal the operator's attentional focus unobtrusively and objectively. Therefore, our aim was to test the feasibility of eye tracking during a simulation of manual docking and to identify links between visual information processing and performance.

METHODS: We hypothesized that duration and number of gazes to specific regions of interest of the simulation (total dwell time and number of dwells) would be associated with docking accuracy. Eye movements were recorded in 10 subjects (30% women, $M = 33.4$ yr old) during the 6° head-down tilt bed rest study AGBRESA during 20 training sessions with the 6df learning program for spacecraft docking.

RESULTS: Subjects' gaze was directed most frequently and longest to the vizor (185 dwells and 22,355 ms per task) followed by the two instrument displays (together 75 dwells and 4048 ms per task). We observed a significant positive relationship between number and duration of visual checks of speed and distance to the docking point and the accuracy of the docking maneuver.

DISCUSSION: In conclusion, eye tracking provides valuable information related to docking accuracy that might prospectively offer the opportunity to improve docking training effectiveness.

KEYWORDS: manually controlled docking, eye tracking, operator performance.

Piechowski S, Johannes B, Pustowalow W, Arz M, Mulder E, Jordan J, Wolf OT, Rittweger J. *Visual attention relates to operator performance in spacecraft docking training*. Aerosp Med Hum Perform. 2022; 93(6):480–486.

Manually controlled spacecraft docking to a space station is a highly safety-relevant maneuver.[3,9] Docking success depends on the ability to control objects with 6 degrees of freedom (DoF). A spacecraft can be navigated along three translational axes and rotated around each axis, which poses substantial challenges to cognitive functioning, motor control, and visual attention. The 6df training tool has been introduced to help the operator acquire and maintain the skill to control 6 DoF autonomously.[16,17] Software was developed by SpaceBit GmbH (Eberswalde, Germany) and hand controls by Koralewski Industrie-Elektronik oHG (Hambuehren, Germany) as a research tool for the German Aerospace Center (DLR) to investigate operational performance based on the Russian TORU manual docking platform. To further improve learning effectiveness and docking reliability, additional information about the underlying information processing would be beneficial.

Given its unobtrusive nature and tight link to cognitive processes, eye tracking is particularly promising in this regard. The observer usually focuses attention at the central direction of gaze, which can be followed through eye tracking.[8,19] Metrics like the number and duration of fixations or dwells are, therefore, used as objective indicators of visual attention.[14] In aviation, eye tracking is an established method to investigate visual

From the German Aerospace Centre (DLR) Institute of Aerospace Medicine, Cologne, North Rhine-Westphalia, Germany.

This manuscript was received for review in September 2021. It was accepted for publication in February 2022.

Address correspondence to: Sarah Piechowski, M.Sc., German Aerospace Centre (DLR) Institute of Aerospace Medicine, Linder Höhe, Cologne, North Rhine-Westphalia 51147, Germany; sarah.piechowski@dlr.de.

DOI: https://doi.org/10.3357/AMHP.6005.2022

information acquisition in the cockpit and train pilots to adhere to optimal scanning patterns.[35] It has been used to detect states of high workload and fatigue[26] or to differentiate between novices and experts.[2] For example, expert pilots show shorter dwells, but more frequent instrument checks than novices,[2,10] suggesting differences in the efficiency of information processing. Also, eye tracking metrics are indicators of situation awareness and useful to determine whether situational changes in the cockpit are actually identified.[31,34] Attention allocation plays an important role for situation awareness, which is predictive of piloting errors.[33] Piloting performance has been associated with more systematic and selective instrument scanning behavior and such successful strategies can be used to improve instructions during training.[4,12,20]

While some eye tracking devices had been developed especially for spaceflight,[5,6] none of them are currently being used in space. Current, state-of-the-art eye trackers are easier to apply and not as bulky as former devices, which makes them attractive for utilization in space. The identification of effective visual scanning behavior during docking might enhance man-machine interaction. However, there are only a few studies of eye tracking during simulated spacecraft docking, aiming at the differentiation between novices and experts[15] and the prediction of performance based on scanning paths.[25] Another research question has been the assessment of an operator's mental workload, fatigue, and attention allocation via eye tracking.[30] As Huemer et al.[15] stated, little is known about information acquisition strategies in spacecraft cockpits and extrapolation from studies of aircraft environments can be problematic, as task requirements differ substantially. Just as in aviation, accuracy and safety of a docking maneuver might be further improved through more efficient training based on optimizing visual scanning profiles. Eye tracking might also facilitate the prediction of operator performance and reliability.

The purpose of this study was to examine whether the analysis of eye movements during 6df docking training provides additional information regarding key factors that contribute to operator performance. Head-down bed rest studies provide a consistent and controlled space analog environment,[11] which allowed collecting feasibility data of eye movements during docking training. As one may infer visual attention allocation from fixations, we expected that certain measurable characteristics of visual fixation are linked to docking performance. This study was performed to identify such indicator characteristics in the context of the 6df tool. Specifically, we explored the relationship between the duration (total dwell time) and frequency (number of dwells) of visual attention to predefined regions of interest (ROI) with the accuracy score of the docking maneuver.

METHODS

Subjects

Subjects were part of the Artificial Gravity Bed Rest Study with the European Space Agency (AGBRESA), which took place in the :envihab facility of the Institute of Aerospace Medicine at the German Aerospace Center (DLR) in Cologne, Germany. After 15 d of familiarization and baseline measurements, subjects spent 60 d in 6° head-down tilt bed rest to simulate the effects of microgravity. After the bed rest phase, subjects stayed in the facility for a 14-d in-house follow-up in order to monitor and facilitate rehabilitation. Our substudy consisted of a training course on how to maneuver an object with 6 DoF using the 6df tool. Because half of the subjects were allocated to a stereoscopic version of the 6df tool that was tested as a learning aid,[27] we were able to collect eye tracking data from 12 of the 24 subjects in two study campaigns. Two subjects had to be excluded due to an insufficient amount of eye tracking data, leading to a final sample of 10 healthy subjects with normal or corrected-to-normal vision. Only glasses were allowed for vision correction. Exclusion criteria involved increased intraocular pressure, acuity correction via laser eye surgery, astigmatism (>3 dpt), myopia (>−6 dpt), hyperopia (>+5 dpt), color vision deficiency, or any other condition with substantial influence on vision. The three women and seven men were 23 to 54 yr old ($M = 33.40$, $SD = 9.01$). Subjects provided written informed consent and were granted monetary compensation for taking part in AGBRESA. The study was prospectively registered with the German register for clinical studies (www.drks.de) with the identifier DRKS00015677 and approved by the ethics committee of the medical association of North-Rhine in Duesseldorf, Germany.

Equipment

The 6df training tool has been described in detail in previous publications.[16,17] In short, 6df is a computer-based and self-sufficient learning program that simulates manual control of an object with 6 DoF or, more specifically, manual spacecraft docking to an abstract space object. Subjects are gradually instructed to control up to 6 DoF. Each task starts with an illustrated instruction text, sometimes including explanatory videos. After each task, performance feedback is given. This includes docking speed as well as yaw, pitch, and bank angles between spacecraft and space station. Based on safety ranges, these parameters are each transformed into a score ranging between zero and one. This performance rating resembles the Russian TORU docking system and is explained in detail by Johannes et al.[18] As an overall docking accuracy score we used the lowest parameter score, because the largest deviation from perfect alignment determines docking success, even if all other parameters should be satisfactory. In this study, all analyses are based on this overall docking accuracy score. A docking accuracy of 0.85 or higher resembles a sufficient performance (docking possible) and a score of 0.95 good (desirable) performance. A docking accuracy lower than 0.85 indicates that the docking maneuver failed. The 6df program adapts to the individual learning speed: if the operator achieves a docking accuracy of at least 0.95, the next (more difficult) task of the training program is presented. An accuracy between 0.85 and 0.95 leads to the repetition of the same task. If the score is below 0.85, not only the actual task has to be repeated, but also the one before. Twelve levels of ascending difficulty are included, which are labeled 1, 2, 5, 10, 15, 20, 21, 22, 23, 25, 50, and 60. Levels 1 (1 DoF) and 2 (2 DoF) familiarize the trainee with each hand control individually. Levels 5 and 10 introduce a predefined flight path (guiding rings) and trainees control 2 DoF, at first consecutively, then simultaneously. Finding the center line (direct line to approach the docking point from safety distance) and stabilization before starting the approach is trained in level 15. In level 20, the spacecraft is already perfectly oriented toward the station, but the trainee has to control the approach speed and first docking contact. Levels 21, 22, and 23 include a linear flight to the center line (flight path is indicated via guiding rings), stabilization, approach, and docking. The number of necessary maneuvers to achieve perfect orientation is increased stepwise. Level 25 is similar, but guiding rings are omitted and subjects have to maintain the proper flight path on their own. For levels 50 and 60 a full standard docking maneuver is required, which is composed of a curved flight around the station to the center line, stabilization, approach, and docking. The flight path is predefined in level 50, but not in the final level 60. Each level comprises several docking tasks of similar difficulty.

Because of the bed rest design, subjects remained supine in 6° head-down tilt without a pillow during all experimental sessions. A 1366 × 768-pixel screen was fixed above the subjects' heads (parallel to the bed) at approximately 60 cm distance to display the 6df docking program. The hand controls for docking (left one for translational movements, right one for rotation) were mounted on a vertically adjustable rack above the subjects' hips so that they could be used conveniently in a lying position with elbows resting on the bed. The remote eye tracking device was attached to the lower edge of the screen. We used Tobii 4C (Tobii Technology, Danderyd, Sweden), a lightweight commercially available eye tracking gear with a binocular sampling rate of 90 Hz. The device uses near-infrared light to create a corneal reflection whose relative position to the pupil's center is measured.[8] This technique allows unobtrusive data collection without head restriction. Light conditions in the laboratory were held constant.

Procedure

Each subject completed at least 20 6df training sessions of approximately 45 min. Sessions were scheduled on average twice a week during the study course, three sessions before bed rest and the remaining sessions during the 60-d bed rest period. Sessions were minimally 1 d and maximally 7 d apart, but the usual interval was every 3 to 4 d. Each single docking task comprised up to 12 min without instructions and feedback, depending on the level (some tasks can be solved faster than others) and the spacecraft's speed (subjects have to follow a speed limit, but they are free to move slower). Therefore, the number of tasks in each session also varied. At the beginning of every session, Tobii's standard 5-point calibration was conducted. In the middle and in each corner of the black screen a red dot appeared and had to be fixated until it vanished. Subsequently, the 6df training started and subjects completed the tasks at their own pace.

Statistical Analysis

Although there are exceptions to this rule, it is generally assumed that the point of attention can be inferred from the position of a fixation.[14] During a fixation, the gaze rests relatively still at a certain point of attention for a short period of time[8] and visual information is obtained.[28,29] A temporal threshold of 100 ms is oftentimes recommended to define a single fixation.[22,23] The Tobii 4C stores 90 gaze coordinates (samples) on the screen per second. Accordingly, we defined fixations to consist of a minimum of nine subsequent gaze samples within an ROI to meet the fixation duration threshold of 100 ms. Four nonoverlapping ROI were specified on the 6df screen as shown in **Fig. 1**: task overview (depicts the task from above), vizor (pentagon that should always be aimed toward the docking cross of the station), standard instruments (speed and distance in relation to the station), and auxiliary instruments. The auxiliary instruments display gives the same information about current speed and remaining distance to the docking point, but is more salient and color-coded: if a subject is too fast or too slow, it will turn from green to yellow and eventually to red. Each ROI is a fixed zone on the screen and has been defined along the outlines of the respective 6df element (vizor, instruments, task overview). As a measure of the amount of attention that each ROI attracted, we computed total dwell times (reported in milliseconds) as the sum of the durations of all fixations recorded within each ROI per task.[14] To track how often information was retrieved from an ROI, we additionally looked at the number of dwells by counting visits to each ROI.

Data processing and statistical analyses were carried out using SPSS Statistics 21 (IBM, Armonk, NY, USA) and R 3.5.3 (The R Foundation, Vienna, Austria)/RStudio (RStudio Inc., Boston, MA, USA). All tests were two-tailed and the level of significance was set to $\alpha = 0.05$. We excluded the first five levels (1, 2, 5, 10, 15) from analysis because these tasks are for familiarization with 6 DoF and did not involve final approach and docking contact. Additionally, tasks that ended early due to collision with guiding rings that mark the ideal flight path were excluded. Hence, visual task structure and the docking accuracy score of all remaining tasks were comparable. Outlier tasks with a docking performance more than two standard deviations below the mean docking accuracy score of all subjects (<0.95) were dropped. Because subjects already received some training in earlier levels and the task difficulty increases throughout the learning program, completely insufficient performance was rare and mostly associated with ring collisions. For a safety relevant operational task like docking, even small differences within the upper bandwidth of performance are crucial. Possible relationships were sought between dwell times and numbers of dwells within the defined ROIs and the 6df docking accuracy score. We computed linear mixed-effects models (LME) with docking accuracy as a dependent variable, a random intercept for subjects, and as fixed factor dwell times on each ROI and number of dwells to each ROI, respectively. Variance components was chosen as covariance structure. Residuals could be accepted as sufficiently normally distributed by inspection of Q-Q plots and histograms.

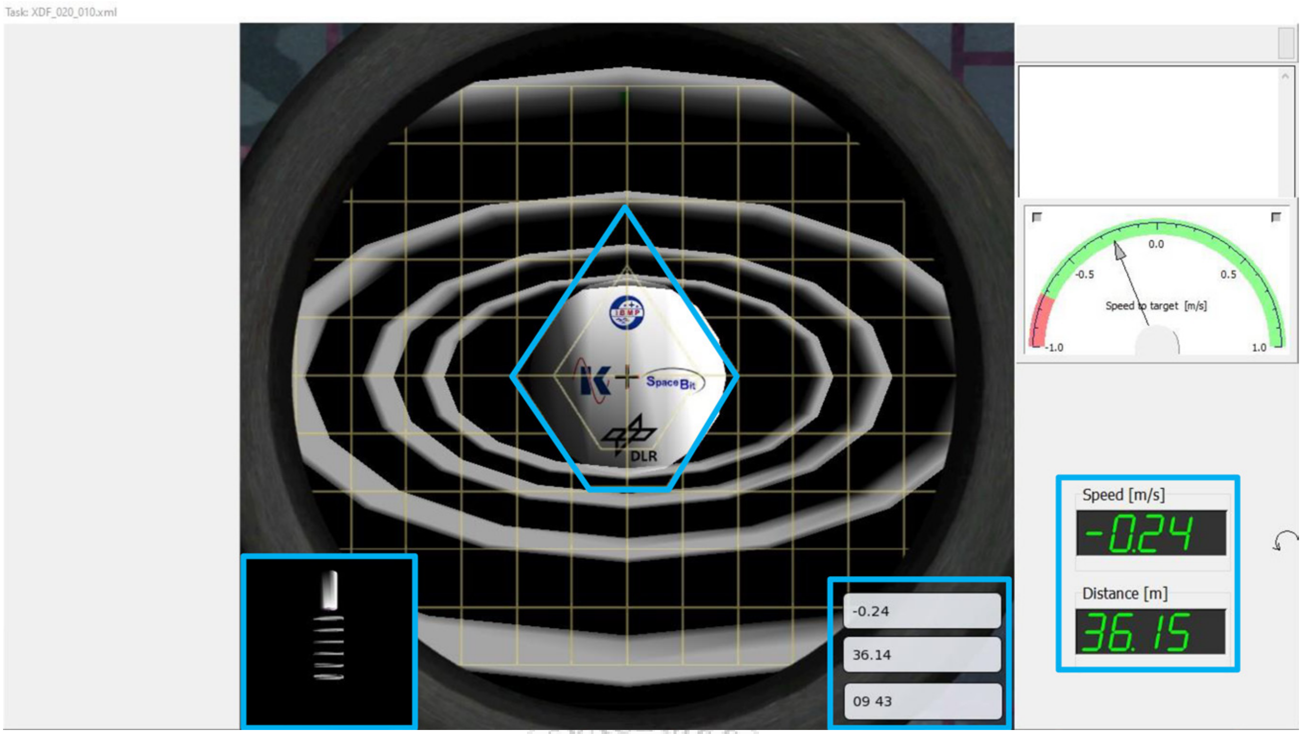


Fig. 1. Docking task with four ROI outlined: task overview (bottom left, sized 3.5% of the screen), vizor (central pentagon, 6.6%), standard instruments (bottom center right, 3.2%), and auxiliary instruments (bottom right, 4.8%).

RESULTS

Taken together, subjects completed 493 docking tasks with eye tracking, with 412 meeting our inclusion criteria. As depicted in **Fig. 2A**, comparing the four ROI, most of the total dwell time per task was dedicated to the vizor (M = 22,355.09 ms, SD = 9018.27). There was one exception for level 20, which is the first introduction to actual docking contact. In level 20, the spacecraft was already centered in front of the station and, therefore, only docking speed had to be controlled. Task overview (M = 128.09 ms, SD = 160.87) and standard instruments (M = 724.59 ms, SD = 1293.99) were rarely attended. The auxiliary instruments display instead attracted substantial total dwell time (M = 3323.24 ms, SD = 1902.50). The number of dwells during each task followed a similar pattern (**Fig. 2B**): subjects most often visited the vizor ROI (M = 185.20, SD = 115.85), but barely the task overview picture (M = 6.53, SD = 7.61). Also, they looked more often to the auxiliary instruments (M = 50.53, SD = 26.01) than to the standard instruments (M = 23.54, SD = 25.57).

Looking at docking performance, we observed a significant effect of dwell time on standard instruments [$F(1, 317.82) = 4.66$, $P = 0.03$] as well as on auxiliary instruments [$F(1, 395.17) = 14.05$, $P < 0.001$] on the accuracy score. The more

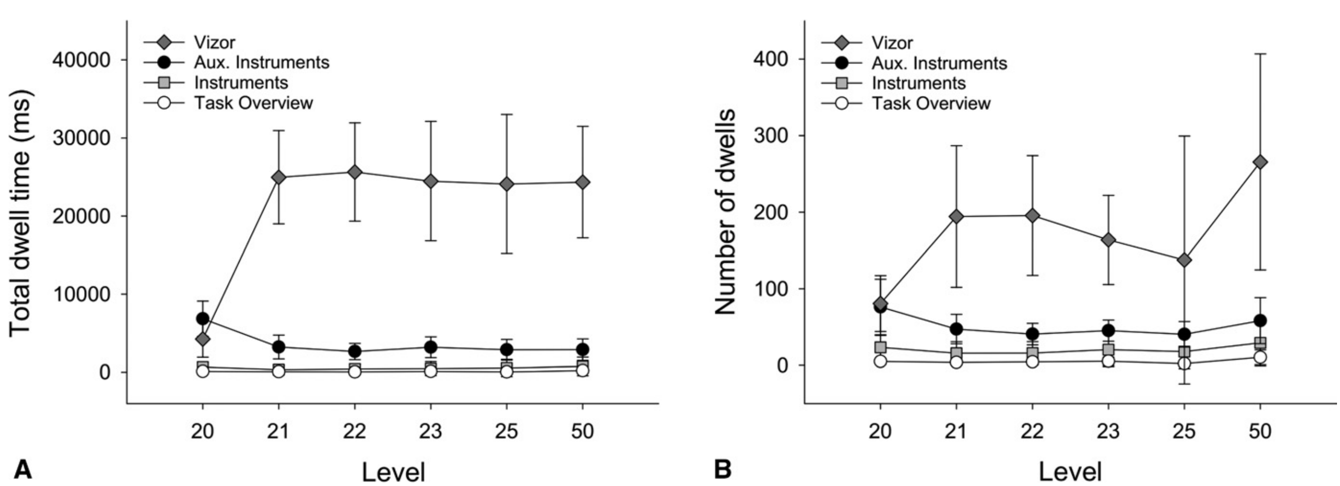


Fig. 2. A) Mean total dwell time and B) mean number of dwells per task on defined ROIs by levels of the training program. Error bars indicate the standard deviation.

dwell time subjects directed to information about actual speed and distance, the higher was their docking accuracy in the learning program. Dwell time on task overview had no significant impact on docking performance [$F(1, 226.04) = 0.42$, $P = 0.52$]. Dwell time on the vizor was significantly negatively related to docking accuracy [$F(1, 407.21) = 5.61$, $P = 0.02$]. We suspected this effect to be caused mainly by level 20, where subjects only had to control approach speed and barely observed the vizor. Indeed, when level 20 was omitted and the analysis repeated, the effect of vizor dwell time on accuracy had a positive direction, but was not significant any more [$F(1, 283.89) = 3.59$, $P = 0.06$]. Considering the number of dwells rather than the amount of time each ROI was observed, docking accuracy was significantly higher the more frequently subjects looked at the auxiliary instruments [$F(1, 372.45) = 8.26$, $P < 0.01$]. We did not observe significant effects of number of dwells on standard instruments [$F(1, 271.68) = 1.98$, $P = 0.16$], vizor [$F(1, 406.85) = 0.29$, $P = 0.59$], or task overview [$F(1, 226.52) = 2.28$, $P = 0.13$].

DISCUSSION

We collected eye tracking data during 6df docking training in a space mission analog environment to explore if visual attention provides additional information about operator performance. Eye tracking technology, which is already appreciated in the aviation domain,[26,35] proved to be feasible and unobtrusive in the spacecraft docking context of our study. In airplane cockpits, areas that are critically important for a given task, or areas that frequently change, attract the most total dwell time.[10] Similarly, during 6df docking, as the vizor's cross should always be directed at the docking point, subjects focused their visual attention mostly on the vizor, and therefore on the space station. This observation is consistent with findings of Tian and colleagues,[30] who reported that about 80% of their subjects' fixation times were directed toward the space station and about 20% to numerical displays. Subjects substantially favored auxiliary instruments over the standard ones, which indicates that learners were making use of the additional information given about actual and required speed. The task overview was generally rarely attended and probably deemed to be expendable.

Subjects who devoted more total dwell time to both instrument displays achieved higher docking accuracy scores. Also, a higher number of dwells to the auxiliary instruments was related to better performance. Frequently and thoroughly checking the concordance of speed and remaining distance seems to be essential for successful docking. In this context, the capability of swiftly detecting deviations from optimal speed might point to differences in situation awareness. This relates to previous literature that emphasizes high situation awareness based on visual attention as an important safety factor in aviation.[31,33,34] Docking accuracy was not related to total dwell time or number of dwells regarding the task overview. This ROI was in general rarely observed by the subjects. The number of dwells to the vizor had no effect on performance; however, total dwell time on the vizor was negatively associated with docking accuracy. The association was explained by level 20, which only demands controlling approach speed. While most of the dwell time and frequency was devoted to the vizor, this was not significantly related to docking accuracy.

Although successful task completion obviously required much dwell time and many dwells on the vizor, gazes to the instruments were the best indicator of performance. Our analyses regarding level 20 even indicate that spending too much time on the station (presumably at the expense of retrieving instrument information) could even hamper performance in some instances. This might be in line with previous observations that experts had better defined scan patterns and concentrated more on the most important instruments.[4,20] Importantly, duration and frequency of visual attention to the instruments were positively related to docking performance. Studies in the aviation domain have frequently linked gaze behavior to piloting performance[4,20,35] and our results confirm the potential of eye movement analysis for spacecraft docking.

Further studies are needed to investigate if—and to what extent—eye tracking data as total dwell time and number of dwells can actually explain performance and predict operator reliability. As a next step, these results could also be useful to improve the 6df learning program, for example, by providing tailored eye tracking-based feedback to remind the trainee to regularly check the instruments during training. This could be implemented with a replay of the docking task that incorporates the operator's scan path over time. In the Soyuz environment as well, instrument information is placed at the margins of the screen. Therefore, if one focuses the vizor at the station, attention must be redirected actively to the instruments. To establish an instrument-checking routine, the display could flash if it is not regarded by the trainee. Understanding how scanning behavior differs between novices and experts would be interesting, as learners could be alerted when their strategy deviates from an optimal expert scan pattern.[15,25] Our study did not include experts, but studies in aircraft cockpits suggest systematic differences from novices in duration and frequency of dwells.[2,10,20] Optimal gaze behavior in a spacecraft could include allocating much attention to the instruments by checking them frequently. However, experts might be able to process information more effectively and therefore have shorter dwells.

Which among the growing number of eye tracking-derived measures are most appropriate for a certain research question is still controversial.[21] In an aviation context, total dwell time and number of dwells have been frequently used to assess visual attention allocation.[10] We aimed at collecting feasibility data to explore the benefit of eye tracking during docking training; therefore, we chose these standard measures. Next, the features of optimal gaze behavior could be analyzed in more detail, e.g., by assessing transition patterns or single fixation durations. Analysis of scanning entropy would allow for the evaluation of performance effectiveness. Unfortunately, the low sampling rate of the Tobii 4C did not allow for the computation of saccadic metrics. Eye tracking has also been used as an indicator for workload,[1,24,26] which is interesting especially for operational tasks, because high workload can facilitate performance decrements and human error.[7,32] But unlike Tian et al.,[30] we were not able to identify pupil dilation as useful to monitor mental workload in the context of docking. Pupil dilation is very sensitive to even small differences in lighting conditions, but in the operational context of docking, sufficient control of the screen as well as environmental brightness and contrasts seemed unrealistic. Because our results were strongly biased by screen luminosity, we did not report pupillometry results for this study and suppose that a more robust eye tracking-based measure might be more promising for the future. A limitation of the eye tracking device used in this study is the lack of verified information on its accuracy and precision. However, Hild et al.[13] evaluated the accuracy of the device in 12 subjects, resulting in a satisfactory accuracy of 0.96° of visual angle. Apart from these methodological considerations, there are significant advantages of using eye tracking to examine the operator's information processing. Eye tracking devices are relatively inexpensive, lightweight, and easy to use. Attention can be assessed objectively and continuously. Importantly, data acquisition is unobtrusive and does not interrupt the operator, which is crucial for the implementation in operational tasks. A limitation of this study, however, is the small sample size. Despite this limitation, eye tracking proved suitable in examining information processing during docking training and deserves to be tested in larger samples.

We conclude that sampling with a small commercial eye tracking device provided valuable insight into information processing during docking with the 6df learning program.[16,17] Visual attention was related to performance in a simulated manual docking maneuver. Specifically, frequency and duration of processing speed and distance information were both associated with higher docking accuracy. Our results are a first step to identify eye-based indicators that can possibly be employed to assess interindividual differences in skill, but also intraindividual fluctuations in manual docking performance. Additionally, performance-associated scanning behavior could contribute to the improvement of docking training, for example, by giving trainees tailored feedback. This might be especially promising for future long duration missions, which require autonomous performance monitoring and training of operational skills.

ACKNOWLEDGMENTS

Financial Disclosure Statement: The AGBRESA study was funded by the German Aerospace Center, the European Space Agency (contract number 4000113871/15/NL/PG), and the National Aeronautics and Space Administration (contract number 80JSC018P0078). This specific work was further supported by the German Aerospace Center (DLR), acting as space agency [DARA-Grants 50WB9128, 50WB93401, 50WB93401-ZA; DLR-Grants 50WB 96220, 50WP0306, 50WP0501, 50WP0602, 50WP0603, 50WP1104, 50WP1304, 50WP1609].

Authors and Affiliation: Sarah Piechowski, M.Sc., Bernd Johannes, Dr., Willi Pustowalow, B.Eng., Michael Arz, Edwin Mulder, Dr., Jens Jordan, Prof. Dr., and Jörn Rittweger, Prof. Dr., German Aerospace Centre (DLR) Institute of Aerospace Medicine, Cologne, North Rhine-Westphalia, Germany; Oliver T. Wolf, Prof. Dr., Department of Cognitive Psychology, Ruhr University Bochum, Bochum, Germany; and Jens Jordan, Chair of Aerospace Medicine, University of Cologne, and Jörn Rittweger, Department of Pediatrics and Adolescent Medicine, University Hospital Cologne, Cologne, Germany.

REFERENCES

1. Beatty J, Lucero-Wagoner B. The pupillary system. In: Cacioppo JT, Tassinary LG, Berntson GG, editors. Handbook of psychophysiology, 2nd ed. Cambridge: University Press; 2000:142–162.
2. Bellenkes AH, Wickens CD, Kramer AF. Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. Aviat Space Environ Med. 1997; 68(7):569–579.
3. Brody AR. Spacecraft flight simulation: a human factors investigation into the man-machine interface between an astronaut and a spacecraft performing docking maneuvers and other proximity operations. Moffett Field (CA): NASA/Ames Research Center; 1988. Report No.: NASA-CR-177502.
4. Chuang LL, Nieuwenhuizen FM, Bülthoff HH. A fixed-based flight simulator study: the interdependence of flight control performance and gaze efficiency. International Conference on Engineering Psychology and Cognitive Ergonomics. New York: Springer Nature; 2013.
5. Clarke AH, Ditterich J, Drüen K, Schönfeld U, Steineke C. Using high frame rate CMOS sensors for three-dimensional eye tracking. Behav Res Methods Instrum Comput. 2002; 34(4):549–560.
6. Clarke AH, Grigull J, Mueller R, Scherer H. The three-dimensional vestibulo-ocular reflex during prolonged microgravity. Exp Brain Res. 2000; 134(3):322–334.
7. Di Nocera F, Camilli M, Terenzi M. A random glance at the flight deck: pilots' scanning strategies and the real-time assessment of mental workload. J Cogn Eng Decis Mak. 2007; 1(3):271–285.
8. Duchowski AT. Eye tracking methodology, 3rd ed. London: Springer; 2017.
9. Ellis SR. Collision in space. Ergon Des. 2000; 8(1):4–9.
10. Glaholt MG. Eye tracking in the cockpit: a review of the relationships between eye movements and the aviator's cognitive state. Toronto: Defence Research and Development Canada; 2014. Report No.: DRDC-RDDC-2014-R153.
11. Hargens AR, Vico L. Long-duration bed rest as an analog to microgravity. J Appl Physiol. 2016; 120(8):891–903.
12. Haslbeck A, Schubert E, Gontar P, Bengler K. The relationship between pilots' manual flying skills and their visual behavior: a flight simulator study using eye tracking. In: Landry SJ, Salvendy G, Karwowski W, editors. Advances in human factors and ergonomics. Advances in human aspects of aviation. Boca Raton: CRC Press; 2012:561–568.
13. Hild J, Peinsipp-Byma E, Voit M, Beyerer J. Suggesting gaze-based selection for surveillance applications. 16th IEEE International Conference on Advanced Video and Signal Based Surveillance. New York: IEEE; 2019.
14. Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J. Eye tracking: a comprehensive guide to methods and measures. Oxford: University Press; 2011.
15. Huemer VA, Hayashi M, Renema F, Elkins S, McCandless JW, McCann RS. Characterizing scan patterns in a spacecraft cockpit simulator: expert vs. novice performance. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Thousand Oaks (CA): Sage Journals; 2005.
16. Johannes B, Bronnikov S, Bubeev Y, Arz M, Piechowski S, Rittweger J. Individual learning curves in manual control of six degrees of freedom. Int J Appl Psychol. 2019; 9(2):47–51.
17. Johannes B, Bronnikov S, Bubeev Y, Dudukin A, Hoermann H-J, et al. A tool to facilitate learning in a complex manual control task. Int J Appl Psychol. 2017; 7(4):79–85.

18. Johannes B, Salnitski V, Dudukin A, Shevchenko L, Bronnikov S. Performance assessment in the PILOT Experiment on board space stations Mir and ISS. Aerosp Med Hum Perform. 2016; 87(6):534–544.
19. Just MA, Carpenter PA. A theory of reading: from eye fixations to comprehension. Psychol Rev. 1980; 87(4):329–354.
20. Kasarskis P, Stehwien J, Hickox J, Aretz A, Wickens C. Comparison of expert and novice scan behaviors during VFR flight. Proceedings of the 11th International Symposium on Aviation Psychology; March 5–8, 2001; Columbus, OH, USA. Columbus (OH): Department of Aviaiton and Aeronautical Engineering, Ohio State University; 2001.
21. Lai M-L, Tsai M-J, Yang F-Y, Hsu C-Y, Liu T-C, et al. A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ Res Rev. 2013; 10:90–115.
22. Majaranta P, Bulling A. Eye tracking and eye-based human–computer interaction. In: Fairclough SH, Gilleade K, editors. Advances in physiological computing. London: Springer; 2014:39–65.
23. Manor BR, Gordon E. Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. J Neurosci Methods. 2003; 128 (1–2):85–93.
24. Marquart G, Cabrall C, de Winter J. Review of eye-related measures of drivers' mental workload. Procedia Manuf. 2015; 3:2854–2861.
25. Matessa M, Remington R. Eye movements in human performance modeling of space shuttle operations. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Los Angeles (CA): Sage Publications; 2005.
26. Peißl S, Wickens CD, Baruah R. Eye-tracking measures in aviation: a selective literature review. International Journal of Aerospace Psychology. 2018; 28(3–4):98–112.
27. Piechowski S, Pustowalow W, Arz M, Rittweger J, Mulder E, et al. Virtual reality as training aid for manual spacecraft docking. Acta Astronaut. 2020; 177:731–736.
28. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull. 1998; 124(3):372–422.
29. Rayner K. Eye movements and attention in reading, scene perception, and visual search. Q J Exp Psychol (Hove). 2009; 62(8): 1457–1506.
30. Tian Y, Zhang S, Wang C, Yan Q, Chen S. Eye tracking for assessment of mental workload and evaluation of RVD interface. International Conference on Man-Machine-Environment System Engineering. New York: Springer; 2018.
31. van de Merwe K, van Dijk H, Zon R. Eye movements as an indicator of situation awareness in a flight simulator experiment. Int J Aviat Psychol. 2012; 22(1):78–95.
32. Wickens CD. Mental workload: assessment, prediction and consequences. International Symposium on Human Mental Workload: Models and Applications. New York: Springer; 2017.
33. Wickens CD, McCarley JS, Alexander AL, Thomas LC, Ambinder M, Zheng S. Attention-situation awareness (A-SA) model of pilot error. In: Foyle DC, Hooey BL, editors. Human performance modeling in aviation. Boca Raton: CRC Press; 2008:213–239.
34. Yu CS, Wang EMY, Li WC, Braithwaite G. Pilots' visual scan patterns and situation awareness in flight operations. Aviat Space Environ Med. 2014; 85(7):708–714.
35. Ziv G. Gaze behavior and visual attention: a review of eye tracking studies in aviation. Int J Aviat Psychol. 2016; 26(3–4):75–104.

4 Study III: Manual Docking Performance after Sleep Deprivation

Sarah Piechowski[1], Lennard J. Kalkoffen[1], Sibylle Benderoth[1], Oliver T. Wolf[2], Jörn Rittweger[1,3], Daniel Aeschbach[1,4*], and Christian Mühl[1*]

Effects of Total Sleep Deprivation on Performance in a Manual Spacecraft Docking Task

Submitted to *npj Microgravity*.

[1] Institute of Aerospace Medicine, German Aerospace Center (DLR), Cologne, Germany

[2] Department of Cognitive Psychology, Faculty of Psychology, Ruhr University Bochum, Germany

[3] Department of Pediatrics and Adolescent Medicine, University Hospital Cologne, Germany

[4] Institute of Experimental Epileptology and Cognition Research, University of Bonn Medical Center, Germany

* The two last authors contributed equally to this work.

Abstract

Sleep deprivation and circadian rhythm disruptions are highly prevalent in shift workers, and also among astronauts. Resulting sleepiness can reduce cognitive performance, lead to catastrophic occupational events, and jeopardize space missions. We investigated whether 24 hours of total sleep deprivation would affect performance not only in the *Psychomotor Vigilance Task* (PVT), but also in a complex operational task, i.e. simulated manual spacecraft docking. Sixty-two healthy participants completed the manual docking simulation *6df* and the PVT once after a night of total sleep deprivation and once after eight hours of scheduled sleep in counterbalanced order. We assessed the impact of sleep deprivation on docking as well as PVT performance and investigated if sustained attention is an essential component of operational performance after sleep loss. The results showed that docking accuracy decreased significantly after sleep deprivation in comparison to the control condition, but only at difficult task levels. PVT performance deteriorated under sleep deprivation. Participants with larger impairments in PVT response speed after sleep deprivation also showed larger impairments in docking accuracy. As a conclusion, sleep deprivation led to impaired *6df* performance, which was partly explained by impairments in sustained attention. Elevated motivation levels due to the novelty and attractiveness of the task may have helped participants to compensate for the effects of sleepiness at easier task levels. Continued testing of manual docking skills could be a useful tool both to detect sleep loss related impairments and assess astronauts' readiness for duty during long duration missions.

Introduction

Insufficient sleep duration and quality are common in many occupational contexts, especially when extended duties or shift work are required (Åkerstedt, 2003; Caruso, 2014; Philibert, 2005). Sleep deprivation affects the performance of people working in health care (Barger et al., 2006), transportation (Åkerstedt, 2000; Dorrian et al., 2007; Sallinen et al., 2020; Vejvoda et al., 2014), or military, but also human space flight. Although the use of sleep-promoting drugs is common, astronauts sleep only about six hours per night, resulting in chronic sleep deprivation over time (Barger et al., 2014; Dijk et al., 2001; Jones et al., 2022). There are numerous reasons why sleep is particularly "lost in space", including environmental (e.g. noise, altered light-dark cycle, hypoxia, and hypercapnia) and psychological (e.g. isolation, confinement, and stress) factors alike (Basner & Dinges, 2014). Moreover, mission demands can require irregular sleep-wake-cycles of astronauts, such as slam-shifts, resulting in a high prevalence of circadian misalignment (Flynn-Evans, Barger, et al., 2016). Sleepiness – in the occupational field often referred to as fatigue – is an important risk factor in work environments that require a continuously high level of performance to avoid potentially catastrophic outcomes (Porcu et al., 1998). In many operational contexts, sleepiness has been shown to impair performance and facilitate human error, increasing the risk of accidents (Balkin et al., 2004; Bendak & Rashid, 2020; Dinges, 1995). Sleep-deprived pilots display degradations in psychomotor control, problem solving, and attention to flight instruments. Moreover, short involuntary lapses into sleep have been the cause of aviation accidents (Caldwell, 2012).

In astronauts, a sleep duration of less than six hours was correlated with impairments in sustained attention and mood on the following day (Jones et al., 2022). One example of the sleep-related risks in space is the crash of the Progress space shuttle with Mir space station in 1997. The life-threatening accident occurred during a manual docking maneuver and fatigue was made out as one of the contributing factors (Ellis, 2000). Manual docking of a spacecraft is a mission-critical operational task in space and requires highly trained cognitive skills as well as motor skills in the control of an object with six degrees of freedom (DoF, Fig. 2). Because docking performance deteriorates fast without continuous training (Salnitski et al., 2001), the *6df* tool has been developed as an autonomous training and maintenance program for controlling six DoF (Johannes et al., 2019; Johannes et al., 2017). Basner and colleagues have recently studied the relationship between cognitive performance in the *Cognition* test battery and operational performance in the *6df* docking task (Basner et al., 2020). The strongest association was observed with performance in the *Digit-Symbol Substitution Task* (DSST), a measure of processing speed, visual tracking, and working memory. These domains are not only crucial to process instrument information during manual docking, but are also known to be sensitive to sleep loss (Goel et al., 2009; Lim & Dinges, 2010). Furthermore, there was a positive correlation between manual docking performance and sustained attention as measured by accuracy in the *Psychomotor Vigilance Task* (PVT). In sleep deprivation research, a decrease in sustained attention has been one of the most reliable effects (Basner & Dinges, 2011). Sleep-deprived individuals are consistently slower in responding to stimuli and more prone to errors of omission and commission (Lim & Dinges, 2008), therefore, the PVT is considered to be one of the most sensitive measures regarding sleep restriction (Balkin et al., 2004). Alertness is a critical factor for space mission safety (Mallis & DeRoshia, 2005), as even small decrements in the ability to timely react to relevant stimuli can compromise task success.

Currently, it is still unclear whether or to what extent sleepiness-related decrements in experimental tests of cognitive and psychomotor performance relate to impairments in more complex operational tasks (Flynn-Evans, Gregory, et al., 2016). The higher demand of complex tasks might as well motivate individuals to apply additional effort to compensate for their sleepiness (Harrison & Horne, 2000). Strangman et al. (2005) for example detected compensatory cerebral responses to sleep deprivation, but no performance impairment in a simulated orbital docking task. Wong et al. (2020) attributed the lack of performance decrements during the course of 28 h of sleep deprivation to the novelty and motivational character of their grappling and docking task. However, there are few studies and only with a small number of participants that looked at the effects of sleep deprivation on space flight-relevant operational performance. Therefore, our study aimed at characterizing the influence of one night of total sleep deprivation (~24 h awake) on manual docking performance. Specifically, we hypothesized that docking accuracy as well as the progression through different levels of task difficulty will deteriorate after sleep deprivation in comparison to performance after normal sleep. Additionally, we assessed if impairment in sustained attention (as measured by the PVT) is a relevant factor to explain docking performance under sleep deprivation.

Methods

Participants

Sixty-six healthy individuals participated in our experiment. Two participants were excluded because they attended only one of the test sessions and two because of technical problems. The final sample consisted of 62 participants, 28 women and 34 men, aged between 18 and 39 (M = 24.84, SD = 4.69) years. Most participants were students recruited via university job web portals. Ahead of the study, participants completed a medical examination as well as questionnaires to rule out presence of sleep problems (*STOP-Bang* questionnaire, *Epworth Sleepiness Scale,* ESS), extreme personality traits (*Freiburger Persönlichkeitsinventar,* FPI) and depression (*Beck Depression Inventory,* BDI). Other exclusion criteria were pregnancy, smoking, drug use, relevant medication, and body mass index above 30 kg/m^2.

Study design

The docking experiment was part of a larger laboratory study that was conducted in the Institute of Aerospace Medicine of the German Aerospace Center (DLR) in Cologne, Germany, from 2019 to 2020. The study followed a randomized counter-balanced cross-over design with a control condition and a sleep deprivation condition. At least one week before arrival at the *Simulation Facility for Occupational Medicine Research* (AMSAN), participants were required to follow a regular sleep protocol (23:00 – 07:00) in order to avoid sleep debt and circadian misalignment. Compliance was ensured via wrist actimetry (Philips Actiwatch Spectrum) and sleep diaries. Caffeine consumption was not allowed in the week before and during the study. Participants were randomly assigned to one of two groups with counterbalanced order of experimental conditions (control, sleep deprivation). Participants spent five days and four nights in the laboratory in three-person teams. An illustration of the study protocol is provided in Fig. 1. Test sessions took place on day three and five. In the control condition, the test session was scheduled from 13:00 to 15:00 and was preceded by an 8-h sleep episode (23:00-07:00). In the sleep deprivation condition the test session took place between 07:00 and 09:00 following approximately 24 h of continuous, monitored wakefulness. Apart from the manual docking simulation task reported here, test sessions consisted of a synthetic operational group task as well as a cognitive test battery including sustained attention, working memory, and decision-making tasks. During the scheduled wake episodes illuminance was maintained at ~100 lux at the horizontal angle of gaze. The study was approved by the ethics committee of the medical association North-Rhine in Dusseldorf, Germany, and participants provided written informed consent prior to their participation.

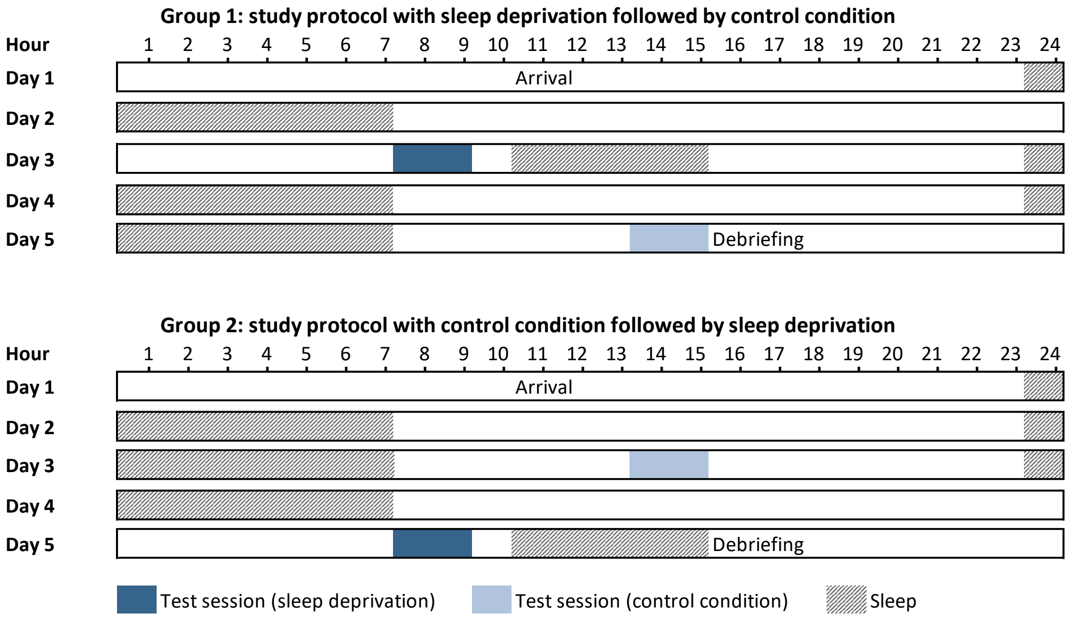


Fig. 1. Study protocol with randomized counterbalanced order of a control and a sleep deprivation condition.

Docking Simulation

The *6df* tool is a computer-based self-learning program that simulates the control of an object with six degrees of freedom (Fig. 2), in this case the manual docking of a spacecraft to a space station (Johannes et al., 2017). Flight dynamics and controller responsiveness are based on the Russian docking training system TORU (Teleoperatiya Ruchnogo Upravleniya – teleoperated manual control) and the actual Soyuz spacecraft. However, the tool is not designed to be a realistic Soyuz simulation, but to teach the principles of the control of any object in space in generic fashion (Fig. 2). The *6df* software was developed by SpaceBit GmbH (Eberswalde, Germany) and hand controls were produced by Koralewski Industrie-Elektronik oHG (Hambühren, Germany). The left hand control operates the translational degrees of freedom, the right hand control all rotational degrees of freedom. In this experiment, we used a newly developed adaptive version of *6df*. It consisted of eleven docking task designs from five difficulty levels ranging from simple control of one degree of freedom up to a standard docking maneuver including a curved flight-around, stabilization, approach, and finally docking.

Participants had no previous experience or training with the docking simulation. Before the first test session, they received a written instruction with general information regarding the *6df* tool, task design, and performance feedback. Each session started with a short instructional film on the use of the hand controls. Additionally, before each docking trial, an illustrated text with specific instructions was presented. After completion of a trial, performance feedback following TORU methodology was given (Johannes et al., 2016). This included single parameters such as forward speed, pitch, bank, and yaw, as well as an aggregated general performance measure (accuracy), with zero being the worst and one the best possible accuracy. Every session started with the same trial on level 4 out of 5; a docking maneuver excluding the curved flight-around to reach the stabilization point. The procedure was adaptive to the performance of the participant and allowed for the application of *6df* without previous training. A docking maneuver was deemed successful if the accuracy was at least .85, and of a good quality if at least .95. Therefore, if the accuracy score was below .85, an easier trial was presented next. For accuracy scores between .85 and .95, the trial was repeated and for scores ≥ .95, a more difficult trial was presented. Each single docking trial lasted up to ten minutes without instructions and feedback, depending on level and participant's speed. After 35 minutes, the session was terminated following completion of the current trial. Therefore, number and level of completed docking trials varied between participants.

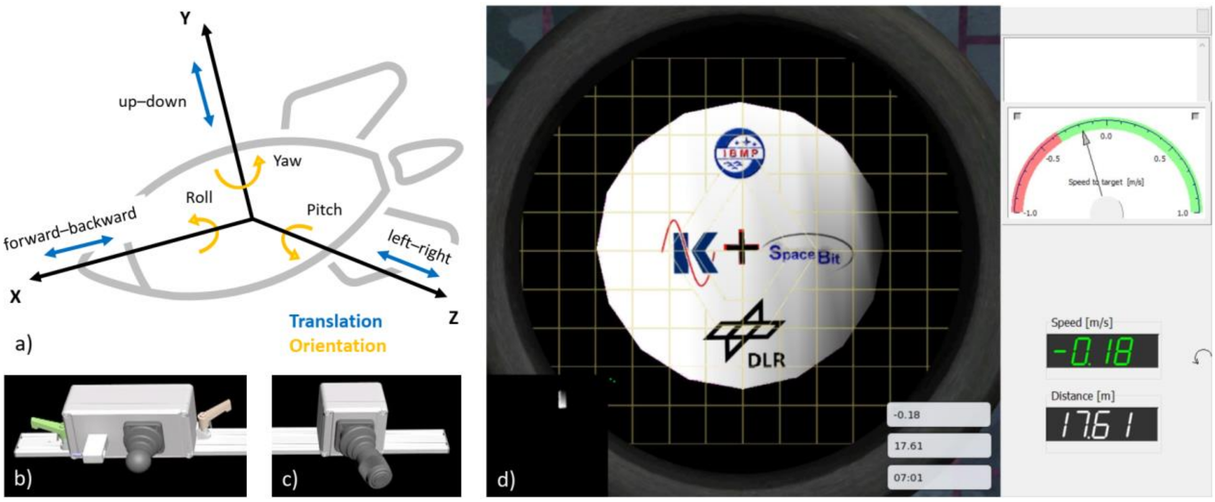


Fig. 2. Six degrees of freedom in spacecraft control (a). Translation is controlled with the left hand control (b), orientation with the right one (c). Screenshot of a level 4 *6df* docking trial showing the view from the spacecraft on the white space station (d). The black cross represents the target/docking port.

Psychomotor Vigilance Task (PVT)

Participants performed a 10-min version of the PVT (Dinges & Powell, 1985; Elmenhorst et al., 2018). The task was to react to the appearance of a millisecond counter as fast as possible by pressing a button with the thumb of the dominant hand (handedness, percent of participants: 6.45% left, 93.55% right). Upon response, the counter stopped and presented the reaction time as feedback for 1 s. Without a response, the counter timed out after 10 s. The inter-stimulus-interval varied pseudo-randomly between 2-10 s. Only responses greater than 130 ms were considered valid and lapses were defined as a response time exceeding 500 ms. Due to its high sensitivity to sleep loss (Basner & Dinges, 2011), the averaged response speed per session, i.e. reciprocal response time (1/RT), and the number of lapses were chosen as outcome.

Statistical Analysis

R 4.0.2/R Studio and IBM SPSS Statistics 26 were used for data processing and analysis. All tests were carried out two-sided and with a significance level of α = .05. When applicable, normal distribution of variables or residuals was verified via visual inspection of Q-Q plots and histograms. A total number of 856 docking trials were absolved. We excluded 56 trials that were discontinued because the participant maneuvered out of the station's reach (distance > 200m), resulting in a final sample of 800 valid docking trials.

For mixed models, conditional and marginal pseudo-R^2 were computed with the MuMIn package for R (Barton, 2020) as an effect size estimate proposed by Nakagawa et al. (2017). Conditional R^2 (R^2_{c}) is interpreted as the variance explained by the entire model including random effects, whereas marginal R^2 (R^2_{m}) represents the variance explained by fixed effects only. For nonparametric analyses, $r = Z \div \sqrt{N}$ was computed as effect size (Rosenthal, 1991).

To check if the sleep deprivation manipulation actually induced sleepiness, we compared self-reported subjective sleepiness between conditions using a linear mixed model. For this purpose, participants completed the *Karolinska Sleepiness Scale* (KSS) (Åkerstedt & Gillberg, 1990; Shahid et al., 2011) at the beginning of each test session. The KSS is a single-item questionnaire that indicates situational sleepiness on a 9-point Likert scale, ranging from 1 (extremely alert) to 9 (extremely sleepy). The effect of sleep deprivation on PVT response speed and number of lapses (following log transformation) was also analyzed using linear mixed models.

To test the effect of sleep deprivation on docking accuracy, we computed a linear mixed model with accuracy as dependent variable and participants as random intercept. As fixed effects we included *6df* level of difficulty, session (first or second) and condition (control or sleep deprivation). For docking accuracy scores an inverse log transformation was used to achieve normal distribution of residuals. Levels 1 and 2 as well as 4 and 5 were merged for this analysis because of similar task demands and the small number of observations in level 1 and 5. In a second step, we added the interaction between condition and the susceptibility to sleep deprivation as measured by PVT response speed and number of lapses to the model. Susceptibility scores were obtained by subtracting control condition (CC) PVT performance from sleep deprivation (SDC) PVT performance. Worse performance after sleep deprivation relative to the control condition is indicated by negative SDC-CC difference values in the case of response speed, and positive SDC-CC difference values in the case of lapses.

Next to the accuracy of the docking maneuvers, we also considered participants' progress through the levels of the adaptive *6df* program during each session. For this purpose, we looked at the level of the highest successfully completed trial (accuracy ≥ .85) in each session. The highest achieved level could vary between 1 (if the only trial a participant was able to solve with sufficient accuracy was on level 1) and 5 (if a participant advanced from level 4 and achieved sufficient accuracy also in a level 5 trial). Because data was not normally distributed, we chose Wilcoxon signed-rank test to compare highest levels between control and sleep deprivation condition as well as between first and second session.

Results

After sleep deprivation, participants reported to feel significantly sleepier (KSS: M = 7.52, SD = 1.54) than in the control condition (M = 3.53, SD = 1.61; $F(1, 61) = 223.15$, $p < .001$, $R^2_m = .62$, $R^2_c = .66$). Additionally, sleep deprivation compromised PVT performance: Response speed deteriorated significantly ($F(1, 61) = 103.985$, $p < .001$, $R^2_m = .19$, $R^2_c = .78$) after sleep deprivation (M = 3.36 s^{-1}, SD = 0.65 s^{-1}) in comparison to control condition performance (M = 3.89, SD = 0.45). Also, the number of lapses was significantly larger ($F(1, 61) = 95.93$, $p < .001$, $R^2_m = .30$, $R^2_c = .61$) after sleep deprivation (M = 9.40, SD = 9.76) than after normal sleep (M = 1.76, SD = 4.02).

Participants completed on average 6.32 docking trials in the baseline session and 6.58 in the sleep-deprived session and achieved an average overall accuracy score of *M* = .76 (*SD* = .33). The linear mixed model revealed significant main effects of level ($F(2, 774.80) = 347.28$, $p < .001$) and session ($F(1, 742.26) = 82.12$, $p < .001$), as well as an interaction between level and session ($F(2, 732.92) = 5.13$, $p = .006$). Participants improved their docking accuracy from the first (*M* = .73, *SD* = .35) to the second session (*M* = .79, *SD* = .32), and this learning effect was evident regardless of condition order. Docking accuracy decreased with increasing difficulty level. Importantly, there was a significant main effect of condition ($F(1, 742.26) = 8.82$, $p = .003$). This result is consistent with our hypothesis that accuracy decreases after sleep deprivation (*M* = .74, *SD* = .35) in comparison to control condition performance (*M* = .78, *SD* = .32). There was no significant interaction of condition with session ($F(1, 53.86) = 1.73$, $p = .19$) or a trifold interaction ($F(2, 774.80) = 0.86$, $p = .42$), but a significant interaction of condition and level ($F(2, 732.92) = 4.35$, $p = .01$). The proportion of explained variance can be described by $R^2_c = .67$ for the whole model and $R^2_m = .44$ for fixed effects only.

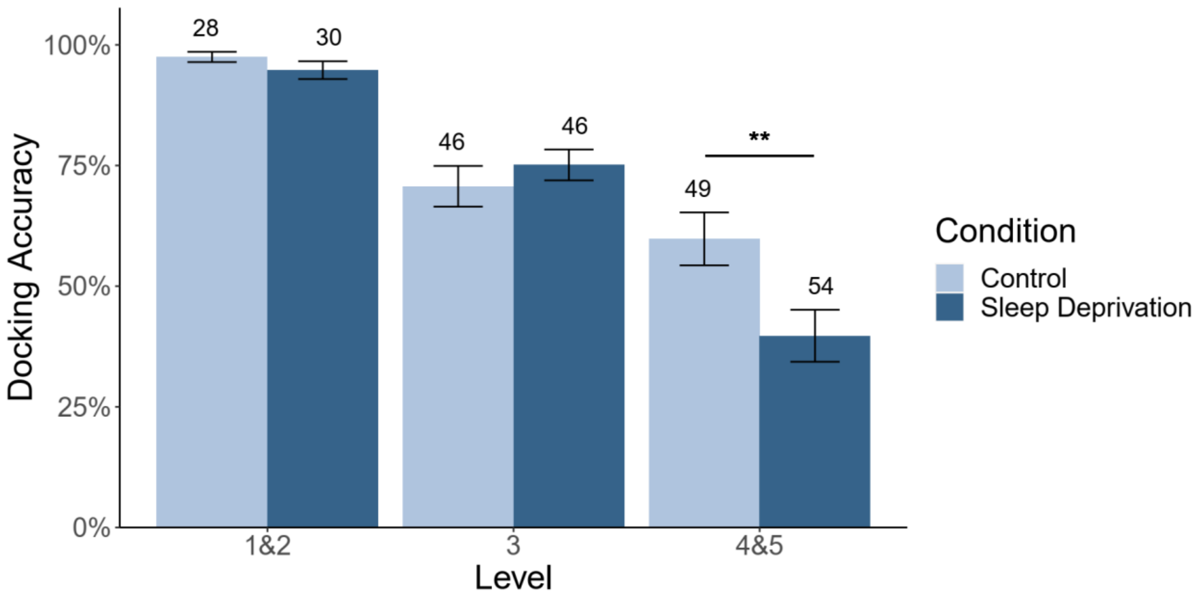


Fig. 3. Average docking accuracy scores under control and sleep deprivation conditions for each level of difficulty (each session started on level 4). Error bars represent the standard error, the number of participants in each category is indicated above the bars. Trials that were discontinued because the participant maneuvered out of the station's reach (distance > 200m) were excluded.

When adding the interaction between condition and the SDC-CC difference in PVT response speed to the model, the significant main effect of condition vanished ($F(1, 726.13) = 0.06$, $p = .81$). However, the effect of condition on docking accuracy remained dependent on the difficulty level ($F(2, 732.17) = 3.92$, $p = .02$). Additionally, there was a significant interaction of condition and the SDC-CC difference in PVT response speed ($F(1, 723.77) = 6.05$, $p = .01$). Participants with higher susceptibility to sleep deprivation (large decrease in PVT response time in the sleep deprivation condition compared with the control condition) also showed a larger difference of docking accuracy between conditions. The proportion of explained variance for this extended model was $R^2_c = .68$ and $R^2_m = .44$ for fixed effects only. When including susceptibility in terms of the number of lapses in the PVT, the main effect of condition on docking accuracy disappeared likewise ($F(1, 728.86) = 1.46$, $p = .23$). The interaction of condition and level persisted ($F(2, 731.75) = 4.01$, $p = .02$), but there was no significant interaction between condition and the SDC-CC difference in the number of PVT lapses ($F(1, 724.11) = 3.22$, $p = .07$). The proportion of explained variance for this second extended model was $R^2_c = .68$ and $R^2_m = .43$ for fixed effects only.

Single linear mixed models were computed separately per category of difficulty level (levels 1 and 2; level 3; levels 4 and 5) to delineate the origin of the interaction between condition and level (Fig. 3). There was no main effect of condition on docking accuracy when only levels 1 and 2 ($F(1, 137.53) = 2.68$, $p = .10$) or level 3 ($F(1, 339.5) = 0.01$, $p = .98$) were considered. Instead, accuracy in level 4 and 5 decreased after sleep deprivation ($F(1, 287.02) = 7.04$, $p = .01$, $R^2_m = .02$, $R^2_c = .33$).

Looking at the progress through the *6df* program instead of accuracy (Fig. 4), participants' performance also improved significantly from the first to the second session ($V = 13.50$, $p < .001$, $r = .76$). The average highest successfully completed level increased from $M = 2.95$ ($SD = 1.06$, $Mdn = 3$) to $M = 3.84$ ($SD = 1.19$, $Mdn = 4$). Participants on average reached a slightly lower level after sleep deprivation ($M = 3.29$, $SD = 1.19$, $Mdn = 3$) than in the control condition ($M = 3.50$, $SD = 1.22$, $Mdn = 3$). However, this difference was not statistically significant ($V = 531.50$, $p = .18$).

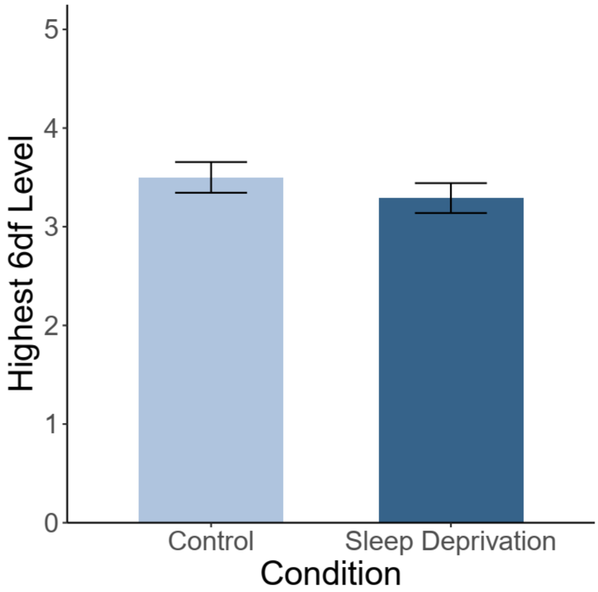


Fig. 4. Average highest successfully completed *6df* level per condition. Error bars indicate the standard error.

Discussion

Astronauts in space usually sleep less (Barger et al., 2014; Dijk et al., 2001) than the seven hours that – according to consensus reports – are needed to maintain health and cognitive performance in adults (Watson et al., 2015). Whereas sleep loss is known to impair performance in a multitude of cognitive tests, the goal of the present study was to investigate the potential effect of sleep deprivation on simulated manual spacecraft docking – a complex and mission-critical operational task – as well as to identify the relationship with performance in the PVT as a standard measure sensitive to sleep loss.

Our results revealed a detrimental effect of sleep deprivation on accuracy in the complex manual docking simulation. Although participants were able to compensate their sleepiness in easier docking trials (levels 1-3), they were significantly impaired during the more difficult trials at levels 4 and 5. Whereas levels 1-3 include only single components of manual control, the difficult trials represent a full docking maneuver as it is required in an operational context. This observation stands in contrast to previous studies that were not able to detect impairments in similarly complex tasks after sleep deprivation, possibly due to small sample sizes (Strangman et al., 2005; Wong et al., 2020) or the lack of a control condition without sleep deprivation (Wong et al., 2020). In contrast to docking accuracy, we found no significant difference between conditions regarding the highest level participants reached. The smaller effect on task progression might be explained by the small variance of the variable due to the limited number of levels in the task.

As expected, participants showed a substantial decline in PVT response speed and an increase in the number of lapses when sleep-deprived. Measures of sustained attention are generally considered to be most sensitive to sleep loss (Lim & Dinges, 2008; Lim & Dinges, 2010). Novelty and attractiveness of the docking simulation are a likely reason for the absence of larger performance decrements across difficulty levels (Wong et al., 2020). Heightened motivation might have helped participants to compensate their sleepiness at least partly. High-level complex skills are generally assumed to be less affected by sleep deprivation compared to monotonous, less demanding tasks, because they enhance motivation and effort (Harrison & Horne, 2000). The PVT however is a very monotonous task (Schleicher et al., 2008). Additionally, speed measures seem to be more susceptible to total sleep deprivation than accuracy measures (Koslowsky & Babkoff, 1992; Lim & Dinges, 2010). If individuals are given sufficient time to complete a motivating task, the slowing of cognitive functioning due to sleep loss can be compensated to a large degree (Harrison & Horne, 2000). Participants in the 3D sensorimotor navigation study by Strangman et al. (2005) displayed no performance decrements, but reported the docking to be more effortful under sleep deprivation. Accordingly, functional magnetic resonance imaging revealed compensatory cerebral responses to sleep deprivation in cortical regions associated with visuospatial processing, memory and attention.

The degree of performance impairment in response to sleep loss is subject to substantial inter-individual differences that are stable and trait-like (Elmenhorst et al., 2017; Elmenhorst et al., 2018; Rupp et al., 2012; Van Dongen et al., 2004; Yamazaki & Goel, 2020). Therefore, the PVT has been used to classify individuals regarding their vulnerability or susceptibility to sleep loss (St Hilaire et al., 2019; St Hilaire et al., 2013). Performance deficits after sleep deprivation are oftentimes attributed to impairments in sustained attention. The latter is seen as prerequisite of more complex cognitive processes (Harrison & Horne, 2000; Lim & Dinges, 2008). In our study, the inclusion of the impairment in sustained attention due to sleep deprivation (measured as the SDC-CC difference in response speed and number of lapses) dissolved the main effect of sleep deprivation on docking accuracy. However, a significant interaction of level and sleep deprivation persisted. For response speed (but not number of lapses), the interaction with condition indicated that participants who reacted to sleep deprivation with a larger decline in PVT response speed also displayed larger performance decrements in docking accuracy.

Whereas sustained attention is indeed an important component of complex operational performance under sleep loss, it can only partly account for impairments in the more complex docking task. The PVT has already been proposed as a short test to evaluate fitness for duty prior to an operational task. For example, the PVT predicted performance decline in a luggage screening task that is based heavily on sustained attention (Basner & Rubinstein, 2011), but was not indicative of performance impairment in a driving simulation (Baulk et al., 2008). On the ISS, the PVT is part of the standard measures used to monitor astronauts' cognitive functioning during their missions. Although the PVT is not sufficient to predict operational performance, it may be useful as a first indicator to timely detect possible performance decrements due to fatigue. However, our results also underline the need for operational task designs to assess performance or readiness for duty in safety-critical contexts.

This study has several limitations. One of them is the lack of previous training with the docking simulation. This resulted in considerable performance variance between participants, which might have partially masked the effects of sleep deprivation. Sleep deprivation itself leads to increased variability within individuals due to state instability as well as between individuals due to differences in vulnerability to sleep loss, which poses the risk of missing performance decrements if sample size is limited (Goel et al., 2009). Although novice students differ from an astronaut population (e.g. professionalized training), they allowed for a larger sample size and a controlled experimental design. In our study, we observed an order effect which likely resulted from the benefits of continued training across the two test sessions. The use of a counterbalanced cross-over design protected against some of these influences masking the sleep deprivation effect. Moreover, our analysis did not reveal a significant interaction between test session and condition. Nonetheless, it will be interesting to investigate the role of training and its potential interaction with the effects of sleep loss in future studies of docking performance. Whereas training diminishes the novelty of the task, exhaustive training is expected to be protective against performance errors (Flynn-Evans, Gregory, et al., 2016). Furthermore, the operational importance of a manual docking maneuver in professionals should have a high impact on motivation, counteracting the decrease in arousal due to sleepiness. Nevertheless, for a highly safety-relevant task like manual docking, even small performance decrements can have serious consequences. In the aviation domain, less than six hours of sleep already pose a substantial risk factor (Bendak & Rashid, 2020). Therefore, further investigation of performance in complex operational tasks is necessary to achieve a comprehensive risk assessment. Our results might not generalize to chronic sleep restriction, which is a highly prevalent state in space (Barger et al., 2014; Jones et al., 2022) and many other occupational contexts (Banks & Dinges, 2007). The effects of partial sleep deprivation on cognitive performance might be even more pronounced compared to total sleep deprivation (Pilcher & Huffcutt, 1996) and it is unclear to what extent motivational factors may be protective under these circumstances. Performance on long-term missions to moon and Mars will likely be more vulnerable, because fewer resources are available to buffer the effects of sleepiness and circadian misalignment (Flynn-Evans, Gregory, et al., 2016).

The aim of this counterbalanced cross-over study was to assess the influence of ~24 h total sleep deprivation on performance in a mission-relevant operational task and a sustained attention task. Our results demonstrate performance decrements in a complex manual docking simulation that were explained partly by decrements in sustained attention. Docking performance was impaired in difficult trials, but not in in easier trials – possibly because the task's novelty and engaging character was enough to overcome impairments during lower task demand. However, even small decrements in accuracy could have catastrophic consequences in safety-critical tasks, especially when various stressors accumulate. Future studies should assess the influence of exhaustive training on the susceptibility to sleep loss. Operational performance measures like those gathered from the *6df* task could be helpful tools to assess readiness for duty under sleep deprivation during long duration missions. The susceptibility to sleep deprivation as measured by the PVT is useful for the early detection of prediction of performance decrements in more complex tasks.

Acknowledgments

This work was supported through funds from the Space Program and the Aeronautics Program of the German Aerospace Center (DLR) and German Space Agency. We would like to thank Marie-Therese Schmitz for statistical advice.

References

Åkerstedt, T. (2000). Consensus statement: fatigue and accidents in transport operations. *Journal of Sleep Research*, *9*(4), 395-395. https://doi.org/10.1046/j.1365-2869.2000.00228.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., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, *52*(1-2), 29-37. https://doi.org/10.3109/00207459008994241

Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., Redmond, D. P., Russo, M., & Wesensten, N. J. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research*, *13*(3), 219-227. https://doi.org/10.1111/j.1365-2869.2004.00407.x

Banks, S., & Dinges, D. F. (2007). Behavioral and physiological consequences of sleep restriction. *Journal of Clinical Sleep Medicine*, *3*(5), 519-528. https://doi.org/10.5664/jcsm.26918

Barger, L. K., Ayas, N. T., Cade, B. E., Cronin, J. W., Rosner, B., Speizer, F. E., & Czeisler, C. A. (2006). Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. *PLoS Medicine*, *3*(12), e487. https://doi.org/10.1371/journal.pmed.0030487

Barger, L. K., Flynn-Evans, E. E., Kubey, A., Walsh, L., Ronda, J. M., Wang, W., Wright Jr, K. P., & Czeisler, C. A. (2014). Prevalence of sleep deficiency and use of hypnotic drugs in astronauts before, during, and after spaceflight: an observational study. *The Lancet Neurology*, *13*(9), 904-912. https://doi.org/10.1016/s1474-4422(14)70122-x

Barton, K. (2020). *MuMIn: Multi-Model Inference* (R package version 1.43.17). https://CRAN.R-project.org/package=MuMIn

Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, *34*(5), 581-591. https://doi.org/10.1093/sleep/34.5.581

Basner, M., & Dinges, D. F. (2014). Lost in space: sleep. *The Lancet Neurology*, *13*(9), 860-862. https://doi.org/10.1016/s1474-4422(14)70176-0

Basner, M., Moore, T. M., Hermosillo, E., Nasrini, J., Dinges, D. F., Gur, R. C., & Johannes, B. (2020). Cognition test battery performance is associated with simulated 6df spacecraft docking performance. *Aerospace Medicine and Human Performance, 91*(11), 861-867. https://doi.org/10.3357/amhp.5602.2020

Basner, M., & Rubinstein, J. (2011). Fitness for duty: A 3 minute version of the Psychomotor Vigilance Test predicts fatigue related declines in luggage screening performance.

Journal of Occupational and Environmental Medicine, *53*(10), 1146. https://doi.org/10.1097/jom.0b013e31822b8356

Baulk, S. D., Biggs, S. N., Reid, K., van den Heuvel, C. J., & Dawson, D. (2008). Chasing the silver bullet: measuring driver fatigue using simple and complex tasks. *Accident Analysis & Prevention*, *40*(1), 396-402. https://doi.org/10.1016/j.aap.2007.07.008

Bendak, S., & Rashid, H. S. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics*, *76*, 102928. https://doi.org/10.1016/j.ergon.2020.102928

Caldwell, J. A. (2012). Crew schedules, sleep deprivation, and aviation performance. *Current Directions in Psychological Science*, *21*(2), 85-89. https://doi.org/10.1177/0963721411435842

Caruso, C. C. (2014). Negative impacts of shiftwork and long work hours. *Rehabilitation Nursing*, *39*(1), 16-25. https://doi.org/10.1002/rnj.107

Dijk, D.-J., Neri, D. F., Wyatt, J. K., Ronda, J. M., Riel, E., Ritz-De Cecco, A., Hughes, R. J., Elliott, A. R., Prisk, G. K., West, J. B., & Czeisler, C. A. (2001). Sleep, performance, circadian rhythms, and light-dark cycles during two space shuttle flights. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 281*, R1647–R166. *https://doi.org/10.1152/ajpregu.2001.281.5.R1647*

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., & 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

Dorrian, J., Hussey, F., & Dawson, D. (2007). Train driving efficiency and safety: examining the cost of fatigue. *Journal of Sleep Research*, *16*(1), 1-11. https://doi.org/10.1111/j.1365-2869.2007.00563.x

Ellis, S. R. (2000). Collision in space. *Ergonomics in Design*, *8*(1), 4-9. https://doi.org/10.1177/106480460000800102

Elmenhorst, D., Elmenhorst, E.-M., Hennecke, E., Kroll, T., Matusch, A., Aeschbach, D., & Bauer, A. (2017). Recovery sleep after extended wakefulness restores elevated A1 adenosine receptor availability in the human brain. *Proceedings of the National Academy of Sciences*, *114*(16), 4243-4248. https://doi.org/10.1073/pnas.1614677114

Elmenhorst, E.-M., Elmenhorst, D., Benderoth, S., Kroll, T., Bauer, A., & Aeschbach, D. (2018). Cognitive impairments by alcohol and sleep deprivation indicate trait characteristics and a potential role for adenosine A1 receptors. *Proceedings of the National Academy of Sciences*, *115*(31), 8009-8014. https://doi.org/10.1073/pnas.1803770115

Flynn-Evans, E. E., Barger, L. K., Kubey, A. A., Sullivan, J. P., & Czeisler, C. A. (2016). Circadian misalignment affects sleep and medication use before and during spaceflight. *npj Microgravity*, *2*(1), 1-6. https://doi.org/10.1038/npjmgrav.2015.19

Flynn-Evans, E. E., Gregory, K., Arsintescu, L., & Whitmire, A. (2016). *Evidence Report: Risk of performance decrements and adverse health outcomes resulting from sleep loss, circadian desynchronization, and work overload* (Report No. JSC-CN-35774). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://ntrs.nasa.gov/citations/20160003864

Goel, N., Rao, H., Durmer, J. S., & Dinges, D. F. (2009). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology, 29*(4), 320-339. https://doi.org/10.1055/s-0029-1237117

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: a review. *Journal of Experimental Psychology: Applied, 6*(3), 236-249. https://doi.org/10.1037/1076-898x.6.3.236

Johannes, B., Bronnikov, S., Bubeev, Y., Arz, M., Piechowski, S., & Rittweger, J. (2019). Individual Learning Curves in Manual Control of Six Degrees of Freedom. *International Journal of Applied Psychology, 9*(2), 47-51.

Johannes, B., Bronnikov, S., Bubeev, Y., Dudukin, A., Hoermann, H.-J., Frett, T., Rittweger, J., & Gaillard, A. (2017). A Tool to Facilitate Learning in a Complex Manual Control Task. *International Journal of Applied Psychology*, *7*(4), 79-85.

Johannes, B., Salnitski, V., Dudukin, A., Shevchenko, L., & Bronnikov, S. (2016). Performance Assessment in the PILOT Experiment On Board Space Stations Mir and ISS. *Aerospace medicine and human performance*, *87*(6), 534-544. https://doi.org/10.3357/amhp.4433.2016

Jones, C. W., Basner, M., Mollicone, D. J., Mott, C. M., & Dinges, D. F. (2022). Sleep deficiency in spaceflight is associated with degraded neurobehavioral functions and elevated stress in astronauts on six-month missions aboard the International Space Station. *Sleep, 45*(3), zsac006. https://doi.org/10.1093/sleep/zsac006

Koslowsky, M., & Babkoff, H. (1992). Meta-analysis of the relationship between total sleep deprivation and performance. *Chronobiology International*, *9*(2), 132-136. https://doi.org/10.3109/07420529209064524

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. Annals of the New York Academy of Sciences, 1129(1), 305-322. https://doi.org/10.1196/annals.1417.002

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. Psychological Bulletin, 136(3), 375-389. https://doi.org/10.1037/a0018883

Mallis, M. M., & DeRoshia, C. (2005). Circadian rhythms, sleep, and performance in space. *Aviation, Space, and Environmental Medicine*, *76*(6), B94-B107.

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Philibert, I. (2005). Sleep loss and performance in residents and nonphysicians: a meta-analytic examination. *Sleep*, *28*(11), 1392-1402. https://doi.org/10.1093/sleep/28.11.1392

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

Porcu, S., Bellatreccia, A., Ferrara, M., & Casagrande, M. J. E. (1998). Sleepiness, alertness and performance during a laboratory simulation of an acute shift of the wake-sleep cycle. *41*(8), 1192-1202. https://doi.org/10.1080/001401398186478

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Sage.

Rupp, T. L., Wesensten, N. J., & Balkin, T. J. (2012). Trait-like vulnerability to total and partial sleep loss. *Sleep*, *35*(8), 1163-1172. https://doi.org/10.5665/sleep.2010

Sallinen, M., van Dijk, H., Aeschbach, D., Maij, A., & Åkerstedt, T. (2020). A large-scale European Union study of aircrew fatigue during long night and disruptive duties. *Aerospace Medicine and Human Performance*, *91*(8), 628-635. https://doi.org/10.3357/AMHP.5561.2020

Salnitski, V., Dudukin, A., & Johannes, B. (2001). Evaluation of operator's reliability in long-term isolation (The "Pilot"-Test). In V. Baranov (Ed.), *Simulation of extended isolation: advances and problems* (pp. 30-50). Slovo.

Schleicher, R., Galley, N., Briest, S., & Galley, L. J. E. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics, 51*(7), 982-1010. https://doi.org/10.1080/00140130701817062

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Karolinska sleepiness scale (KSS). In *STOP, THAT and one hundred other sleep scales* (pp. 209-210). Springer. https://doi.org/10.1007/978-1-4419-9893-4_47

St Hilaire, M. A., Kristal, B. S., Rahman, S. A., Sullivan, J. P., Quackenbush, J., Duffy, J. F., Barger, L. K., Gooley, J. J., Czeisler, C. A., & Lockley, S. W. (2019). Using a single daytime performance test to identify most individuals at high-risk for performance impairment during extended wake. *Scientific Reports, 9*(1), 1-13. https://doi.org/10.1038/s41598-019-52930-y

St Hilaire, M. A., Sullivan, J. P., Anderson, C., Cohen, D. A., Barger, L. K., Lockley, S. W., & Klerman, E. B. (2013). Classifying performance impairment in response to sleep loss using pattern recognition algorithms on single session testing. *Accident Analysis & Prevention*, *50*, 992-1002. https://doi.org/10.1016/j.aap.2012.08.003

Strangman, G. E., Thompson, J. H., Strauss, M. M., Marshburn, T. H., & Sutton, J. P. (2005). Functional brain imaging of a complex navigation task following one night of total sleep deprivation: a preliminary study. *Journal of Sleep Research*, *14*(4), 369-375. https://doi.org/10.1111/j.1365-2869.2005.00488.x

Van Dongen, P., Baynard, M. D., Maislin, G., & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep*, *27*(3), 423-433. https://doi.org/10.1093/sleep/27.3.423

Vejvoda, M., Elmenhorst, E. M., Pennig, S., Plath, G., Maass, H., Tritschler, K., Basner, M., & Aeschbach, D. (2014). Significance of time awake for predicting pilots' fatigue on short-haul flights: implications for flight duty time regulations. *Journal of Sleep Research*, *23*(5), 564-567. https://doi.org/10.1111/jsr.12186

Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., Gangwisch, J., Grandner, M. A., Kushida, C., Malhotra, R. K., Martin, J. L., Patel, S. R., Quan, S. F., & Tasali, E. (2015). Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *Journal of Clinical Sleep Medicine*, *11*(6), 591-592. https://doi.org/10.5664/jcsm.4758

Wong, L., Pradhan, S., Karasinski, J., Hu, C., Strangman, G. E., Ivkovic, V., Arsintescu, L., & Flynn-Evans, E. (2020). Performance on the Robotics On-Board Trainer (ROBoT-r) Spaceflight Simulation During Acute Sleep Deprivation. *Frontiers in Neuroscience*, *14*, 697. https://doi.org/10.3389/fnins.2020.00697

Yamazaki, E. M., & Goel, N. (2020). Robust stability of trait-like vulnerability or resilience to common types of sleep deprivation in a large sample of adults. *Sleep*, *43*(6), zsz292. https://doi.org/10.1093/sleep/zsz292

5 General Discussion

Space is a challenging working environment that demands high levels of cognitive and sensorimotor performance under adverse conditions. Although mission success heavily depends on the reliability of human performance, the influence of space flight-related stressors on operational tasks has rarely been studied systematically. Especially in the context of future long-duration missions, more efficient training methods and objective measures to predict performance will be necessary. To face these gaps, three studies were conducted using a desktop-based simulation of manual spacecraft docking, which is a complex mission-relevant task that involves cognitive domains vulnerable to the stressors of space flight. The first study aimed at the validation of a stereoscopic visualization to facilitate spatial orientation during the learning process. The second study applied eye tracking to identify measures of gaze behavior related to docking performance. Finally, the third study assessed the influence of sleep deprivation, a stressor highly prevalent in space, on docking performance.

5.1 Summary of main results

5.1.1 Study I: Stereoscopic learning aid for manual docking

Due to increased information density, future long-duration missions will require optimized training efficiency. For manual docking, one key challenge is to control an object in 3D on the basis of a 2D monitor. During training, a mental representation of position and orientation in relation to the docking point must be established. To facilitate this process, a stereoscopic version of *6df* was developed. Participants absolved the full *6df* learning program in the course of 20 sessions during an HDT bed rest study. Twelve participants were allocated to the standard 2D program and the other twelve to the 3D visualization that supplemented the first part of the acquisition process. Positive effects of the 3D visualization on learning speed and learning success (performance in the higher-fidelity docking simulation TORU) were hypothesized.

The results indicated no persistent advantages of the stereoscopic visualization compared to the standard 2D learning program. In earlier difficulty levels (5-15), 3D participants required less task repetitions to ascend to the next level. However, there was no main effect of the 3D presentation on learning speed – neither until the end of the stereoscopic support (level 15) nor until the end of the whole learning program. The small positive effect of the stereoscopic view on learning speed did not carry over into the second phase of the training, when all participants continued with the standard 2D presentation. For learning success, results were similar. 3D participants had no general advantage in the more complex TORU simulation. They performed slightly better during the first trials of the new task, but there was no difference left between 2D and 3D group at the end of the TORU session. Possibly, the 3D participants accustomed slightly faster to the new task demands. In conclusion, stereoscopic visualization only led to partial, but non-persistent, increments in learning speed and learning success.

5.1.2 Study II: Eye tracking as performance indicator for manual docking

In the 2D group from study I, eye tracking could be implemented into the *6df* learning program. Manual docking is a visually demanding task that requires high levels of situation awareness. The aim of study II was to identify gaze behavior associated with docking performance. We hypothesized that attention allocation to relevant regions of the simulation display would be related to the accuracy of the docking maneuver.

Eye tracking proved to be a feasible and unobtrusive method to gain insight into information processing during spacecraft docking. Participants concentrated most of their visual attention on the vizor that was used to orientate the spacecraft in the direction of the docking point. The task overview picture was rarely attended. Participants preferred the auxiliary instruments (providing color-coded information on the actual and allowed speed) over the standard ones. Importantly, gaze behavior was significantly related to docking performance. Longer total dwell time on both instrument displays and a higher number of dwells on the auxiliary instruments were associated with higher docking accuracy. In summary, the number and duration of visual instrument checks are suitable candidates to predict performance in manual docking.

5.1.3 Study III: Manual docking performance after sleep deprivation

The aim of the third study was to assess the impact of sleep deprivation on manual docking performance. Disturbances of sleep duration and quality are highly prevalent in space. Whereas laboratory studies have shown diverse decrements in cognitive performance due to sleep deprivation, the picture is less clear when it comes to operational performance. Therefore, we tested the effect of 24 h sleep deprivation on manual docking performance. Additionally, performance in the PVT was considered as a measure of cognitive performance, specifically of sustained attention, that is highly sensitive to sleep loss and has gold standard status in sleep research. We expected decreases in sustained attention and operational performance following sleep deprivation. The second aim of this study was to assess if decrements in sustained attention as measured by the PVT were associated with sleep-deprived docking performance.

Compared to the control condition, total sleep deprivation led to an increase in subjective sleepiness and to performance decrements in the PVT. The latter was evident in a decrease in response speed as well as an increase in the number of lapses. Sleep deprivation had no significant impact on the highest difficulty level participants were able to successfully complete in the adaptive manual docking task. However, docking accuracy significantly decreased after sleep deprivation. This effect was dependent on the difficulty level: the impairment in docking accuracy occurred only in the more difficult levels that resembled a standard docking maneuver, but not in the easier tasks that comprised only single aspects of manual control in six DoF. The susceptibility to sleep loss as measured by the difference in PVT performance between conditions explained part of the sleep deprivation effect on docking accuracy. Participants with larger decrements in PVT response speed following sleep deprivation also exhibited worse docking accuracy when sleep-deprived.

5.2 Integrative discussion and future directions

More basic knowledge on complex operational performance is needed to tackle the challenges that will arise during future long-term missions. The three studies of this dissertation investigated aspects of operational performance, aiming at the enhancement of training efficiency and the prediction of performance under the influence of space flight-related stressors.

5.2.1 Virtual reality in manual docking training

Training of manual control in six DoF is complex and time-consuming. Technical innovations might help to support and shorten skill acquisition, especially in the context of heightened autonomy during future long-term missions. VR has been proposed as a tool to facilitate a more intuitive understanding for control tasks that rely on 2D screens (Pirker, 2022). The 3D glasses and stereoscopic presentation investigated in study I proved to be feasible additions to the *6df* tool. Although stereoscopic viewing may provoke motion sickness (Hwang & Peli, 2014), none of our participants reported major discomfort. However, stereoscopic presentation shortened the training time only temporarily. The positive effect of the 3D visualization did not persist after switching to the standard *6df* version. The real maneuver is done via a 2D screen; therefore, an effective training aid should facilitate spatial orientation also in 2D conditions. Conclusions regarding learning success were similarly inconsistent. 3D participants seemed to accustom faster to the TORU tasks that are used in cosmonaut training and incorporate more complex flight characteristics than *6df*. Possibly, they acquired a more robust spatial orientation that helped to generalize the skill to control six DoF. However, 2D and 3D participants performed equally in the second half of the TORU test session. For the present, the advantages of this VR technology were not substantial enough to replace the standard learning program that requires less hardware. We decided to use a non-immersive desktop-based approach to reduce the likelihood of motion sickness and to allow for an unhindered view on the hand controls. Our focus was the clarification of spatial relations in space, but there are also fully immersive and graphically realistic simulations of spacecraft control (Bosch Bruguera et al., 2019). Although immersive VR is not necessarily superior in training contexts (Aoki et al., 2008), it remains to be tested if immersive VR might have more advantages for the development of spatial orientation in manual control than our non-immersive approach.

Moreover, larger studies are needed to further evaluate the effects of stereoscopic viewing as a training aid. Interindividual differences in learning speed were exceptionally high, some participants required almost twice as much task repetitions than the fastest individuals. These large differences might have masked some effects of the 3D presentation in our small sample. We were also not able to make subgroup analyses to identify which participants benefitted most from the training intervention. More knowledge on the predictors of learning speed could help to adjust the learning program to the individual's needs. For example, stereoscopic presentation might help especially individuals that have difficulties with spatial orientation to start with, whereas the additional value of 3D might be restricted when these abilities are already pronounced. In future studies, spatial abilities should be assessed because they are an integral part of controlling six DoF (Menchaca-Brandan et al., 2007; Wang et al., 2014), especially in novices (Du et al., 2015). In space, spatial disorientation (De la Torre, 2014; Glasauer & Mittelstaedt, 1998) and spatial illusions (Kornilova, 1997; Kornilova et al., 1996) due to microgravity pose additional challenges that can lead to errors in the estimation of distances and acceleration (Clément et al., 2013; Steinberg, 2019). Under these extreme conditions, the effect of VR on spatial orientation might be more pronounced. This would be relevant for on-board training and skill maintenance that is needed during long-term missions (Rector et al., 2021). Because some participants initially reported having some visual difficulties with the stereoscopic image, prior familiarization with 3D glasses might also play a role. Although stereoscopic presentation has not proven to be a sustainable training aid yet, the identification of additional predictors of learning success could help to locate individuals that benefit most from VR support. In a more extreme environment, e.g. around gravitational transitions and very high workload, the small positive effect of stereoscopic presentation might be more clearly pronounced.

5.2.2 Eye tracking in manual docking training

Eye tracking devices allow for an objective and continuous assessment of visual attention that is unobtrusive and does not interfere with the operational task. Whereas eye tracking has been frequently used to investigate pilots' state and instrument monitoring strategies (Peißl et al., 2018; Ziv, 2016), data from spacecraft cockpits are scarce (Huemer et al., 2005). Study II offered the opportunity to collect first data on gaze behavior during skill acquisition with the *6df* tool. Participants focused most of their visual attention on the vizor and thereby on the target space station. This observation is in line with previous data from Tian et al. (2018), who reported that 80% of operators' fixation times were directed at the space station and only 20% on the simulated spacecraft's instruments. In general, most dwell time is directed to areas that are especially important for the task or that change frequently (Glaholt, 2014). According to the SEEV model, visual attention in complex operational environments, such as instrument scanning in the cockpit, is driven by four components: salience (e.g. luminance, color), effort (e.g. distance that the eye has to be moved), expectancy (regarding the frequency of events in an area), and the value of an area for task completion (Wickens et al., 2008). The descriptive eye tracking data also indicated that participants indeed used the additional color-coded information on the actual and prescribed speed that was provided by the auxiliary display. In turn, the standard instruments and the task overview were rarely attended. This descriptive view on gaze behavior was already valuable to gain insight into how operators used the *6df* tool. Such information can be used to further adapt the simulation's design to the learners' needs.

But more importantly, we were able to identify associations between gaze behavior and docking performance. Although most attention was devoted to the vizor and this focus on the target was necessary for task completion, the amount of attention to the vizor was not predictive of docking accuracy. Instead, frequently and thoroughly monitoring the instruments turned out to be crucial for performance. More total dwell time on both instruments and a higher number of dwells to the auxiliary instruments were related to higher docking accuracy. Instrument monitoring is important, because this information is needed during approach to continuously update the speed to the remaining distance. Our results are in line with accident analyses in aviation, showing that piloting errors can often be attributed to inadequate instrument monitoring (Lounis et al., 2019). Individuals who devote more attention to the instruments might gain higher situation awareness and thereby timely detect deviations from optimal speed. Situation awareness can be defined as the “perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (Endsley, 1988). Attention is usually focused at the central direction of gaze (Just & Carpenter, 1980) and it can be assumed that the situation awareness of an operator is high as long as they visually attend to the system (Johnson et al., 2017). Accordingly, the percentage of time a relevant region is fixated, dwell times, and fixation rates have been shown to be highly predictive of situation awareness (Moore & Gugerty, 2010; van de Merwe et al., 2012). The lack of situation awareness and resulting misjudgments have often been involved in flight accidents (Wickens et al., 2008).

Previous studies have linked gaze behavior to performance in a variety of manual control tasks, mainly in aircraft control. In a flight simulator, selective and focused gaze behavior on key instruments was associated with control performance (Chuang et al., 2013). Simulated Canadarm2 control performance (also relying on six DoF) was correlated with fixation duration, saccade duration, and pupil diameter (Guo et al., 2021). Lefrancois et al. (2016) associated ineffective gaze patterns (i.e. failure to sufficiently monitor critical instruments) with errors in performing a stabilized manual aircraft approach. The high percentage of errors in instrument monitoring was also attributed to the high reliance on automation. This is a critical issue also in spacecraft control, where automated docking is the normal situation. Many studies investigating the relationship between visual scanning and performance also compared novices and experts (Huemer et al., 2005; Matessa & Remington, 2005; Peißl et al., 2018). One observation was that experts check relevant instruments more frequently (Bellenkes et al., 1997; Glaholt, 2014; Kasarskis et al., 2001), which matches the gaze behavior that was favorable in our study. Experts also check the instruments even if no change is expected and might need shorter individual dwells to retrieve the information (Ziv, 2016). Similar findings also stem from entirely different occupational fields, such as surgical expertise: In contrast to junior surgeons, experts devoted more attention to task-relevant areas in terms of fixation rates and dwell times (Tien et al., 2015). Knowledge on expert gaze behavior can be used to deduce successful gaze behavior and facilitate the understanding of novices regarding important task features. Studies with manual docking experts in realistic settings are a next step to validate successful scanning behavior and the predictive value of gaze behavior for operator reliability.

Certainly, the results from study II already provide various starting points for future *6df* developments. We have shown that eye tracking can provide additional information on operator performance. More available information facilitates the individual and specific improvement of performance (Hockey et al., 2011). Additionally, operators are often unaware of their scanning strategies (Robinski & Stein, 2013), although these are safety-critical. Peysakhovich et al. (2018) defined four stages of eye tracking integration in the field of aviation. In the first stage, eye tracking is used during training to estimate the pilot’s monitoring skills, give feedback on current gaze behavior, and teach optimal gaze behavior based on expert scan patterns. But the authors also see future potential for the implementation of eye tracking in the actual operational setting. In the second stage, the pilots’ gaze during flight would be recorded to have additional information to aid the investigation of incidents. In the third stage, eye tracking could serve as an operational support system, e.g. the pilots could be alerted if their gaze behavior deviates from a defined safety range or specific information in the cockpit could be highlighted to aid information processing. Eventually, in stage four, automation might be allowed to take over control based on the information on operator status obtained via eye tracking. Although some of these considerations look far into the future, at least stage one integration of eye tracking is within reach for manual spacecraft control. For example, operators might receive tailored feedback on their scanning behavior. This could be implemented by a replay of the task with overlaid scan paths or a numeric feedback that includes the number of dwells and the percentage of dwell time allocated to the different regions of interest (ROI). With more information on optimal scanning behavior, the system could give individual recommendations to improve gaze behavior and performance, as well as track strategy changes throughout the training progress. Such eye tracking based learning software has been shown to improve the performance of novices by teaching successful gaze strategies, e.g. in training for laparoscopic surgery (Chetwood et al., 2012; Vine et al., 2012; Wilson et al., 2011). Muehlethaler and Knecht (2016) introduced an eye tracking-supported situation awareness training for pilots in a flight simulator. The training included the discussion of scanning methods after the first flight, the joint development of new gaze strategies with a trainer, and an application phase during the second flight. This approach was supposed to engage the trainee in self-exploration and facilitate skill transfer, while giving the instructor objective data for individual feedback. Another approach established a database of standard visual scanning patterns and alerts the pilot if their gaze behavior deviates from this norm (Lounis et al., 2019). In aviation, certain visual scanning rules that foster safety are already part of the training routine (Colvin et al., 2005). In the case of manual docking, successful scanning behavior will likely include frequently checking the instruments as a key factor. Instead of feedback after the task, this could also be fostered during the task. Because attention is usually focused at the vizor and attention must be redirected actively towards the periphery to retrieve instrument information, a monitoring routine must be established during training. To facilitate such a routine, the instruments could flash if they are not regarded or an acoustic signal could catch the trainee's attention. Thereby, eye tracking is a promising method for the support of an efficient autonomous docking training that is needed for long-term missions, when no experienced instructor will be available.

To predict performance or an operator's status in the working environment, the robustness of a measure regarding confounding factors is crucial. There is a plethora of eye tracking metrics that have been used in research (Rahal & Fiedler, 2019) and more specifically in the aviation domain (Peißl et al., 2018). Because there is no consensus on the most appropriate measures for specific research questions or applications (Lai et al., 2013), comparisons between studies and the selection of suitable metrics are difficult. We focused on the number of dwells and total dwell time, because these metrics have been frequently used in aviation studies and were related to performance (Glaholt, 2014). Due to the low sampling rate of the eye tracking device used, we were not able to consider saccadic metrics. A next step would be to include a more in-depth analysis of scan patterns (Lounis et al., 2021). Another promising application of eye tracking that should be considered is the monitoring of the operator's state during a maneuver (Scannella et al., 2018), e.g. the assessment of workload (Marquart et al., 2015; Peißl et al., 2018). Mental overload facilitates human error (Wickens, 2017) and, therefore, should be detected timely in safety-critical operational tasks. Tian et al. (2018) for example used pupil dilation to assess operator workload during different phases of a simulated manual docking maneuver. Pupillary responses have long been linked to arousal and mental workload (Beatty, 1982; Beatty & Lucero-Wagoner, 2000). We conducted pupil dilation analyses on the data from study II as well, but found the results to be highly confounded by the varying screen luminosity during approach. In an operational context, the control of screen luminosity and environmental lighting would be highly problematic. Because methods must be applicable not only in the laboratory, but also under the harsh conditions in space, more robust workload metrics will have to be tested in the future. For example, the spatial distribution or randomness of fixations has been proposed as an alternative indicator of workload (Di Nocera et al., 2006, 2007; Diaz-Piedra et al., 2019), as well as the frequency of pupil oscillations (Duchowski et al., 2018; Guo et al., 2021). Future studies are needed to identify the measures that are most suitable to predict performance and detect critical operator states in a manual docking context. In conclusion, eye tracking is a promising addition to manual docking that can provide additional information on operator performance. In comparison with many other objective measures, eye tracking is unobtrusive and easy to apply. These advantages could contribute to training and monitoring and thereby to the safety and reliability of manual docking.

5.2.3 Sleep deprivation as a risk factor for operational performance

Stress levels in space are expected to impair performance in astronauts (Morphew, 2001; Patel et al., 2020), although the associations between stressors and cognitive performance are not yet established (Strangman et al., 2014). Sleep loss is a prevalent stressor in space (Barger et al., 2014; Dijk et al., 2003) and has been associated with performance decrements in laboratory tasks (Belenky et al., 2003; Van Dongen et al., 2003). However, the impact of sleep deprivation on performance in the more complex operational tasks astronauts have to master is still a matter of debate. We could demonstrate that sleep loss significantly impaired docking accuracy in the *6df* simulation, a safety-relevant operational task in space. This stands in contrast to previous studies that investigated manual docking tasks, but found no significant effect of sleep deprivation (Strangman et al., 2005; Wong et al., 2020). The reason for the absence of effects in previous studies might have methodological reasons, specifically small sample sizes and, in the study by Wong et al. (2020), learning confounds due to the lack of a control group. Task complexity might be a relevant factor determining the vulnerability of a task to the effects of sleepiness. Whereas the participants were able to compensate for their sleepiness in easy levels that incorporated only single components of manual control in six DoF, their performance was significantly impaired in more difficult levels that resembled a standard docking maneuver. The latter is usually demanded in an operational context; therefore, sleep loss might have serious consequences for the success and safety of a space mission.

Sustained attention is the cognitive domain deemed to be most sensitive to sleep loss (Lim & Dinges, 2008; Lim & Dinges, 2010). In our study as well, sleep deprivation led to decrements in sustained attention as measured by the PVT and a significant increase in subjective sleepiness as measured by the Karolinska Sleepiness Scale. Whereas we were able to identify an effect of sleep deprivation on docking accuracy, this was not the case for our second dependent variable, the progression through the difficulty levels in the adaptive task. This variable had a rather small variance due to the limited number of levels included in the paradigm and the short session duration. For the testing of more experienced operators, this methodological issue might have no serious consequences, because docking accuracy in difficult tasks would be the outcome of relevance (instead of level progression). A possible reason for the absence of a larger effect of sleep deprivation on *6df* across difficulty levels and also on task progression has already been discussed by Wong et al. (2020) and is founded in the novelty and attractiveness of the docking task. Whereas the PVT is very monotonous and itself may induce fatigue (Schleicher et al., 2008), complex tasks facilitate motivation and offer various opportunities for compensational strategies (Harrison & Horne, 2000). In contrast to the PVT, *6df* was novel, challenging, and motivating due to its relevance for space flight. This might have led to the partial compensation of sleepiness in *6df* performance. However, additional effort is a limited resource and may entail physiological or behavioral costs, such as strain, impairments in secondary tasks, narrowing of attention, or risky decisions (Hockey et al., 2011). According to the compensatory control model, more subtle performance decrements can occur, even if primary performance is protected (Hockey & Robert, 1997). The inclusion of secondary tasks during *6df* might help to investigate if such hidden performance decrements appear. Eye tracking data has been obtained also for study III and could help to assess whether sleep deprivation leads to sloppy information processing, a potential latent risk for performance. Moreover, sleep deprivation leads to a slowing in cognitive functioning that can be compensated if time is not strictly limited (Harrison & Horne, 2000). *6df* is a self-paced task and focuses on accuracy instead of speed. On the contrary, speed measures such as PVT response time are expected to be more sensitive to sleep loss (Koslowsky & Babkoff, 1992; Lim & Dinges, 2010). Compensational processes under sleep deprivation in a manual docking task have been reported by Strangman et al. (2005), in terms of increased effort and compensatory cerebral responses in cortical regions related to visuospatial processing and attention. In space, even small or rare deviations from optimal manual control performance can be meaningful, because error margins are small, consequences can be life-threatening, and resources necessary for compensation are typically scarce (Ivkovic et al., 2019).

In our study, sustained attention as measured by the PVT was able to explain at least part of the sleep deprivation effect on docking accuracy. Sustained attention can be regarded as a prerequisite of more complex cognitive processes (Harrison & Horne, 2000; Lim & Dinges, 2008). Sleep loss consistently impairs sustained attention, which will in turn have negative consequences also for operational tasks that demand high levels of attention. During docking for example, changes in speed and remaining distance require permanent attention. Participants with large performance decrements in response speed after sleep loss in comparison with the control condition also showed larger performance decrements in docking accuracy. In other words, these participants had a higher susceptibility to sleep loss that was also indicative of impairments in the more complex operational task. Previous studies have shown that this susceptibility to performance decrements due to sleep loss varies strongly between individuals, but is stable and trait-like within individuals (Elmenhorst et al., 2018; Rupp et al., 2012; Van Dongen et al., 2004; Yamazaki & Goel, 2020). With this in mind, the PVT has been used to classify individuals regarding their susceptibility to sleep loss (St Hilaire et al., 2019; St Hilaire et al., 2013). Based on this reasoning, the PVT has been proposed as a short fitness for duty test prior to an operational task – with mixed results, depending on the actual demands of the operational task that was studied (Basner & Rubinstein, 2011; Baulk et al., 2008). The PVT is a practical candidate for a fitness of duty test, because it is relatively short (typically ten minutes; three minutes for the short version (Basner et al., 2011)), only minor instructions are needed, and the learning curve is minimal (Dorrian et al., 2005). Although PVT performance cannot be extrapolated to performance in everyday tasks, its ecological validity is considered high, because most operational tasks require high levels of sustained attention and a timely response to changes in the environment (Dorrian et al., 2005). Moreover, the PVT is already part of the *Cognition* test battery (Basner et al., 2015) that is applied regularly to monitor astronauts' cognitive functioning before, during, and after missions to ISS. Hence, the use of the PVT as a fitness for duty test prior to operational tasks in space would be an efficient way to identify potential risks due to sleepiness. The PVT is tightly associated with sleep loss and reference values are available due to the repeated administration. Our results confirm that sustained attention is an important factor in complex operational performance after sleep deprivation. However, sustained attention was of course not sufficient to explain performance that also involves higher-order spatial and sensorimotor processes. Furthermore, motivational effects might be problematic in a monotonous task like the PVT, especially if the test is frequently repeated during a long-term mission. This is an obstacle for the validity of results that usually does not occur in operational tasks. As a conclusion, this underlines the importance of developing realistic operational measures that allow for the assessment of fitness for duty.

Future studies should investigate the effects of training and their interaction with sleep loss. The study by Wong et al. (2020), and also our results reported here, have shown that learning the manual control of six DoF continues despite sleep deprivation. But if and to what extent learning is attenuated by sleepiness remains an open question. The answer could be of interest for long-duration missions that will likely require skill acquisition to continue inflight. It would be also important to assess whether training, especially in professionals, can be a protective factor against sleep deprivation. Although novelty will disappear as a motivating factor, the high operational relevance of the task should preserve high levels of motivation. Astronauts have proven a high capability of protecting mission critical goals against stressors (Hockey et al., 2011). In general, exhaustive training is supposed to be protective against human error (Flynn-Evans, Gregory, et al., 2016), though even small remaining impairments may lead to catastrophic outcomes. Additionally, sleep deprivation has been shown to negatively affect flight simulator performance even in experienced pilots (Caldwell et al., 2004). Next to the effects of prolonged training, chronic sleep restriction is another interesting domain for future research. Partial but chronic sleep deprivation is a reality in space (Barger et al., 2014) and similar sleep durations have been identified as a risk factor for safety in aviation (Bendak & Rashid, 2020). Chronic sleep restriction might be even more detrimental than 24 h total sleep deprivation (Pilcher & Huffcutt, 1996), because compensational resources exhaust over time. If there is no recovery from prolonged sleep restriction, performance decrements tend to accumulate over time in a dose-dependent manner (Kanas & Manzey, 2008). This could potentially increase the risks associated with sleepiness over the course of long-duration missions.

In conclusion, the shortened *6df* paradigm used in study III proved to be a feasible tool to enable research on operational task performance in a relatively large sample of novices and for short-duration studies. We were able to provide evidence for the detrimental effect of sleep deprivation on manual control accuracy in six DoF. These were explained partly by decrements in sustained attention. Therefore, the PVT might be a suitable screening tool for the early detection of performance risks due to sleepiness in operational tasks. Because sleep-restricted individuals tend to subjectively underestimate their cognitive impairment (Van Dongen et al., 2003; Zhou et al., 2012), objective cognitive and operational tasks are important to gauge readiness for duty.

5.2.4 Outlook: further development of the *6df* tool

6df was developed as a self-learning program to enable the acquisition and maintenance of manual control skills with less dependence on instructor-based training in a high-fidelity simulator (Johannes et al., 2017). Studies I and II showed that successful learning with *6df* was possible, even under conditions of simulated microgravity. Although there were large differences in the time needed, all participants were able to complete the program in the scope of twenty 45-min sessions. Looking towards future long-duration missions, it is important to develop autonomous training tools to acquire and maintain critical skills over long periods of time and with limited ground support (Stuster, 2010). With respect to restrictions in crew time (Barshi & Dempsey, 2016), the relevance of insufficient training as the cause of accidents (Ellis, 2000), and potential learning impairments in isolation (Bosch Bruguera et al., 2021; Hainley Jr et al., 2013), efforts to enhance training efficiency should continue. Although stereoscopic presentation was not convincing as a training aid yet, future research may identify which individuals can profit from a specific intervention. In general, training aids should be customized to fit an individual's needs. Eye tracking might be able to provide such tailored feedback and training. During more autonomous missions, self-monitoring will be crucial for astronauts (Manzey et al., 1995). Incorporating eye tracking-based feedback into *6df* could help to identify current gaze strategies as well as potentials for improvement. Additionally, performance decrements could be detected timely, so that individual countermeasures can be applied.

To assess operator reliability, not only performance should be considered, but also the psychophysiological state of the operator. For example, eye tracking could be used to obtain an unobtrusive real-time assessment of operator workload (Di Nocera et al., 2006, 2007; Marquart et al., 2015). The operator should be able to perform a safety-critical task while still having free cognitive capacity to react to unforeseen situations. And even if performance is stabilized at a high level, effort indicators may reveal more subtle performance impairments (Hockey et al., 2011; Johannes, Bronnikov, et al., 2021). A prior approach to the inclusion of workload indicators into *6df* made use of P300 event-related potentials (Johannes, Bubeev, et al., 2021). If a robust eye tracking measure could be implemented, this would be easier to apply in an operational context. In comparison with electroencephalography, eye tracking is not dependent on secondary tasks, less obtrusive, and less time-consuming. Future studies are needed to identify appropriate measures that are insensitive or allow to control for the differences in screen brightness that occur during the docking tasks. Next to the detection of high workload, eye tracking has also been used to detect critical states of fatigue (Di Stasi et al., 2016; Diaz-Piedra et al., 2016; McKinley et al., 2011). Under sleep deprivation, decrements in complex motor performance have been associated with decrements in visual perception (Russo et al., 2005). After sleep deprivation, increased latency in pupil constriction and decreased saccadic velocity have been predictive of accidents in a driving simulator (Rowland et al., 2005). Our results from study III confirmed that sleep loss is indeed a risk for manual docking performance. Eye tracking could help to estimate this risk prior to the execution of an operational task and alert the operator if necessary. Sustained attention as measured by the PVT and eye tracking metrics have been used in a complementary manner to investigate flight performance under sleep-deprived conditions (Naeeri et al., 2019). If such fitness for duty tests are combined with an operational task design like *6df*, the timely identification of critical operator states may enhance the safety of the maneuver.

Sensorimotor deficits after gravity transitions can have detrimental impact on manual control performance (Jones, 2010). The skill to control six DoF deteriorates quickly without appropriate maintenance training (Salnitski et al., 2001). To preserve performance even under the extreme conditions of long-duration missions, tools are needed that allow for autonomous and motivating training as well as self-monitoring of performance (Flynn, 2005). Operational tasks like *6df* have the advantage of assessing relevant performance while engaging work motivation. Taken together, the three studies presented in this dissertation helped to identify possibilities for the further development of *6df* as a training tool that can support operators during long-duration missions. Prospectively, *6df* operators could be provided with additional information on their information processing strategies and physiological state (e.g. fatigue and workload) to gauge their fitness for duty and improve training efficiency.

5.3 Limitations

When interpreting the results, some limitations have to be considered. First, the study populations investigated differed from an astronaut population. Whereas astronauts receive hundreds of hours of training, e.g. before controlling a robotic arm (Bloomberg et al., 2015), our participants were non-professionals with only very limited training in the docking task. For such complex skills, it may be expected that expert performance is qualitatively different from that of a novice (Johannes et al., 2011). Demographical differences, e.g. in educational level and age, might also influence performance. Age and gender effects on *6df* learning progress have already been demonstrated (Johannes et al., 2019; Piechowski et al., 2020). Although participants in analog studies are usually selected to be as comparable as possible, the generalizability of results to the strictly selected astronaut population remains problematic. Related to the use of surrogate samples, also analog studies come with limitations. Due to astronaut availability and mission constraints, space flight-related research relies on analog or simulation studies, for example isolation and bed rest. However, analog environments can only mimic some specific factors relevant to space flight, but they do not represent the space environment as a whole (Hockey et al., 2011; Kanas & Manzey, 2008). An HDT bed rest study can reproduce the fluid shifts associated with microgravity, but other stressors are omitted, for example high risk and isolation. On the one hand, this allows to investigate individual space flight stressors while controlling for others, but on the other hand, effects may be different when astronauts are exposed to the interaction of various stressors in space. Due to the lack of a control group without HDT bed rest in studies I and II, no conclusion on the effects of simulated microgravity on manual docking were possible. Additionally, sample sizes are usually small in analog studies (although oftentimes large compared to studies conducted in space), because of the complex and expensive design.

Moreover, high inter-individual variability is often observed during the learning phase of complex tasks (Schneider, 1985) and also in research on cognitive performance in space (Strangman et al., 2014). In the case of the *6df* tool, this has been documented before by Johannes et al. (2019). In the studies of this thesis, large differences in learning progress and performance were evident in the short version as well as in the complete *6df* program. Small effects might have been masked by extreme values, especially in connection with small sample size. Sleep deprivation additionally increases the variability between individuals due to trait-like differences in susceptibility to sleep loss as well as within individuals due to increased state instability (Goel et al., 2009), which exacerbates the risk of missing a significant effect in a sleep deprivation study. Next to the high variability in training progress, learning effects in complex manual control tasks usually continue for a prolonged time until a plateau can be reached (Ivkovic et al., 2019). Learning progress was the outcome of interest in the first study, therefore, differential learning effects are mostly a problem of increasing variance between participants during the course of the study. Regarding the second study, gaze behavior might have changed with increasing proficiency. We excluded the early acquisition phases (levels 1-15) and observed a similar distribution of total dwell time and number of dwells to the different ROI in all higher levels (except for level 20, which had slightly different task demands).

Nonetheless, future studies should investigate changes in gaze behavior and their link to performance throughout the learning process as well as differences between novices and experts. Previous studies in aviation (Bellenkes et al., 1997; Kasarskis et al., 2001) as well as manual docking (Matessa & Remington, 2005) showed that experts' scan patterns differ systematically from less experienced operators. In the third study, continued learning might have been most problematic, because participants were exposed to quite difficult tasks without any previous training. Although condition order was balanced to control for learning effects, this particular *6df* paradigm likely exacerbated the high inter-individual variability that might have partly masked the effect of sleep deprivation. Learning effects are generally hard to eliminate in tasks of such high complexity and during reasonable observation periods, even in professional samples.

5.4 Implications beyond space flight

Although living and working in space is quite unique, it is by far not the only highly demanding working environment. Some of the stressors introduced by Kanas and Manzey (2008) are specific to space flight, for example microgravity. But most other stressors occur in various operational settings and are as relevant on Earth as in space. Fluctuations between monotony and high workload as well as team and leadership conflicts are almost universal issues. Many occupations involve working under high risk and danger, while errors easily result in fatal consequences (e.g. rescue workers). Because stressors and problem areas overlap, the results of human factors research in space are often also applicable on Earth.

The generalizability of the skills acquired with the *6df* tool is still to be tested, but similarities with other manual control tasks exist. For example, the control of robotic arms usually also involves six DoF. The significance of robotic systems will likely continue to grow in the future, also in medicine. Telemedicine is an emerging field that involves robotic systems that can be controlled remotely for diagnostics (Huang & Lan, 2019) or surgery. These systems could not only support astronauts in the case of a medical emergency during remote missions (Haidegger & Benyo, 2008; Haidegger et al., 2011), but are also valuable on Earth (Bogue, 2021; Diana & Marescaux, 2015), for example in areas that lack specialized personnel. In minimally invasive surgery, operators are confronted with a similar problem than in the manual control of a spacecraft: depth perception is complicated by the use of a 2D screen (Diana & Marescaux, 2015). VR trainings have been developed as a countermeasure, and some systems are already equipped with a 3D camera that supports the surgeon (Diana & Marescaux, 2015). In conclusion, complex manual control skills play an important role in emerging occupational fields that involve human-machine interaction – not only in space, but also in “ground-based” operational contexts. Abstract simulations that teach a general skill relevant for diverse applications, such as *6df*, are a valuable testbed for new developments that can aid the efficient acquisition and maintenance of complex skills.

The implementation of eye tracking technology in training contexts is another topic of growing interest in many occupational domains (Rosch & Vogel-Walcutt, 2013). Human error is involved in most accidents in high-risk industries, such as aviation, transportation, and construction (Martinez-Marquez et al., 2021). Eye tracking has been discovered as one method to support operational safety, e.g. by detecting critical operator states or assessing fitness for duty. Especially indicators of fatigue and workload have been investigated in different occupational groups, such as pilots (Di Nocera et al., 2007; Diaz-Piedra et al., 2016; Scannella et al., 2018), air traffic controllers (Martin et al., 2011), truck drivers (Morad et al., 2009), surgeons (Bednarik et al., 2018; Wu et al., 2020), and control room operators in process plants (Bhavsar, 2022) or nuclear power plants (Choi & Seong, 2020). Eye tracking is also used to guide an operator’s attention during learning and individualize training protocols (Rosch & Vogel-Walcutt, 2013). For example, flight instructors can be supported with additional insight into information processing and situation awareness of the trainee (Muehlethaler & Knecht, 2016; Robinski & Stein, 2013). This helps to identify if occurring errors are rooted in insufficient instrument monitoring (Li et al., 2020; Niehorster et al., 2020). Examples from the medical domain include the assessment of surgical skills during training (Ahmidi et al., 2012; Dilley et al., 2020) and the projection of the supervisor’s eye gaze onto the trainee’s screen during laparoscopic tasks to aid target identification (Chetwood et al., 2012). Eye tracking can provide real-time, objective, and continuous data while not interfering with the task – advantages that are compelling in many operational contexts. Research on suitable eye tracking metrics and efficient training strategies is not confined to single use cases.

Regarding the stressors that are typical for space flight, this dissertation focused on sleep loss as one important influencing factor for human performance. The absence of sufficient sleep is obviously not a problem restricted to space flight. Survey data obtained in the US indicates that about 30% of adults sleep six hours or less per night (Luckhaupt et al., 2010) – a duration comparable to the current situation in space. For some occupations, this percentage is even higher, e.g. in military personnel (Troxel et al., 2015). More than 20% of individuals engage in some sort of shift work and especially night shifts have been associated with shorter and more disturbed sleep (Åkerstedt & Wright, 2009; Harrington, 2001). Many night workers even report nodding off during their shifts (Åhsberg et al., 2000; Åkerstedt & Wright, 2009). Long working hours can also lead to sleep deprivation, which in turn negatively impacts health and performance (Caruso, 2006). If working hours counteract the circadian rhythm, performance efficiency is at risk (Harrington, 2001). Accordingly, long working hours (Dembe et al., 2005; Dong, 2005; Folkard & Lombardi, 2006) and shift work (Fransen et al., 2006; Loudoun & Allan, 2008) have been associated with an increased likelihood of human error and accidents. Fatigue due to shift work or long working hours can jeopardize occupational safety in various domains (Papadopoulos et al., 2010), e.g. transport (McCartt et al., 2000), medical care (Winwood et al., 2006), and aviation (Gregory et al., 2010). As already mentioned, even in the occupational domain many studies investigated the impact of sleep deprivation on simple tasks of sustained attention (Anderson et al., 2012; Harrison & Horne, 2000), whereas studies on complex operational performance are scarce. However, accident analyses indicate the high operational risk that comes along with sleep loss. For example, sleep deprivation is one of the leading causes of motor vehicle accidents (Czeisler et al., 2016) and a sleep duration of six hours per night can already pose a significant risk factor (Gottlieb et al., 2018). Residents in emergency medicine had a higher risk of being involved in motor vehicle crashes while driving home after a night shift compared to other shifts (Steele et al., 1999). Also in residents, long working hours have been shown to result in performance decrements in sustained attention and simulated driving tasks – these were even comparable to the influence of 0.04 to 0.05 g% blood alcohol concentration (Arnedt et al., 2005). These results highlight the large impact of sleep deprivation and shift work on vehicle control performance. Another high-risk group are pilots. In a survey study, more than 84% of pilots reported incidents of impaired flight performance due to fatigue and almost 28% reported falling asleep during flight (Gregory et al., 2010). Pilots are especially at risk for sleepiness in the context of night flights, jetlag, or prolonged duty hours (Bourgeois-Bougrine et al., 2003). Sleep deprivation and circadian misalignment due to inappropriate work schedules have been associated with aviation accidents (Price & Holley, 1990) and decreases in flight simulator performance (Caldwell et al., 2004). Our results from study III emphasize the detrimental effect of sleep deprivation not only on sustained attention, but also on more complex sensorimotor tasks. Whereas it is not sufficient to measure performance impairments only by tasks of simple attention, the PVT might be useful as a short screening tool to detect risks for performance prior to an operational task (Benderoth et al., 2021). In conjunction with an increased use of appropriate operational performance measures, this could increase safety not only in manual docking, but also in other occupational domains that require high levels of performance under stressful circumstances.

5.5 Conclusion

Working in an extreme environment poses multiple challenges to optimal cognitive and operational performance. In human space flight, highly autonomous long-duration missions will likely aggravate the risk of performance decrements. The presented studies contributed to the understanding of operational performance in a complex manual control skill that is relevant in space flight, but also in other contexts of human-machine interaction. Firstly, we aimed at the improvement of autonomous training efficiency. Introducing a stereoscopic presentation of the *6df* tasks led to initial increments in training progress, however, these were not persistent over time. Secondly, we explored eye tracking to gain additional insight into visual information processing during docking training. Frequency and duration of instrument monitoring were significantly associated with docking performance. These results are a first step to predict operator performance and develop eye tracking-based training interventions for *6df*. Thirdly, we investigated the influence of sleep loss – a frequent stressor in space – on operational performance. In contrast to previous studies, we demonstrated performance decrements due to sleep deprivation in a six DoF manual control task. Impairments in sustained attention were able to explain the impairments in docking accuracy at least partly. Hence, possibilities arise to comprehensively assess an operator’s fitness for duty following sleep loss. Taken together, these findings show several promising roads for a continued development of the *6df* tool as an autonomous training program that is able to support the safety and reliability of manual control operations.

Literature

Abich, J., Parker, J., Murphy, J. S., & Eudy, M. (2021). A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality, 25*(4), 919-933. https://doi.org/10.1007/s10055-020-00498-8

Ahmidi, N., Ishii, M., Fichtinger, G., Gallia, G. L., & Hager, G. D. (2012). An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye-tracking and tool-motion data. *International Forum of Allergy & Rhinology, 2*(6), 507-515. https://doi.org/10.1002/alr.21053

Åhsberg, E., Kecklund, G., Åkerstedt, T., & Gamberale, F. (2000). Shiftwork and different dimensions of fatigue. *International Journal of Industrial Ergonomics*, *26*(4), 457-465. https://doi.org/10.1016/S0169-8141(00)00007-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. https://doi.org/10.1016/j.jsmc.2009.03.001

Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric Disease and Treatment, 3*(5), 553-567.

Anderson, C., Sullivan, J. P., Flynn-Evans, E. E., Cade, B. E., Czeisler, C. A., & Lockley, S. W. (2012). Deterioration of neurobehavioral performance in resident physicians during repeated exposure to extended duration work shifts. *Sleep*, *35*(8), 1137-1146. https://doi.org/10.5665/sleep.2004

Aoki, H., Oman, C. M., Buckland, D. A., & Natapoff, A. (2008). Desktop-VR system for preflight 3D navigation training. *Acta Astronautica*, *63*(7-10), 841-847. https://doi.org/10.1016/j.actaastro.2007.11.001

Aoki, H., Oman, C. M., & Natapoff, A. (2007). Virtual-reality-based 3D navigation training for emergency egress from spacecraft. *Aviation, Space, and Environmental Medicine, 78*(8), 774-783.

Arnedt, J. T., Owens, J., Crouch, M., Stahl, J., & Carskadon, M. A. (2005). Neurobehavioral performance of residents after heavy night call vs after alcohol ingestion. *JAMA, 294*(9), 1025-1033. https://doi.org/10.1001/jama.294.9.1025

Arshad, I., & Ferré, E. R. (2022). Cognition in zero gravity: Effects of non-terrestrial gravity on human behaviour. *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218221113935

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775-779. https://doi.org/10.1016/0005-1098(83)90046-8

Barger, L. K., Flynn-Evans, E. E., Kubey, A., Walsh, L., Ronda, J. M., Wang, W., Wright Jr, K. P., & Czeisler, C. A. (2014). Prevalence of sleep deficiency and use of hypnotic drugs in astronauts before, during, and after spaceflight: an observational study. *The Lancet Neurology*, *13*(9), 904-912. https://doi.org/10.1016/S1474-4422(14)70122-X

Barshi, I., & Dempsey, D. L. (2016). *Evidence report: Risk of performance errors due to training deficiencies* (Report No. JSC-CN-35755). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://ntrs.nasa.gov/citations/20160003868

Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, *34*(5), 581-591. https://doi.org/10.1093/sleep/34.5.581

Basner, M., & Dinges, D. F. (2014). Lost in space: sleep. *The Lancet Neurology*, *13*(9), 860-862. https://doi.org/10.1016/S1474-4422(14)70176-0

Basner, M., Dinges, D. F., Mollicone, D., Ecker, A., Jones, C. W., Hyder, E. C., Di Antonio, A., Savelev, I., Kan, K., Goel, N., Morukov, B. V., & Sutton, J. P. (2013). Mars 520-d mission simulation reveals protracted crew hypokinesis and alterations of sleep duration and timing. *Proceedings of the National Academy of Sciences*, *110*(7), 2635-2640. https://doi.org/10.1073/pnas.1212646110

Basner, M., Dinges, D. F., Mollicone, D. J., Savelev, I., Ecker, A. J., Di Antonio, A., Jones, C. W., Hyder, E. C., Kan, K., Morukov, B. V., & Sutton, J. P. (2014). Psychological and behavioral changes during confinement in a 520-day simulated interplanetary mission to mars. *PLoS One*, *9*(3), e93298. https://doi.org/10.1371/journal.pone.0093298

Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica*, *69*(11-12), 949-959. https://doi.org/10.1016/j.actaastro.2011.07.015

Basner, M., Moore, T. M., Hermosillo, E., Nasrini, J., Dinges, D. F., Gur, R. C., & Johannes, B. (2020). Cognition test battery performance is associated with simulated 6df spacecraft docking performance. *Aerospace Medicine and Human Performance*, *91*(11), 861-867. https://doi.org/10.3357/AMHP.5602.2020

Basner, M., & Rubinstein, J. (2011). Fitness for duty: A 3 minute version of the Psychomotor Vigilance Test predicts fatigue related declines in luggage screening performance. *Journal of Occupational and Environmental Medicine*, *53*(10), 1146-1154. https://doi.org/10.1097/jom.0b013e31822b8356

Basner, M., Savitt, A., Moore, T. M., Port, A. M., McGuire, S., Ecker, A. J., Nasrini, J., Mollicone, D. J., Mott, C. M., McCann, T., Dinges, D. F., & Gur, R. C. (2015). Development and validation of the cognition test battery for spaceflight. *Aerospace Medicine and Human Performance*, *86*(11), 942-952. https://doi.org/10.3357/AMHP.4343.2015

Baulk, S. D., Biggs, S. N., Reid, K., van den Heuvel, C. J., & Dawson, D. (2008). Chasing the silver bullet: measuring driver fatigue using simple and complex tasks. *Accident Analysis & Prevention*, *40*(1), 396-402. https://doi.org/10.1016/j.aap.2007.07.008

Beard, B. L. (2019). *The Cognition and Fine Motor Skills Test Batteries: Normative Data and Interdependencies* (Report No. NASA/TM—20205008023). National Aeronautics and Space Administration: Ames Research Center. https://ntrs.nasa.gov/citations/20205008023

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276-292. https://doi.org/10.1037/0033-2909.91.2.276

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (2 ed., pp. 142-162). Cambridge University Press.

Bechtel, R. B., & Berning, A. (1991). The third-quarter phenomenon: do people experience discomfort after stress has passed? In A. A. Harrison, Y. A. Clearwater, & C. P. McKay (Eds.), *From Antarctica to outer space* (pp. 261-265). Springer. https://doi.org/10.1007/978-1-4612-3012-0_24

Bednarik, R., Koskinen, J., Vrzakova, H., Bartczak, P., & Elomaa, A.-P. (2018, June 18-21). Blink-based estimation of suturing task workload and expertise in microsurgery. *31st IEEE International Symposium on Computer-Based Medical Systems*, Karlstad, Sweden. https://doi.org/10.1109/CBMS.2018.00048

Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., Russo, M. B., & Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of Sleep Research*, *12*(1), 1-12. https://doi.org/10.1046/j.1365-2869.2003.00337.x

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, *68*(7), 569-579.

Bendak, S., & Rashid, H. S. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics*, *76*, 102928. https://doi.org/10.1016/j.ergon.2020.102928

Benderoth, S., Hörmann, H.-J., Schießl, C., & Elmenhorst, E.-M. (2021). Reliability and validity of a 3-min psychomotor vigilance task in assessing sensitivity to sleep loss and alcohol: fitness for duty in aviation and transportation. *Sleep*, *44*(11), zsab151. https://doi.org/10.1093/sleep/zsab151

Berger, M., Mescheriakov, S., Molokanova, E., Lechner-Steinleitner, S., Seguer, N., & Kozlovskaya, I. (1997). Pointing arm movements in short-and long-term spaceflights. *Aviation, Space, and Environmental Medicine*, *68*(9), 781-787.

Bhavsar, P. (2022). Context-dependent cognitive workload monitoring using pupillometry for control room operators to prevent overload. *IISE transactions on occupational ergonomics and human factors*, *10*(2), 91-103. https://doi.org/10.1080/24725838.2022.2076177

Bloomberg, J. J., Reschke, M. F., Clement, G. R., Mulavara, A. P., & Taylor, L. C. (2015). *Evidence Report: Risk of impaired control of spacecraft/associated systems and decreased mobility due to vestibular/sensorimotor alterations associated with space flight* (Report No. JSC-CN-34446). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://ntrs.nasa.gov/citations/20150018603

Bock, O. (1998). Problems of sensorimotor coordination in weightlessness. *Brain Research Reviews*, *28*(1-2), 155-160. https://doi.org/10.1016/S0165-0173(98)00035-6

Bock, O., Fowler, B., & Comfort, D. (2001). Human sensorimotor coordination during spaceflight: an analysis of pointing and tracking responses during the "Neurolab" Space Shuttle mission. *Aviation, Space, and Environmental Medicine*, *72*(10), 877-883.

Bock, O., Weigelt, C., & Bloomberg, J. J. (2010). Cognitive demand of human sensorimotor performance during an extended space mission: a dual-task study. *Aviation, Space, and Environmental Medicine*, *81*(9), 819-824. https://doi.org/10.3357/asem.2608.2010

Bogue, R. (2021). The rise of surgical robots. *Industrial Robot*, *48*(3), 335-340. https://doi.org/10.1108/ir-01-2021-0020

Bosch Bruguera, M., López Bermúdez, S., Detrell, G., & Ewald, R. (2022, September 18-22). *Development of a Virtual Reality space docking simulator for research and training–A case application in the space analogue SIRIUS-21* [Paper presentation]. 73rd International Astronautical Congress, Paris, France.

Bosch Bruguera, M., Fink, A., Schröder, V., López Bermúdez, S., Dessy, E., van den Berg, F. P., Lawson, G., Dangoisse, C., Possnig, C., Albertsen, N., Pattyn, N., & Ewald, R. (2021). Assessment of the effects of isolation, confinement and hypoxia on spaceflight piloting performance for future space missions-The SIMSKILL experiment in Antarctica. *Acta Astronautica*, *179*, 471-483. https://doi.org/10.1016/j.actaastro.2020.11.019

Bosch Bruguera, M., Ilk, V., Ruber, S., & Ewald, R. (2019). *Use of Virtual Reality for astronaut training in future space missions-Spacecraft piloting for the Lunar Orbital Platform-Gateway (LOP-G)* [Paper presentation]. 70th International Astronautical Congress, Washington D.C., USA.

Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblentz, A. (2003). Perceived fatigue for short- and long-haul flights: a survey of 739 airline pilots. *Aviation, Space, and Environmental Medicine*, *74*(10), 1072-1077.

Brody, A. R. (1988). *Spacecraft flight simulation: A human factors investigation into the man-machine interface between an astronaut and a spacecraft performing docking maneuvers and other proximity operations* (Report No. NASA-CR-177502). National Aeronautics and Space Administration: Ames Research Center. https://ntrs.nasa.gov/citations/19890019649

Brown, C., Hicks, J., Rinaudo, C. H., & Burch, R. (2021). The use of augmented reality and virtual reality in ergonomic applications for education, aviation, and maintenance. *Ergonomics in Design*. Advance online publication. https://doi.org/10.1177/10648046211003469

Bryson, S. (2013). *Virtual reality: A definition history – a personal essay*. arXiv. https://doi.org/10.48550/arXiv.1312.4322

Caldwell, J. A. (2012). Crew schedules, sleep deprivation, and aviation performance. *Current Directions in Psychological Science*, *21*(2), 85-89. https://doi.org/10.1177/0963721411435842

Caldwell, J. A., Caldwell, J. L., Brown, D. L., & Smith, J. K. (2004). The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots. *Military Psychology*, *16*(3), 163-181. https://doi.org/10.1207/s15327876mp1603_2

Carriot, J., Mackrous, I., & Cullen, K. E. (2021). Challenges to the vestibular system in space: how the brain responds and adapts to microgravity. *Frontiers in Neural Circuits*, *15*, 760313. https://doi.org/10.3389/fncir.2021.760313

Caruso, C. C. (2006). Possible broad impacts of long work hours. *Industrial Health*, *44*(4), 531-536. https://doi.org/10.2486/indhealth.44.531

Cheng, Y., Jianliang, Z., Yingli, D., & Wei, Z. (2015, May 23-25). *Research on intuitive controlling of unmanned Lunar Rover* [Paper presentation]. *27th* Chinese Control and Decision Conference, Qingdao, China. https://doi.org/10.1109/CCDC.2015.7162804

Chetwood, A. S., Kwok, K.-W., Sun, L.-W., Mylonas, G. P., Clark, J., Darzi, A., & Yang, G.-Z. (2012). Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surgical Endoscopy*, *26*(7), 2003-2009. https://doi.org/10.1007/s00464-011-2143-x

Chintamani, K., Cao, A., Ellis, R. D., & Pandya, A. K. (2010). Improved telemanipulator navigation during display-control misalignments using augmented reality cues. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *40*(1), 29-39. https://doi.org/10.1109/tsmca.2009.2030166

Chintamani, K., Cao, A., Ellis, R. D., Tan, C.-A., & Pandya, A. K. (2011). An analysis of teleoperator performance in conditions of display-control misalignments with and without movement cues. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 139-155. https://doi.org/10.1177/1555343411409348

Choi, M. K., & Seong, P. H. (2020). A methodology for evaluating human operator's fitness for duty in nuclear power plants. *Nuclear Engineering and Technology*, *52*(5), 984-994. https://doi.org/10.1016/j.net.2019.10.024

Chuang, L. L., Nieuwenhuizen, F. M., & Bülthoff, H. H. (2013). A fixed-based flight simulator study: the interdependence of flight control performance and gaze efficiency. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics: Applications and Services. Proceedings of the 10th International Conference Part II* (pp. 95-104). Springer. https://doi.org/10.1007/978-3-642-39354-9_11

Clément, G., Skinner, A., & Lathan, C. (2013). Distance and size perception in astronauts during long-duration spaceflight. *Life*, *3*(4), 524-537. https://doi.org/10.3390/life3040524

Colvin, K., Dodhia, R., & Dismukes, R. (2005). Is pilots' visual scanning adequate to avoid mid-air collisions? *International Symposium on Aviation Psychology*, 141-146.

Creech, S., Guidi, J., & Elburn, D. (2022, March 5-12). *Artemis: An Overview of NASA's Activities to Return Humans to the Moon* [Paper presentation]. IEEE Aerospace Conference, Big Sky, MT, USA. https://doi.org/10.1109/AERO53065.2022.9843277

Currie, N. J., & Peacock, B. (2002). International space station robotic systems operations-a human factors perspective. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(1), 26-30. https://doi.org/10.1177/154193120204600106

Czeisler, C. A., Wickwire, E. M., Barger, L. K., Dement, W. C., Gamble, K., Hartenbaum, N., Ohayon, M. M., Pelayo, R., Phillips, B., Strohl, K., Tefft, B., Rajaratnam, S. M. W., Malhotra, R., Whiton, K., & Hirshkowitz, M. (2016). Sleep-deprived motor vehicle operators are unfit to drive: a multidisciplinary expert consensus statement on drowsy driving. *Sleep Health*, *2*(2), 94-99. https://doi.org/10.1016/j.sleh.2016.04.003

Davis, J. R., Vanderploeg, J. M., Santy, P. A., Jennings, R. T., & Stewart, D. F. (1988). Space motion sickness during 24 flights of the space shuttle. *Aviation, Space, and Environmental Medicine*, *59*(12), 1185-1189.

Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. *Nature*, *388*(6639), 235-235. https://doi.org/10.1038/40775

De la Torre, G. (2014). Cognitive neuroscience in space. *Life*, *4*(3), 281-294. https://doi.org/10.3390/life4030281

Dembe, A. E., Erickson, J. B., Delbos, R. G., & Banks, S. M. (2005). The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. *Occupational and Environmental Medicine*, *62*(9), 588-597. https://doi.org/10.1136/oem.2004.016667

Demertzi, A., Van Ombergen, A., Tomilovskaya, E., Jeurissen, B., Pechenkova, E., Di Perri, C., Litvinova, L., Amico, E., Rumshiskaya, A., Rukavishnikov, I., Sijbers, J., Sinitsyn, V., Kozlovskaya, I. B., Sunaert, S., Parizel, P. M., van de Heyning, P. H., Laureys, S., & Wuyts, F. L. (2016). Cortical reorganization in an astronaut's brain after long-duration spaceflight. *Brain Structure and Function, 221*(5), 2873-2876. https://doi.org/10.1007/s00429-015-1054-3

Demontis, G. C., Germani, M. M., Caiani, E. G., Barravecchia, I., Passino, C., & Angeloni, D. (2017). Human pathophysiological adaptations to the space environment. *Frontiers in Physiology*, *8*, 547. https://doi.org/10.3389/fphys.2017.00547

Dempsey, D. L., & Barshi, I. (2020). Applying Research-Based Training Principles: Toward Crew-Centered, Mission-Oriented Space Flight Training. In L. Blackwell Landon, K. J. Slack, & E. Salas (Eds.), *Psychology and human performance in space programs* (1st ed., pp. 63-80). CRC Press.

Di Nocera, F., Camilli, M., & Terenzi, M. (2006). Using the distribution of eye fixations to assess pilots' mental workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(1), 63-65. https://doi.org/10.1177/154193120605000114

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*, *1*(3), 271-285. https://doi.org/10.1518/155534307X255627

Di Stasi, L. L., McCamy, M. B., Martinez-Conde, S., Gayles, E., Hoare, C., Foster, M., Catena, A., & Macknik, S. L. (2016). Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. *Physiology & Behavior*, *153*, 91-96. https://doi.org/10.1016/j.physbeh.2015.10.024

Diana, M., & Marescaux, J. (2015). Robotic surgery. *British Journal of British Surgery*, *102*(2), e15-e28. https://doi.org/10.1002/bjs.9711

Diaz-Piedra, C., Rieiro, H., Cherino, A., Fuentes, L. J., Catena, A., & Di Stasi, L. L. (2019). The effects of flight complexity on gaze entropy: An experimental study with fighter pilots. *Applied Ergonomics*, *77*, 92-99. https://doi.org/10.1016/j.apergo.2019.01.012

Diaz-Piedra, C., Rieiro, H., Suárez, J., Rios-Tejada, F., Catena, A., & Di Stasi, L. L. (2016). Fatigue in the military: towards a fatigue detection test based on the saccadic velocity. *Physiological Measurement*, *37*(9), N62. https://doi.org/10.1088/0967-3334/37/9/N62

Dijk, D.-J., Neri, D. F., Wyatt, J. K., Ronda, J. M., Riel, E., Ritz-De Cecco, A., Hughes, R. J., & Czeisler, C. A. (2003). Sleep, circadian rhythms, and performance during space shuttle missions. In J. C. Buckey & J. L. Homick (Eds.), *The Neurolab Spacelab Mission: Neuroscience Research in Space. Results from the STS-90, Neurolab Spacelab Mission* (pp. 211-221). National Aeronautics and Space Administration. https://ntrs.nasa.gov/citations/20030068197

Dijk, D.-J., Neri, D. F., Wyatt, J. K., Ronda, J. M., Riel, E., Ritz-De Cecco, A., Hughes, R. J., Elliott, A. R., Prisk, G. K., West, J. B., & Czeisler, C. A. (2001). Sleep, performance, circadian rhythms, and light-dark cycles during two space shuttle flights. *American Journal of Physiology – Regulatory, Integrative, and Comparative Physiology, 281*(5), R1647–R1664. https://doi.org/10.1152/ajpregu.2001.281.5.R1647

Dilley, J., Singh, H., Pratt, P., Omar, I., Darzi, A., & Mayer, E. (2020). Visual behaviour in robotic surgery—Demonstrating the validity of the simulated environment. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *16*(2), e2075. https://doi.org/10.1002/rcs.2075

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

Dismukes, R., Goldsmith, T. E., & Kochan, J. A. (2015). *Effects of acute stress on aircrew performance: literature review and analysis of operational aspects* (Report No. NASA/TM-2015-218930). National Aeronautics and Space Administration: Ames Research Center. https://ntrs.nasa.gov/citations/20190002685

Dong, X. (2005). Long workhours, work scheduling and work-related injuries among construction workers in the United States. *Scandinavian Journal of Work, Environment & Health*, *31*(5), 329-335. https://doi.org/10.5271/sjweh.915

Dorrian, J., Rogers, N. L., & Dinges, D. F. (2005). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In C. A. Kushida (Ed.), *Sleep Deprivation – Clinical Issues, Pharmacology, and Sleep Loss Effects* (pp. 39-70). CRC Press. https://doi.org/10.3109/9780203998007-4

Du, X., Zhang, Y., Tian, Y., Huang, W., Wu, B., & Zhang, J. (2015). The influence of spatial ability and experience on performance during spaceship rendezvous and docking. *Frontiers in Psychology*, *6*, 955. https://doi.org/10.3389/fpsyg.2015.00955

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice* (2nd ed.). Springer. https://doi.org/10.1007/978-1-84628-609-4

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M., & Giannopoulos, I. (2018). The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 282. https://doi.org/10.1145/3173574.3173856

Duda, K. R., Prasov, Z., York, S. P., West, J. J., Robinson, S. K., & Handley, P. M. (2015). *Development of an integrated simulation platform for real-time task performance assessment* [Paper presentation]. IEEE Aerospace Conference, Big Sky, MT, USA. https://doi.org/10.1109/AERO.2015.7118974

Durmer, J. S., & Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology, 25*(1), 117-129. https://doi.org/10.1055/s-2005-867080

Ellis, S. R. (2000). Collision in space. *Ergonomics in Design, 8*(1), 4-9. https://doi.org/10.1177/106480460000800102

Elmenhorst, E.-M., Elmenhorst, D., Benderoth, S., Kroll, T., Bauer, A., & Aeschbach, D. (2018). Cognitive impairments by alcohol and sleep deprivation indicate trait characteristics and a potential role for adenosine A1 receptors. *Proceedings of the National Academy of Sciences, 115*(31), 8009-8014. https://doi.org/10.1073/pnas.1803770115

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society Annual Meeting, 32*(2), 97-101. https://doi.org/10.1177/154193128803200221

Ennis, S., Rometsch, F., Saling, F., Fischer, B., Mittler, P., Ferra, L., Vizzi, C., Casini, A. E. M., Marnat, M., Thevenot, C., & Boyer, L. (2021, October 25-29). *Astronaut training on-board the International Space Station using a standalone Virtual Reality headset* [Paper presentation]. 72nd International Astronautical Congress, Dubai, United Arab Emirates.

Finseth, T., Dorneich, M. C., Keren, N., Franke, W. D., & Vardeman, S. (2020, October 11-15). *Designing Training Scenarios for Stressful Spaceflight Emergency Procedures* [Paper presentation]. AIAA/IEEE 39th Digital Avionics Systems Conference, Antonio, TX, USA. https://doi.org/10.1109/DASC50938.2020.9256403

Flynn-Evans, E. E., Barger, L. K., Kubey, A. A., Sullivan, J. P., & Czeisler, C. A. (2016). Circadian misalignment affects sleep and medication use before and during spaceflight. *npj Microgravity, 2*, 15019. https://doi.org/10.1038/npjmgrav.2015.19

Flynn-Evans, E. E., Gregory, K., Arsintescu, L., & Whitmire, A. (2016). *Evidence Report: Risk of performance decrements and adverse health outcomes resulting from sleep loss, circadian desynchronization, and work overload* (Report No. JSC-CN-35774). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://ntrs.nasa.gov/citations/20160003864

Flynn, C. F. (2005). An operational approach to long-duration mission behavioral health and performance factors. *Aviation, Space, and Environmental Medicine, 76*(6), B42-B51.

Folkard, S., & Lombardi, D. A. (2006). Modeling the impact of the components of long work hours on injuries and “accidents”. *American Journal of Industrial Medicine, 49*(11), 953-963. https://doi.org/10.1002/ajim.20307

Fong, T., Rochlis Zumbado, J., Currie, N., Mishkin, A., & Akin, D. L. (2013). Space telerobotics: unique challenges to human–robot collaboration in space. *Reviews of Human Factors and Ergonomics, 9*(1), 6-56. https://doi.org/10.1177/1557234X13510679

Frank, J., Spirkovska, L., McCann, R., Wang, L., Pohlkamp, K., & Morin, L. (2013, March 2-9). *Autonomous mission operations.* IEEE Aerospace Conference, Big Sky, MT, USA. https://doi.org/10.1109/AERO.2013.6496927

Fransen, M., Wilsmore, B., Winstanley, J., Woodward, M., Grunstein, R., Ameratunga, S., & Norton, R. (2006). Shift work and work injury in the New Zealand blood donors' health study. *Occupational and Environmental Medicine*, *63*(5), 352-358. https://doi.org/10.1136/oem.2005.024398

Freina, L., & Canessa, A. (2015). Immersive vs desktop virtual reality in game based learning. In R. Munkvold & L. Kolas (Eds.), *Proceedings of the 9th European Conference on Games Based Learning* (pp. 195-202). Academic Conferences and Publishing International.

Gabriel, G., van Baarsen, B., Ferlazzo, F., Kanas, N., Weiss, K., Schneider, S., & Whiteley, I. (2012). Future perspectives on space psychology: recommendations on psychosocial and neurobehavioural aspects of human spaceflight. *Acta Astronautica*, *81*(2), 587-599. https://doi.org/10.1016/j.actaastro.2012.08.013

Garrett-Bakelman, F. E., Darshi, M., Green, S. J., Gur, R. C., Lin, L., Macias, B. R., McKenna, M. J., Meydan, C., Mishra, T., Nasrini, J., Piening, B. D., Rizzardi, L. F., Sharma, K., Siamwala, J. H., Taylor, L., Hotz Vitaterna, M., Afkarian, M., Afshinnekoo, E., Ahadi, S., …Turek, F.W. (2019). The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. *Science*, *364*(6436), eaau8650. https://doi.org/10.1126/science.aau8650

Gibb, R., Ercoline, B., & Scharff, L. (2011). Spatial disorientation: decades of pilot fatalities. *Aviation, Space, and Environmental Medicine*, *82*(7), 717-724. https://doi.org/10.3357/ASEM.3048.2011

Glaholt, M. G. (2014). *Eye tracking in the cockpit: a review of the relationships between eye movements and the aviators cognitive state* (Report No. DRDC-RDDC-2014-R153). Defence Research and Development Canada, Toronto Research Centre. https://apps.dtic.mil/sti/citations/AD1000097

Glasauer, S., & Mittelstaedt, H. (1998). Perception of spatial orientation in microgravity. *Brain Research Reviews*, *28*(1-2), 185-193. https://doi.org/10.1016/S0165-0173(98)00038-1

Goecks, V. G., Chamitoff, G. E., Borissov, S., Probe, A., McHenry, N. G., Cluck, N., Paddock, E., Schweers, J. P., Bell, B. N., Hoblit, J., Graffagnino, F., & Miralles, E. (2017, January 9-13). *Virtual reality for enhanced 3D astronaut situational awareness during robotic operations in space* [Paper presentation]. AIAA Information Systems-AIAA Infotech@ Aerospace, Grapevine, TX, USA. https://doi.org/10.2514/6.2017-0883

Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Circadian rhythms, sleep deprivation, and human performance. *Progress in Molecular Biology and Translational Science*, *119*, 155-190. https://doi.org/10.1016/B978-0-12-396971-2.00007-5

Goel, N., Rao, H., Durmer, J. S., & Dinges, D. F. (2009). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, *29*(4), 320-339. https://doi.org/10.1055/s-0029-1237117

Goemaere, S., Vansteenkiste, M., & Van Petegem, S. (2016). Gaining deeper insight into the psychological challenges of human spaceflight: the role of motivational dynamics. *Acta Astronautica*, *121*, 130-143. https://doi.org/10.1016/j.actaastro.2015.12.055

Gottlieb, D. J., Ellenbogen, J. M., Bianchi, M. T., & Czeisler, C. A. (2018). Sleep deficiency and motor vehicle crash risk in the general population: a prospective cohort study. *BMC Medicine*, *16*(1), 1-10. https://doi.org/10.1186/s12916-018-1025-7

Gregory, K. B., Winn, W., Johnson, K., & Rosekind, M. R. (2010). Pilot fatigue survey: exploring fatigue factors in air medical operations. *Air Medical Journal*, *29*(6), 309-319. https://doi.org/10.1016/j.amj.2010.07.002

Guo, Y., Freer, D., Deligianni, F., & Yang, G.-Z. (2021). Eye-tracking for performance evaluation and workload estimation in space telerobotic training. *IEEE Transactions on Human-Machine Systems*, *52*(1), 1-11. https://doi.org/10.1109/thms.2021.3107519

Gushchin, V.I., Vinokhodova, A., Komissarova, D., Belakovsky, M., & Orlov, O. (2019). Experiments with Isolation: Past, Present, and Future. *Human Physiology*, *45*(7), 730-739. https://doi.org/10.1134/s0362119719070077

Haidegger, T., & Benyo, Z. (2008). Surgical robotic support for long duration space missions. *Acta Astronautica*, *63*(7-10), 996-1005. https://doi.org/10.1016/j.actaastro.2008.01.005

Haidegger, T., Sándor, J., & Benyó, Z. (2011). Surgery in space: the future of robotic telesurgery. *Surgical Endoscopy*, *25*(3), 681-690. https://doi.org/10.1007/s00464-010-1243-3

Hainley Jr, C. J., Duda, K. R., Oman, C. M., & Natapoff, A. (2013). Pilot performance, workload, and situation awareness during lunar landing mode transitions. *Journal of Spacecraft and Rockets*, *50*(4), 793-801. https://doi.org/10.2514/1.a32267

Hambuchen, K., Marquez, J., & Fong, T. (2021). A review of NASA human-robot interaction in space. *Current Robotics Reports*, *2*(3), 265-272. https://doi.org/10.1007/s43154-021-00062-5

Hancke, T. (2020). Ironies of automation 4.0. *IFAC-PapersOnLine*, *53*(2), 17463-17468. https://doi.org/10.1016/j.ifacol.2020.12.2122

Hargens, A. R., & Vico, L. (2016). Long-duration bed rest as an analog to microgravity. *Journal of Applied Physiology*, *120*(8), 891-903. https://doi.org/10.1152/japplphysiol.00935.2015

Harrington, J. M. (2001). Health effects of shift work and extended hours of work. *Occupational and Environmental Medicine*, *58*(1), 68-72. https://doi.org/10.1136/oem.58.1.68

Harrison, A. A., & Fiedler, E. R. (2011). Introduction: Psychology and the U.S. Space Program. In D. A. Vakoch (Ed.), *Psychology of Space Exploration: Contemporary Research in Historical Perspective* (pp. 1-16). National Aeronautics and Space Administration.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: a review. *Journal of Experimental Psychology: Applied*, *6*(3), 236-249. https://doi.org/10.1037/1076-898x.6.3.236

Haslbeck, A., Schubert, E., Gontar, P., & Bengler, K. (2012). The relationship between pilots' manual flying skills and their visual behavior: a flight simulator study using eye tracking. In S. J. Landry (Ed.), *Advances in Human Aspects of Aviation* (1st ed., pp. 561-568). CRC Press. https://doi.org/10.1201/b12321-63

Heuer, H., Manzey, D., Lorenz, B., & Sangals, J. (2003). Impairments of manual tracking performance during spaceflight are associated with specific effects of microgravity on visuomotor transformations. *Ergonomics*, *46*(9), 920-934. https://doi.org/10.1080/0014013031000107559

Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Katz, E. S., Kheirandish-Gozal, L., Neubauer, D. N., O'Donnell, A. E., Ohayon, M., Peever, J., Rawding, R., Sachdeva, R. C., Setters, B., Vitiello, M V., Catesby Ware, J., & Adams Hillard, P. J. (2015). National Sleep Foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health*, *1*(1), 40-43. https://doi.org/10.1016/j.sleh.2014.12.010

Hockey, G., Åkerstedt, T., Gaillard, A., Manzey, D., Mulder, L., Pattyn, N., & Tattersall, A. (2011). *Human performance in extended space operations* (Report of ESA Topical Team in Psychology). European Space Agency. http://viper.rma.ac.be/pub/Nathalie/hockey11.pdf

Hockey, G., & Robert, J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1-3), 73-93. https://doi.org/10.1016/S0301-0511(96)05223-4

Holland, A. W. (2000). Psychology of spaceflight. *Journal of Human Performance in Extreme Environments*, *5*(1), 1. https://doi.org/10.7771/2327-2937.1002

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Huang, Q., & Lan, J. (2019). Remote control of a robotic prosthesis arm with six-degree-of-freedom for ultrasonic scanning and three-dimensional imaging. *Biomedical Signal Processing and Control*, *54*, 101606. https://doi.org/10.1016/j.bspc.2019.101606

Huemer, V. A., Hayashi, M., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005). Characterizing scan patterns in a spacecraft cockpit simulator: expert vs. novice performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(1), 83-87. https://doi.org/10.1177/154193120504900119

Hufenbach, B., Reiter, T., & Sourgens, E. (2014). ESA strategic planning for space exploration. *Space Policy*, *30*(3), 174-177. https://doi.org/10.1016/j.spacepol.2014.07.009

Hupfeld, K., McGregor, H., Reuter-Lorenz, P., & Seidler, R. (2021). Microgravity effects on the human brain and behavior: dysfunction and adaptive plasticity. *Neuroscience & Biobehavioral Reviews, 122,* 176-189. https://doi.org/10.1016/j.neubiorev.2020.11.017

Hwang, A. D., & Peli, E. (2014). Instability of the perceived world while watching 3D stereoscopic imagery: a likely source of motion sickness symptoms. *i-Perception*, *5*(6), 515-535. https://doi.org/10.1068/i0647

Ivkovic, V., Sommers, B., Cefaratti, D. A., Newman, G., Thomas, D. W., Alexander, D. G., & Strangman, G. E. (2019). Operationally relevant behavior assessment using the robotic on-board trainer for research (ROBoT-r). *Aerospace Medicine and Human Performance*, *90*(9), 819-825. https://doi.org/10.3357/amhp.5324.2019

Jackson, M. L., Gunzelmann, G., Whitney, P., Hinson, J. M., Belenky, G., Rabat, A., & Van Dongen, H. P. (2013). Deconstructing and reconstructing cognitive performance in sleep deprivation. *Sleep Medicine Reviews*, *17*(3), 215-225. https://doi.org/10.1016/j.smrv.2012.06.007

Jillings, S., Van Ombergen, A., Tomilovskaya, E., Rumshiskaya, A., Litvinova, L., Nosikova, I., Pechenkova, E., Rukavishnikov, I., Kozlovskaya, I. B., Manko, O., Danilichev, S., Sunaert, S., Parizel, P. M., Sinitsyn, V., Petrovichev, V., Laureys, S., zu Eulenburg, P., Sijbers, J., Wuyts, F. L., & Jeurissen, B. (2020). Macro-and microstructural changes in cosmonauts' brains after long-duration spaceflight. *Science Advances, 6*(36), eaaz9488. https://doi.org/10.1126/sciadv.aaz9488

Johannes, B., Bronnikov, S., Bubeev, Y., Arz, M., Piechowski, S., & Rittweger, J. (2019). Individual Learning Curves in Manual Control of Six Degrees of Freedom. *International Journal of Applied Psychology, 9*(2), 47-51.

Johannes, B., Bronnikov, S., Bubeev, Y., Dudukin, A., Hoermann, H.-J., Frett, T., Rittweger, J., & Gaillard, A. (2017). A Tool to Facilitate Learning in a Complex Manual Control Task. *International Journal of Applied Psychology*, *7*(4), 79-85.

Johannes, B., Bronnikov, S. V., Bubeev, J. A., Kotrovskaya, T. I., Shastlivtseva, D. V., Piechowski, S., Hoermann, H.-J., & Jordan, J. (2021). Operator's Reliability During Spacecraft Docking Training on Board Mir and ISS. *Aerospace Medicine and Human Performance*, *92*(7), 541-549. https://doi.org/10.3357/amhp.5745.2021

Johannes, B., Bubeev, J. A., Kotrovskaya, T. I., Bronnikov, S. V., Piechowski, S., Elmenhorst, E.-M., Rittweger, J., Jordan, J., & Gaillard, A. W. (2021). Assessing Cognitive Capacity by P3 During a Complex Manual Control Task. *Journal of Psychophysiology*, *35*(1), 43-50. https://doi.org/10.1027/0269-8803/a000260

Johannes, B., Salnitski, V., Dudukin, A., Shevchenko, L., & Bronnikov, S. (2016). Performance Assessment in the PILOT Experiment On Board Space Stations Mir and ISS. *Aerospace Medicine and Human Performance*, *87*(6), 534-544. https://doi.org/10.3357/amhp.4433.2016

Johannes, B., Salnitski, V. P., Artyukhova, A., Savchenko, E., Dudukin, A., & Bubeev, Y. (2011, October 3-7). *The Mars500-experiment "6df" - A teaching and testing approach - First Results* [Paper presentation]. 62nd International Astronautical Congress, Cape Town, South Africa.

Johnson, A. W., Duda, K. R., Sheridan, T. B., & Oman, C. M. (2017). A closed-loop model of operator visual attention, situation awareness, and performance across automation mode transitions. *Human Factors*, *59*(2), 229-241. https://doi.org/10.1177/0018720816665759

Jones, C. W., Basner, M., Mollicone, D. J., Mott, C. M., & Dinges, D. F. (2022). Sleep deficiency in spaceflight is associated with degraded neurobehavioral functions and elevated stress in astronauts on six-month missions aboard the International Space Station. *Sleep*, *45*(3), zsac006. https://doi.org/10.1093/sleep/zsac006

Jones, P. M. (2010). Human performance in space. *Reviews of Human Factors and Ergonomics*, *6*(1), 172-197. https://doi.org/10.1518/155723410x12849346788787

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329-354. https://doi.org/10.1037/0033-295x.87.4.329

Kanas, N. (2011). From Earth's orbit to the outer planets and beyond: Psychological issues in space. *Acta Astronautica*, *68*(5-6), 576-581. https://doi.org/10.1016/j.actaastro.2010.04.012

Kanas, N., & Manzey, D. (2008). *Space psychology and psychiatry* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4020-6770-9

Kanas, N., Salnitskiy, V., Gushin, V., Weiss, D. S., Grund, E. M., Flynn, C., Kozerenko, O., Sled, A., & Marmar, C. R. (2001). Asthenia—does it exist in space? *Psychosomatic Medicine*, *63*(6), 874-880. https://doi.org/10.1097/00006842-200111000-00004

Kanas, N., Sandal, G. M., Boyd, J. E., Gushin, V. I., Manzey, D., North, R., Leon, G. R., Suedfeld, P., Bishop, S. L., Fiedler, E. R., Inoue, N., Johannes, B., Kealey, D. J., Kraft, N. O., Matsuzaki, I., Musson, D., Palinkas, L. A., Salnitski, V., Sipes, W., . . . Wang, J. (2013). Psychology and culture during long-duration space missions. In D. A. Vakoch (Ed.), *On Orbit and Beyond* (pp. 153-184). Springer.

Kane, R. L., Short, P., Sipes, W., & Flynn, C. F. (2005). Development and validation of the spaceflight cognitive assessment tool for windows (WinSCAT). *Aviation, Space, and Environmental Medicine*, *76*(6), B183-B191.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Wickens, C. (2001, March 5-8). *Comparison of expert and novice scan behaviors during VFR flight* [Paper presentation]. 11th International Symposium on Aviation Psychology, Columbus, OH, USA.

Kelly, S. (2017). *Endurance: My year in space, a lifetime of discovery*. Alfred A. Knopf.

Koppelmans, V., Bloomberg, J. J., Mulavara, A. P., & Seidler, R. D. (2016). Brain structural plasticity with spaceflight. *npj Microgravity, 2*(1), 1-8. https://doi.org/10.1038/s41526-016-0001-9

Kornilova, L. (1997). Orientation illusions in spaceflight. *Journal of Vestibular Research*, *7*(6), 429-439. https://doi.org/10.3233/ves-1997-7602

Kornilova, L., Mueller, C., & Chernobyl'skii, L. (1996). Phenomenology of spatial illusory reactions under conditions of weightlessness. *Human Physiology*, *21*(4), 344-351.

Koslowsky, M., & Babkoff, H. (1992). Meta-analysis of the relationship between total sleep deprivation and performance. *Chronobiology International*, *9*(2), 132-136. https://doi.org/10.3109/07420529209064524

Krikalev, S. K., Kalery, A. Y., & Sorokin, I. V. (2010). Crew on the ISS: Creativity or determinism? *Acta Astronautica*, *66*(1-2), 70-73. https://doi.org/10.1016/j.actaastro.2009.05.022

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90-115. https://doi.org/10.1016/j.edurev.2013.10.001

Lefrancois, O., Matton, N., Gourinat, Y., Peysakhovich, V., & Causse, M. (2016, September 26-30). *The role of Pilots' monitoring strategies in flight performance* [Paper presentation]. European Association for Aviation Conference, Cascais, Portugal.

Li, W.-C., Jakubowski, J., Braithwaite, G., & Zhang, J. (2020). Did you see what your trainee pilot is seeing? Integrated eye tracker in the simulator to improve instructors' monitoring performance. *Proceedings of the International Workshop on Eye-Tracking in Aviation*, *1*, 39-46. https://doi.org/10.3929/ethz-b-000407643

Li, Y., & Qu, L. (2021). Stress (Including Oxidative Stress). In L. R. Young & J. P. Sutton (Eds.), *Handbook of Bioastronautics* (pp. 239-253). Springer. https://doi.org/10.1007/978-3-319-12191-8_94

Lieberman, H. R., Tharion, W. J., Shukitt-Hale, B., Speckman, K. L., & Tulley, R. (2002). Effects of caffeine, sleep loss, and stress on cognitive performance and mood during US Navy SEAL training. *Psychopharmacology*, *164*(3), 250-261. https://doi.org/10.1007/s00213-002-1217-9

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1129*(1), 305-322. https://doi.org/10.1196/annals.1417.002

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, *136*(3), 375-389. https://doi.org/10.1037/a0018883

Liu, P., Jiang, G., Liu, Y., & An, M. (2016, September 26). *The utility evaluation of the astronaut virtual training system in layout familiarization training* [Paper presentation]. 6th International Conference on IT Convergence and Security, Prague, Czech Republic. https://doi.org/10.1109/ICITCS.2016.7740376

Lonchakov, Y. V., Kuritsyn, À. A., & Yaropolov, V. I. (2017, September 28). *Information Technologies of Training Cosmonauts for a Space Flight* [Paper presentation]. International Workshop on Engineering Technologies and Computer Science, Moscow, Russia. https://doi.org/10.1109/EnT.2017.24

Loudoun, R., & Allan, C. (2008). The effect of time of day on injury patterns amongst adolescents in Australia. *Applied Ergonomics*, *39*(5), 663-670. https://doi.org/10.1016/j.apergo.2007.12.005

Lounis, C., Peysakhovich, V., & Causse, M. (2019). Flight eye tracking assistant (FETA): Proof of concept. In Stanton, N. (Ed.), *Advances in Human Factors of Transportation. Proceedings of the AHFE International Conference. Advances in Intelligent Systems and Computing, Vol. 964.* Springer. https://doi.org/10.1007/978-3-030-20503-4_66

Lounis, C., Peysakhovich, V., & Causse, M. (2021). Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study. *PLoS One*, *16*(2), e0247061. https://doi.org/10.1371/journal.pone.0247061

Love, S. G., & Reagan, M. L. (2013). Delayed voice communication. *Acta Astronautica, 91*, 89-95. https://doi.org/10.1016/j.actaastro.2013.05.003

Luckhaupt, S. E., Tak, S., & Calvert, G. M. (2010). The prevalence of short sleep duration by industry and occupation in the National Health Interview Survey. *Sleep, 33*(2), 149-159. https://doi.org/10.1093/sleep/33.2.149

Maida, J. C., Bowen, C. K., & Pace, J. (2007). Improving robotic operator performance using augmented reality. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 51*(27), 1635-1639. https://doi.org/10.1177/154193120705102703

Manzey, D. (2000). Monitoring of mental performance during spaceflight. *Aviation, Space, and Environmental Medicine*, *71*(9), A69–A75.

Manzey, D., Lorenz, B., & Poljakov, V. (1998). Mental performance in extreme environments: results from a performance monitoring study during a 438-day spaceflight. *Ergonomics*, *41*(4), 537-559. https://doi.org/10.1080/001401398186991

Manzey, D., Lorenz, B., Schiewe, A., Finell, G., & Thiele, G. (1993). Behavioral aspects of human adaptation to space analyses of cognitive and psychomotor performance in space during an 8-day space mission. *The Clinical Investigator, 71*(9), 725-731. https://doi.org/10.1007/BF00209727

Manzey, D., Schiewe, A., & Fassbender, C. (1995). Psychological countermeasures for extended manned spaceflights. *Acta Astronautica*, *35*(4-5), 339-361. https://doi.org/10.1016/0094-5765(95)98736-S

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing, 3,* 2854-2861. https://doi.org/10.1016/j.promfg.2015.07.783

Marshburn, T. H., Strangman, G. E., Strauss, M. S., & Sutton, J. P. (2003). *SpaceDock: A performance task platform for spaceflight operations* (Report No. JSC-CN-8062). National Aeronautics and Space Administration. https://ntrs.nasa.gov/citations/20100039485

Martin, C., Cegarra, J., & Averty, P. (2011). Analysis of mental workload during en-route air traffic control task execution based on eye-tracking technique. In D. harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. 9th International Conference, EPCE. Lecture Notes in Computer Science, Vol. 6781.* Springer. https://doi.org/10.1007/978-3-642-21741-8_63

Martinez-Marquez, D., Pingali, S., Panuwatwanich, K., Stewart, R. A., & Mohamed, S. (2021). Application of eye tracking technology in aviation, maritime, and construction industries: a systematic review. *Sensors, 21*(13), 4289. https://doi.org/10.3390/s21134289

Matessa, M., & Remington, R. (2005). Eye movements in human performance modeling of space shuttle operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(12), 1114-1118. https://doi.org/10.1177/154193120504901203

McCartt, A. T., Rohrbaugh, J. W., Hammer, M. C., & Fuller, S. Z. (2000). Factors associated with falling asleep at the wheel among long-distance truck drivers. *Accident Analysis & Prevention*, *32*(4), 493-504. https://doi.org/10.1016/S0001-4575(99)00067-6

McKinley, R. A., McIntire, L. K., Schmidt, R., Repperger, D. W., & Caldwell, J. A. (2011). Evaluation of eye metrics as a detector of fatigue. *Human Factors*, *53*(4), 403-414. https://doi.org/10.1177/0018720811411297

Meir Drexler, S., & Wolf, O. T. (2017). Stress and memory consolidation. In N. Axmacher & B. Rasch (Eds.), *Cognitive neuroscience of memory consolidation* (pp. 285-300). Springer. https://doi.org/10.1007/978-3-319-45066-7_17

Menchaca-Brandan, M. A., Liu, A. M., Oman, C. M., & Natapoff, A. (2007). Influence of perspective-taking and mental rotation abilities in space teleoperation. *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, 271–278. https://doi.org/10.1145/1228716.1228753

Mhatre, S. D., Iyer, J., Puukila, S., Paul, A. M., Tahimic, C. G., Rubinstein, L., Lowe, M., Alwood, J. S., Sowa, M. B., Bhattacharya, S., Globus, R. K., & Ronca, A. E. (2021). Neuro-consequences of the spaceflight environment. *Neuroscience & Biobehavioral Reviews, 132,* 908-935. https://doi.org/10.1016/j.neubiorev.2021.09.055

Milstead, L. (2022). *Human Research Program Integrated Research Plan* (Report No. HRP-47065 Revision N). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://humanresearchroadmap.nasa.gov/

Moore, K., & Gugerty, L. (2010). Development of a novel measure of situation awareness: The case for eye movement analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(19), 1650-1654. https://doi.org/10.1177/154193121005401961

Moore, S. T., Dilda, V., Morris, T. R., Yungher, D. A., MacDougall, H. G., & Wood, S. J. (2019). Long-duration spaceflight adversely affects post-landing operator proficiency. *Scientific Reports*, *9*(1), 1-14. https://doi.org/10.1038/s41598-019-39058-9

Moore, T. M., Basner, M., Nasrini, J., Hermosillo, E., Kabadi, S., Roalf, D. R., McGuire, S., Ecker, A. J., Ruparel, K., Port, A. M., Jackson, C. T., Dinges, D. F., & Gur, R. C. (2017). Validation of the cognition test battery for spaceflight in a sample of highly educated adults. *Aerospace Medicine and Human Performance*, *88*(10), 937-946. https://doi.org/10.3357/AMHP.4801.2017

Morad, Y., Barkana, Y., Zadok, D., Hartstein, M., Pras, E., & Bar-Dayan, Y. (2009). Ocular parameters as an objective tool for the assessment of truck drivers fatigue. *Accident Analysis & Prevention*, *41*(4), 856-860. https://doi.org/10.1016/j.aap.2009.04.016

Morphew, E. (2001). Psychological and human factors in long duration spaceflight. *McGill Journal of Medicine*, *6*(1), 74-80. https://doi.org/10.26443/mjm.v6i1.555

Muehlethaler, C. M., & Knecht, C. P. (2016). Situation awareness training for general aviation pilots using eye tracking. *IFAC-PapersOnLine*, *49*(19), 66-71. https://doi.org/10.1016/j.ifacol.2016.10.463

Naeeri, S., Mandal, S., & Kang, Z. (2019). Analyzing pilots’ fatigue for prolonged flight missions: Multimodal analysis approach using vigilance test and eye tracking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 111-115. https://doi.org/10.1177/1071181319631092

National Aeronautics and Space Administration (2018). *National Space Exploration Campaign Report.* https://www.nasa.gov/sites/default/files/atoms/files/nationalspaceexplorationcampaign.pdf

Newman, R. L., & Rupert, A. H. (2020). The magnitude of the spatial disorientation problem in transport airplanes. *Aerospace Medicine and Human Performance*, *91*(2), 65-70. https://doi.org/10.3357/AMHP.5442.2020

Niehorster, D. C., Hildebrandt, M., Smoker, A., Jarodzka, H., & Dahlströhm, N. (2020). Towards eye tracking as a support tool for pilot training and assessment. In V. Peysakhovich, D. Rudi, M. Caussee, P. Kiefer, F. Dehais, & M. Raubal (Eds.), *Eye-Tracking in Aviation. Proceedings of the 1st International Workshop on Eye-Tracking in Aviation* (pp. 17-28). ETH Zurich. https://doi.org/10.3929/ethz-b-000407625

Oberg, J. (1998). Shuttle-Mir's lessons for the international space station. *IEEE Spectrum*, *35*(6), 28-37. https://doi.org/10.1109/6.681969

Olbrich, M., Graf, H., Keil, J., Gad, R., Bamfaste, S., & Nicolini, F. (2018). Virtual reality based space operations–a study of ESA's potential for VR based training and simulation. In J. Chen, & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation. 10th International Conference, VAMR. Lecture Notes in Computer Science, Vol. 10909.* Springer. https://doi.org/10.1007/978-3-319-91581-4_33

Palinkas, L. A. (2007). Psychosocial issues in long-term space flight: overview. *Gravitational and Space Research*, *14*(2), 25-33.

Papadopoulos, G., Georgiadou, P., Papazoglou, C., & Michaliou, K. (2010). Occupational and public health and safety in a changing work environment: An integrated approach for risk assessment and prevention. *Safety Science*, *48*(8), 943-949. https://doi.org/10.1016/j.ssci.2009.11.002

Patel, Z. S., Brunstetter, T. J., Tarver, W. J., Whitmire, A. M., Zwart, S. R., Smith, S. M., & Huff, J. L. (2020). Red risks for a journey to the red planet: The highest priority human health risks for a mission to Mars. *npj Microgravity*, *6*(1), 1-13. https://doi.org/10.1038/s41526-020-00124-6

Peißl, S., Wickens, C. D., & Baruah, R. (2018). Eye-tracking measures in aviation: a selective literature review. *The International Journal of Aerospace Psychology*, *28*(3-4), 98-112. https://doi.org/10.1080/24721840.2018.1514978

Petit, G., Cebolla, A. M., Fattinger, S., Petieau, M., Summerer, L., Cheron, G., & Huber, R. (2019). Local sleep-like events during wakefulness and their relationship to decreased alertness in astronauts on ISS. *npj Microgravity*, *5*(1), 1-9. https://doi.org/10.1038/s41526-019-0069-0

Peysakhovich, V., Lefrançois, O., Dehais, F., & Causse, M. (2018). The neuroergonomics of aircraft cockpits: the four stages of eye-tracking integration to enhance flight safety. *Safety*, *4*(1), 8. https://doi.org/10.3390/safety4010008

Piechowski, S., Johannes, B., Pustowalow, W., Arz, M., Mulder, E., Jordan, J., Wolf, O.T., & Rittweger, J. (2022). Visual attention relates to operator performance in spacecraft docking training. Aerospace Medicine and Human Performance, 93(6), 480–486. https://doi.org/10.3357/AMHP.6005.2022

Piechowski, S., Pustowalow, W., Arz, M., Rittweger, J., Mulder, E., Wolf, O. T., Johannes, B., & Jordan, J. (2020). Virtual reality as training aid for manual spacecraft docking. *Acta Astronautica*, *177*, 731-736. https://doi.org/10.1016/j.actaastro.2020.08.017

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

Pirker, J. (2022, March 5-12). *The Potential of Virtual Reality for Aerospace Applications* [Paper presentation]. IEEE Aerospace Conference, Big Sky, MT, USA. https://doi.org/10.1109/AERO53065.2022.9843324

Porcu, S., Bellatreccia, A., Ferrara, M., & Casagrande, M. J. E. (1998). Sleepiness, alertness and performance during a laboratory simulation of an acute shift of the wake-sleep cycle. *Ergonomics*, *41*(8), 1192-1202. https://doi.org/10.1080/001401398186478

Previc, F. H., Lopez, N., Ercoline, W. R., Daluz, C. M., Workman, A. J., Evans, R. H., & Dillon, N. A. (2009). The effects of sleep deprivation on flight performance, instrument scanning, and physiological arousal in pilots. *The International Journal of Aviation Psychology*, *19*(4), 326-346. https://doi.org/10.1080/10508410903187562

Price, W., & Holley, D. (1990). Shiftwork and safety in aviation. *Occupational Medicine (Philadelphia, Pa.)*, *5*(2), 343-377.

Raaijmakers, J. G. (1990). *Decision making under mental and physical stress.* TNO Institute for Human Factors. https://pure.uva.nl/ws/files/2190559/34113_decision2.pdf

Rahal, R.-M., & Fiedler, S. (2019). Understanding cognitive and affective mechanisms in social psychology through eye-tracking. *Journal of Experimental Social Psychology*, *85*, 103842. https://doi.org/10.1016/j.jesp.2019.103842

Rector, T., Cripe, C., & Casler, J. (2021). Training and Learning for Long Duration Spaceflight. In Stephanidis, C., Antona, M., & Ntoa, S. (Eds.), *HCI International 2021 - Late Breaking Posters. HCII 2021. Communications in Computer and Information Science, Vol. 1499* (pp. 564-571). Springer. https://doi.org/10.1007/978-3-030-90179-0_73

Robinski, M., & Stein, M. (2013). Tracking visual scanning techniques in training simulation for helicopter landing. *Journal of Eye Movement Research*, *6*(2), 1-17. https://doi.org/10.16910/jemr.6.2.3

Roma, P. G., Hursh, S. R., Hienz, R. D., Emurian, H. H., Gasior, E. D., Brinson, Z. S., & Brady, J. V. (2011). Behavioral and biological effects of autonomous versus scheduled mission management in simulated space-dwelling groups. *Acta Astronautica*, *68*(9-10), 1581-1588. https://doi.org/10.1016/j.actaastro.2009.09.034

Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, Technology & Work*, *15*(3), 313-327. https://doi.org/10.1007/s10111-012-0234-7

Rovetto, R. J. (2013). The essential role of human spaceflight. *Space Policy*, *29*(4), 225-228. https://doi.org/10.1016/j.spacepol.2013.08.001

Rowland, L. M., Thomas, M. L., Thorne, D. R., Sing, H. C., Krichmar, J. L., Davis, H. Q., Balwinski, S. M., Peters, R. D., Kloeppel-Wagner, E., Redmond, D. P., Alicandri, E., & Belenky, G. (2005). Oculomotor responses during partial and total sleep deprivation. *Aviation, Space, and Environmental Medicine*, *76*(7), C104-C113.

Roy-O'Reilly, M., Mulavara, A., & Williams, T. (2021). A review of alterations to the brain during spaceflight and the potential relevance to crew in long-duration space exploration. *npj Microgravity*, *7*(1), 1-9. https://doi.org/10.1038/s41526-021-00133-z

Rupp, T. L., Wesensten, N. J., & Balkin, T. J. (2012). Trait-like vulnerability to total and partial sleep loss. *Sleep*, *35*(8), 1163-1172. https://doi.org/10.5665/sleep.2010

Russo, M. B., Kendall, A. P., Johnson, D. E., Sing, H. C., Thorne, D. R., Escolas, S. M., Santiago, S., Holland, D. A., Hall, S. W., & Redmond, D. P. (2005). Visual perception, psychomotor performance, and complex motor performance during an overnight air refueling simulated flight. *Aviation, Space, and Environmental Medicine*, *76*(7), C92-C103.

Salnitski, V., Dudukin, A., & Johannes, B. (2001). Evaluation of operator's reliability in long-term isolation (The "Pilot"-Test). In V. Baranov (Ed.), *Simulation of extended isolation: advances and problems* (pp. 30-50). Slovo.

Sandal, G., & Leon, G. R. (2011). From the Past to the Future. In D. A. Vakoch (Ed.), *Psychology of Space Exploration: Contemporary Research in Historical Perspective* (pp. 195-204). National Aeronautics and Space Administration.

Sauer, J., Hockey, G. R. J., & Wastell, D. G. (1999a). Maintenance of complex performance during a 135-day spaceflight simulation. *Aviation, Space, and Environmental Medicine*, *70*(3), 236-244.

Sauer, J., Hockey, G. R. J., & Wastell, D. G. (1999b). Performance evaluation in analogue space environments: adaptation during an 8-month Antarctic wintering-over expedition. *Aviation, Space, and Environmental Medicine*, *70*(3), 230-235.

Scannella, S., Peysakhovich, V., Ehrig, F., Lepron, E., & Dehais, F. (2018). Assessment of ocular and physiological metrics to discriminate flight phases in real light aircraft. *Human Factors*, *60*(7), 922-935. https://doi.org/10.1177/0018720818787135

Scheer, F. A., Wright Jr, K. P., Kronauer, R. E., & Czeisler, C. A. (2007). Plasticity of the intrinsic period of the human circadian timing system. *PLoS One*, *2*(8), e721. https://doi.org/10.1371/journal.pone.0000721

Schleicher, R., Galley, N., Briest, S., & Galley, L. J. E. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, *51*(7), 982-1010. https://doi.org/10.1080/00140130701817062

Schneider, S., Brümmer, V., Carnahan, H., Dubrowski, A., Askew, C. D., & Strüder, H. K. (2008). What happens to the brain in weightlessness? A first approach by EEG tomography. *Neuroimage*, *42*(4), 1316-1323. https://doi.org/10.1016/j.neuroimage.2008.06.010

Schneider, S., Brümmer, V., Göbel, S., Carnahan, H., Dubrowski, A., & Strüder, H. K. (2007). Parabolic flight experience is related to increased release of stress hormones. *European*

Journal of Applied Physiology, *100*(3), 301-308. https://doi.org/10.1007/s00421-007-0433-8

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors*, *27*(3), 285-300. https://doi.org/10.1177/001872088502700305

Schoofs, D., Preuß, D., & Wolf, O. T. (2008). Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology*, *33*(5), 643-653. https://doi.org/10.1016/j.psyneuen.2008.02.004

Schoofs, D., Wolf, O. T., & Smeets, T. (2009). Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. *Behavioral Neuroscience*, *123*(5), 1066-1075. https://doi.org/10.1037/a0016980

Seidler, R. D., & Mulavara, A. P. (2021). Sensorimotor Adaptation, Including SMS. In L. R. Young & J. P. Sutton (Eds.), *Handbook of Bioastronautics* (pp. 197-203). Springer. https://doi.org/10.1007/978-3-319-12191-8_22

Seymour, N. E. (2008). VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World Journal of Surgery*, *32*(2), 182-188. https://doi.org/10.1007/s00268-007-9307-9

Shelhamer, M. (2017). Why send humans into space? Science and non-science motivations for human space flight. *Space Policy*, *42*, 37-40. https://doi.org/10.1016/j.spacepol.2017.10.001

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews*, *68*, 651-668. https://doi.org/10.1016/j.neubiorev.2016.06.038

Slack, K. J., Williams, T. J., Schneiderman, J. S., Whitmire, A. M., Picano, J. J., Leveton, L. B., Schmidt, L. L., & Shea, C. (2016). *Risk of Adverse Cognitive or Behavioral Conditions and Psychiatric Disorders: Evidence Report* (Report No. JSC-CN-35772). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://ntrs.nasa.gov/citations/20160004365

Smith, M. G., Kelley, M., & Basner, M. (2020). A brief history of spaceflight from 1961 to 2020: An analysis of missions and astronaut demographics. *Acta Astronautica*, *175*, 290-299. https://doi.org/10.1016/j.actaastro.2020.06.004

St Hilaire, M. A., Kristal, B. S., Rahman, S. A., Sullivan, J. P., Quackenbush, J., Duffy, J. F., Barger, L. K., Gooley, J. J., Czeisler, C. A., & Lockley, S. W. (2019). Using a single daytime performance test to identify most individuals at high-risk for performance impairment during extended wake. *Scientific Reports*, *9*(1), 1-13. https://doi.org/10.1038/s41598-019-52930-y

St Hilaire, M. A., Sullivan, J. P., Anderson, C., Cohen, D. A., Barger, L. K., Lockley, S. W., & Klerman, E. B. (2013). Classifying performance impairment in response to sleep loss using pattern recognition algorithms on single session testing. *Accident Analysis & Prevention*, *50*, 992-1002. https://doi.org/10.1016/j.aap.2012.08.003

Staal, M. A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework* (Report No. NASA/TM-2004-212824). National Aeronautics and Space Agency: Ames Research Center. https://ntrs.nasa.gov/citations/20060017835

Stankovic, A. S., Pryputniewicz, A., Holder, S., York, S. P., Handley, P. M., Karasinski, J. A., Robinson, S. K., West, J. J., & Duda, K. R. (2022). Longitudinal Impacts of Simulated Long-Duration Spaceflight Missions on Operationally Relevant Measures of Human Performance Using a Portable Simulation Platform. *Human Factors*. Advance online publication. https://doi.org/10.1177/00187208221113629

Starcke, K., & Brand, M. (2016). Effects of stress on decisions under uncertainty: A meta-analysis. *Psychological Bulletin*, *142*(9), 909-933. https://doi.org/10.1037/bul0000060

Steele, M. T., Ma, O. J., Watson, W. A., Thomas Jr, H. A., & Muelleman, R. L. (1999). The occupational risk of motor vehicle collisions for emergency medicine residents. *Academic Emergency Medicine*, *6*(10), 1050-1053. https://doi.org/10.1111/j.1553-2712.1999.tb01191.x

Steinberg, F., Kalicinski, M., Dalecki, M., & Bock, O. (2015). Human performance in a realistic instrument-control task during short-term microgravity. *PLoS One*, *10*(6), e0128992. https://doi.org/10.1371/journal.pone.0128992

Steinberg, S. (2019). *Integrated Research Plan to Assess the Combined Effects of Space Radiation, Altered Gravity, and Isolation and Confinement on Crew Health and Performance: Implementation Strategy.* National Aeronautics and Space Administration. https://ntrs.nasa.gov/citations/20205003665

Strangman, G. E., Gur, R., & Basner, M. (2021). Cognitive performance in space. In L. R. Young & J. P. Sutton (Eds.), *Handbook of Bioastronautics* (pp. 399-405). Springer. https://doi.org/10.1007/978-3-319-12191-8_31

Strangman, G. E., Sipes, W., & Beven, G. (2014). Human cognitive performance in spaceflight and analogue environments. *Aviation, Space, and Environmental Medicine*, *85*(10), 1033-1048. https://doi.org/10.3357/ASEM.3961.2014

Strangman, G. E., Thompson, J. H., Strauss, M. M., Marshburn, T. H., & Sutton, J. P. (2005). Functional brain imaging of a complex navigation task following one night of total sleep deprivation: a preliminary study. *Journal of Sleep Research*, *14*(4), 369-375. https://doi.org/10.1111/j.1365-2869.2005.00488.x

Stuster, J. W. (2010). *Behavioral issues associated with long-duration space expeditions: review and analysis of astronaut journal*s (Report No. JSC-CN-21128). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://ntrs.nasa.gov/citations/20100026549

Stuster, J. W., Adolf, J. A., Byrne, V. E., & Greene, M. (2018). *Human exploration of Mars: Preliminary lists of crew tasks* (Report No. NASA/CR-2018-220043). National Aeronautics and Space Administration: Lyndon B. Johnson Space Center. https://ntrs.nasa.gov/citations/20190001401

Stuster, J. W., Bachelard, C., & Suedfeld, P. (2000). The relative importance of behavioral issues during long-duration ICE missions. *Aviation, Space, and Environmental Medicine*, *71*(9), A17-A25.

Suedfeld, P. (2005). Invulnerability, coping, salutogenesis, integration: four phases of space psychology. *Aviation, Space, and Environmental Medicine*, *76*(6), B61-B66.

Summers, R. L., Johnston, S. L., Marshburn, T. H., & Williams, D. R. (2005). Emergencies in space. *Annals of Emergency Medicine*, *46*(2), 177-184. https://doi.org/10.1016/j.annemergmed.2005.02.010

Tian, Y., Zhang, S., Wang, C., Yan, Q., & Chen, S. (2018). Eye tracking for assessment of mental workload and evaluation of RVD interface. In S. Long & B. S. Dhillon (Eds.), *Lecture Notes in Electrical Engineering: Vol. 527. Man-Machine-Environment System Engineering. Proceedings of the 18th International Conference on MMESE* (pp. 11-17). Springer. https://doi.org/10.1007/978-981-13-2481-9_2

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2015). Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surgical Endoscopy*, *29*(2), 405-413. https://doi.org/10.1007/s00464-014-3683-7

Troxel, W. M., Shih, R. A., Pedersen, E. R., Geyer, L., Fisher, M. P., Griffin, B. A., Haas, A. C., Kurz, J., & Steinberg, P. S. (2015). Sleep in the military: Promoting healthy sleep among US servicemembers. *Rand Health Quarterly*, *5*(2), 19.

van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *International Journal of Aviation Psychology*, *22*(1), 78-95. https://doi.org/10.1080/10508414.2012.635129

Van Dongen, H., 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

Van Dongen, P., Baynard, M. D., Maislin, G., & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep*, *27*(3), 423-433. https://doi.org/10.1093/sleep/27.3.423

Van Ombergen, A., Laureys, S., Sunaert, S., Tomilovskaya, E., Parizel, P. M., & Wuyts, F. L. (2017). Spaceflight-induced neuroplasticity in humans as measured by MRI: what do we know so far? *npj Microgravity, 3*(1), 1-12. https://doi.org/10.1038/s41526-016-0010-8

van Schie, M. K., Lammers, G. J., Fronczek, R., Middelkoop, H. A., & van Dijk, J. G. (2021). Vigilance: discussion of related concepts and proposal for a definition. *Sleep Medicine, 83*, 175-181. https://doi.org/10.1016/j.sleep.2021.04.038

Vine, S. J., Masters, R. S., McGrath, J. S., Bright, E., & Wilson, M. R. (2012). Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery, 152*(1), 32-40. https://doi.org/10.1016/j.surg.2012.02.002

Vine, S. J., Moore, L. J., & Wilson, M. R. (2016). An integrative framework of stress, attention, and visuomotor performance. *Frontiers in Psychology, 7*, 1671. https://doi.org/10.3389/fpsyg.2016.01671

Wang, C., Tian, Y., Chen, S., Tian, Z., Jiang, T., & Du, F. (2014). Predicting performance in manually controlled rendezvous and docking through spatial abilities. *Advances in Space Research*, *53*(2), 362-369. https://doi.org/10.1016/j.asr.2013.10.031

Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., Gangwisch, J., Grandner, M. A., Kushida, C., Malhotra, R. K., Martin, J. L., Patel, S. R., Quan, S. F., & Tasali, E. (2015). Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *Journal of Clinical Sleep Medicine, 11*(6), 591-592. https://doi.org/10.5664/jcsm.4758

Watt, D. (1997). Pointing at memorized targets during prolonged microgravity. *Aviation, space, and environmental medicine*, *68*(2), 99-103. https://europepmc.org/article/med/9125097

Weigmann, D. A., & Shappell, S. A. (1997). Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology*, *7*(1), 67-81. https://doi.org/10.1207/s15327108ijap0701_4

Welch, R. B., Hoover, M., & Southward, E. F. (2009). Cognitive performance during prismatic displacement as a partial analogue of "space fog". *Aviation, Space, and Environmental Medicine*, *80*(9), 771-780. https://doi.org/10.3357/ASEM.2415.2009

Whiteley, I., & Bogatyreva, O. (2018). *Toolkit for a space psychologist: To support astronauts in exploration missions to the Moon and Mars*. Cosmic Baby Books.

Wickens, C. D. (2017). Mental workload: assessment, prediction and consequences. In L. Longo & M. C. Leva (Eds.), *Communications in Computer and Information Science: Vol. 726. 1st International Symposium on Human Mental Workload: Models and Applications* (pp. 18-29). Springer. https://doi.org/10.1007/978-3-319-61061-0_2

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2008). Attention-situation awareness (A-SA) model of pilot error. In D. C. Foyle & B. L. Hooey (Eds.), *Human performance modeling in aviation* (pp. 213-239). CRC Press.

Wilson, M. R., Vine, S. J., Bright, E., Masters, R. S., Defriend, D., & McGrath, J. S. (2011). Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. *Surgical Endoscopy*, *25*(12), 3731-3739. https://doi.org/10.1007/s00464-011-1802-2

Winwood, P. C., Winefield, A. H., & Lushington, K. (2006). Work-related fatigue and recovery: the contribution of age, domestic responsibilities and shiftwork. *Journal of Advanced Nursing*, *56*(4), 438-449. https://doi.org/10.1111/j.1365-2648.2006.04011.x

Wolf, O. T. (2017). Stress and memory retrieval: mechanisms and consequences. *Current Opinion in Behavioral Sciences*, *14*, 40-46. https://doi.org/10.1016/j.cobeha.2016.12.001

Wollseiffen, P., Klein, T., Vogt, T., Abeln, V., Strüder, H. K., Stuckenschneider, T., Sanders, M., Claassen, J. A., Askew, C. D., Carnahan, H., & Schneider, S. (2019). Neurocognitive performance is enhanced during short periods of microgravity—part 2. *Physiology & Behavior*, *207*, 48-54. https://doi.org/10.1016/j.physbeh.2019.04.021

Wollseiffen, P., Vogt, T., Abeln, V., Strüder, H. K., Askew, C. D., & Schneider, S. (2016). Neuro-cognitive performance is enhanced during short periods of microgravity. *Physiology & Behavior*, *155*, 9-16. https://doi.org/10.1016/j.physbeh.2015.11.036

Wong, L., Pradhan, S., Karasinski, J., Hu, C., Strangman, G. E., Ivkovic, V., Arsintescu, L., & Flynn-Evans, E. (2020). Performance on the Robotics On-Board Trainer (ROBoT-r) Spaceflight Simulation During Acute Sleep Deprivation. *Frontiers in Neuroscience*, *14*, 697. https://doi.org/10.3389/fnins.2020.00697

Wonsick, M., & Padır, T. (2021, March 6-13). *Human-humanoid robot interaction through virtual reality interfaces* [Paper presentation]. IEEE Aerospace Conference (50100), Big Sky, MT, USA. https://doi.org/10.1109/AERO50100.2021.9438400

Wu, C., Cha, J., Sulek, J., Zhou, T., Sundaram, C. P., Wachs, J., & Yu, D. (2020). Eye-tracking metrics predict perceived workload in robotic surgical skills training. *Human Factors*, *62*(8), 1365-1386. https://doi.org/10.1177/0018720819874544

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., & Akdere, M. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, *2*, 645153. https://doi.org/10.3389/frvir.2021.645153

Yaden, D. B., Iwry, J., Slack, K. J., Eichstaedt, J. C., Zhao, Y., Vaillant, G. E., & Newberg, A. B. (2016). The overview effect: Awe and self-transcendent experience in space flight. *Psychology of Consciousness: Theory, Research, and Practice*, *3*(1), 1-11. https://doi.org/10.1037/cns0000086

Yamazaki, E. M., & Goel, N. (2020). Robust stability of trait-like vulnerability or resilience to common types of sleep deprivation in a large sample of adults. *Sleep*, *43*(6), zsz292. https://doi.org/10.1093/sleep/zsz292

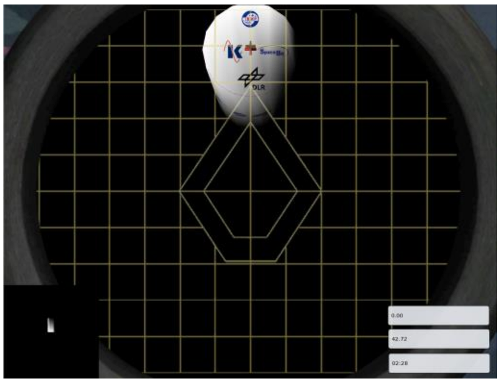
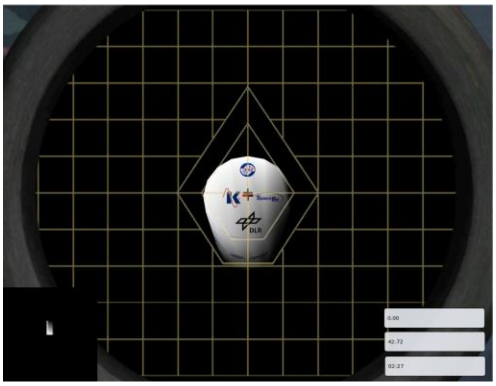
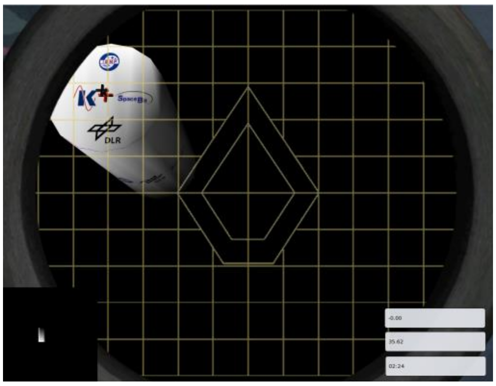
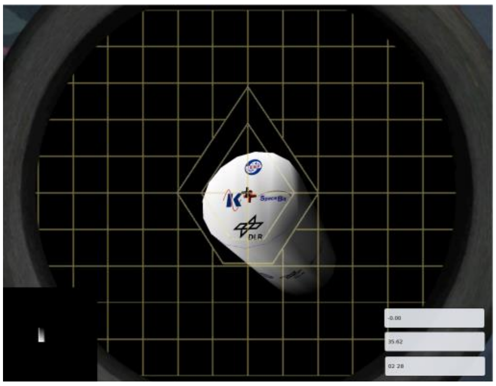
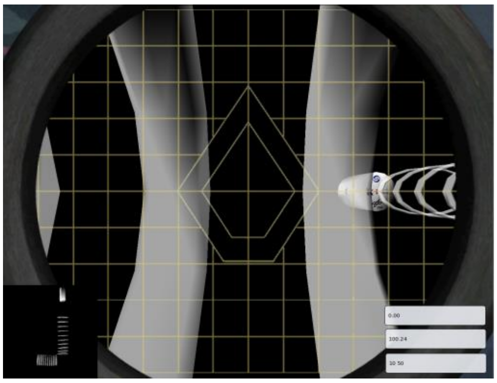
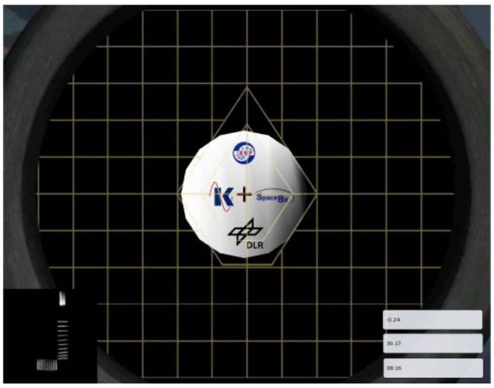
Zhou, X., Ferguson, S. A., Matthews, R. W., Sargent, C., Darwent, D., Kennaway, D. J., & Roach, G. D. (2012). Mismatch between subjective alertness and objective performance under sleep restriction is greatest during the biological night. *Journal of Sleep Research*, *21*(1), 40-49. https://doi.org/10.1111/j.1365-2869.2011.00924.x

Ziv, G. (2016). Gaze behavior and visual attention: A review of eye tracking studies in aviation. *The International Journal of Aviation Psychology*, *26*(3-4), 75-104. https://doi.org/10.1080/10508414.2017.1313096

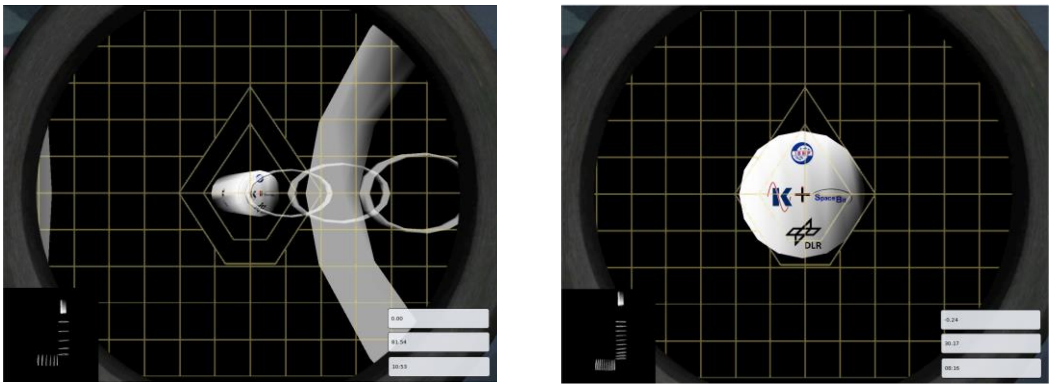
Appendix

Table 2

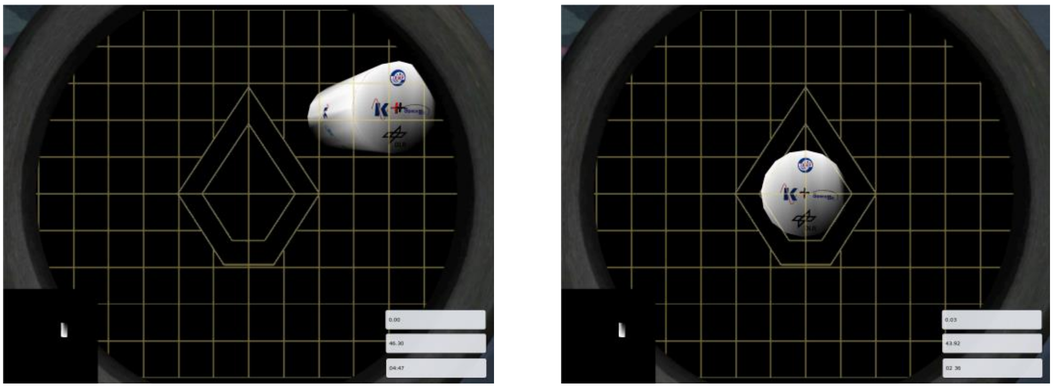
Difficulty levels included in the 6df docking tool

Level	Start position	Final position
1 (1)	Hand control familiarization, only one DoF per task.	
		
2 (2)	Hand control familiarization, two DoF per task.	
		
5	Linear flight to center line (one DoF), stabilization, and approach (guiding rings).	
		

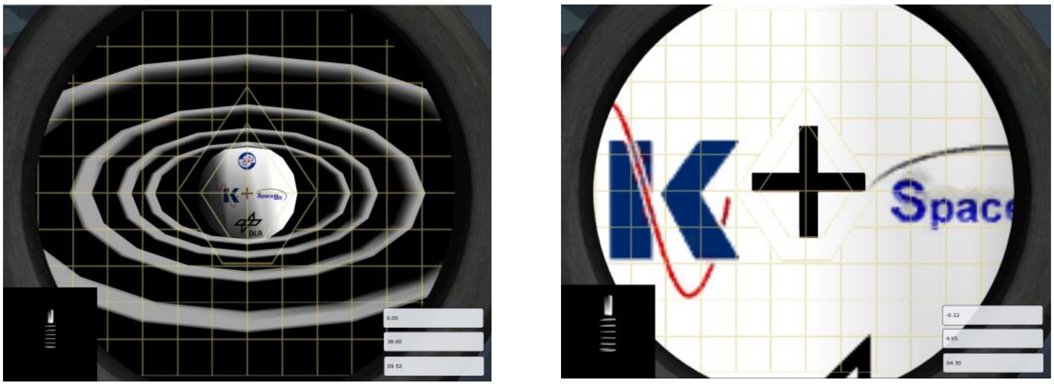
10 Linear flight to center line (two DoF), stabilization, and approach (guiding rings).



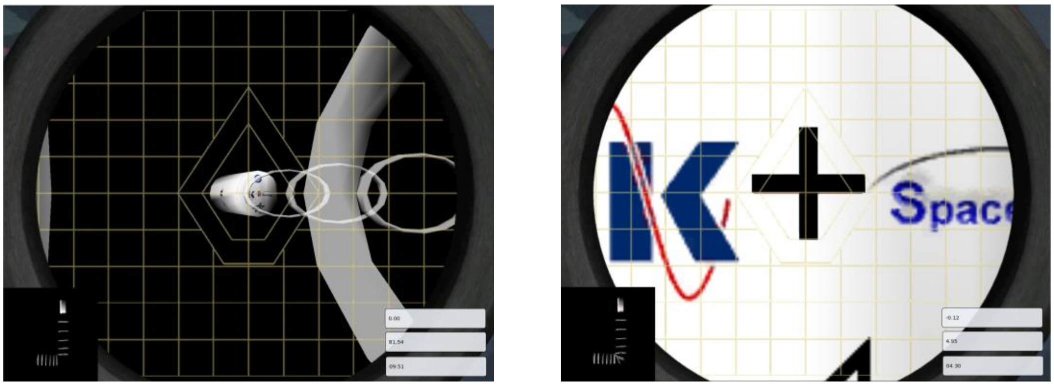
15 (3) Only stabilization on the center line, no approach and docking.



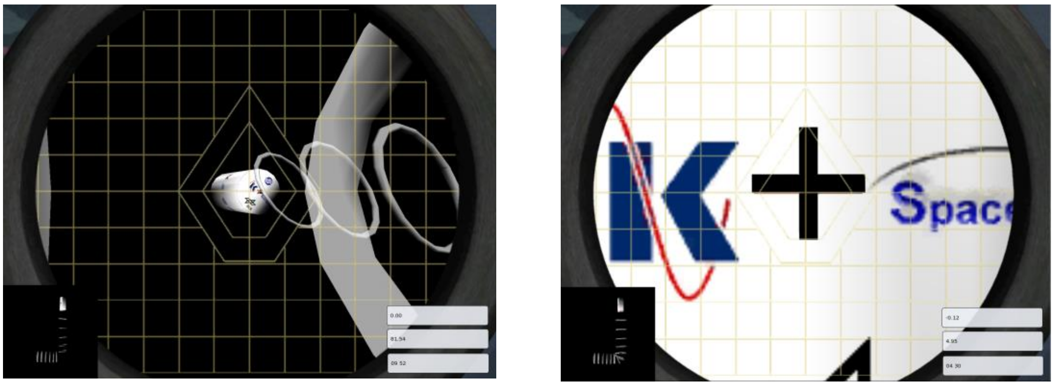
20 Approach and docking contact, only speed is controlled (with and without guiding rings).



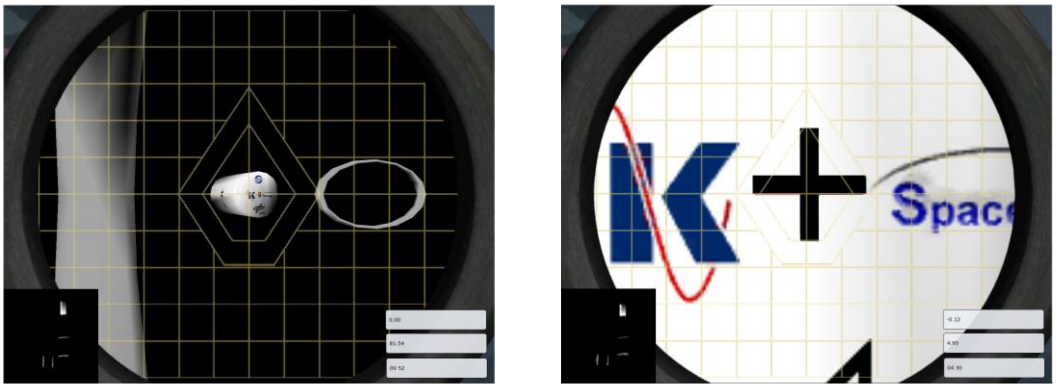
21 Linear flight to center line, stabilization, and docking (guiding rings).



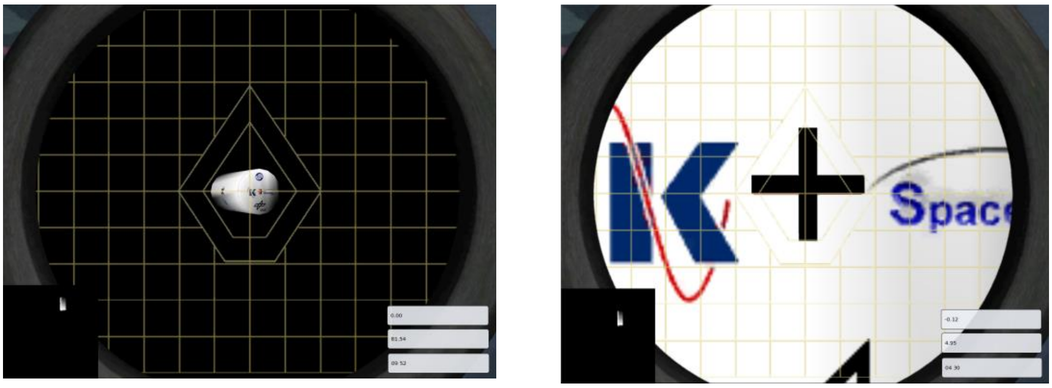
22 Linear flight to center line, stabilization, bank correction, and docking (guiding rings).



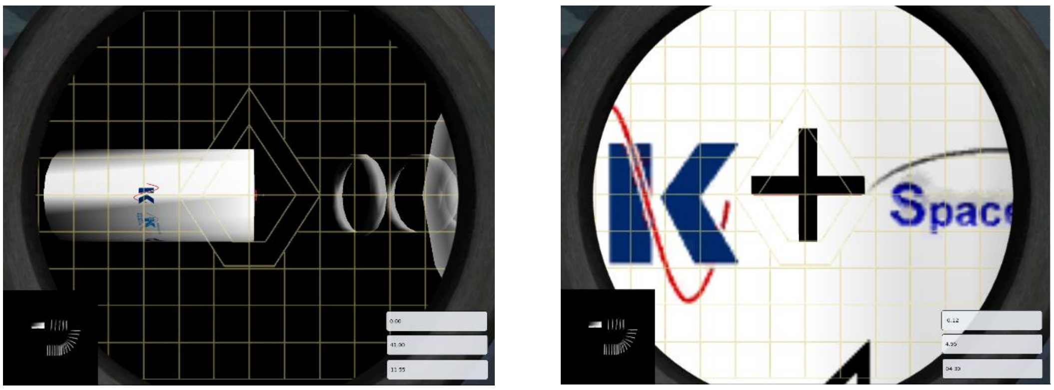
23 Linear flight to center line, stabilization, and docking (fewer guiding rings).



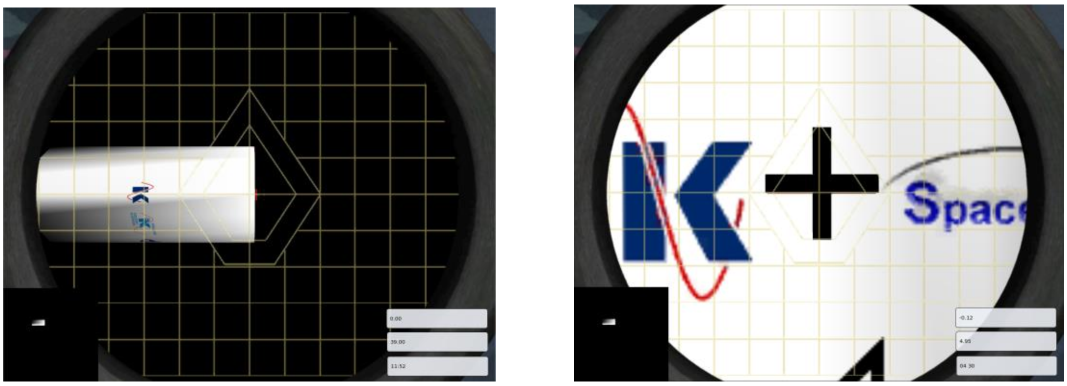
25 (4) Linear flight to center line, stabilization, and docking (no guiding rings).



50 Standard docking maneuver including curved flight-around (guiding rings).



60 (5)	Standard docking maneuver including curved flight-around (omission of guiding rings).



Note. In studies I and II, participants absolved all listed levels. Levels used for the shortened paradigm of study III are indicated in brackets. Each level included 2-12 single docking tasks of similar difficulty.

Table 3

Performance parameters computed by the 6df docking tool

Performance measure	Unit	Description
Accuracy /Quality	%	Aggregate measure indicating overall success of the docking maneuver based on safety ranges; the lowest of the individual performance parameters beneath transformed into a value ranging from 0 to 1
Phi 1	°	Accuracy of yaw angle at docking contact (orientation)
Teta 1	°	Accuracy of pitch angle at docking contact (orientation)
Phi 2	°	Accuracy of horizontal position at docking contact (translation)
Teta 2	°	Accuracy of vertical position at docking contact (translation)
Gamma	°	Accuracy of bank/roll angle at docking contact (orientation)
Distance	m	Distance to docking point (translation)
Forward Speed	m/s	Spacecraft speed during docking (in relation to the docking point - negative scores indicate approach, positive scores distancing)
Stabilization	°	Accuracy of spacecraft stabilization on the center line in safety distance from the station
Approach Trajectory	°	Accuracy of final approach trajectory on the center line
Shear Rate	°/s	Undesired forces on the docking point due to late angle corrections shortly before docking contact
Flight Orientation	°	Accuracy of constantly orienting the spacecraft's vizor to the docking point

Author Contributions

Study I: Piechowski, S., Pustowalow, W., Arz, M., Rittweger, J., Mulder, E., Wolf, O. T., Johannes, B., & Jordan, J. (2020). Virtual reality as training aid for manual spacecraft docking. *Acta Astronautica, 177,* 731-736.

Concept and design of the *6df* study were developed by Bernd Johannes and Sarah Piechowski. Michael Arz and Willi Pustowalow developed the stereoscopic version of the *6df* software. The bed rest study AGBRESA was managed by Edwin Mulder. Data collection was carried out by Sarah Piechowski and Bernd Johannes. Data analysis was conducted by Sarah Piechowski and Bernd Johannes, data visualization by Sarah Piechowski. The original draft was written by Sarah Piechowski. All authors reviewed and edited the manuscript.

Study II: Piechowski, S., Johannes, B., Pustowalow, W., Arz, M., Mulder, E., Jordan, J., Wolf, O. T., & Rittweger, J. (2022). Visual Attention Relates to Operator Performance in Spacecraft Docking Training. *Aerospace Medicine and Human Performance, 93(6),* 480-486.

Concept and design of the *6df* study were developed by Sarah Piechowski and Bernd Johannes. Willi Pustowalow and Michael Arz contributed to eye tracker integration into the *6df* software. The bed rest study AGBRESA was managed by Edwin Mulder. Data collection was carried out by Sarah Piechowski and Bernd Johannes. Data analysis was conducted by Sarah Piechowski and Bernd Johannes, data visualization by Sarah Piechowski. The original draft was written by Sarah Piechowski. All authors reviewed and edited the manuscript.

Study III: Piechowski, S., Kalkoffen, L. J., Benderoth, S., Wolf, O.T., Rittweger, J., Aeschbach, D., & Mühl, C. (2023). Effects Total Sleep Deprivation on Performance in a Manual Spacecraft Docking Task. *Submitted to npj Microgravity.*

Concept and study design were developed by Sarah Piechowski, Christian Mühl, Sibylle Benderoth, and Daniel Aeschbach. Data collection was carried out by Christian Mühl, Sibylle Benderoth, and Sarah Piechowski, with support from student assistants. Data was curated by Sarah Piechowski and Lennard J. Kalkoffen. Data analysis and visualization was conducted by Sarah Piechowski. The original draft was written by Sarah Piechowski. All authors reviewed and edited the manuscript.

Statement

Ich versichere an Eides statt, dass ich die eingereichte Dissertation selbstständig und ohne unzulässige fremde Hilfe verfasst, andere als die in ihr angegebene Literatur nicht benutzt und dass ich alle ganz oder annähernd übernommenen Textstellen sowie verwendete Grafiken, Tabellen und Auswertungsprogramme kenntlich gemacht habe. Es wurde keine kommerzielle Vermittlung oder Beratung in Anspruch genommen. Außerdem versichere ich, dass die vorgelegte elektronische mit der schriftlichen Version der Dissertation übereinstimmt und die Abhandlung in dieser oder ähnlicher Form noch nicht anderweitig als Promotionsleistung vorgelegt und bewertet wurde. Es handelt sich bei der eingereichten Dissertation um vier in Wort und Bild völlig übereinstimmende Exemplare und digitale Abbildungen enthalten ausschließlich die originalen Daten.

Sarah Piechowski

Curriculum Vitae

Personal Information

Name	Sarah Piechowski
Date of birth	20.01.1993 in Bonn
Work address	German Aerospace Center (DLR) Institute of Aerospace Medicine Department of Sleep and Human Factors Linder Höhe 51147 Cologne
E-Mail	sarah.piechowski@dlr.de

Employment

since 2019	Doctoral candidate German Aerospace Center (DLR), Institute of Aerospace Medicine, Cologne
2015 – 2018	Research assistant Max Planck Institute for Research on Collective Goods, Research Group on Cognitive Processes Underlying Economic Decision Making, Bonn

Education

2016 – 2018	Master of Science Psychology University of Bonn Thesis: *#instaneidisch – Eine experimentelle Studie zu Neid, Rumination und Selbstwert im Kontext sozialer Medien* (supervised by Dr. Helmut Appel and Prof. Dr. Jutta Backhaus)
2013 – 2016	Bachelor of Science Psychology University of Bonn Thesis: *Der Zusammenhang von Impulsivität, genetischen Faktoren und psychophysiologischen Maßen der Emotionsregulation: Der 5-HTTLPR und die elektrodermale Aktivität bei Betrachtung emotionaler Stimuli* (supervised by Dr. Thomas Plieger and Dr. Martin Melchers)
2012 – 2013	Studies of Law and Economics University of Bonn
2012	Abitur Gymnasium Aloisiuskolleg, Bonn

Peer-reviewed Publications

Piechowski, S., Johannes, B., Pustowalow, W., Arz, M., Mulder, E., Jordan, J., Wolf, O.T., & Rittweger, J. (2022). Visual attention relates to operator performance in spacecraft docking training. *Aerospace Medicine and Human Performance, 93*(6), 480–486. https://doi.org/10.3357/AMHP.6005.2022

Johannes, B., Bubeev, J.A., Kotrovskaya, T.I., Bronnikov, S.V., Piechowski, S., Elmenhorst, E.-M., Rittweger, J., Jordan, J., & Gaillard, A.W.K. (2021). Assessing Cognitive Capacity by P3 During a Complex Manual Control Task. *Journal of Psychophysiology, 35*(1), 43-50. https://doi.org/10.1027/0269-8803/a000260

Johannes, B., Bronnikov, S. V., Bubeev, J. A., Kotrovskaya, T. I., Shastlivtseva, D. V., Piechowski, S., Hoermann, H.-J., & Jordan, J. (2021). Operator's Reliability During Spacecraft Docking Training on Board Mir and ISS. *Aerospace Medicine and Human Performance*, *92*(7), 541-549. https://doi.org/10.3357/amhp.5745.2021

Piechowski, S., Pustowalow, W., Arz, M., Rittweger, J., Mulder, E., Wolf, O. T., Johannes, B., & Jordan, J. (2020). Virtual reality as training aid for manual spacecraft docking. *Acta Astronautica*, *177*, 731-736. https://doi.org/10.1016/j.actaastro.2020.08.017

Johannes, B.W., Bronnikov, S.V., Bubeev, Y., Kotrovskaya, T.I., Shastlivtseva, D.V., Piechowski, S., Hoermann, H.-J., Rittweger, J., & Jordan, J. (2019). Operational and Experimental Tasks, Performance, and Voice in Space. *Aerospace Medicine and Human Performance, 90*(7), 624-631. https://doi.org/10.3357/AMHP.5351.2019

Johannes, B.W., Bronnikov, S.V., Bubeev, Y., Arz, M., Piechowski, S., & Rittweger, J. (2019). Individual Learning Curves in Manual Control of Six Degrees of Freedom. *International Journal of Applied Psychology*, *9*(2), 47-51.

Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.-M., Fiedler, S., & Nosek, B.A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLoS Biology, 14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Conference Contributions

Piechowski, S., Kalkoffen, L.J., Benderoth, S., Rittweger, J., Aeschbach, D., & Mühl, C. (2022). Effects of total sleep deprivation on manual spacecraft docking performance. *Journal of Sleep Research, 31*(S1), e13740. 26th Conference of the European Sleep Research Society, Athens, Greece. https://doi.org/10.1111/jsr.13738

Piechowski, S., Kalkoffen, L.J., Benderoth, S., Rittweger, J., Aeschbach, D., & Mühl, C. (2022, February 7-10). *Effects of total sleep deprivation on manual spacecraft docking performance* [Poster presentation]. Human Research Program Investigators' Workshop, online.

Piechowski, S., Benderoth, S., Rittweger, J., Aeschbach, D., & Mühl, C. (2021, December 3-5). *Die Effekte von Schlafentzug auf die operationelle Leistung im manuellen Docking* [Conference presentation]. 59. Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Luft- und Raumfahrtmedizin e.V., online.

Piechowski, S., Pustowalow, W., Arz, M., Rittweger, J., Mühl, C., Mulder, E., Wolf, O.T., & Jordan, J. (2021, April 5-8). *Visual Dwell Time on instruments is associated with Manual Docking Performance [Conference presentation]*. 23rd IAA Humans in Space Symposium, online.

Piechowski, S., Pustowalow, W., Arz, M., Rittweger, J., Mulder, E., & Johannes, B. (2019, December 7). *Virtual Reality and Eyetracking during Docking Training [Conference presentation]*. 4th Human Physiology Workshop, Cologne, Germany.

Piechowski, S., Pustowalow, W., Arz, M., Rittweger, J., Mulder, E., & Johannes, B. (2019, October 24-26). *Virtual Reality und Eyetracking beim Docking Training [Conference presentation]*. 57. Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Luft- und Raumfahrtmedizin e.V., Berlin-Schönhagen, Germany.

Acknowledgements

First of all, I would like to express my gratitude to my supervisors Prof. Dr. Oliver T. Wolf and Prof. Dr. Jörn Rittweger for giving me the wonderful opportunity to conduct this research and for guiding me along this academic journey.

This endeavor would not have been possible without the encouraging support of Dr. Bernd Johannes, the originator of the *6df* simulation. Thank you for sharing your wealth of experience and deep passion for space psychology with me.

I would also like to thank Dr. Christian Mühl for always answering my questions and providing feedback, and in particular for the many fruitful discussions we had throughout the last years.

Many thanks to all co-authors for their valuable input and advice. I would also like to acknowledge the ongoing technical support of our external partners Spacebit and Kora, who have been essential to the *6df* simulation.

A special thank you is devoted to all participants involved in the studies of this dissertation – especially to the “terrestrial astronauts” that challenged themselves in the bed rest study. Without their dedication, this research would not have been possible.

I am deeply grateful for the unconditional moral support, encouragement, and patience I received from my family and Sebastian. Thanks to everyone who accompanied me on this journey.