Working Memory Load and Visuospatial Demands During Driving: A behavioral and eye-tracking analysis

Bachelor's Project Thesis

Milan de Mooij, s3513092, m.de.mooij@student.rug.nl,
Supervisors: Dr. J.P. Borst & M. Held, M.Sc.

Abstract: To combat human-error in driving and increase car safety, a tremendous amount of research has been conducted in the field of automated driving. However, there is evidence that complete autonomous control is not desirable when it comes to the safest driving experience. Adaptive automation is a system where the level of automation is adjusted to the state of the operator. In the context of driving, such a system could change its level of automation to the cogntive load of the driver to counteract the negative effects of cognitive overload or underload. For such a system to function, a robust method of measuring cognitive load is required.

This study investigated whether cognitive load, here defined as a combination of working memory load (WML) and visuospatial demands, has an influence on working memory performance and driving performance. Furthermore, we also used eye-tracking to find out whether pupil size and eye-fixations are influenced by cogntive load.

A simulated-driving experiment with eye-tracking was conducted in which WML and visuospatial demands were manipulated separately. In the simulation participants drove on a straight highway for 60 minutes. WML was manipulated by an n-back task (n = 0, 1, 2, 3, 4), by doing a speed-regulation-task. Visuospatial demands were manipulated by a change in the driving environment: a construction site with reduced lane width, increasing driving difficulty.

Results indicate that working memory performance is only influenced by WML. Driving performance is only affected by an increase in visuospatial demands. Furthermore, pupil-size predicted WML, but not visuospatial demands. Lastly, the number of eye-fixations on the speedometer decreased when WML increased.

1 Introduction

In today's society, practically everyone uses a car to get around. Only in the United States, each household has an average of 1.88 vehicles at home (I. Wagner, 2021). In 2017, only 9% did not possess a car or light truck (I. Wagner, 2019).

When you look at the leading causes of death of Americans, the third position, below hearth disease and cancer, is 'accidents' (Murphy et al., 2018). 21.9%, of these accidents are deaths related to motor vehicle crashes (J. Elflein, 2021). de Waard (1996) states that most car accidents can be attributed to human error. This emphasizes that research relating to car-safety and autonomous driving is of utmost importance.

The act of driving a car is a cognitively demanding task. A driver has to carry out multiple tasks simultaneously, such as monitoring information from inside and outside the vehicle, controlling the car, and adhering to traffic rules. Moreover, the driver is located in a highly dynamic environment, where unforeseen events can take place at any moment. All these sub-tasks demand cognitive resources, which are in limited supply (Salvucci and Taatgen, 2008). An important concept in this context is mental workload. It is described as the ratio of cognitive demands to allocated resources (de Waard, 1996). This means that mental workload is a measure of how much a task demands from the driver's cognition in relation to the cognitive resources that are available. Cognitive load can be categorized in two regions (Wickens, 2008). The first one being cognitive underload, where the cognitive demand is less than the cognitive resources available. The other one is cognitive overload, where the cognitive demand is higher than the cognitive resources available. The latter poses a problem, which is described by Meister (1976) in a relational model that maps task demand to task performance. In this model, a region is described where extreme levels of load, that is overload, lead to diminished performance in the task at hand.

In driving, two cognitive components are essential; working memory and visual attention (de Waard, 1996). A large demand is made on working memory, as the complex representation of the environment has to be retained and updated constantly. It has been shown that the ability to focus decreases under conditions where cognitive load on working memory is high (Lavie, 2010). Also, a study claims that variability in growth and capacity of working memory in adolescents is associated with a higher rate of car crashes (Walshe et al., 2019). Driving is largely a visual task, as the driver has to constantly make estimations of the visuospatial relation of surrounding objects, such as cars, lane indicators, and other obstacles. Furthermore, it is required to perceive information from inside and outside the vehicle. Brooks et al. (2018) have linked a decrease in driving performance to an increase in peristimulus alpha activity, which indicates deficient visuospatial attention.

To account for undesired levels of cognitive load, 'adapative automation' can be used. It refers to a system where the level of automation is adjusted to the state of the operator (Byrne and Parasuraman, 1996). One might wonder why a fully automated system would not be preferable. A study in the aviation industry showed that automation can have a negative result on situational awareness in terms of over-dependence on the system, decreased vigilance and lack of understanding the system's capabilities (Endsley and Kiris, 1995). Also, a human factors study found that driving performance was worse when participants had to regain control of an automated vehicle while they were distracted by a secondary task (Merat et al., 2012). These instances indicate that automation can lead to cognitive underload, which in turn can lead to a decrease in performance. This is where adaptive automation is preferable, a system that will only increase automation when cognitive load is too high.

For such a system to function, a method that measures cognitive load must be developed. Examples of physiological measures of cognitive load are EEG (Antonenko et al., 2010), heart rate (Paas et al., 1994) and pupil dilation (van Gerven et al., 2004). Findings of Scheunemann et al. (2019) show that there is an interaction between visuospatial demands and working memory on the brain level and in task performance. To further consolidate an interaction between these two cognitive systems, this research tries to find an effect of working memory load and visuospatial demands on pupil dilation. We therefore ask the question *"Can pupillometry predict cognitive load while driving?"*. Furthermore, we investigate the relationship between eye-fixations on a speedometer and cognitive load. An inverse relationship is to be expected, due to the fact there is a limited supply of cognitive resources (Salvucci and Taatgen, 2008). This means that when cognitive load increases, a lack of cognitive resources arises, which in turns could mean that the participant has less capacity to check the speedometer. For this, we try to answer the question *"Does cognitive load have an influence on the frequency of speedometer checking?"*.

Additionaly, we also focus on driving behavior. To effectively estimate when cognitive load is too high, we use two aforementioned measurements that could suffer from cognitive overload, namely working memory performance and driving behavior. Scheunemann et al. (2019) found that an increase in cognitive load in one domain could lead to a decrease in performance in the other domain. They showed that an increase in working memory load, decreased driving performance, where the measurement of driving performance is related to a visual task. Also, they showed that an increase in visuospatial demands led to a decrease in working memory performance. They claim that this interaction between two cognitive systems is caused by a competition of limited cognitive resources (Salvucci and Taatgen, 2008). Scheunemann et al. (2019) proposed some improvements in their paper. Mainly, they addressed that their way of increasing working memory load was not perfect: increasing levels of working memory load were tested on a shorter time frame. This study accounts for this. We are therefore interested if we can replicate these results. To investigate this we try to answer the following two questions: *"Does cognitive load have an influence on working memory performance?"* and *"Does cognitive load have an influence on driving performance?"*.

To be able to answer the two questions related to eye-tracking and the two question related to driving behavior, a simulated driving experiment with eye-tracking is conducted where working memory load and visuospatial demands are manipulated. The experiment is mostly replicated from two other studies, namely Unni et al. (2017) and Scheunemann et al. (2019), which both tried to predict varying working memory loads and visuospatial demands, and corresponding brain areas.

The experiment consisted of the participant driving on a simulated highway. The working memory load was manipulated by conducting a modified n-back task during driving (Unni et al., 2017). The n-back task is a way to gradually increase cognitive working memory load (Kirchner, 1958). This n-back task was performed in terms of a speed-regulation task, as the participant has to adjust its speed to the speed sign that was seen n speed signs back.

Visuospatial demands were manipulated by having two highway conditions; a *non-construction* condition and a *construction* condition. The *non-construction* condition was a normal three-lane highway, while the *construction* condition was a three-lane highway with narrower lanes and the left-most lane being obstructed. This increases the visuospatial demands which increases the driving difficulty (Scheunemann et al., 2019).

2 Methods

2.1 Participants

A total of 38 volunteers (23 male, 12 female, 3 other) aged 20-36 ($M = 23.1 \pm 3.0$). All drivers possessed a driver's license. All participants signed an informed consent form prior to the experiment and were compensated €12 for their participation.

2.2 Experimental Set-up

The experiment took place on a simulated straight three-lane highway, see Figure A.1 (see Appendix). The features of the environment were minimal. Either side of the road was coloured green, signifying grass. There were no median strips dividing the road from the rest of the environment. Other traffic consisted of a single car (referred to as *autocar*), represented by a blue rectangle. The *autocar* would stick to traffic rules such as overtaking from the left, staying on the right lane as much as possible, and following the current speed limit. Additionally, the car was programmed to stay in a certain range from the participants car, ensuring that the simulated car would stay relevant.

The participant could see a black dashboard that filled the bottom of the screen. The speed of the car was shown in the center of the dashboard, represented by an integer. When the left or right indicators were pressed, they would appear on the dashboard in the respective sides as yellow blinking arrows. The simulation had three rear-view mirrors: one on the top, one on the left, and one on the right. The *autocar* was visible in the corresponding mirrors depending on the distance from the car.

In the construction condition the leftmost lane was closed off by a row of pylons. The lanes were separated by a full yellow line and were narrower than the non-construction condition, see Figure A.2 (see Appendix). Precisely, the non-construction condition had lanes that were 3.5 meters wide, following the widths of Germany's national highways. The construction site lanes have a smaller width of 2.5 meters.

Speed signs that passed were identical to general speed signs in The Netherlands; black digits enclosed by a red circle.

Within each trial the participants were presented with at least nine speed signs at intervals of 20 seconds. The first speed sign was presented after 5 seconds. For n-back tasks with $n > 0$, there was a build-up phase of n speed signs preceding the nine speed signs where the participant would perform the task. For example, for $n = 4$, the build-up phase would be the first four speed signs. After the build-up phase, the task of regulating speed would start. Due to a difference in length of build-up phases per n-back trial, each trial differed in number of speed signs; for each n-back, n speed signs were added. For a schematic overview of the n-back task, see Figure 2.1.

Participants interacted with the simulation using a steering wheel with indicators and a throttle and brake pedal (Driving Force GT by Logitech). The steering wheel was secured to the table in front of the screen and remained in the same location for all participants. The pedals were placed on the floor such that participants could move it closer

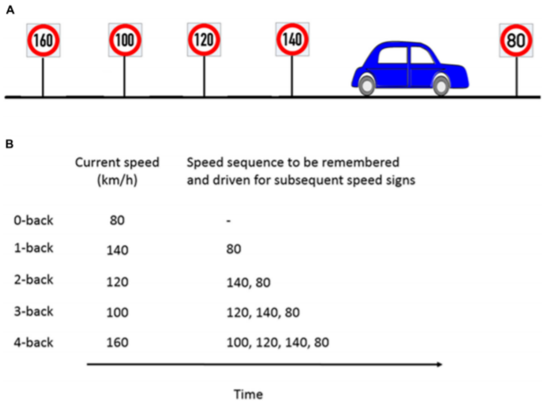

Figure 2.1: Example of n-back experimental paradigm to manipulate working memory load. (A) Consider a scenario where the participant is about to pass the 80 km/h speed sign and the previous four speed signs were as shown in the schematic. (B) For the corresponding n-back task, participants had to memorize the last n speed signs and drive at the n-th speed sign which occurred previously. For example, at 1-back, the participant's target speed is the previous sign (140 km/h) and has to keep the current speed sign in memory (80 km/h). Figure taken from Unni et al. (2017), caption adapted from Scheunemann et al. (2019).

or further depending on their level of comfort. An eye-tracking camera (EyeLink Portable Duo by SR Research), placed between the screen and the steering wheel, was used to continuously record the eye movements and pupil size of participants. The method of tracking that we employed was remote tracking using a target sticker on the participant's forehead. This method was chosen because stabilizing the head using a head rest was not feasible considering the set-up with the steering wheel.

2.3 Experimental Procedure

The procedure of the experiment follows that of Scheunemann et al. (2019) closely. The experiment consisted of 20 trials in total, divided by a short break into two blocks of 10 trials each. The participant was allowed to take a short brake between the blocks. Within a block, each n-back trial ($n = 0, 1, 2, 3, 4$) appeared twice: once with a construction site and once without. The order of the trials was determined pseudorandomly with a few conditions. Firstly, no n-back level could appear twice in a row. Secondly, the construction/non-construction conditions were alternated from trial to trial. These constraints on the randomization were incorporated with the aim of avoiding habituation effects for the memory task and the visuospatial demands. Finally, the order of the trials in the first block was reversed to form the order of trials in the second block.

Prior to performing the experiment, the participant was given instructions about the driving and the memory task. They then performed a practice round (one 2-back trial with no construction and a total of 5 speed signs) to get accustomed to the simulation and the steering wheel. Next, the eye-tracker was calibrated. This involved the participant following a target around the computer screen with their eyes. This procedure was repeated twice: once to calibrate and once to validate whether that calibration was accurate. If the validation was inaccurate, calibration was performed again.

After calibration, the experiment began. Every trial (excluding the very first one) was preceded by an eye-tracking drift correction. This required the participant to look at a target at the center of the screen. If the measured eye position deviated too far from the position of the target, calibration was performed again. Otherwise the deviation was automatically taken into account with recording of the eye position. After drift correction, a pop-up message appeared telling the participant which n-back task they should perform in the next trial. The percentage of total trials they had already completed was also shown in the message. The participant had to press an OK button on the steering wheel to start the trial. Unlike in the experiment by Scheunemann et al. (2019), we did not include warning messages telling participants to change their speed when it was incorrect, as this could instigate an unwanted effect of the participant trying to guess the correct speed.

2.4 Data collection

Behavioral data was recorded to track the participant's driving behavior and performance on the n-back task. The raw variables were recorded every 5 milliseconds. To assess the participant's driving behavior, the steering angle of the steering wheel was recorded. Also, how much the brake-and accelerate-pedals were pressed was recorded. The position of the participant's car and the *autocar* was recorded as well. The speed of the car was also recorded and used to determine the error rate of the speed regulation task. The moment when a speed sign appears was also recorded, which is useful to determine when the participant is expected to change his/her speed. Lastly, a variable was recorded that tracked when the indicators were used. This helps determine when a lane change was initiated and how long it took.

The eye-tracker recorded a number of raw variables at a rate of 500 Hz, two of which are relevant for the current study. Eye positions were measured in x and y coordinates relative to the PC monitor (1920 × 1080 px). Pupil size was measured in terms of diameter in arbitrary units. The eyetracker recorded only one eye (specifically the left) as this is most common in eye-tracking experiments (Hutton, 2019).

2.5 Data analysis

A number of participants were excluded for each analysis. 16 participants were excluded for all analyses either due to incorrect trial lengths or the participant having outlying error-rates in 0-back and 1-back trials, which were regarded as the task performance not being indicative of a real attempt. For the eye-tracking analysis, 6 more participants were excluded because of missing data, resulting in a sample group of 16. For the speed regulation task performance and the lane deviation, no more participants were excluded, resulting in a sample group of 22. For the steering reversal rate, 15 participants were excluded due to missing data, resulting in a sample group of 7. For the number of collisions, 3 participants were excluded, resulting in a sample group of 19.

Driving behavior

To analyse the working memory performance in the speed regulation task, error-rates were calculated. The error-rate was calculated in terms of a proportion; it is the percentage of target speeds that the participant failed to reach, expressed as a value from 0 to 1, where 0 is 0% error-rate and 1 is 100% error-rate. Each trial was scored manually by checking, for each speed sign, if the participant reached a target speed and stayed on this target speed for a significant time; *error rate = incorrect target speed in trial / number of speed signs in trial.*

To analyse the driving performance, lane deviation was calculated for each trial over all participant. The lane deviation measure is used to compare driving performance between different levels of working memory performance and visuospatial demands. Lane deviation is defined as the average deviation from the lane center over a trial. Lane changing manoeuvres were excluded as they are an intended deviation from the lane center and should therefore not be included. Lane deviation is denoted in an arbitrary unit, as the values are based on the specifics of the lanes and cars.

Additionally, the number of collisions were analysed as a measure of driving performance. It is expressed as the proportion between number of collisions and number of interactinos. A collision is defined as the simulated car and the participant's car touching. An interaction is defined as the participants car overtaking the simulated car, where a possible collision can occur. The proprtion is calculated as follows; *proportion collisions = number of collisions / number of interactions*

To support both measures of driving performance, another measure is used to validate the difference in driving difficulty. To measure driving difficulty, steering reversal rate is calculated. Steering reversal rate is a frequency measure, which provide an indication of driving difficulty, rather than driving performance (McLean and Hoffman, 1975).

Steering reversal rate was defined as the number of times the center of the steering wheel was crossed. Steering reversal rate generally increases as driving difficulty increases, as more steering correction are necessary to driving correctly (MacDonald and Hoffman, 1980). The steering reversal rate is expressed per second; *steering reversal rate = total steer reversals in trial / number of seconds in trial*

Eye-tracking

The eye-tracking data consisted of fixations, saccades and blinks. Fixations where used, as they are the most reliable measurements of both pupil size and fixation location.

To effectively compare pupil size measurements, baseline correction must be applied. It accounts for fluctuations in pupil size during the experiment

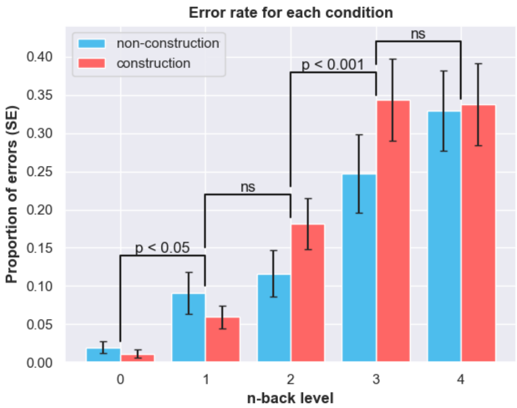

Figure 3.1: Error-rate in the speed regulation task for every combination of n-back level and visuospatial demand across all participants. Black vertical lines indicate standard error of the mean ($n = 22$). Black horizontal lines indicate significant difference in means between subsequent n-back levels of paired sample t-test.

n-back level		p-value	Bonferroni
0-back	1-back	$\mathbf{3.80e^{-4}}$	$\mathbf{4.00e^{-3}}$
0-back	2-back	$\mathbf{2.52e^{-7}}$	$\mathbf{2.52e^{-6}}$
0-back	3-back	$\mathbf{1.15e^{-9}}$	$\mathbf{1.15e^{-8}}$
0-back	4-back	$\mathbf{3.04e^{-11}}$	$\mathbf{3.04e^{-10}}$
1-back	2-back	**0.006**	0.062
1-back	3-back	$\mathbf{1.06e^{-6}}$	$\mathbf{1.06e^{-5}}$
1-back	4-back	$\mathbf{1.01e^{-8}}$	$\mathbf{1.01e^{-7}}$
2-back	3-back	$\mathbf{1.35e^{-5}}$	$\mathbf{1.35e^{-4}}$
2-back	4-back	$\mathbf{4.27e^{-7}}$	$\mathbf{4.27e^{-6}}$
3-back	4-back	0.311	1.00

Table 3.1: Results of paired sample t-test, which tests if there is a significant difference in mean error-rate between every n-back condition. Bonferroni correction is applied. Values in bold indicate a significant difference.

(Mathôt et al., 2018). The baseline period we chose was between the start of the trial and the appearance of the first speed sign, which is a 5 second window. This mean pupil size during this time frame is used to correct the pupil size of that trial. In our study, subtractive baseline correction is used; *corrected pupil size = pupil size - baseline pupil size* The result is an arbitrary unit.

Lastly, the fixations on the speedometer are analysed. To measure this, an area of interest must be chosen, that is, when a fixation is on the speedometer. As this area appeared to differ within trials and between participants, the bounds of this area were chosen manually. Fixations on the speedometer were expressed as a percentage of the total number of fixations during a trial; *fixations on speedometer = fixations on speedometer in trial / total fixation in trial*

3 Results

3.1 Driving Behavior

Error-Rate in Speed Regulation Task

Figure 3.1 shows the performance on the speed regulation task in terms of error-rate. For 2-back and 3-back, driving in the construction condition results in a higher error-rate. For 0-back and 4-back, there appears to be no evident difference. 1-back results in a slight increase of error-rate in the non-construction site. Altogether, no consistent pattern can be observed between the two visuospatial conditions. However, there seems to be an effect between n-back level and error-rate; as n-back level increases, error-rate increases. Only between 3-back and 4-back, the error-rates appear quite similar.

A two-way repeated measures ANOVA with factors visuospatial demands and n-back level showed a main effect only for n-back level [$F(4, 84) = 26.46, p < 0.001$] and a marginal interaction effect between the two factors [$F(4, 84) = 2.51, p = 0.048$].

To further investigate the main effect of n-back level on error-rate, a paired sample t-test is performed. When interpreting Table 3.1, the Bonferroni p-values are used, as it corrects for a potentially inflated type-1 error. Table 3.1 shows that there is a significant difference in mean error-rate for all n-back pairs, except for the pair 1-back - 2-back and 3-back - 4-back. This is also shown by the horizontal lines in Figure 3.1. It creates a grouping effect of the n-back levels: 0-back (lower n-back level), 1-back and 2-back (middle n-back level) and 3-back and 4-back (higher n-back level).

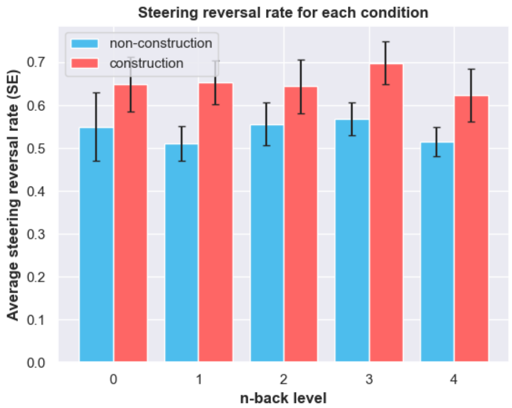

Figure 3.2: Steering reversal rate for every combination of n-back level and visuospatial demand across all participants. Black vertical lines indicate standard error of the mean ($n = 7$).

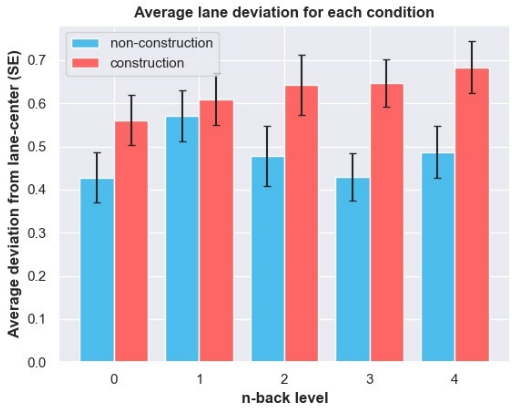

Figure 3.3: Lane deviation from the center of the lane for every combination of n-back level and visuospatial demand across all participants. Black vertical lines indicate standard error of the mean ($n = 22$).

Steering Reversal Rate

Figure 3.2 shows the steering reversal rate for every condition. When we look at steering reversal rate between n-back levels, there appears to be no increase or decrease in steering reversal rate when n-back level increases. However, between construction and non-construction, there seems to be a clear difference, where the construction condition has a higher steering reversal rate than the non-construction condition.

A two-way repeated measures ANOVA with factors visuospatial demands and n-back level showed a main effect only for visuospatial demands $[F(1,6) = 28.93, p < 0.001]$.

The main effect of visuospatial demands on steering reversal rate indicates that the construction site increases driving difficulty.

Lane deviation

Figure 3.3 shows the lane deviation from the center for each condition. First, we look at the difference between the construction condition and non-construction. For each n-back level except 1-back, we see a higher lane deviation in the construction site. When we look at the difference in lane deviation between n-back levels, there seems to be an increase in lane deviation, which is only visible in the construction condition.

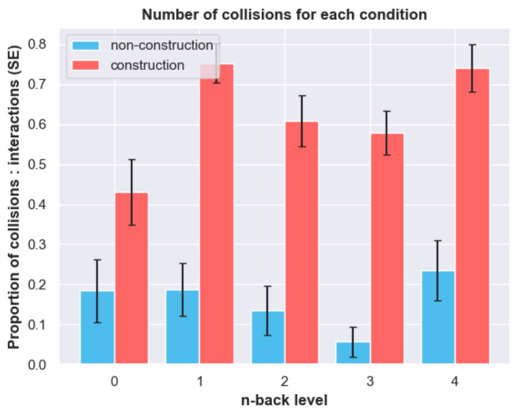

Figure 3.4: Number of collisions relative to number of interactions for every combination of n-back level and visuospatial demand across all participants. Black vertical lines indicate standard error of the mean ($n = 19$).

A two-way repeated measures ANOVA with factors visuospatial demands and n-back level showed a main effect only for visuospatial demands $[F(1,21) = 14.52, p < 0.001]$.

The main effect of visuospatial demands on the lane deviation indicates that the construction site decreases driving performance, as participants deviate more in the construction site.

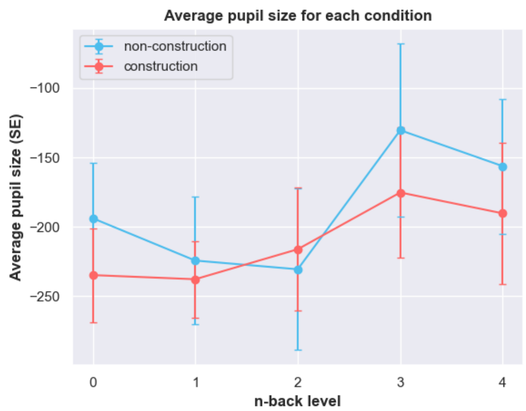

Figure 3.5: Mean pupil size for every condition. Pupil size was correct by using subtractive baseline correction. Vertical lines indicate standard error of the mean ($n = 16$).

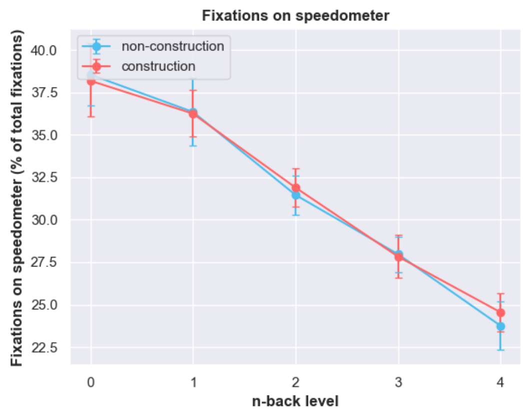

Figure 3.6: Fixations on the speedometer as a percentage of total fixations for every conditions over all participants. Bars indicate standard error of the mean ($n = 16$).

Car collisions

Figure 3.4 shows the number car collisions in proportion to the number of interaction for every condition. When we look at the difference between visuospatial demands for each n-back level, we see a clear difference. For each n-back level, the construction condition has a higher number of collisions than the non-construction condition. Looking at the the difference in collisions between n-back levels, no coherent pattern can be observed.

A two-way repeated measures ANOVA with factors visuospatial demands and n-back level showed a main effect only for visuospatial demands $[F(1, 18) = 124.32, p < 0.001]$.

The main effect of visuospatial demands on the proportion of collisions : interactions indicates that the construction site decreases driving performance.

3.2 Pupil size

Figure 3.5 shows the mean pupil size for each condition in terms of n-back level and visuospatial demands. When we look at pupil size over n-back levels, the figure shows that n-back level has no consistent effect on pupil size in the non-construction condition, as from 0-back to 2-back, the pupil size decreases, to then increase on 3-back and then drops again on 4-back. However, for the construction condition we see a more stable pattern, when n-back increases, pupil size seem to increase as well, only for 4-back there is a small drop, but still higher than 0,1,2-back. When looking at the differences between visuospatial demands for each n-back level, pupil sizes seem to be lower for the construction site in 0,3,4-back, but no real difference is observed for 1-back and 2-back.

A two-way repeated measures ANOVA showed a main effect between n-back level and pupil size $[F(4, 60) = 2.97, p < 0.05]$ and no effect of visuospatial demands or an interaction effect between the two factors.

3.3 Eye fixations

Figure 3.6 shows the number of fixations on the speedometer. It shows a negative correlation between n-back level and fixations on the speedometer. Evidently, there is no effect between visuospatial demands, as the lines are nearly identical.

As expected, performing a two-way repeated measures ANOVA showed a main effect between n-back level and pupil size $[F(4, 60) = 47.68, p < 0.001]$ and no effect of visuospatial demands or an interaction effect between the two factors.

4 Discussion

In this research, we tried to answer multiple questions related to cognitive load during driving. The first two questions are answered by analysing eye-tracking data and the second two questions are answered by analysing behavioral driving data.

First, we focus on the eye-tracking findings. Starting with the question *"Can pupil size predict cognitive load?"*.

We found that there was a main effect of n-back level on pupil size. This is also supported when we look at Figure 3.5, we see that for the higher n-back levels, namely 3-back and 4-back, the pupil size is higher than that of the lower n-back levels. This means that an increase in n-back level, and therefore an increase in WML leads to increasing pupil size. As for visuospatial demands, Figure 3.5 shows that there is no coherent pattern between construction and non-construction in terms of pupil size; for 0-back, 3-back and 4-back the pupil size seems lower in construction site, but for 1-back and 2-back, no real difference is observed. Additionally, no significant effect of visuospatial demands is found on pupil size.

We can therefore conclude that pupil size only predicts one component of cognitive load, namely working memory load, and not visuospatial demands.

The second question we asked relating to eye-tracking was, *"Does cognitive load have an effect on the frequency of speedometer checking?"*

The results in Figure 3.6 gave us a clear picture; the number of fixations on the speedometer decreased as n-back level increased. Moreover, a main effect of n-back level on the number of fixations was found. As for visuospatial demands, both the plot as well as the statistical test indicated that there was no effect on the number of fixations on the speedometer.

From this we can conclude that only working memory load has an effect on the frequency of speedometer checking. Increased working memory load leads to a decrease in speed-keeping attempts, possibly due to a lack of cognitive resources.

In the context of practical applications, we have shown that an important component of cognitive load, namely working memory load, can be predicted by pupil size. One could think of a setting where eye-tracking is used to indicate when the working memory might be overloaded during driving. It also further consolidates the consensus that pupil dilation is a viable measure of cognitive load (van Gerven et al., 2004). Furthermore, we have shown the effect of working memory load on speed-keeping, which shows that an overloaded working memory decreases focus on speed-regulation. This is consistent with findings of Lavie (2010), she showed that the ability to focus decreased when working memory load increased.

In general, incorporating eye-tracking equipment in a car seems like a feasible approach towards adaptive automation. As shown, pupil dilation can be used to measure cognitive load to some degree, giving us a possible time-frame of higher risks in human-driving. Additionally, eye-fixations might give us an indication of how loaded our working memory is, which might also provide us with points in time where additional automation is required.

Next, we will discuss the questions related to driving behavior. Starting with the question *"Does cognitive load have an influence on working memory performance?"*

As Figure 3.1 shows, working memory performance decreases as working memory load increases. This is also consistent with the statistical tests. However, there was no difference in working memory performance between visuospatial demands; there was no coherent pattern in Figure 3.1 and no effect of visuospatial demands in the statistical tests.

We can conclude that working memory performance is only influenced by one of the two components of cognitive load in driving, namely working memory load. This is not in line with what we expected. As Scheunemann et al. (2019) argued for a competition of cognitive resources, where the two relevant domains in driving, namely visuospatial attention and working memory, could negatively impact one another. They showed that an increase in demands for the cognitive domain associated with visuospatial attention led to decrease in working memory performance, which is correspondent with the cognitive system of working memory. Our results do not show this inter-domain interaction. They only show a decrease in working memory performance, when working memory load increases.

A possible explanation for the difference in results might be that our participants were not instructed which of the two tasks, namely driving and the speed regulation task, to prioritize. Participants might have prioritized the n-back task over driving correctly, diminishing the effect of the visuospatial demands. However, Scheunemann et al. (2019) also stated that this might have happened in their study. In their experiment, the participant only received feedback (warning messages that indicated that they were driving at an incorrect speed) on the working memory task. They believe that this might have shifted the focus of the experiment to the speed regulation task. It is hard to determine whether this effect actually occurred, and if it explains the difference. Investigating when participants prioritize a specific task, in a dual-task paradigm, might be an interesting idea for future studies.

The remaining question we still have to answer was *"Does cognitive load have an influence on driving behavior?"*

For this we first validated when task difficulty was increased. We did this by adopting a frequency measure, which is usually a measure of task difficulty (McLean and Hoffman, 1975). The used measure was steering reversal rate.

Figure 3.2 showed that for steering reversal rate, the construction condition seemed to increase driving difficulty, as the steering reversal rate was higher than that of the non-construction condition. However, we did not see an increase in driving difficulty for increased working memory load. Both findings were consistent with the statistical tests.

As the steering reversal rate indicated that the construction site increased driving difficulty, we could then look at the measures of driving performance.

In Figure 3.3, lane deviation was calculated. Once again, apart from the 1-back level, each n-back level had more lane deviation in the construction site. Supported by the statistical, which found that visuospatial demands had a main effect on lane deviation, we can conclude that visuospatial demands indeed have an effect on driving performance. Furthermore, Figure 3.3 showed a slight effect of increased lane deviation as n-back levels increase for the construction site. Statistical tests showed otherwise, as no significant effect of n-back level was found on driving performance.

Figure 3.4 showed the other measure of driving performance, namely number of collisions. It was in line with the findings for lane deviation; the construction condition had more collisions and was therefore associated with decreased driving performance. And again, no effect was found for increased working memory load.

We can conclude that, once again, only one component of cognitive load has an influence on driving performance, namely visuospatial demands. This is not consistent with the findings of Scheunemann et al. (2019). They found an interaction between two cognitive domains; they showed that an increase in working memory performance led to a decrease in driving performance, where performance was related to a highly visual measure, namely lane deviation.

A possible reason for the difference might be that Scheunemann et al. (2019) used a highly realistic virtual reality simulation, when our study used a relatively simple visualization of a car dashboard. It is probable that it was more difficult in our simulation to perceive when the car was correctly centered, and as both lane deviation and number of collisions require a good perception of lane centering, this might have caused a difference in results.

For both findings relating to driving behavior there are practical implications. We have shown that working memory performance can predict the amount of working memory load in driving, which is a surface level observation. However, a malfunctioning working memory does lead to more risks in driving (Walshe et al., 2019), therefore our findings emphasize the importance of adaptive automation to notice and react when working memory load is too high. Furthermore, driving performance is an indicator of a visually more demanding driving task. It gives a better grasp of when a driver might need additional support, possibly in lane-keeping.

It is important to note that our study has some limitations. First of all, a number of participants were excluded in our study. As this differed for basically each analysis, the sample group is inconsistent across analyses. As already mentioned, the driving simulation that we used lacked realism. The traffic consisted of a single other car and the control of the car is not likely to be a realistic counterpart of an actual car. These limitations reduce the statistical power of our results.

Future studies could try to improve the realism of the driving simulation. By increasing the number of other traffic participants, as well as a more realistic car-control. Additionally, one could combine this with our more complete version of the speed-regulation task and see what results they find.

5 Conclusion

Our study has shown that eye-tracking is a valuable asset in the field of adaptive automation. Pupil size in driving proved to be an indicator of working memory load. Additionally, eye-fixations on the speedometer showed that as working memory load increased, the number of fixations decreased. This shows that as WML increases, our speed-regulation capabilities decrease. Both findings could be used to find valuable time-frames where additional automation is required. We also sought to find a possible effect of visuospatial demands on eye-fixations and pupil size, but no significant effect was observed. Furthermore, driving performance and working memory performance gave us more valuable insights. We found that driving performance in terms of lane deviation and number of collisions was a predictor of visuospatial demands. Similar to the eye-tracking results, adaptive automation can use driving performance to detect when a driver might be visually overloaded. Also, the speed regulation task showed that increasing WML decreased the working memory performance, which in itself is not so valuable, but it emphasizes the importance of finding better ways to detect different levels of cognitive load.

We believed that there were two components of cognitive load that were most important, namely working memory load and visuospatial attention. As can be seen in our results, we did not find a measure that predicted both components. This shows there is still a lot to discover about the interactions of different components of cognitive load during driving. When the effect of each subtask in driving can be figured out independently, it provides us with valuable information of how to perfect adaptive automation and increase overall car-safety.

References

P. Antonenko, F. G. Paas, R. Grabner, and T. van Gog. Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22:425–438, 2010.

J. R. Brooks, A. D. Passaro, S. E. Kerick, J. O. Garcia, P. J. Franaszczuk, and J. M. Vettel. Overlapping brain network and alpha power changes suggest visuospatial attention effects on driving performance. *Behavioral Neuroscience*, 132(1): 23–33, 2018.

E. A. Byrne and R. Parasuraman. Psychophysiology and adaptive automation. *Biological Psychology*, 42(3):249–268, 1996.

D. de Waard. *The measurement of drivers' mental workload*. PhD thesis, University of Groningen, 1996.

M. R. Endsley and E. O. Kiris. The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2):381–394, 1995.

S. B. Hutton. Eye tracking methodology. In C. Klein and U. Ettinger, editors, *Eye Movement Research*, pages 277–308. Cham: Springer, 2019.

I. Wagner. Number of households with one or more vehicles in u.s. 2017, 2019. URL `https://www.statista.com/statistics/184082/vehicles-per-household-in-the-usa-in-2001/`.

I. Wagner. Car ownership: number of vehicles per u.s. household 2001-2017, 2021. URL `https://www.statista.com/statistics/551403/number-of-vehicles-per-household-in-the-united-states/#:~:text=On%20average%2C%20there%20are%201.88,disposal%20in%20that%20same%20year`.

J. Elflein. Deaths by motor vehicle-related injuries in the u.s. 1950-2018, 2021. URL `https://www.statista.com/statistics/184607/deaths-by-motor-vehicle-related-injuries-in-the-us-since-1950/`.

W. K. Kirchner. Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4):352–358, 1958.

N. Lavie. Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, 19(3):143–148, 2010.

W. A. MacDonald and E. R. Hoffman. Review of relationships between steering wheel reversal rate and driving task demand. *Human Factors*, 22: 733–739, 1980.

S. Mathôt, J. Fabius, E. van Heusden, and S. van der Stigchel. Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, 50(1):94–106, 2018.

J. R. McLean and E. R. Hoffman. Steering reversals as a measure of driver performance and steering task difficulty. *Human Factors*, 17(3):248–256, 1975.

D. Meister. *Behavioral foundations of system development.* New York: Wiley, 1976.

N. Merat, A. H. Jamson, F. C. H. Lai, and O. Carsten. Highly automated driving, secondary task performance, and driver state. *Human Factors*, 54(5):762–771, 2012.

S. L. Murphy, J. Xu, K. D. Kochanek, and E. Arias. Mortality in the united states, 2017. *NCHS Data Brief*, 328:1-8, 2018.

F. G. Paas, J. J. Merriënboer, and J. J. Adam. Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79:419–430, 1994.

D. D. Salvucci and N. A. Taatgen. Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1):101–130, 2008.

J. Scheunemann, A. Unni, K. Ihme, M. Jipp, and J. W. Rieger. Demonstrating brain-level interactions between visuospatial attentional demands and working memory load while driving using functional near-infrared spectroscopy. *Frontiers in Human Neuroscience*, 12, 2019.

A. Unni, K. Ihme, M. Jipp, and J. W. Rieger. Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in Human Neuroscience*, 11, 2017.

P. W. van Gerven, F. G. Paas, J. J. Merriënboer, and H. G. Schmidt. Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41:167–174, 2004.

E. A. Walshe, F. K. Winston, L. M. Betancourt, A. Khurana, K. Arena, and D. Romer. Working memory development and motor vehicle crashes in young drivers. *Jama Network Open*, 2(9), 2019.

C. D. Wickens. Multiple resources and mental workload. *Human Factors*, 50(3):449–455, 2008.

A Appendix

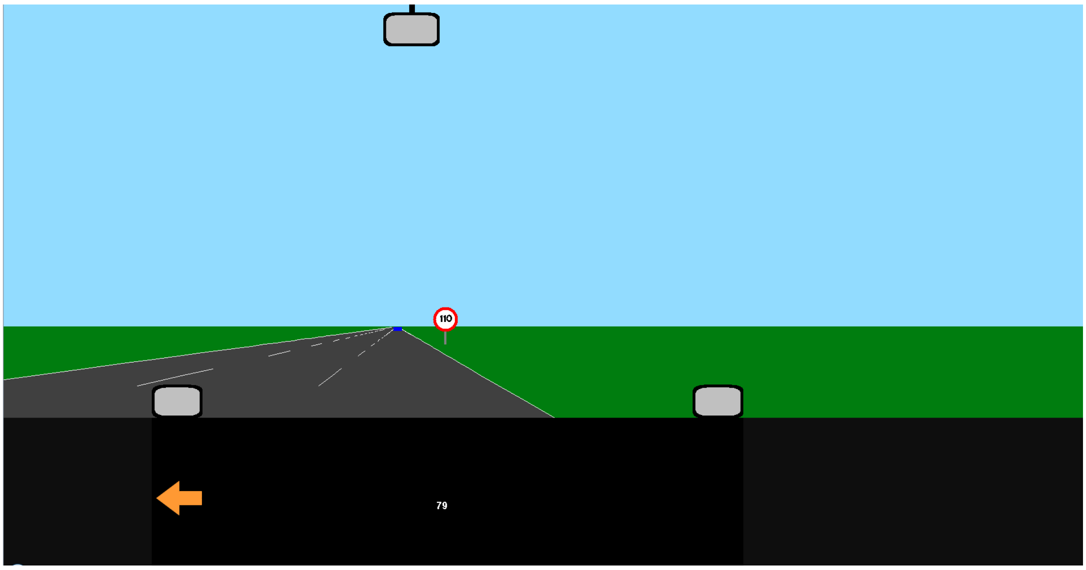

Figure A.1: The driving simulation for the non-construction condition. The orange arrow is the indicator that blinks three times once the participant presses the blinker. The number at the center of the dashboard shows the current speed of the participant's car. At the top and on both sides of the screen are mirrors which are used to see the simulated car approaching from behind.

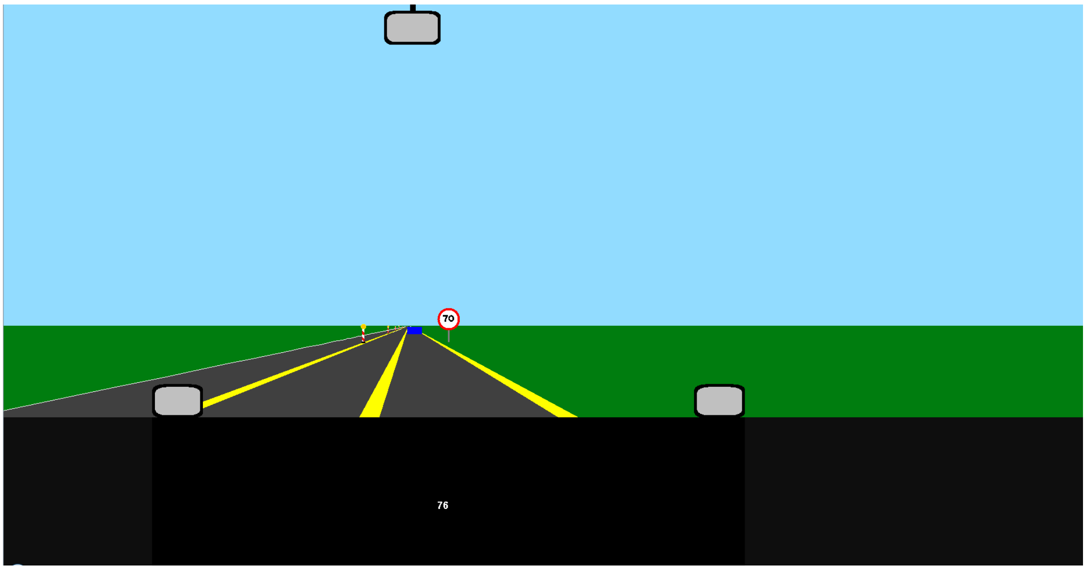

Figure A.2: The driving simulation for the construction condition. All lanes are more narrow than in the non-construction condition and the leftmost lane is closed off by a row of pylons. The blue rectangle on the right-most lane is the simulated car.