Does Aging Affect Sustained Attention? A Comparative Study Using Attentional Tasks

Master's Thesis

Andrei Suhani

Faculty of Behavioral, Management and Social Sciences, University of Twente

Human Factors and Engineering Psychology

dr. Cesco Willemse

dr. Rob van der Lubbe

03.07.2025

Abstract

As global populations age, workforce shortages in cognitively demanding professions like air traffic control (ATC) are generating discussions about increasing the retirement age. However, such policy shifts raise concerns about the potential impact of aging on core cognitive functions, particularly sustained attention. This study investigates whether sustained attention declines with age, using two tasks: the Sustained Attention to Response Task (SART) and Mackworth's Clock Test (MCT), administered to 29 participants across three age groups (18–44, 45–54, 55–65). Performance was measured using d-prime (d′) scores, alongside secondary measures including eye-tracking data and self-reported measurements. Contrary to initial hypotheses and background neuroscientific expectations, results revealed no significant decline in sustained attention across age groups. In fact, older participants performed comparably or slightly better than their younger counterparts in some cases. Regression analyses also indicated that perceived task difficulty significantly predicted performance. These findings challenge assumptions about cognitive aging and suggest that older adults may retain the capacity to perform attentionally demanding tasks, particularly when bolstered by experience or adaptive strategies. The study has implications for retirement policy, workforce planning, and our broader understanding of cognitive aging.

Does Aging Affect Sustained Attention? A Comparative Study Using Attentional Tasks

Aging Workforce and ATCO Challenges

The global population is experiencing a significant shift as the proportion of aged individuals continues to rise, reshaping societal structures and policies, including retirement age (Baumgartner et al., 2024). This shift is driven by modern trends where longevity and a greater life expectancy mean that people are living healthier and longer lives (Royal NLR - Netherlands Aerospace Centre et al., 2022). In response, many countries are raising the retirement age or even removing mandatory retirement altogether. These policy changes are not only a reflection of extended life expectancy but also serve as a strategic response to mounting staff shortages across various sectors (Chichkanov et al., 2019). In response, there has been a global trend toward increasing retirement age, with some countries even eliminating mandatory retirement. These policy adjustments reflect not only longer lifespans but also aim to address staff shortages (Chichkanov et al., 2019).

Staff shortages, however, are influenced by more than demographic changes alone. Beyond aging populations and declining fertility rates, other factors are also contributing to the shortfall. Technological advancements have altered the demand for specific skills, and in some cases, there is a lack of motivation of potential professionals to fill available roles due to job expectations and changing generational priorities (Bello, 2024). However, some professions are especially sensitive to staff shortages, with particular emphasis on roles where human expertise and cognitive acuity are paramount, such as Air Traffic Control Officers (ATCOs) (NATS Public, 2022).

ATCOs play an essential role in the functioning of the airspace network, where their availability directly impacts network efficiency, resilience, and safety. A decline in the number of ATCOs due to retirement or other factors could have significant repercussions, including reduced resilience in service and potential discomfort or inconvenience for passengers. Moreover, a shortage of trained ATCOs complicates the adoption of new technology, as any new system or procedure requires rigorous testing and acclimatization periods (NATS Public, 2022).

Given these challenges, maintaining a well-staffed and experienced ATCO workforce is essential. However, as experienced ATCOs retire, organizations face a considerable loss of valuable knowledge and expertise. Training new ATCOs to reach the level of seasoned professionals requires time and resources, creating a gap that may impact airspace operations in the short term (NATS Public, 2022). Consequently, several countries, including Switzerland, have proposed raising the retirement age specifically for ATCOs (Baumgartner et al., 2024).

While extending the working lives of ATCOs may solve immediate staffing shortages, this approach raises valid concerns Air traffic control is a profession that demands high levels of cognitive performance and rapid decision-making. Aging is naturally accompanied by various declines in perceptual and cognitive functions, prompting questions about the ability of older individuals to safely perform in such demanding roles (Baumgartner et al., 2024; Boyd & Stolzer, 2024; Murman, 2015).

Dimensions of Aging

To fully understand these implications, it is important to recognize that aging is a complex and multidimensional process. It does not affect everyone in the same way, nor does it affect all cognitive functions in the same way. Age itself encompasses several dimensions: chronological, biological, psychological, and functional age (Straeter et al., 2003). Chronological age, the simplest to measure and usually connected to the idea of age, refers to the number of years a person has lived (Rollandi et al., 2019; Straeter et al., 2003). Biological age reflects the functional state of an individual's body and its alignment with their natural lifespan (Rollandi et al., 2019; Straeter et al., 2003), while psychological age refers to the cognitive and emotional adaptability of an individual, including learning and memory. Finally, functional age considers a person's ability to actively engage and function within society (Straeter et al., 2003).

These distinctions are important, as they underscore that aging is not a one-dimensional process. While some abilities decline, others may remain resilient or even improve with age. For instance, vocabulary and long-term memory tend to show resilience, often staying stable or strengthening over time. Conversely, aspects like processing speed, response time, and short-term memory typically exhibit a gradual decline with age (Boot & Royal NLR - Netherlands Aerospace Centre, 2024).

Crystallized and Fluid Abilities

To better understand how aging affects cognitive abilities, it is helpful to distinguish between two main types of intelligence: crystallized and fluid intelligence. According to Anstey and Low (2004) and Murman (2015), multiple studies reveal that these two types of intelligence develop differently over the lifespan. Crystallized intelligence is grounded in accumulated knowledge and skills drawn from past cognitive experiences and is closely associated with long-term memory (Murman, 2015; Anstey & Low, 2004). This form of intelligence includes vocabulary, factual knowledge, and abstract reasoning and is often tested to assess the depth and stability of an individual’s knowledge. Research suggests that in these crystallized domains, older adults typically perform at levels equal to or even surpassing those of younger adults, reflecting the experience and knowledge gained over a lifetime (Cabeza et al., 2002; Harada et al., 2013).

Fluid intelligence, on the other hand, is more closely linked to the cognitive processes essential for handling new information and solving novel problems. It includes problem-solving, spatial reasoning, and processing speed, all of which require mental agility and depend on working memory capacity. Unfortunately, fluid intelligence tends to decline with age, making tasks that rely heavily on these skills more challenging for older adults (Anstey & Low, 2004; Boot & Royal NLR - Netherlands Aerospace Centre, 2024; Murman, 2015). These findings suggest that while older adults often retain a rich base of knowledge and experience (crystallized intelligence), their ability to quickly process and react to new information (fluid intelligence) may diminish over time. This distinction has important implications for fields like air traffic control, where the cognitive demands include both the application of learned knowledge and the rapid processing of new information in high-stakes situations.

Cognitive Aging and Sustained Attention

To better understand why fluid abilities may decline with age, it is helpful to examine the biological background of cognitive aging. Specifically, age-related changes in the central nervous system (CNS), which comprises the brain and spinal cord, can help explain these cognitive shifts. The CNS is primarily composed of two types of tissue: grey matter and white matter. Grey matter, which appears grey due to its high concentration of neuronal cell bodies, dendrites, and synapses, is essential for processing sensory information, voluntary movement, perception, speech, learning, and cognition (Boot & Royal NLR - Netherlands Aerospace Centre, 2024; Royal NLR - Netherlands Aerospace Centre et al., 2022).

White matter, in contrast, appears white because of its high lipid content, produced by oligodendrocytes, which insulates nerve fibres (axons) and are a type of glial cells. This insulation is crucial for efficient communication between grey matter areas, as well as between the grey matter and the rest of the body. White matter, therefore, facilitates the transmission of information throughout the brain. Around the age of 20, grey matter begins to decline, particularly in the prefrontal cortex, due to a reduction in synaptic connections between neurons. White matter changes occur later, typically beginning around age 40, but decline more rapidly, impacting communication with hippocampal structures and contributing to age-associated memory decline (Royal NLR - Netherlands Aerospace Centre et al., 2022). These anatomical shifts help to explain why cognitive abilities such as conceptual reasoning, memory, and processing speed tend to decrease with age.

Importantly, these structural changes also impact the function of key cortical networks involved in attention and executive control. Staub et al. (2013) highlight that multiple neuropsychological and functional imaging studies have demonstrated activation of the fronto-parietal network during sustained attention tasks. Specifically, the anterior cingulate cortex, dorsolateral prefrontal cortex, and parietal regions, particularly in the right hemisphere, play a central role in mediating sustained attention, though this function is not limited exclusively to these areas. Zanto and Gazzaley (2019) further support the idea that sustained attention relies heavily on the prefrontal cortex, a region known to undergo age-related decline. This decline contributes to reductions in sustained attention performance among older adults. While not all cognitive functions governed by the prefrontal cortex are equally impacted by aging, sustained attention appears to be particularly vulnerable. In fact, performance deficits observed in healthy older adults often resemble those seen in individuals with frontal lobe lesions (Hedden, 2007).

Sustained Attention and Air Traffic Control

Among the cognitive abilities essential to successfully perform the air traffic control activity, attention plays a foundational role. For ATCOs, attention is critical, as their responsibilities demand constant alertness and strong working memory capacity. (Wium &

Eaglestone, 2022). To ensure safety, ATCOs must sustain prolonged attention on radar displays and communication channels while simultaneously being able to rapidly shift their focus between dynamic events, showing both sustained and transient attention (McFarlane, 2024; Zhou et al., 2024). This ability to continuously monitor, prioritize, and respond to stimuli in a high-stakes environment makes attention an essential cognitive resource in air traffic control. However, research shows that attention declines with age (Cabeza et al., 2002; Harada et al., 2013; Murman, 2015; Royal NLR - Netherlands Aerospace Centre et al., 2022; Straeter et al., 2003). As a cognitive process with limited capacity, attention encompasses various forms, including selective, divided, and sustained attention, each impacted by aging in different ways (Straeter et al., 2003). This paper focuses specifically on the aging of sustained attention, also known as vigilance, defined as the capacity to maintain focus on relevant stimuli while ignoring irrelevant distractions in a given context (McAvinue et al., 2012; Straeter et al., 2003).

Sustained attention is crucial for tasks requiring prolonged vigilance, where individuals must detect subtle, often unpredictable signals over extended periods. It involves a persistent state of readiness, or vigilance, to detect these signals, though performance in sustained attention tasks tends to decrease over time. Sustained attention is not only a core aspect of attentional function but also influences other forms of attention, such as selective attention, and contributes to general cognitive capacity (Sarter et al., 2001). Neuroscientific research reveals that sustained attention tasks consistently activate regions in the right hemisphere, particularly the prefrontal and parietal areas (Sarter et al., 2001).

As sustained attention is classified as a fluid cognitive ability, skills that tend to decline with age, it is closely related to functions in the prefrontal cortex, an area known to lose grey matter with age (Royal NLR - Netherlands Aerospace Centre et al., 2022). Consequently, it is inferred that sustained attention performance also declines with age, though studies on this topic yield mixed results. Some studies report a decline in sustained attention among older individuals compared to younger ones, while others find no significant difference or even an advantage for older adults, possibly due to enhanced self-control strategies (McAvinue et al., 2012; Royal NLR - Netherlands Aerospace Centre et al., 2022; Staub et al., 2013; Straeter et al., 2003). The inconsistency in findings indicates that the relationship between aging and sustained attention is still inconclusive, underscoring the importance of further investigation, especially in role of ATCOs, where sustained attention is critical. As existing studies often use small samples, fail to replicate operational conditions, or rely on brief task durations that may not capture the long-term attention demands faced by ATCOs. As policy shifts increasingly favour extended careers, understanding the empirical relationship between age and sustained attention becomes vital for informed regulatory decisions. Moreover, while certain cognitive capacities do decline with age, research suggests that the extensive experience accumulated over the course of an ATCO's career may compensate for these declines, enabling older controllers to maintain high levels of job performance (Nunes & Kramer, 2009).

This paper, therefore, seeks to contribute to the existing body of work by exploring the effects of aging on sustained attention, a topic still lacking in conclusive evidence and recommended for further study by existing research (Staub et al., 2013). Additionally, as the ATCO profession faces an aging workforce and staff shortages, the need to understand the impact of aging on sustained attention becomes even more pressing. Given that prior experience can significantly influence sustained attention performance, potentially hiding age-related effects, this study will focus on individuals without professional experience in ATC roles. To better isolate the effect of age itself, participants will be part of a general population of office workers rather than trained ATCOs. The central research question for this study is: *"What is the effect of aging on sustained attention?"*

To investigate this, an experiment was developed using a between-subject design. While a longitudinal within-subject design would have provided deeper insight into age-related changes over time, it was not feasible due to time constraints. Participants were divided into three age groups and completed two tasks commonly used to measure sustained attention: the Sustained Attention to Response Task (SART) and Mackworth's Clock Test (MCT). The first age group spans a broader range (27 years), while the latter two cover shorter intervals (10 years each). This grouping was chosen to allow more precise analysis of attentional performance in older adults, given the study's aim of exploring whether increasing the retirement age is justifiable from a cognitive standpoint. Task performance was quantified through correct and incorrect responses, translated into hits and false alarm rates in order to calculate a performance coefficient aligned with signal detection theory (SDT). Additionally, subjective questionnaires were administered to support data triangulation and exploratory analysis, capturing self-reported measures of sleepiness, task difficulty, boredom, and perceived performance. Based on prior studies, the hypothesis is that "Older people will have a lower performance in sustained attention tasks compared to their younger counterparts", who are expected to exhibit better performance in sustained attention tasks.

Methods

Design

This study employed a between-subjects design, dividing participants into three distinct age groups: 18–45 (age group 1), 45–55 (age group 2), and 55–65 (age group 3). This design was selected to examine the influence of age on sustained attention performance. The primary dependent variable was the performance coefficient, calculated based on task responses. Age group served as the main independent variable, alongside the subjective measures such as perceived sleepiness, task difficulty, and boredom. **Participants**

In the study 29 NLR (Koninklijk Nederlands Lucht- en Ruimtevaartcentrum) and LVNL (Luchtverkeersleiding Nederland) employees with mean Age (Mage) = 45.3 (SD = 12.3) participated. Age group 1 had Mage = 31.4 (SD = 6.22), age group 2 had a Mage = 50.4 (SD = 2.70), and age group 3 had a Mage = 57.3 (SD = 2.85). Out of the 29 participants, 25 were Dutch, one German, one Belgian, and two Romanians. Additionally, 19 participants were male, and 10 female. The study received Ethical Approval from the University of Twente Ethical Committee (application number 241053) and NLR's Review Board for Human-Subjects Research

Materials

Two primary tasks were used in this experiment: the Sustained Attention to Response Task (SART) and Mackworth's Clock Test (MCT). The SART, developed to assess sustained attention, has been widely used in attentional research and validated by several studies (Smilek et al., 2010). The MCT, originally designed to simulate the continuous monitoring required of British radar operators, has been utilized in a variety of attentional research contexts, including investigations on sleep deprivation (Lichstein et al., 2000; Ustun et al., 2019).

The SART is a computerized task designed to assess sustained attention. In this task, participants respond to non-target numbers (1–9, excluding the target number 3) (Figure 1) and withhold their response when the target number appears. This repetitive response pattern necessitates occasional inhibition, providing a measure of lapses in attention and response control.

The MCT is a well-established measure of sustained attention. Participants monitor a clock face with a moving hand, which typically advances one position at a time but occasionally "double jumps" forward by two positions. Participants must detect these infrequent double jumps by pressing a button, with missed detections and false alarms recorded as errors (Figure 2).

For exploratory purposes, eye-tracking data was also collected using a Tobii Eye Tracker 4C. To ensure proper functionality, the Tobii Experience Driver (version 1.81) was installed. This setup allowed for the collection of additional physiological data to complement the behavioural and subjective measures. **Figure 1**

Difference between normal jump and double jump (depicted by red bar) from starting point in MCT task

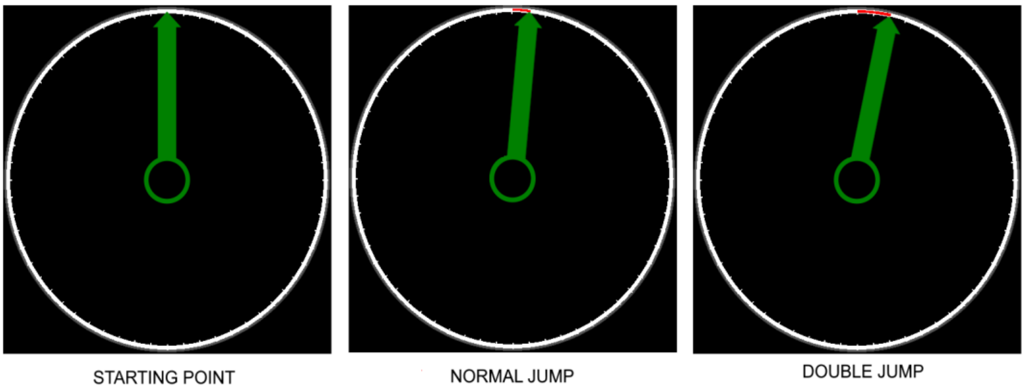
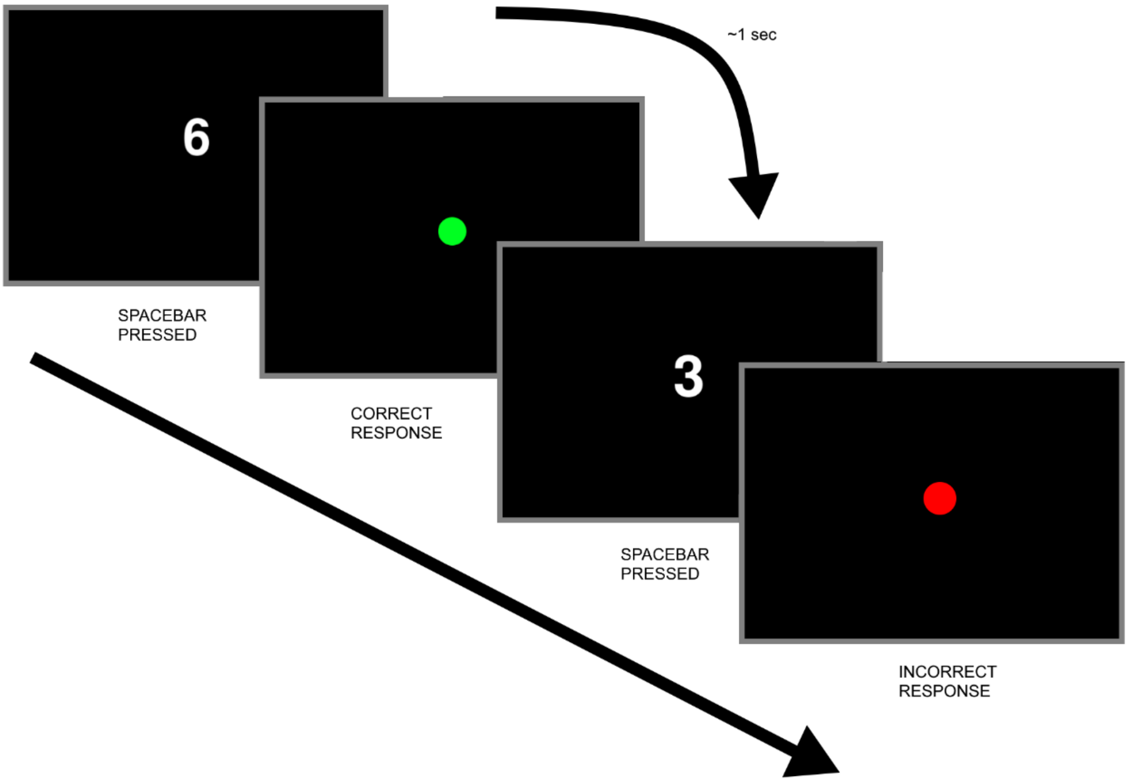


Figure 2

Progression of SART task, with correct and incorrect response



The study included both primary and secondary outcome measures. D-prime (d') is a sensitivity index derived from SDT, which quantifies an individual's ability to distinguish targets (signals) from non-targets (noise). It is calculated using the hit rate (the proportion of correct responses to target stimuli) and the false alarm rate (the proportion of incorrect responses to non-targets). Secondary outcomes were included to triangulate the results and provide additional insights. These secondary measures consisted of eye-tracking data and self-reported survey responses.

The collected data were categorized into objective and subjective measures. Objective measures included task performance and eye-tracking metrics, while subjective measures consisted of participants' survey responses. The eye-tracking data included blink rate, percentage of eye closure (PERCLOS), and the Index of Pupillary Activity (IPA). Survey responses assessed perceived task difficulty, boredom, and sleepiness using standardized scales: one item from the Multidimensional State Boredom Scale (MSBS; Fahlman et al., 2011), the Karolinska Sleepiness Scale (KSS; Shahid et al., 2011), and the Modified Cooper-Harper Scale (Wierwille et al., 1986) (see Appendix C).

Each survey measure was conceptually paired with an eye-tracking metric. Perceived task difficulty was measured using the Modified Cooper-Harper Scale and IPA, both of which assess cognitive workload (Fehringer, 2021; Wierwille et al., 1986). Boredom was measured subjectively using the MSBS and objectively through blink rate, as both are commonly used to assess boredom levels (Esposito et al., 2022; Fahlman et al., 2011). Fatigue was quantified subjectively with the KSS and objectively using PERCLOS, as the KSS is a validated fatigue assessment tool, and PERCLOS has demonstrated efficacy in detecting fatigue based on facial metrics (Rinaldi et al., 2021; Shahid et al., 2011).

Procedure

Before the experiment, participants were seated and provided with a briefing document outlining the study's purpose, tasks, and questionnaires. In addition to reviewing the document, participants received verbal instructions from the experimenter to ensure comprehension. Afterward, they reviewed and signed an informed consent form.

Following consent, participants were introduced to the experimental timeline. Each task lasted 15 minutes, divided into three 5-minute sessions. After each session, participants completed the survey measures. Once the first task was completed, the second task was introduced, following the same structure. The order of task presentation was fully randomized to control for order effects. Before beginning the experimental phase, eye-tracker calibration was conducted to ensure accurate data collection.

After completing both tasks, participants had the opportunity to ask any questions regarding the experiment. As a token of appreciation, they received merchandise from NLR, such as umbrellas, hats, mugs, or books.

Data Analysis

The analysis was performed in RStudio using R 4.4.2. (for used code see Appendix B).

To assess performance in the vigilance tasks (SART and MCT), SDT was employed. This approach accounts for both correct detections (hits) and incorrect responses (false alarms), allowing for a bias-free evaluation of participants' sensitivity to target stimuli. For each participant and session, the number of hits, misses, false alarms, and correct rejections was calculated. These values were then used to compute hit and false alarm rates as follows:

$$Hit\ Rate = \frac{Hits + 0.5}{Hits + Misses + 1}$$

$$False\ Alarm\ Rate\ = \frac{False\ Alarms + 0.5}{False\ Alarms + Correct\ Rejections + 1}$$

These adjusted rates help to avoid extreme values (i.e., 0 or 1). Sensitivity (d′) was then computed using the standard normal transformation:

$$d' = \Phi^{-1}(Hit\ Rate) - \Phi^{-1}(False\ Alarm\ Rate)$$

where Φ^{-1} represents the inverse cumulative normal distribution function.

Survey responses were used as subjective indicators of task engagement and fatigue and were directly related to the study's hypotheses. Participants rated their level of sleepiness, perceived task difficulty, and boredom at multiple points during the tasks. These ratings were used to explore whether perceived performance varied across sessions or age groups and whether they were associated with performance. Descriptive statistics (means and standard deviations) were computed for each of the subjective variables to examine general patterns across the sample.

Prior to transforming the eye-tracking data, it was visualized to assess quality. It was determined that the quality was suboptimal for all measurements except for blink rate. The data necessary to compute PERCLOS were not recorded, likely due to limitations in the data collection software. Additionally, pupil size data included a quality variable, which was frequently poor, making the measurements unreliable. A substantial amount of missing data was also observed for IPA. Consequently, both PERCLOS and IPA were excluded from further analysis.

Only blink rate was retained for analysis. Total blinks per session were counted and divided by five to obtain the mean blink rate per minute per session. Following prior literature, which shows that a typical person blinks between 5 and 45 times per minute depending on cognitive state (Bentivoglio et al., 1997; Biondi et al., 2022), extreme values were excluded by setting a minimum threshold of 2 and a maximum of 50 blinks per minute. Descriptive statistics were used to examine blink rate across sessions and participants.

To investigate the impact of age on performance and potential variables, a linear mixed-effects model was applied. Additionally, Pearson correlation analyses were conducted between subjective ratings, eye-tracking measures, and performance. Several visualizations were created to enhance interpretation and identify trends in the data. Finally, although an analysis of performance over time was initially planned, differences in task structure and score ranges between SART and MCT made it impractical to conduct this analysis across both tasks simultaneously. As a result, the timeline analysis was conducted separately for each task.

Results

Descriptive Analysis

The results of the descriptive analysis are presented in Table 1. Information about the performance measured by d' and survey scores by age group are presented.

Table 1. Descriptive Statistics for Performance and Subjective Measures Across Age Groups

Age Group	Mean d'	SD d'	Mean Sleepiness	SD Sleepiness	Mean Difficulty	SD Difficulty	Mean Boredom	SD Boredom
1	2.09	1.30	5.82	1.98	4.29	1.68	3.05	1.17
2	2.67	1.38	5.43	1.50	4.25	1.28	2.89	0.78
3	2.73	1.32	4.63	1.89	4.74	2.23	3.31	1.50

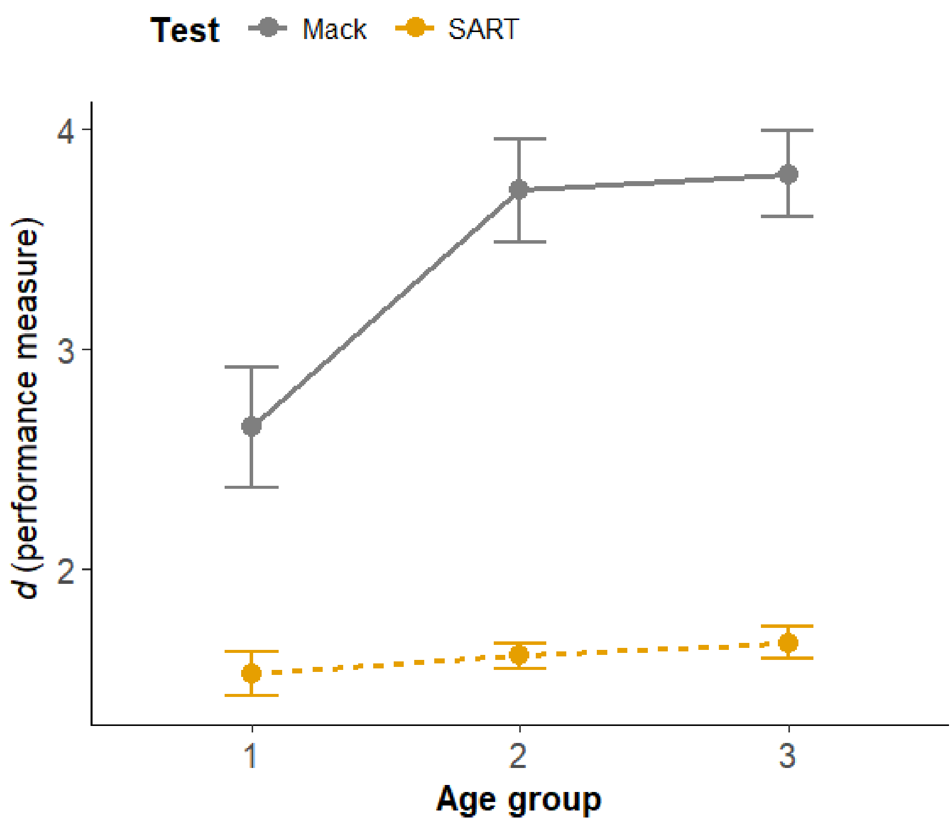
Note. Age group 1 = 18-44 years, age group 2 = 45-54 years, and age group 3 = 55-65 years. Sleepiness, difficulty, and boredom were measured on subjective scales. Sleepiness and difficulty were rated on a 1-10 scale, while boredom was rated on a 1-7 scale. Higher values indicate higher sleepiness, difficulty, or boredom.

In Figure 3, a line plot illustrates the evolution of d' values across age groups. The plot is split by test. It is evident that one of the tests resulted in higher d' values. Figure 4 also shows the d' values across sessions, but it does for each participant. A repeated measures analysis of variance (ANOVA) indicated a statistically significant effect of test type on d', $F(1, 28) = 47.90$, $p < .001$. Pairwise comparisons revealed that d' scores were significantly higher for the Mackworth test ($M = 3.34$, $SE = 0.26$) compared to the SART ($M = 1.60$, $SE =$

0.07), with a mean difference of -1.74, $t(28) = -6.92$, $p < .001$. This difference is further supported by descriptive statistics: the Mackworth task had a higher mean d' score and greater variability ($SD = 1.41$) than the SART ($SD = 0.439$). Additionally, the distributions differed in shape, with the Mackworth task showing a slight leftward skew (-0.594) and a platykurtic distribution (1.76), while the SART task was positively skewed (1.10) and leptokurtic (4.19), indicating a concentration of scores around lower d' values. Given these substantial differences, standardization across tasks does not appear feasible, as their underlying distributions and difficulty levels are distinct. As a result, analyses were conducted separately for each task to accurately assess the effects of age and session on sustained attention performance. Consequently, it was not possible to examine performance trends across all six sessions combined, as each task only included three sessions and their scores could not be standardized across tasks.

Figure 3

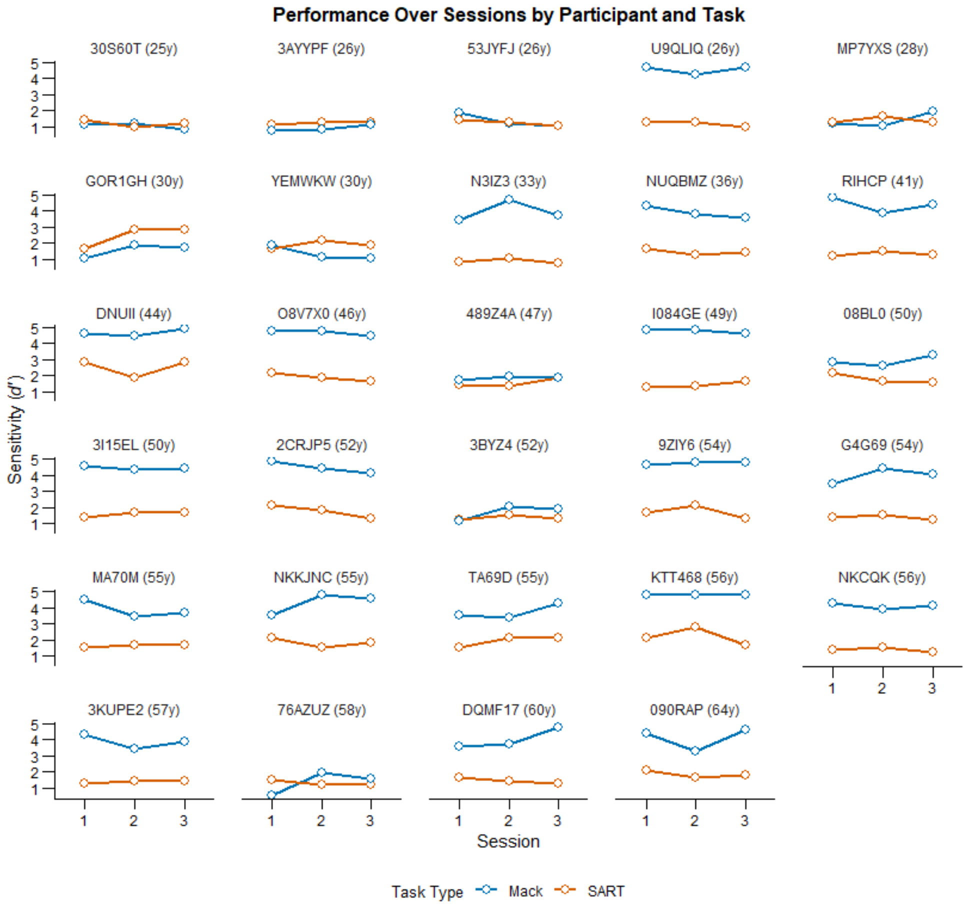
Mean d' performance scores across the three age groups. The solid grey line represents the MCT; the dashed orange line represents the SART.



Note: Error bars show SE.

Figure 4

D' performance scores across the sessions plotted for each participant. The blue line represents the MCT; the orange line represents the SART.



Statistical analysis

Assumption Testing and transformation

Shapiro-Wilk tests were conducted to assess the normality of d' scores across age groups for both tasks. In the Mackworth test, results indicated significant deviations from normality across all age groups: age group 1 ($W = 0.821$, $p < .001$), age group 2 ($W = 0.810$, $p < .001$), and age group 3 ($W = 0.820$, $p < .001$). Similarly, in the SART task, the 18–44 years group showed a significant departure from normality (*W* = 0.824, $p < .001$), while the 45–54 years (*W* = 0.902, $p = .015$) and 55–65 years (*W* = 0.890, $p = .008$) groups also showed significant deviations from normality.

Levene's test for equality of variances was also conducted for each task. For the Mackworth test, the results were significant, $F(2, 84) = 4.64$, $p = .012$, indicating that the assumption of homogeneity of variance was violated. In contrast, for the SART task, Levene's test was not significant, $F(2, 84) = 1.42$, $p = .247$, suggesting that variance was homogenous across age groups.

To assess and potentially improve the normality of d′ scores, various transformations were applied to d' scores for the Mackworth test. A log transformation was selected as the most appropriate alternative. Even after the adjustment, Shapiro-Wilk tests indicated that d' scores remained non-normally distributed across all age groups: 18–44 years (*W* = 0.848, $p < .001$), 45–54 years (*W* = 0.778, $p < .001$), and 55–65 years (*W* = 0.613, $p < .001$). Levene's test for homogeneity of variance on the log-transformed values was significant, $F(2, 84) = 9.82$, $p < .001$, indicating that the assumption of equal variance was still not valid. For the SART test, the log transformation proving to be the most effective transformation. Following this transformation, Shapiro-Wilk tests suggested an improvement in normality. While the 18–44 years group remained significantly non-normal (*W* = 0.923, $p = .022$), the 45–54 years (*W* = 0.927, $p = .060$) and 55–65 years (*W* = 0.939, $p = .112$) groups showed non-significant results, indicating that normality assumptions were met. Additionally, Levene's test for equality of variances was not significant, $F(2, 84) = 2.39$, $p = .098$, confirming that the assumption of homogeneity of variance was satisfied.

Although, the transformations improved on the assumptions, the fix was not complete. Therefore, for the main analysis, the original values were used due to the robustness of the linear mixed-effects model. Outlier analyses were conducted on the d′ (sensitivity) scores using both the z-score and interquartile range (IQR) methods. The z-score method, which identifies values exceeding ±3 standard deviations from the mean, revealed no outliers (all $|z| < 3$). Similarly, the IQR method, which flags values below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR, also identified no outliers. Therefore, all d′ values were retained for subsequent analyses.

Main analysis

A linear mixed-effects model was conducted to examine the effects of age group and session on d' scores for the Mackworth task, with participants being included as a random effect. The analysis revealed that age group did not significantly affect d' scores ($b = 1.26$, $SE = 0.66$, $t(25.10) = 1.90$, $p = .069$). Compared to session 1 (the reference category), session number did not significantly affect d′ scores, neither in session 2 ($b = -0.11$, $SE = 0.24$, $t(36.37) = -0.46$, $p = .645$) nor in session 3 ($b = -0.11$, $SE = 0.24$, $t(36.37) = -0.44$, $p = .663$). In addition, the interaction between session and age group was not significant (all $p > .05$).

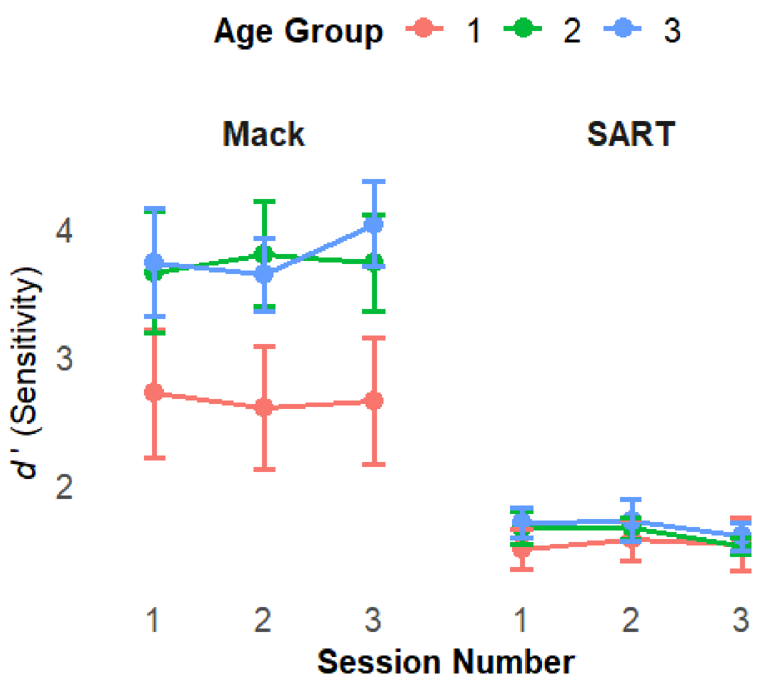
A linear mixed-effects model was conducted to examine the effects of age group and session on d' scores for the SART task, with participants being included as a random effect. The analysis revealed that age group did not significantly affect d' scores, ($b = 0.09$, $SE = 0.23$, $t(39.41) = 0.40$, $p = .693$). Session number also did not significantly affect d' scores for session 2 ($b = 0.01$, $SE = 0.18$, $t(35.44) = 0.04$, $p = .967$), or session 3 ($b = 0.08$, $SE = 0.19$, $t(36.47) = 0.42$, $p = .676$). The interaction between session and age group was not significant for session 2 ($b = -0.04$, $SE = 0.24$, $t(35.44) = -0.18$, $p = .856$, or session 3, ($b = -0.17$, $SE = 0.25$, $t(36.36) = -0.70$, $p = .490$. In addition, the interaction between session and age group was not significant (all $p > .05$).

These findings suggest that age group and session number did not significantly influence performance in the any of the two tasks. Additionally, any interaction between the two variables also did not significantly influence the performance. In Figure 5, the d’ score is plotted by session number and age group. It is also split by task for a better view.

Furthermore, a blink-rate analysis was conducted for both tasks. For the SART task, a linear mixed-effects model revealed no significant effect of blink rate per minute ($\beta = -0.005$, $SE = 0.004$, $t(58.97) = -1.08$, $p = .287$), on performance. For the Mackworth task, a separate linear mixed-effects model was performed to examine the effects of blink rate per minute on the d' performance. The analysis found no significant effect of blink rate per minute ($\beta = -0.012$, $SE = 0.009$, $t(54.07) = -1.40$, $p = .166$). The results for both tasks are suggesting that the blink rate has no effect on the performance of the participants.

Figure 5

Performance plotted by age group, per session number. The plot is split by task



Correlation analysis

To examine predictors of sustained attention performance, separate multiple linear regression analyses were conducted for the Mackworth and SART tasks. Predictor variables were task difficulty, subjective sleepiness (KSS), and blink rate.

For the Mackworth task, the multiple regression model results indicated that task difficulty significantly predicted performance, ($\beta = -0.28$, $SE = 0.12$, $t(51.60) = -2.40$, $p = .019$), meaning that higher perceived difficulty was associated with lower d′ scores. However, neither subjective sleepiness ($p > .05$) nor blink rate ($p > .05$) were significant predictors (see Table A1 in Appendix A). Similarly, for the SART task, difficulty remained the only significant predictor of d′ scores ($\beta = -0.09$, $SE = 0.04$, $t(28.20) = -2.15$, $p = .04$). Sleepiness and blink rate did not significantly predict performance ($p > .05$) (see Table A2 in Appendix A).

Furthermore, a multiple regression analysis was conducted for both tasks, including the previously analysed variables and age. For the Mackworth task, the multiple regression revealed that task difficulty and age were significant predictors of performance ($p < .05$), while sleepiness and blink rate were not ($p > .05$). In contrast, for the SART task, only task difficulty significantly predicted d′ scores ($p < .05$), with age, sleepiness, and blink rate showing no significant effects ($p > .05$) (see Table A3 and A4 in Appendix A).

To further explore these relationships, Pearson correlation coefficients were computed between each predictor and *d′* scores for both tasks. For the Mackworth task, *d′* scores were negatively correlated with task difficulty ($r = -.09$), sleepiness ($r = -.02$), and blink rate ($r = -.01$), and positively correlated with age ($r = .06$). Among these, only the correlations with task difficulty and age were statistically significant ($p < .05$), suggesting that lower subjective difficulty and older age were associated with higher performance. For the SART task, *d′* scores were negatively correlated with task difficulty ($r = -.11$) and blink rate ($r = -.01$), while correlations with sleepiness ($r = .02$) and age ($r = .01$) were positive. However, only the correlation with task difficulty reached statistical significance ($p < .05$).

Discussion

The current research aimed to further explore the impact of age on sustained attention, as previous findings were inconclusive. Previous research has demonstrated contradicting results. Some studies showed that sustained attention declines with age, while others have found that sustained attention performance increases with age or remains the same throughout one's lifetime (McAvinue et al., 2012; Royal NLR - Netherlands Aerospace Centre et al., 2022; Straeter et al., 2003). Based on this mixed evidence, the central research question guiding this study was: *"What is the effect of aging on sustained attention?"* To address this question, it was hypothesized that older individuals would show lower performance in sustained attention tasks compared to their younger counterparts. However, the final analysis did not support this hypothesis. The results indicated no significant differences in performance between the different age groups. Therefore, based on the current data, it can be said that there is no effect of aging on sustained attention.

This outcome is consistent with several earlier studies that similarly found no difference in the development of sustained attention performance across age groups (Berardi et al., 2001; Bunce & Sisa, 2002; Davies & Griew, 1963; Gridley et al., 1986; Griew & Davies, 1962; Neal & Pearson, 1966). However, it is important to recognize that while foundational studies do not present a clear answer on the expected pattern, the broader neuroscientific literature tends to suggest that sustained attention should decline with age. This alternative viewpoint is supported by findings that associate aging with diminished cognitive performance (Murman, 2015; Royal Netherlands Aerospace Centre, 2022; Straeter et al., 2003). These sources reinforce the expectation that age would play a role in task performance.

Nevertheless, the analysis revealed a counterintuitive result regarding performance measures across age groups. Initially, the results indicated that older individuals demonstrated better performance in sustained attention tasks than younger individuals, which is the opposite of the hypothesized direction. While this finding contradicts the expected trend, it does not suggest that aging has no effect; rather, it points toward an effect in the reverse direction. This is also supported by several other studies that have reported comparable results (Brache et al., 2010; Carriere et al., 2010; Jackson & Balota, 2011). Specifically, the descriptive analysis indicated that younger participants performed worse than older participants over time, but this pattern was only evident in the Mackworth Clock Task (MCT) sessions. For the Sustained Attention to Response Task (SART), performance did not differ across age groups. In the SART, sustained attention appeared stable and balanced between age groups, without clear advantages or disadvantages associated with age.

Following the descriptive phase, the analysis was extended using more advanced statistical methods. These more robust analyses confirmed that no significant difference in sustained attention performance existed between the younger and older age groups. Although some influencing factors, such as subjective difficulty ratings, were identified, these did not involve age as a key modifying variable when age was treated categorically. In other words, whether a participant was assigned to the younger or an older group had no significant effect on their overall performance. However, when age was analysed as a continuous variable rather than in discrete categories, it did show an effect. This effect supported the trend observed in the descriptive analysis, namely, performance improving with increasing age, although once again, this was limited to the MCT task. The SART task remained unaffected by age, regardless of whether age was treated as a continuous or categorical variable.

These findings provide further support for studies that suggest no clear decline in sustained attention performance with increasing age. Even though the current study incorporated additional variables, including subjective self-assessments and eye-tracking data, the main outcome remained consistent with the literature that disputes the idea of age-related decline in sustained attention. Therefore, based on the current findings, it can reasonably be concluded that aging, in itself, does not negatively affect performance on sustained attention tasks. As such, these results could contribute to ongoing discussions regarding policies on the retirement age. Since sustained attention appears unaffected by age in this context, raising the retirement age may be supported from a cognitive performance standpoint.

That said, this interpretation is complicated by the current findings where age, treated as a continuous variable, did show a significant effect in one of the tasks. This suggests that while age might influence sustained attention, it might not be the only factor involved, such as task familiarity or domain-specific experience. For instance, although older individuals may experience a slight decline in raw cognitive performance, this may be offset by greater experience and the use of well-developed heuristics. As a result, their overall performance could match that of younger individuals who possess higher cognitive efficiency but less experience. Other moderating variables may contribute, such as it happened with the self-perceived survey, as the perceived difficulty did seem to have an effect on the performance, with the participants that considered a task harder on the self-assessment scale also performed worse than the participants considering the task easier.

These results can be attributed to a few possible reasons, primarily the inability to analyse the performance throughout the entire experiment. As the scores differed greatly between the tasks and the tasks had different attributes, they could not be combined, leading to a halved analysis time. Additionally, although pre-tests showed that a training phase was not necessary, some people performed better during the second session of the tasks. The improvement can be attributed to the novelty of the task and to the learning effect. Another factor that could have influenced the results is the time of testing. As this was an attention test, the time of day when the test was administered could be an important factor, even though the sleepiness scale did not show any influence on the scores. Controlling for this effect is recommended for future research. It is also possible that age is not a strong or consistent predictor of sustained attention performance. Although the second analysis showed a significant effect of age when treated as a continuous variable, it was expected that age groups, particularly in interaction with time, would also show significant effects. The absence of such findings suggests that age alone may not reliably predict performance in sustained attention tasks, or that its influence is more complex and intertwined with other variables.

Limitations and future research

This study had several limitations that should be considered. One main concern of this study is the chosen tasks to measure sustained attention, both the SART and Mackworth Clock Test having certain constraints that could influence the findings. For the SART, there is an ongoing debate regarding the validity of its measures. Although previous studies have demonstrated good validity, emerging evidence suggests that its validity may not be as strong as initially believed (Mensen et al., 2022). Similarly, validity concerns exist for the Mackworth Clock Test. The original version of the test was never publicly released, meaning that all subsequent versions are adaptations rather than exact replications. As a result, formal validation studies on these versions are limited (Lichstein et al., 2000).

Another key concern in this study is the choice of framework used to interpret performance on the MCT. Signal Detection Theory was applied to calculate a performance index based on hits and false alarms. However, due to the limited number of opportunities for sustained false alarms in the MCT, the resulting d' values are often inflated and approach the upper limit of the scale. While this framework has been used in prior research, such as Lichstein et al. (2000), who employed it with even fewer target events (12 over 30 minutes, compared to approximately 15 in 5 minutes in the current study), this study identifies it as a limitation. SDT was chosen partly to ensure consistency across tasks for analysis, but the low event rate in the MCT compromises the reliability of the index. It is therefore recommended that future studies either adopt a more suitable framework for this task or increase the number of possible target events to allow for more accurate estimation of performance.

Additionally, the study was constrained by material limitations related to eye-tracking. The Tobii eye tracker used in this study was designed for casual applications, such as gaming and general computer use, rather than precise measurement of ocular features like pupil size. Consequently, the quality of the collected eye-tracking data was suboptimal, and data had to be discarded. Future research should consider using a professional-grade eye tracker, such as the Smart Eye Pro, to improve measurement accuracy. Furthermore, due to time constraints, a cross-sectional design was employed, incorporating both between- and within-subjects factors. Participants from different age groups were compared rather than tracking the same individuals over time. Although a cross-sectional approach was the most feasible option, a longitudinal design would provide more comprehensive insights by assessing sustained attention performance across an individual's lifespan. Such an approach would yield more valid and detailed data but would also increase both time and financial costs.

The experimental setup may have introduced additional limitations. The inability to fully control the lighting conditions in the room affected eye-tracker performance. Additionally, the fixed positioning of the eye tracker posed challenges due to variations in participants' heights. Adjustments to desk height, chair positioning, or monitor inclination were necessary to accommodate different participants, but these modifications sometimes resulted in poor eye-tracking data, especially when participants moved or leaned out of the calibrated area. To mitigate these issues, future studies should utilize an eye tracker that can dynamically adjust to participants' positions. Another limitation was the participant sample. Although efforts were made to recruit individuals with specialized expertise, the participants were not air traffic control officers (ATCOs). While sustained attention is expected to be relatively stable across healthy individuals, the exclusion of specialized ATCOs can have impacted the internal validity of the study. Thus, for future research, it would be recommended that the study should employ ATCOs, as their training and experience may influence sustained attention through the development of heuristics specific to their work

Conclusion

The present study aimed to explore the impact of age on sustained attention, using two popular tasks: the Sustained Attention to Response Task and Mackworth's Clock Test. Although previous research has often suggested that sustained attention declines with age, the current results did not support this expectation. Interestingly, older participants appeared to perform as well as younger participants, and in specific cases even better. Other expected influences, such as the effect of time spent on task, blink rate, and sleepiness, did not clearly impact performance. The only notable factor affecting outcomes was how difficult participants perceived the task to be, with those who found the task harder generally performing worse.

These findings may be explained by several factors. Differences in task design and scoring prevented a unified analysis, and some participants seemed to perform better simply due to becoming more familiar with the task over time. Additionally, practical limitations such as the use of a basic eye tracker, variations in lighting, and the physical setup of the test environment may have influenced data quality. The absence of trained ATCOs in the sample also limited the applicability of the findings to professional contexts. Despite these limitations, the study offers useful insights for future research. Longitudinal designs that follow individuals over time, rather than comparing separate age groups, would help provide clearer evidence on how attention changes with age. Including more specialized populations, such as ATCOs, and using professional-grade eye-tracking equipment would also enhance the quality of the data and the relevance of the findings.

Importantly, the results of this study may have implications for ongoing discussions about retirement age. Since older adults did not show poorer performance in sustained attention tasks compared to younger adults, the assumption that aging necessarily reduces cognitive performance in this area is not supported. In fact, some older participants performed better, possibly due to greater experience or learned strategies that help maintain focus. While these findings alone do not justify raising the retirement age, they suggest that older individuals can continue to perform well in tasks that rely on sustained attention. Thus, from a cognitive standpoint, increasing the retirement age may be reasonable in professions that require vigilance, provided that individual differences and task-specific skills are also taken into account.

Overall, while the current study did not find strong evidence that age negatively affects sustained attention, it highlights the complexity of the topic and the need for more refined approaches. These results suggest that policies on retirement should consider cognitive performance alongside other factors, rather than relying solely on age as a determining criterion.

References

Anstey, K. J., & Low, L.-F. (2004). Normal cognitive changes in aging. *Australian Family Physician*, *33*(10). https://pubmed.ncbi.nlm.nih.gov/15532151/

Baumgartner, M., Cabon, P., Drogoul, F., Causse, M., Imbert, J., & Mouratille, D. (2024). Scientific Evaluation of the impact of an increase in the retirement age on the Cognitive Functions and Well-Being of Air Traffic Controllers (ATCOs). *AHFE International*. https://doi.org/10.54941/ahfe1005189

Bello, C. (2024, January 31). In data: The EU faces a major demographic decline with 27.3 million fewer people by 2100. *Euronews*. https://www.euronews.com/next/2023/04/04/china-sees-first-population-decline-in-six-decades-where-does-the-eu-stand

Berardi, A., Parasuraman, R., & Haxby, J. V. (2001). Overall vigilance and sustained attention decrements in healthy aging. *Experimental Aging Research*, *27*(1), 19–39. https://doi.org/10.1080/036107301750046124

Boot, M. J. B. & Royal NLR - Netherlands Aerospace Centre. (2024). *Measuring cognitive aging in the aviation domain*.

Boyd, D. D., & Stolzer, A. J. (2024). Is declined cognitive function predictive for fatal accidents involving aging pilots? *Safety*, *10*(3), 71. https://doi.org/10.3390/safety10030071

Brache, K., Scialfa, C., & Hudson, C. (2010). Aging and vigilance: Who has the inhibition deficit? *Experimental Aging Research*, *36*(2), 140–152. https://doi.org/10.1080/03610731003613425

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

Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging Gracefully: Compensatory Brain Activity in High-Performing Older Adults. *NeuroImage*, *17*(3), 1394–1402. https://doi.org/10.1006/nimg.2002.1280

Carriere, J. S. A., Cheyne, J. A., Solman, G. J. F., & Smilek, D. (2010). Age trends for failures of sustained attention. *Psychology and Aging*, *25*(3), 569–574. https://doi.org/10.1037/a0019363

Chichkanov, V., Chistova, E., Tyrsin, A., & Stepanov, A. (2019). Consequences of raising the retirement age for the labor market in the regions of Russia. *MONTENEGRIN JOURNAL OF ECONOMICS*, *15*(1), 31–45. https://doi.org/10.14254/1800-5845/2019.15-1.3

Davies, D. R., & Griew, S. (1963). A Further Note on the Effect of Aging on Auditory Vigilance Performance: the Effect of Low Signal Frequency. *Journal of Gerontology*, *18*(4), 370–371. https://doi.org/10.1093/geronj/18.4.370

Gridley, M. C., Mack, J. L., & Gilmore, G. C. (1986). Age effects on a nonverbal auditory sustained attention task. *Perceptual and Motor Skills*, *62*(3), 911–917. https://doi.org/10.2466/pms.1986.62.3.911

Griew, S., & Davies, D. R. (1962). The effect of aging on auditory vigilance performance. *Journal of Gerontology*, *17*(1), 88–90. https://doi.org/10.1093/geronj/17.1.88

Harada, C. N., Love, M. C. N., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine*, *29*(4), 737–752. https://doi.org/10.1016/j.cger.2013.07.002

Hedden, T. (2007). Imaging cognition in the aging human brain. In D. Riddle (Ed.), *Brain Aging: Models, Methods, and Mechanisms*. CRC Press/Taylor & Francis. https://www.ncbi.nlm.nih.gov/books/NBK3868/

Jackson, J. D., & Balota, D. A. (2011). Mind-wandering in younger and older adults: Converging evidence from the sustained attention to response task and reading for comprehension. *Psychology and Aging*, *27*(1), 106–119. https://doi.org/10.1037/a0023933

Lichstein, K. L., Riedel, B. W., & Richman, S. L. (2000). The Mackworth Clock Test: a computerized version. *The Journal of Psychology*, *134*(2), 153–161. https://doi.org/10.1080/00223980009600858

Mackworth, N. H. (1948). The Breakdown of Vigilance during Prolonged Visual Search. *Quarterly Journal of Experimental Psychology*, *1*(1), 6–21. https://doi.org/10.1080/17470214808416738

McAvinue, L. P., Habekost, T., Johnson, K. A., Kyllingsbæk, S., Vangkilde, S., Bundesen, C., & Robertson, I. H. (2012). Sustained attention, attentional selectivity, and attentional capacity across the lifespan. *Attention Perception & Psychophysics*, *74*(8), 1570–1582. https://doi.org/10.3758/s13414-012-0352-6

McFarlane, J. M. (2024, August 22). *SC.2: Case Study – Air Traffic control*. Pressbooks. https://opentextbc.ca/psychologymtdisupplement/chapter/air-traffic-control/

Mensen, J. M., Dang, J. S., Stets, A. J., & Helton, W. S. (2022). The effects of real-time performance feedback and performance emphasis on the sustained attention to response task (SART). *Psychological Research*, *86*(6), 1972–1979. https://doi.org/10.1007/s00426-021-01602-6

Murman, D. (2015). The impact of age on cognition. *Seminars in Hearing*, *36*(03), 111–121. https://doi.org/10.1055/s-0035-1555115

NATS Public. (2022). ATC Staffing Resilience Plan. In *NATS Public* (pp. 2–16). https://www.caa.co.uk/media/0qglsd50/nats-atc-staffing-resilience-plan-january-2022.pdf

Neal, G. L., & Pearson, R. G. (1966). Comparative Effects of Age, Sex, and Drugs upon Two Tasks of Auditory Vigilance. *Perceptual and Motor Skills*, *23*(3), 967–974. https://doi.org/10.1177/003151256602300360

Nunes, A., & Kramer, A. F. (2009). Experience-based mitigation of age-related performance declines: Evidence from air traffic control. *Journal of Experimental Psychology Applied*, *15*(1), 12–24. https://doi.org/10.1037/a0014947

Rollandi, G. A., Chiesa, A., Sacchi, N., Castagnetta, M., Puntoni, M., Amaro, A., Banelli, B., & Pfeffer, U. (2019). Biological Age versus Chronological Age in the Prevention of Age Associated Diseases. *OBM Geriatrics*, *3*(2), 1. https://doi.org/10.21926/obm.geriatr.1902051

Rothaug, J. & EUROPEAN ORGANISATION FOR THE SAFETY OF AIR NAVIGATION EUROCONTROL EUROPEAN AIR TRAFFIC MANAGEMENT PROGRAMME. (2003). Age, experience and automation in European air traffic control. In *EATMP Infocentre* (1.0) [Report]. https://skybrary.aero/sites/default/files/bookshelf/5446.pdf

Royal NLR - Netherlands Aerospace Centre, Van Drongelen, A., Roelen, A. L. C., Zon, G. D. R., Smeltink, J. W., & Maij, A. (2022). *The Aging Operator: Causes, consequences and Possibilities for Sustainable Deployment.* (NLR-TR-2022-087).

Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Research Reviews*, *35*(2), 146–160. https://doi.org/10.1016/s0165-0173(01)00044-3

Staub, B., Doignon-Camus, N., Després, O., & Bonnefond, A. (2013). Sustained attention in the elderly: What do we know and what does it tell us about cognitive aging? *Ageing Research Reviews*, *12*(2), 459–468. https://doi.org/10.1016/j.arr.2012.12.001

Straeter, O., Woldring, M., & Rothaug, J. (2003). *Age, experience and automation in European air traffic control*. https://www.semanticscholar.org/paper/Age%2C-Experience-and-Automation-in-European-Air-Straeter-Woldring/c1b746bcfac0b1767323998ec99f5b2682db50f4

Ustun, I., Ozer, E., Habib, E., Tatliesme, B., & Akin, A. (2019, November 21). *Vigilance overload measured by computerized Mackworth Clock test*. arXiv.org. https://arxiv.org/abs/1912.04823

Wium, J., & Eaglestone, J. (2022). Job analysis for air traffic control. *Aviation Psychology and Applied Human Factors*, *12*(1), 31–49. https://doi.org/10.1027/2192-0923/a000218

Zanto, T. P., & Gazzaley, A. (2019). Aging of the frontal lobe. *Handbook of Clinical Neurology*, 369–389. https://doi.org/10.1016/b978-0-12-804281-6.00020-3

Zhou, K., Wang, C., Xie, S., Zhou, Y., Zhang, X., Wang, Y., & Tang, H. (2024). Effect of task interruption on the situation awareness of air traffic controllers. *PLoS ONE*, *19*(11), e0314183. https://doi.org/10.1371/journal.pone.0314183

Appendix A

Regression Tables

The following tables summarize the results of regression analyses conducted to examine the predictors of performance (*d′* scores) in the Mackworth and SART tasks.

Table A1

Regression Analysis for Predicting d' Scores in the Mackworth Task

Effect	Estimate (β)	SE	df	t	p
Difficulty					
Intercept	4.56	0.53	43.40	8.66	< 0.05
Difficulty	-0.23	0.12	51.60	-2.40	0.02
Sleepiness					
Intercept	4.24	0.49	61.04	8.60	< 0.05
Sleepiness	-0.15	0.08	62.30	-1.92	0.06
Blink rate					
Intercept	3.73	0.34	43.56	10.99	< 0.05
Blink rate	-0.01	0.01	54.06	-1.40	1.17

Note. Results from single-variable regression analyses examining the relationship between potential predictors and d' scores in the Mackworth task. β = standardized regression coefficient; SE = standard error; df = degrees of freedom; t = t-value; p = p-value. $p < .05$.

Table A2

Regression Analysis for Predicting d' Scores in the SART Task

Effect	Estimate (β)	SE	df	t	p
Difficulty					
Intercept	2.08	0.20	27.09	10.10	< 0.05
Difficulty	-0.09	0.04	28.20	-2.15	0.04
Sleepiness					
Intercept	1.83	0.18	44.74	10.22	< 0.05
Sleepiness	-0.03	0.03	50.11	-0.99	0.32
Blink rate					
Intercept	1.76	0.12	52.97	14.98	< 0.05
Blink rate	-0.01	0.01	58.97	-1.08	0.28

Note. Results from single-variable regression analyses examining the relationship between potential predictors and d' scores in the SART task. β = standardized regression coefficient; SE = standard error; df = degrees of freedom; t = t-value; p = p-value. $p < .05$.

Table A3

Multiple Regression Analysis for Predicting d' Scores in the Mackworth Task

Effect	Estimate (β)	SE	df	t	p
Intercept	2.00	0.92	29.42	2.16	< 0.05
Difficulty	-0.33	0.11	38.12	-2.84	< 0.05
Sleepiness	-0.01	0.09	59.73	-0.19	0.84
Blink rate	-0.01	0.01	59.64	-0.02	0.98
Age	0.06	0.01	22.28	3.60	< 0.05

Note. Results from a multiple regression analysis examining the relationship between potential predictors and d' scores in the Mackworth task. β = standardized regression coefficient; SE = standard error; df = degrees of freedom; t = t-value; p = p-value. $p < .05$.

Table A4

Multiple Regression Analysis for Predicting d' Scores in the SART Task

Effect	Estimate (β)	SE	df	t	p
Intercept	1.83	0.41	22.69	4.42	< 0.05
Difficulty	-0.11	0.05	29.03	-2.10	< 0.05
Sleepiness	0.02	0.04	49.05	0.55	0.58
Blink rate	-0.01	0.01	56.67	-1.22	0.23
Age	0.01	0.01	17.67	0.99	0.33

Note. Results from a multiple regression analysis examining the relationship between potential predictors and d' scores in the SART task. β = standardized regression coefficient; SE = standard error; df = degrees of freedom; t = t-value; p = p-value. $p < .05$.

Appendix B
R Code

All statistical analyses reported in this thesis were performed using R. The following code was used for data analysis.

```
df1 <- read.csv("final_data_merged_new.csv")
library(dplyr)
library(tidyverse)
#data handling

df2 <- df1 %>%
  mutate(
    hit = case_when(
      Test_Type == "SART" & Correct == TRUE ~ 1,
      Test_Type == "Mackworth" & Response_Type == "hit" ~ 1,
      TRUE ~ 0
    ),
    false_alarm = case_when(
      Test_Type == "SART" & Correct == FALSE ~ 1,
      Test_Type == "Mackworth" & Response_Type == "false_alarm" ~ 1,
      TRUE ~ 0
    ),
    miss = case_when(
      Test_Type == "Mackworth" & Response_Type == "miss" ~ 1,
      Test_Type == "SART" & Correct == TRUE ~ 0,  # If it's a hit, it's not a miss
      TRUE ~ 0
    ),
    correct_rejection = case_when(
```

```
    Test_Type == "Mackworth" & Response_Type == "correct_rejection" ~ 1,
    Test_Type == "SART" & Correct == FALSE ~ 0,  # If it's a false alarm, it's not a correct
rejection
    TRUE ~ 0
   )
  )

dprime_data <- df2 %>%
  group_by(Participant_ID, Session) %>%
  summarise(
    hits = sum(hit),
    misses = sum(miss),
    false_alarms = sum(false_alarm),
    correct_rejections = sum(correct_rejection),
    hit_rate = hits / (hits + misses),
    fa_rate = false_alarms / (false_alarms + correct_rejections),
    .groups = "drop"
  )

dprime_data <- dprime_data %>%
  mutate(
    hit_rate = (hits + 0.5) / (hits + misses + 1),
    fa_rate = (false_alarms + 0.5) / (false_alarms + correct_rejections + 1),
    d_prime = qnorm(hit_rate) - qnorm(fa_rate)
  )

dprime_data <- dprime_data %>%
  mutate(
    d_prime = qnorm(hit_rate) - qnorm(fa_rate)
```

```
  )
print(dprime_data)
df2$Bored <- as.numeric(as.character(df2$Bored))
df2$KSS <- as.numeric(as.character(df2$KSS))
df2$Difficulty <- as.numeric(as.character(df2$Difficulty))
aggregated_data_2 <- df2 %>%
  group_by(Participant_ID, Session) %>%
  summarise(
    Total_Blinks = mean(Total_Blinks),
    Blink_Rate = mean(Blink_Rate),
    Bored = mean(Bored),
    KSS = mean(KSS),
    Difficulty = mean(Difficulty),
    Age = mean(Age),
    Nationality = first(Nationality),
    Gender = first(Gender),
    ATCO = first(ATCO),
    Presentation = first(Presentation)
  )

merged_data <- merge( dprime_data, aggregated_data_2,  by = c("Participant_ID",
"Session"))
write_csv(merged_data, "final_usable_data_last.csv")


library(tidyverse)
library(ggplot2)
library(tidyr)
library(lme4)
library(lmerTest)
```

```
library(car)
library(emmeans)
library(gridExtra)
library(psych)
library(afex)  # For repeated measures ANOVA

# DATA IMPORT
final <- read.csv("final_usable_data_last.csv")
# DATA HANDLING
final$Test <- ifelse(
  (final$Presentation == 1 & final$Session %in% 1:3) |
    (final$Presentation == 2 & final$Session %in% 4:6),
  "SART",
  ifelse(
    (final$Presentation == 1 & final$Session %in% 4:6) |
      (final$Presentation == 2 & final$Session %in% 1:3),
    "Mack",
    NA
  )
)
final$blink_pm <- 60 / final$Blink_Rate
final <- final %>%
  mutate(agegroup = case_when(
    Age < 45 ~ 1,
    Age >= 45 & Age < 55 ~ 2,
    Age >= 55 ~ 3
  ))
```

```
final <- final %>%
  mutate(
    KSS = round((KSS / 10) * 9 + 1),
    Bored = round((Bored / 10) * 6 + 1),
    Difficulty = round((Difficulty / 10) * 9 + 1)
  )
# Session only from 1 to 3
final <- final %>%
  group_by(Participant_ID) %>%
  mutate(Session = case_when(
    Session == 4 ~ 1,
    Session == 5 ~ 2,
    Session == 6 ~ 3,
    TRUE ~ Session
  )) %>%
  ungroup()
#Make new datasets to do analysis per task

SART_data <- final %>%
  filter(Test == "SART")
Mackworth_data <- final %>%
  filter(Test == "Mack")
## SART blink filtered
# Define reasonable blink rate range (adjust as needed)
lower_threshold <- 2   # Minimum acceptable blink rate
upper_threshold <- 50  # Maximum acceptable blink rate
```

```
# Filter dataset while grouping by participant and session
final_filtered_SART <- SART_data %>%
  group_by(Participant_ID, Session) %>%  # Group by participant and session
  filter(blink_pm >= lower_threshold & blink_pm <= upper_threshold) %>%
  ungroup()  # Ungroup after filtering

## Mack blink filtered
# Define reasonable blink rate range (adjust as needed)
lower_threshold <- 2   # Minimum acceptable blink rate
upper_threshold <- 50  # Maximum acceptable blink rate

# Filter dataset while grouping by participant and session
final_filtered_Mack <- Mackworth_data %>%
  group_by(Participant_ID, Session) %>%  # Group by participant and session
  filter(blink_pm >= lower_threshold & blink_pm <= upper_threshold) %>%
  ungroup()  # Ungroup after filtering
## 1. Descriptive Statistics
# 1. DESCRIPTIVE STATISTICS BY AGE GROUP
desc_stats <- final %>%
  group_by(agegroup) %>%
  summarise(
    # Performance metrics
    mean_dprime = mean(d_prime, na.rm = TRUE),  # Correct placement of parentheses
    sd_dprime = sd(d_prime, na.rm = TRUE),

    # Subjective measures
    mean_KSS = mean(KSS, na.rm = TRUE),
    sd_KSS = sd(KSS, na.rm = TRUE),
```

```
    mean_Difficulty = mean(Difficulty, na.rm = TRUE),
    sd_Difficulty = sd(Difficulty, na.rm = TRUE),
    mean_Bored = mean(Bored, na.rm = TRUE),
    sd_Bored = sd(Bored, na.rm = TRUE),

    # Physiological measure
    mean_blink = mean(blink_pm, na.rm = TRUE),
    sd_blink = sd(blink_pm, na.rm = TRUE)
  )

print(desc_stats)
final <- final %>%
  group_by(agegroup,Test) %>%
  mutate(meand = mean(d_prime, na.rm = TRUE))

final$agegroup <- as.factor(final$agegroup)
final$Session <- as.factor(final$Session)
final %>%
  group_by(agegroup) %>%
  summarise(mean_age = mean(Age, na.rm = TRUE))
plotfinal <- final %>%
  group_by(agegroup, Test) %>%
  summarise(
    meand = mean(d_prime, na.rm = TRUE),  # Calculate mean d'
    se = sd(d_prime, na.rm = TRUE) / sqrt(n())  # Compute standard error
  ) %>%
  ungroup()
```

```
plotfinal <- plotfinal %>%
  bind_rows(
    final %>%
      group_by(agegroup) %>%
      summarise(
        Test = "Mean",
        meand = mean(meand, na.rm = TRUE),  # Average of meand across Test types
        se = sqrt(sum(se^2) / 2)  # Pooled SE
      ) %>%
      ungroup()
  )
glimpse(plotfinal)  # Check if 'se' exists in the dataset
colnames(plotfinal)  # List all column names
ggplot(plotfinal, aes(x = agegroup, y = meand, group = Test, color = Test)) +
  geom_line(aes(linetype = Test), size = 1) +  # Line plot for each Test
  geom_point(size = 3) +  # Points for individual Test lines

  # Error bars
  geom_errorbar(aes(ymin = meand - se, ymax = meand + se), width = 0.2, size = 0.8) +
 # Mean line across both Test types

  labs(
    x = "Age group",
    y = expression(italic(d) ~ "(performance measure)"),
    color = "Test type"  # Adjust legend label to sentence case
  ) +
  theme_classic(base_size = 12) +  # APA prefers a clean theme, TNR/Arial font
```

```
  scale_color_manual(values = c("Mackworth" = "#0072B2", "SART" = "#E69F00")) +
  theme(
    legend.position = "top",
    legend.title = element_text(face = "bold"),  # Slight emphasis on legend title
    axis.title = element_text(face = "bold"),  # Bold axis labels for readability
    axis.text = element_text(size = 12),  # Ensure text is readable
    panel.grid.major = element_blank(),  # Remove major gridlines (APA style)
    panel.grid.minor = element_blank(),  # Remove minor gridlines (APA style)
    plot.title = element_blank()  # Remove title; APA requires a caption below the figure
  )
## check differences in Test performance
# Perform Repeated Measures ANOVA
anova_test <- aov_ez(id = "Participant_ID",  # Unique participant ID
                     dv = "d_prime",  # Dependent variable
                     within = "Test",  # Within-subject factor
                     data = final)
# Print results
print(anova_test)

pairwise_test <- emmeans(anova_test, pairwise ~ Test, adjust = "bonferroni")
print(pairwise_test)
## 2. Pre-tests
### 2.1 Pre-Test Mackworth

# ----- Shapiro-Wilk Test for Normality per Age Group -----
# We run the Shapiro-Wilk test for normality on 'd_prime' for each Age_Group
shapiro_results <- Mackworth_data %>%
  group_by(agegroup) %>%
```

```
  summarise(
    Shapiro_W = shapiro.test(d_prime)$statistic,  # Extracts W statistic
    Shapiro_p = shapiro.test(d_prime)$p.value     # Extracts p-value
  )

print("Shapiro-Wilk Test results for each Age Group:")
print(shapiro_results)

# ----- Levene's Test for Homogeneity of Variances -----
# Levene's Test compares variances of 'd_prime' across Age Groups.
levene_model <- leveneTest(d_prime ~ as.factor(agegroup), data = Mackworth_data)
print("Levene's Test for Homogeneity of Variances:")
print(levene_model)
# ----- Visual Inspection of Data Distributions -----
# Histogram of d_prime for each Age Group
hist_plot <- ggplot(Mackworth_data, aes(x = d_prime)) +
  geom_histogram(color = "black", fill = "lightblue", bins = 30) +
  facet_wrap(~ agegroup) +
  labs(title = "Histogram of d_prime by Age Group", x = "d_prime", y = "Frequency")

print(hist_plot)
# Q-Q Plots to check normality for each Age Group
qq_plot <- ggplot(Mackworth_data, aes(sample = d_prime)) +
  stat_qq(color = "darkblue") +
  stat_qq_line(color = "red") +
  facet_wrap(~ agegroup) +
  labs(title = "Q-Q Plot of d_prime by Age Group", x = "Theoretical Quantiles", y = "Sample
Quantiles")
```

```
print(qq_plot)
### 2.2 Pre-Test SART
# ----- Shapiro-Wilk Test for Normality per Age Group -----
# We run the Shapiro-Wilk test for normality on 'd_prime' for each Age_Group
shapiro_results <- SART_data %>%
  group_by(agegroup) %>%
  summarise(
    Shapiro_W = shapiro.test(d_prime)$statistic,  # Extracts W statistic
    Shapiro_p = shapiro.test(d_prime)$p.value     # Extracts p-value
  )

print("Shapiro-Wilk Test results for each Age Group:")
print(shapiro_results)

# ----- Levene's Test for Homogeneity of Variances -----
# Levene's Test compares variances of 'd_prime' across Age Groups.
levene_model <- leveneTest(d_prime ~ as.factor(agegroup), data = SART_data)
print("Levene's Test for Homogeneity of Variances:")
print(levene_model)

# ----- Visual Inspection of Data Distributions -----
# Histogram of d_prime for each Age Group
hist_plot <- ggplot(SART_data, aes(x = d_prime)) +
  geom_histogram(color = "black", fill = "lightblue", bins = 30) +
  facet_wrap(~ agegroup) +
  labs(title = "Histogram of d_prime by Age Group", x = "d_prime", y = "Frequency")

print(hist_plot)
```

```
# Q-Q Plots to check normality for each Age Group
qq_plot <- ggplot(SART_data, aes(sample = d_prime)) +
  stat_qq(color = "darkblue") +
  stat_qq_line(color = "red") +
  facet_wrap(~ agegroup) +
  labs(title = "Q-Q Plot of d_prime by Age Group", x = "Theoretical Quantiles", y = "Sample
Quantiles")

print(qq_plot)
## DATA TRANSFORMATION HOMOGENITY
# Try different transformations and check normality
Mackworth_data <- Mackworth_data %>%
  mutate(d_prime_log = log(d_prime))
# Test after transformation
# ----- Shapiro-Wilk Test for Normality per Age Group -----
# We run the Shapiro-Wilk test for normality on 'd_prime' for each Age_Group
shapiro_results <- Mackworth_data %>%
  group_by(agegroup) %>%
  summarise(
    Shapiro_W = shapiro.test(d_prime_log)$statistic,  # Extracts W statistic
    Shapiro_p = shapiro.test(d_prime_log)$p.value     # Extracts p-value
  )

print("Shapiro-Wilk Test results for each Age Group:")
print(shapiro_results)
# ----- Levene's Test for Homogeneity of Variances -----
# Levene's Test compares variances of 'd_prime' across Age Groups.
levene_model <- leveneTest(d_prime_log ~ as.factor(agegroup), data = Mackworth_data)
```

```
print("Levene's Test for Homogeneity of Variances:")
print(levene_model)

# ----- Visual Inspection of Data Distributions -----
# Histogram of d_prime for each Age Group
hist_plot <- ggplot(Mackworth_data, aes(x = d_prime_log)) +
  geom_histogram(color = "black", fill = "lightblue", bins = 30) +
  facet_wrap(~ agegroup) +
  labs(title = "Histogram of d_prime by Age Group", x = "d_prime", y = "Frequency")

print(hist_plot)

# Q-Q Plots to check normality for each Age Group
qq_plot <- ggplot(Mackworth_data, aes(sample = d_prime_log)) +
  stat_qq(color = "darkblue") +
  stat_qq_line(color = "red") +
  facet_wrap(~ agegroup) +
  labs(title = "Q-Q Plot of d_prime by Age Group", x = "Theoretical Quantiles", y = "Sample
Quantiles")

print(qq_plot)

##TRANSFORMATION SART
# Try different transformations and check normality
SART_data <- SART_data %>%
  mutate(d_prime_log =log(d_prime))
## Test after transformation
```

```
# ----- Shapiro-Wilk Test for Normality per Age Group -----
# We run the Shapiro-Wilk test for normality on 'd_prime' for each Age_Group
shapiro_results <- SART_data %>%
  group_by(agegroup) %>%
  summarise(
    Shapiro_W = shapiro.test(d_prime_log)$statistic,  # Extracts W statistic
    Shapiro_p = shapiro.test(d_prime_log)$p.value     # Extracts p-value
  )
print("Shapiro-Wilk Test results for each Age Group:")
print(shapiro_results)

# ----- Levene's Test for Homogeneity of Variances -----
# Levene's Test compares variances of 'd_prime' across Age Groups.
levene_model <- leveneTest(d_prime_log ~ as.factor(agegroup), data = SART_data)
print("Levene's Test for Homogeneity of Variances:")
print(levene_model)

# ----- Visual Inspection of Data Distributions -----
# Histogram of d_prime for each Age Group
hist_plot <- ggplot(SART_data, aes(x = d_prime_log)) +
  geom_histogram(color = "black", fill = "lightblue", bins = 30) +
  facet_wrap(~ agegroup) +
  labs(title = "Histogram of d_prime by Age Group", x = "d_prime", y = "Frequency")

print(hist_plot)

# Q-Q Plots to check normality for each Age Group
qq_plot <- ggplot(SART_data, aes(sample = d_prime_log)) +
```

```
  stat_qq(color = "darkblue") +
  stat_qq_line(color = "red") +
  facet_wrap(~ agegroup) +
  labs(title = "Q-Q Plot of d_prime by Age Group", x = "Theoretical Quantiles", y = "Sample Quantiles")
print(qq_plot)

###  3. MODELING
## 3.1 LMM

# Convert agegroup to factor if it's not already
SART_data$agegroup <- as.factor(SART_data$agegroup)
Mackworth_data$agegroup <- as.factor(Mackworth_data$agegroup)
SART_data$Session <- as.factor(SART_data$Session)
Mackworth_data$Session <- as.factor(Mackworth_data$Session)
## SART MODELING
# Basic model with random intercept for subjects

model1 <- lmer(d_prime ~ agegroup*Session + (1|Participant_ID), data = SART_data)
# Then check assumptions on residuals  new assumptions
residuals_model <- residuals(model1)
fitted_values <- fitted(model1)

# Normality of residuals
qqnorm(residuals_model)
qqline(residuals_model)
hist(residuals_model)
# Homoscedasticity
plot(fitted_values, residuals_model)
```

```
abline(h = 0, col = "red")


# Print model summary
summary(model1)
anova(model1)
anova_results <- aov_ez(
  id = "Participant_ID",      # Subject ID
  dv = "d_prime",             # Dependent variable
  within = "Session",         # Repeated measure factor
  between = "agegroup",       # Between-subjects factor
  data = SART_data
)


print(anova_results)


# Step 2: Add interaction between Age Group and Session
model2 <- lmer(d_prime_log ~ agegroup + Session + blink_pm * Session + Session * KSS +
Difficulty + (1|Participant_ID), data = SART_data)
# Print model summary
summary(model2)
# Step 3: Compare models to see if the interaction improves fit
anova(model1, model2)
# Step 5: Check model diagnostics
# Residual plots
plot(model1)


# Normality of residuals
qqnorm(residuals(model1))
qqline(residuals(model1))
```

```
# Step 6: Post-hoc comparisons if main effects or interactions are significant
# For age group main effect
emmeans_age <- emmeans(model1, specs = "agegroup")
pairs(emmeans_age, adjust = "tukey")

# For session main effect
emmeans_session <- emmeans(model1, specs = "Session")
pairs(emmeans_session, adjust = "tukey")

# Step 7: Visualize the results
# Predicted values by age group and session
SART_data$predicted <- predict(model1)

# Plot predicted values
ggplot(SART_data, aes(x = Session, y = predicted, color = agegroup, group = agegroup)) +
  stat_summary(fun = mean, geom = "line") +
  stat_summary(fun = mean, geom = "point", size = 3) +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2) +
  labs(title = "Predicted Performance (log d') by Age Group and Session",
       x = "Session", y = "Predicted log(d')", color = "Age Group") +
  theme_minimal()

# If you want to back-transform to original scale for interpretation
SART_data$predicted_original <- exp(SART_data$predicted)

# Plot on original scale
ggplot(SART_data, aes(x = Session, y = predicted_original, color = agegroup, group =
agegroup)) +
```

```
  stat_summary(fun = mean, geom = "line") +
  stat_summary(fun = mean, geom = "point", size = 3) +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2) +
  labs(title = "Predicted Performance (d') by Age Group and Session",
       x = "Session", y = "Predicted d'", color = "Age Group") +
  theme_minimal()

## Mackworth MODELING
# Basic model with random intercept for subjects
model1_mack <- lmer(d_prime ~ agegroup * Session + (1|Participant_ID), data =
Mackworth_data)
# Print model summary
summary(model1_mack)
anova(model1_mack)
anova_results <- aov_ez(
  id = "Participant_ID",       # Subject ID
  dv = "d_prime",           # Dependent variable
  within = "Session",       # Repeated measure factor
  between = "agegroup",     # Between-subjects factor
  data = final_filtered_SART
)
print(anova_results)
anova_results$sphericity
anova_results$sphericity.correction
# Step 2: Add interaction between Age Group and Session
model2_mack <- lmer(d_prime_log ~ agegroup * Session + (1|Participant_ID), data =
Mackworth_data)

# Print model summary
```

```
summary(model2_mack)
# Step 3: Compare models to see if the interaction improves fit
anova(model1_mack, model2_mack)

# Step 5: Check model diagnostics
# Residual plots
plot(model1_mack)

# Normality of residuals
qqnorm(residuals(model1_mack))
qqline(residuals(model1_mack))

# Step 6: Post-hoc comparisons if main effects or interactions are significant
# For age group main effect
emmeans_age <- emmeans(model1_mack, specs = "agegroup")
pairs(emmeans_age, adjust = "tukey")

# For session main effect
emmeans_session <- emmeans(model1_mack, specs = "Session")
pairs(emmeans_session, adjust = "tukey")

# Step 7: Visualize the results
# Predicted values by age group and session
Mackworth_data$predicted <- predict(model1_mack)

# Plot predicted values
ggplot(Mackworth_data, aes(x = Session, y = predicted, color = agegroup, group =
agegroup)) +
  stat_summary(fun = mean, geom = "line") +
```

```
  stat_summary(fun = mean, geom = "point", size = 3) +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2) +
  labs(title = "Predicted Performance (log d') by Age Group and Session",
       x = "Session", y = "Predicted log(d')", color = "Age Group") +
  theme_minimal()

# If you want to back-transform to original scale for interpretation
Mackworth_data$predicted_original <- exp(Mackworth_data$predicted)
# Plot on original scale
ggplot(Mackworth_data, aes(x = Session, y = predicted_original, color = agegroup, group = agegroup)) +
  stat_summary(fun = mean, geom = "line") +
  stat_summary(fun = mean, geom = "point", size = 3) +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2) +
  labs(title = "Predicted Performance (d') by Age Group and Session",
       x = "Session", y = "Predicted d'", color = "Age Group") +
  theme_minimal()
## CORRELATION analysis for SART
#both valid
cor_test_result <- cor.test(SART_data$d_prime, SART_data$Age, method = "pearson")
print(cor_test_result)


cor_test_result <- cor.test(final_filtered_SART$d_prime, final_filtered_SART$blink_pm, method = "spearman")
print(cor_test_result)
#all these 3 valid
cor_test_result <- cor.test(final_filtered_SART$d_prime, final_filtered_SART$KSS, method = "pearson")
print(cor_test_result)
```

```
cor_test_result <- cor.test(final_filtered_SART$d_prime, final_filtered_SART$Difficulty,
method = "pearson")

print(cor_test_result)

cor_test_result <- cor.test(SART_data$d_prime_log, SART_data$Bored, method =
"pearson")

print(cor_test_result)


cor_test_result <- cor.test(SART_data$Age, SART_data$KSS, method = "pearson")  #THis
valid

print(cor_test_result)

cor_test_result <- cor.test(SART_data$Age, SART_data$Difficulty, method = "pearson")

print(cor_test_result)

cor_test_result <- cor.test(SART_data$Age, SART_data$Bored, method = "pearson")

print(cor_test_result)


## ANALYSIS FOR AFTER CORRELATION SART


# Check for multicollinearity first

cor(SART_data[, c("KSS", "Difficulty", "Bored", "blink_pm")], use = "complete.obs")


# Model with all subjective measures as covariates

model_full <- lmer(d_prime ~ agegroup + Session + KSS + blink_pm + (1|Participant_ID),
data = SART_data)

summary(model_full)

anova(model_full)


# Fit the original model (model_full_mack)

model_full_sart <- lmer(d_prime ~  KSS + Difficulty +

                      (1 | Participant_ID),

                     data = final_filtered_SART)
```

```
summary(model_full_sart)
anova(model_full_sart)

# Fit the extended model with added interactions for Session with KSS, Difficulty, and blink_pm
model_full_interactions_sart <- lmer(d_prime_log ~ agegroup + Session + Difficulty +
                                    KSS * Session + blink_pm * Session +
                                    (1 | Participant_ID),
                                  data = final_filtered_SART)

summary(model_full_sart)
anova(model_full_sart)
anova(model_full_interactions_sart, model_full_sart)
# Then check assumptions on residuals  new assumptions
residuals_model <- residuals(model_full)
fitted_values <- fitted(model_full)

# Normality of residuals
qqnorm(residuals_model)
qqline(residuals_model)
hist(residuals_model)

# Homoscedasticity
plot(fitted_values, residuals_model)
abline(h = 0, col = "red")

## CORRELATION analysis for MACK
#First valid
```

```
cor_test_result <- cor.test(Mackworth_data$d_prime_log, Mackworth_data$Age, method =
"pearson")

print(cor_test_result)


cor_test_result <- cor.test(Mackworth_data$d_prime_log, Mackworth_data$blink_pm,
method = "spearman")

print(cor_test_result)

#all these 3 valid

cor_test_result <- cor.test(Mackworth_data$d_prime_log, Mackworth_data$KSS, method =
"pearson")

print(cor_test_result)

cor_test_result <- cor.test(Mackworth_data$d_prime_log, Mackworth_data$Difficulty,
method = "pearson")

print(cor_test_result)

cor_test_result <- cor.test(Mackworth_data$d_prime_log, Mackworth_data$Bored, method =
"pearson")

print(cor_test_result)


cor_test_result <- cor.test(Mackworth_data$Age, Mackworth_data$KSS, method =
"pearson")

print(cor_test_result)

cor_test_result <- cor.test(Mackworth_data$Age, Mackworth_data$Difficulty, method =
"pearson") #this and next valid

print(cor_test_result)

cor_test_result <- cor.test(Mackworth_data$Age, Mackworth_data$Bored, method =
"pearson")

print(cor_test_result)


## ANALYSIS FOR AFTER CORRELATION MACK
```

```
# Check for multicollinearity first
cor(Mackworth_data[, c("KSS", "Difficulty", "Bored", "blink_pm")], use = "complete.obs")

# Model with all subjective measures as covariates
# Fit the original model (model_full_mack)
model_full_mack <- lmer(d_prime_log ~ agegroup + Session + Difficulty + KSS + blink_pm
+
                    (1 | Participant_ID),
                  data = final_filtered_Mack)
summary(model_full_mack)

# Fit the extended model with added interactions for Session with KSS, Difficulty, and
blink_pm
model_full_interactions <- lmer(d_prime ~ Age +
                              (1 | Participant_ID),
                            data = final_filtered_Mack)
summary(model_full_interactions)
anova(model_full_interactions)

anova(model_full_interactions, model_full_mack)

# Fit a linear model ignoring random effects
lm_model <- lm(d_prime_log ~ agegroup * Session + Difficulty * Session +
             KSS * Session + blink_pm * Session, data = final_filtered_Mack)

## Check VIF for overfitting
vif(lm_model)

final_filtered_Mack$Session <- as.numeric(final_filtered_Mack$Session)
```

```
final_filtered_Mack$agegroup <- as.numeric(final_filtered_Mack$agegroup)


cor_matrix <- cor(final_filtered_Mack[, c("agegroup", "Session", "blink_pm", "KSS",
"Difficulty")])

print(cor_matrix)

# Then check assumptions on residuals  new assumptions

residuals_model <- residuals(model_full_mack)

fitted_values <- fitted(model_full_mack)


# Normality of residuals

qqnorm(residuals_model)

qqline(residuals_model)

hist(residuals_model)


# Homoscedasticity

plot(fitted_values, residuals_model)

abline(h = 0, col = "red")

#create dataset for blink_rate SART

# Check the result

summary(final_filtered_SART$blink_pm)

# Run LMM

lmm_model_SART1 <- lmer(d_prime_log ~ agegroup * Session + Difficulty + KSS *
Session +

                        blink_pm * Session + Session * Difficulty + (1 | Participant_ID), data =
final_filtered_SART)


summary(lmm_model_SART1)

anova(lmm_model_SART1)
```

```
# Run LMM

lmm_model_SART2 <- lmer(d_prime_log ~ agegroup + Session + Difficulty + KSS +
blink_pm + (1 | Participant_ID), data = final_filtered_SART)

summary(lmm_model_SART2)

anova(lmm_model_SART2)

anova(lmm_model_SART1,lmm_model_SART2)


lmm_model_SART3 <- lmer(d_prime_log ~ agegroup + Session + Difficulty + KSS +
blink_pm + (1 | Participant_ID), data = final_filtered_SART)


# Fit a linear model ignoring random effects

lm_model_sart <- lm(d_prime_log ~ agegroup + Session + Difficulty + KSS + blink_pm ,
data = final_filtered_SART)

## Check VIF for overfitting

vif(lm_model_sart)

final_filtered_SART$Session <- as.numeric(final_filtered_SART$Session)


cor(final_filtered_SART$Session, final_filtered_SART$KSS)

cor(final_filtered_SART$Session, final_filtered_SART$blink_pm)


# Then check assumptions on residuals  new assumptions

residuals_model <- residuals(lmm_model_SART)

fitted_values <- fitted(lmm_model_SART)

# Normality of residuals

qqnorm(residuals_model)

qqline(residuals_model)

hist(residuals_model)
```

```
# Homoscedasticity
plot(fitted_values, residuals_model)
abline(h = 0, col = "red")

# Basic scatter plot with regression line
ggplot(final_filtered_SART, aes(x = blink_pm, y = d_prime_log)) +
  geom_point(alpha = 0.5, color = "blue") +  # Scatter points
  geom_smooth(method = "lm", color = "red", se = TRUE) +  # Linear regression line
  labs(
    title = "Relationship Between d_prime and Blink Rate",
    x = "Blink Rate per Minute",
    y = "d_prime"
  ) +
  theme_minimal()

ggplot(final_filtered_SART, aes(x = blink_pm, y = d_prime_log)) +
  geom_point(alpha = 0.5, color = "blue") +  # Scatter points
  geom_smooth(method = "lm", color = "red", se = TRUE) +  # Linear regression line
  labs(
    title = "Relationship Between d_prime and Blink Rate by Age Group",
    x = "Blink Rate per Minute",
    y = "d_prime"
  ) +
  theme_minimal() +
  facet_wrap(~ agegroup)  # Facet by age group
```

```
ggplot(final_filtered_SART, aes(x = blink_pm, y = d_prime_log)) +
  geom_point(alpha = 0.5, color = "blue") +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  facet_wrap(~ Session) +  # Separate plots by session
  labs(
    title = "d_prime vs. Blink Rate per Minute (by Session)",
    x = "Blink Rate per Minute",
    y = "d_prime"
  ) +
  theme_minimal()

#create dataset for blink_rate Mackworth
# Check the result
summary(final_filtered_Mack$blink_pm)

# Run LMM
lmm_model_Mack <- lmer(d_prime ~ Age*Session  + (1 | Participant_ID), data = final_filtered_SART)

lmm_model_Mack <- lmer(d_prime ~ agegroup*Session + Difficulty*Session  + (1 | Participant_ID), data = final_filtered_Mack)
summary(lmm_model_Mack)
# Basic scatter plot with regression line
ggplot(final_filtered_Mack, aes(x = blink_pm, y = d_prime_log)) +
  geom_point(alpha = 0.5, color = "blue") +  # Scatter points
  geom_smooth(method = "lm", color = "red", se = TRUE) +  # Linear regression line
  labs(
    title = "Relationship Between d_prime and Blink Rate",
    x = "Blink Rate per Minute",
```

```
    y = "d_prime"
  ) +
  theme_minimal()


ggplot(final_filtered_Mack, aes(x = blink_pm, y = d_prime_log)) +
  geom_point(alpha = 0.5, color = "blue") +  # Scatter points
  geom_smooth(method = "lm", color = "red", se = TRUE) +  # Linear regression line
  labs(
    title = "Relationship Between d_prime and Blink Rate by Age Group",
    x = "Blink Rate per Minute",
    y = "d_prime"
  ) +
  theme_minimal() +
  facet_wrap(~ agegroup)  # Facet by age group

ggplot(final_filtered_Mack, aes(x = blink_pm, y = d_prime_log)) +
  geom_point(alpha = 0.5, color = "blue") +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  facet_wrap(~ Session) +  # Separate plots by session
  labs(
    title = "d_prime vs. Blink Rate per Minute (by Session)",
    x = "Blink Rate per Minute",
    y = "d_prime"
  ) +
  theme_minimal()
```

```
## Correlations new and multiple regressions

#MACK
mrMack<- lmer(d_prime ~ KSS + (1 | Participant_ID), data = final_filtered_Mack)
summary(mrMack)
mrMack<- lmer(d_prime ~ Difficulty + KSS + Age + blink_pm + (1 | Participant_ID), data =
final_filtered_Mack) # valid
summary(mrMack)
mrMack<- lmer(d_prime ~ blink_pm + (1 | Participant_ID), data = final_filtered_Mack)
summary(mrMack)
mrMack<- lmer(d_prime ~ Difficulty + (1 | Participant_ID), data = final_filtered_Mack) #
valid
summary(mrMack)
mrMack<- lmer(d_prime ~ Age + Difficulty + (1 | Participant_ID), data =
final_filtered_Mack) # valid
summary(mrMack)
mrMack<- lmer(d_prime ~ Difficulty + KSS + blink_pm + Age + (1 | Participant_ID), data =
final_filtered_Mack)
summary(mrMack)

#Correlations

#all these 3 valid
cor_test_result <- cor.test(final_filtered_Mack$d_prime, final_filtered_Mack$Difficulty,
method = "pearson") #valid
print(cor_test_result)
cor_test_result <- cor.test(final_filtered_Mack$d_prime, final_filtered_Mack$Age, method =
"pearson") #valid
print(cor_test_result)
```

```
cor_test_result <- cor.test(final_filtered_Mack$d_prime, final_filtered_Mack$KSS, method = "pearson")

print(cor_test_result)

cor_test_result <- cor.test(final_filtered_Mack$d_prime, final_filtered_Mack$blink_pm, method = "pearson")

print(cor_test_result)


#SART

mrSART<- lmer(d_prime ~ KSS + (1 | Participant_ID), data = final_filtered_SART)

summary(mrSART)


mrSART<- lmer(d_prime ~ Difficulty + (1 | Participant_ID), data = final_filtered_SART) # valid

summary(mrSART)


mrSART<- lmer(d_prime ~ blink_pm + (1 | Participant_ID), data = final_filtered_SART)

summary(mrSART)


mrSART<- lmer(d_prime ~ Age + (1 | Participant_ID), data = final_filtered_SART)

summary(mrSART)


mrSART<- lmer(d_prime ~ Age + Difficulty + (1 | Participant_ID), data = final_filtered_SART)

summary(mrSART)


mrSART<- lmer(d_prime ~ Difficulty + KSS + blink_pm + Age + (1 | Participant_ID), data = final_filtered_SART)

summary(mrSART)


#Correlationsd
```

```
#all these 3 valid

cor_test_result <- cor.test(final_filtered_SART$d_prime, final_filtered_SART$Difficulty, method = "pearson")

print(cor_test_result)

cor_test_result <- cor.test(final_filtered_SART$d_prime, final_filtered_SART$Age, method = "pearson")

print(cor_test_result)

cor_test_result <- cor.test(final_filtered_Mack$blink_pm, final_filtered_Mack$KSS, method = "pearson")

print(cor_test_result)

cor_test_result <- cor.test(final_filtered_SART$d_prime, final_filtered_SART$blink_pm, method = "pearson")

print(cor_test_result)


plotfinal2 <- final %>%

  group_by(Session, agegroup, Test) %>%  # Here, 'Test' is used as Task variable

  summarise(mean_dprime = mean(d_prime, na.rm = TRUE),

            se = sd(d_prime, na.rm = TRUE) / sqrt(n()),

            .groups = "drop")


p <- ggplot(plotfinal2, aes(x = Session, y = mean_dprime, color = agegroup, group = agegroup)) +

  geom_line(size = 1) +  # Line plot for each agegroup with proper size

  geom_point(size = 3) +  # Points for individual d' values

  geom_errorbar(aes(ymin = mean_dprime - se, ymax = mean_dprime + se), width = 0.2, size = 1) +  # Error bars with appropriate width and size

  facet_wrap(~ Test) +  # Facet by Test type

  theme_minimal(base_family = "Times New Roman", base_size = 12) +  # APA requires Times New Roman font and font size 12

  labs(x = "Session Number",
```

```
    y = expression(italic(d) ~ "' (Sensitivity)"),  # Correct use of italic for d'
    color = "Age Group") +  # Adjust legend label
  theme(
   legend.position = "top",  # Legend at the top of the plot
   legend.title = element_text(face = "bold", size = 12),  # Bold and size for the legend title
   legend.text = element_text(size = 12),  # Set text size for the legend
   axis.title = element_text(face = "bold", size = 12),  # Bold axis titles
   axis.text = element_text(size = 12),  # Axis text size
   strip.text = element_text(face = "bold", size = 12),  # Bold facet labels
   panel.grid.major = element_blank(),  # Remove major gridlines
   panel.grid.minor = element_blank(),  # Remove minor gridlines
   plot.margin = unit(c(1, 1, 1, 1), "cm")  # Margin adjustments to avoid clipping
  )

print(p)
```

Appendix C
Subjective surveys

All following surveys were used to measure the subjective performance of participants.

AGE:

GENDER:

NATIONALITY:

How sleepy are you?	Extremely alert	Very alert	Alert	Rather alert	Neither alert nor sleepy	Some signs of sleepiness	Sleepy, but no effort to keep awake	Sleepy, but some effort to keep awake	Very sleepy, great effort to keep awake, fighting sleep	Extremely sleepy, can't keep awake
No1	1	2	3	4	5	6	7	8	9	10
No2	1	2	3	4	5	6	7	8	9	10
No3	1	2	3	4	5	6	7	8	9	10
No4	1	2	3	4	5	6	7	8	9	10
No5	1	2	3	4	5	6	7	8	9	10
No6	1	2	3	4	5	6	7	8	9	10

I feel bored	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree
No1	1	2	3	4	5	6	7
No2	1	2	3	4	5	6	7
No3	1	2	3	4	5	6	7
No4	1	2	3	4	5	6	7
No5	1	2	3	4	5	6	7
No6	1	2	3	4	5	6	7

How difficult is the exercise?	Minimal Effort	Low Effort	Acceptable Effort	Moderate Effort	High effort	Maximum Effort for adequate performance	Maximum Effort for moderate errors	Maximum Effort to avoid large errors	Intense Effort: Many Errors left	Impossible
No1	1	2	3	4	5	6	7	8	9	10
No2	1	2	3	4	5	6	7	8	9	10
No3	1	2	3	4	5	6	7	8	9	10
No4	1	2	3	4	5	6	7	8	9	10
No5	1	2	3	4	5	6	7	8	9	10
No6	1	2	3	4	5	6	7	8	9	10