MASTER THESIS

Compliance behaviour and workload during simulated driving with in-vehicle information messages.

Author: Bente E. Rootmensen (S1686569)

Master Human Factors and Engineering Psychology

Faculty of Behavioural, Management, and Social Sciences

Graduation project commissioned by Keypoint Consultancy, Enschede

Supervisor Keypoint Consultancy: Dr. ir. Rick Schotman

First supervisor University of Twente: Dr. Simone Borsci

Second supervisor University of Twente: Prof. dr. Jan Maarten Schraagen

APRIL 2021

KEYPOINT CONSULTANCY

UNIVERSITY OF TWENTE.

Abstract

Introduction. In the Netherlands, in-vehicle information systems (IVIS) are more and more incorporated in the traffic environment. Although collection of traffic-related data is quite well organized, the way this data is translated into information and presented to the driver is not well determined yet. This study aimed to advise about the appropriate amount of information presented within in-vehicle messages, to make drivers comply correctly with the messages, but without reducing road safety. ***Method.*** Different in-vehicle messages were presented during the experiment in a driving simulator. Two levels of scenario type ('Traffic jam' and 'Alternative destination') and three levels of message complexity (low, intermediate, and high information quantity) were varied in a within-subjects design. Preference, compliance behaviour, eye movements and subjective mental workload were measured during the experiment. ***Results.*** On average 73.6% of the participants preferred the message with intermediate information quantity. Participants complied the least with the low information quantity message, while the message with high information quantity was watched the longest. Scores of subjective workload did not result in overload but increased with more information quantity. ***Conclusion.*** The message with intermediate information quantity resulted in the best option during this study. These findings are in line with literature, which states that drivers need information that is easy to comprehend but does not contain non-relevant information that increases workload. This study recommends that at least advice for behaviour and a reason for presentation of the message should be present in in-vehicle messages. Further extensions are not recommended.

Keywords: Compliance Behaviour, Mental Workload, In-Vehicle Information Systems, Driving Simulator

Table of contents

1. Introduction

The cognitive ability to focus on multiple tasks is used many times in daily life, e.g., the ability to control a car. Driving is an example of an activity in which the distribution of perceptual, motor and cognitive tasks is used. An example of an extra stimulus during driving is the use of in-vehicle information systems (IVIS). These devices are being used more and more by road users. IVISs can present the driver information he or she normally does not have access to, such as information about an accident or road construction. Furthermore, IVISs decrease anxiety on the road, because of the created overview of the traffic (Vashitz, Shinar, & Blum, 2008). However, these devices require some extra or different cognitive attention in comparison with driving without IVIS. It is an additional task added to the driving situation (Rosenberger, 2015), which may cause different behaviour.

In-vehicle information requires much more complex and precise data about the current traffic situation than, for example, signs because they are aimed to present traffic information meant for the individual driver. Furthermore, these messages present the state of the actual traffic situation instead of a non-changeable static situation when physical signs are used. These complex data should be collected and processed to understandable information for the road user. A collective name of all service providers in collaboration with the already existing physical infrastructure in the Netherlands, which enable digital services, is the so-called 'digital infrastructure' (Stichting Digitale Infrastructuur Nederland, 2020). The data are collected by road managers and information services. They exchange and process it to understandable information for the road user. The information services determine which information will be handed over to the driver at a specific point in time.

To give an overview of the scope of this digital infrastructure, the so-called Data Top 15 is set up. This is a list of 15 data categories in traffic that are potentially involved in the new digital infrastructure (Felici, 2019; Partnership Talking Traffic, 2019). Road managers and information services collect data from the traffic based on these categories, convert this to understandable information and present this to the driver. The driver is minimally or not involved at the start of this process. If it is known in which situation the driver needs a specific amount of information, road managers or information services can make much more efficient choices in data collection.

Together with national and local road managers, two use-cases were collected from practice and used during the current study as realistic traffic scenarios. The use-cases were chosen since these are currently relevant in practice. For each of these traffic scenarios, it is the goal to advise what is the appropriate amount of information that can be conveyed to drivers. Causing that drivers comply with this traffic information, but without reducing road safety. To reach this goal, the extent of compliance behaviour and mental workload were tested during driving with an IVIS. A driving simulator was used in which different kinds of in-vehicle messages were presented with different levels of information quantity in two traffic scenarios. Preference for one of the messages was asked before starting the driving experiment. In the simulated driving environment, the compliance behaviour and mental workload were tested for each of these messages. Furthermore, eye-tracking data were used to measure the time drivers watched the display and the road.

1.1. Theoretical background

1.1.1. Theoretical model

A model that describes the distribution of cognitive workload during a driving situation is the multiple resource model by Wickens (2002) (Figure 1). This is an empirically validated model showing how the level of cognitive load influences people's performance. It is a model that proposes four important dimensions that determine the variance in time-sharing performance. Each dimension contains two separate levels. When two tasks use one level of a different dimension (e.g., when both use auditory perception), it is difficult for the user to separate the two tasks, which may lead to masking or overlapping of one another. When two tasks both use different dimensions (e.g., one using visual resources and one using auditory resources), the performance on these tasks is better (Wickens, 2008). When driving a car, spatial perception/cognition and manual responding are intensively used. If an extra stimulus is added, which also uses these dimensions, problems can arise in driving performance. Driving the car remains the task with the highest priority (Levy & Pashler, 2008).

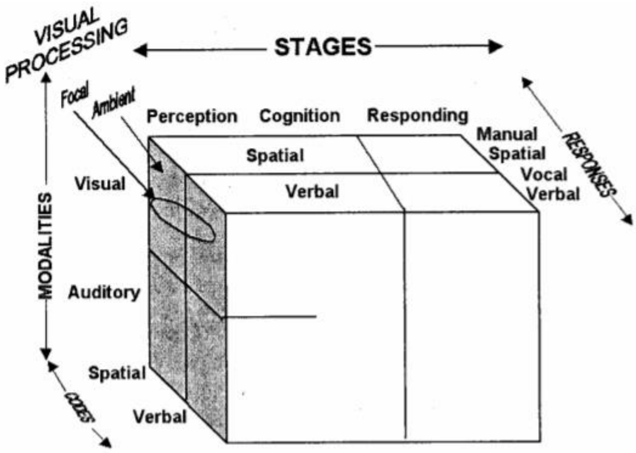


Figure 1. The multiple resource model (Wickens, 2002)

1.1.2. Driver behaviour

As previously explained, drivers can perceive information and stimuli in many different forms and modalities (Wickens, 2002). It is their job to prioritise these and respond to them if necessary. The final driver response or driver behaviour is visible in the control of the car. Many external factors outside the vehicle influence driver behaviour, such as the design of the road itself (de Waard et al., 2007; Foy, & Chapman, 2018), the complexity of the traffic (Teh, Jamson, Carsten, & Jamson, 2014), but also traffic information (Lyu, Xie, Wu, Fu, & Deng, 2017). The task of reading and processing traffic information is a secondary task or embedded task (de Waard, 1996) next to the primary task of driving.

Compliance with the given information will cost a certain amount of effort based on watching the display, cognitively processing the information, and acting according to the information. This effort is based on, among other things, the novelty and information quantity of the message (Lyu et al., 2017) and may affect driving performance (Lansdown, Brook-Carter, & Kersloot, 2004) and compliance. Driver behaviour is very task-specific for different road scenarios (de Waard, 1996). For each road scenario, the driver uses different cognitive resources, which in the end, may result in the specific behaviour and driving performance.

1.1.3. Mental workload

The information of an in-vehicle message needs to be processed by the driver within a specific time span. The relevant information is processed by drivers' working memory capacity, which is in a sense the span to keep items active and to bind these together into structures that are relevant to the task (Uus, Seitlinger, & Ley, 2020). The effort or 'demand for cognitive control imposed by a task' (ISO, 2016) results in the specific cognitive load of an individual, which differs per task and situation. For presenting traffic information, the mental workload is an important factor to keep in mind. The amount of information on traffic signs (Lyu, et al., 2017), and presentation type of information are just a few examples of how this could influence the mental workload. As previously explained, driver performance or behaviour on activities during a ride can be influenced by the effort or mental workload of a driver. For example, the performance of the primary task, that is, the driving task itself (Foy, & Chapman, 2018; Vidulich, & Tsang, 2012). When a new task should be executed, such as reading an in-vehicle message, the driver's attention is diverted from the driving task. This diversion of attention, caused by the complexity or novelty of the message may cause cognitive overload, which could cause major incidents (da Silva, 2014; Strayer, et al., 2013; Figure 2).

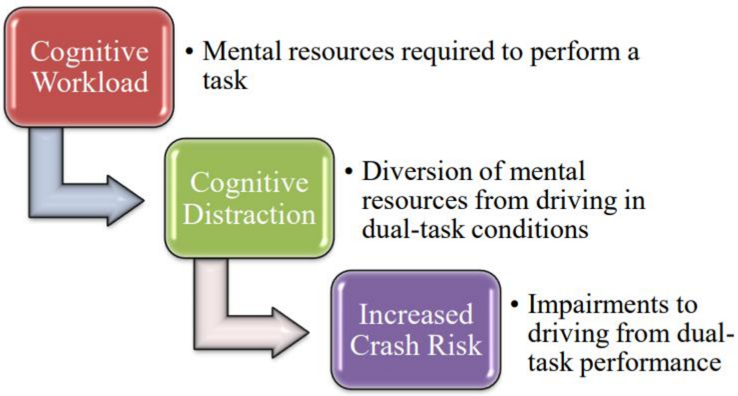


Figure 2. Association between cognitive workload and driving from dual-task performance (Strayer, et al., 2013).

The mental workload is not only influenced by the amount of information but also by how the message is conveyed. Therefore, the design of the message should meet some design principles which requires as little mental workload as possible when reading the message (Harms, Dijksterhuis, Jelijs, de Waard, & Brookhuis, 2018). The form and modality of the presentation are different for every message, causing the possibility of a large difference in process abilities by the driver per message. This will be elaborated further in section 1.3.

1.2. Research questions

Behaviour measures are very context-specific and differ during changing factors such as the urgency of a situation. Furthermore, the amount of traffic information also influences driver behaviour (Guo, Wei, & Wang, 2016; Lyu, et al., 2017). The best-quantified measure to investigate the driver response on the amount of traffic information is compliance behaviour (Williams & Noyes, 2007). For the current study, compliance behaviour is defined based on the interaction with in-vehicle messages and the quantity of information presented in these messages. Compliance is high when the driver supports the underlying reason of the message (Kroon, et al. 2019). For every traffic scenario, it is aimed to reach a certain target behaviour. Following this, compliance behaviour is defined as the accordance between the targeted behaviour and the behavioural outcome based on the specifically given instructions. To increase compliance from the driver, Kroon et al. (2019) advise that arguments about why the message is presented should be added to the in-vehicle message. At the same time, when a message gets too complex, because of the high quantity of information, this may negatively affect behaviour and comprehension (NHTSA, 2016). According to this literature, differences in information quantity influence driver behaviour in different ways. To investigate which level of information quantity is appropriate, so that drivers will comply properly, RQ1 is formulated and tested:

RQ1: Do different levels of information quantity of in-vehicle messages result in different compliance behaviour?

Furthermore, drivers have a preference about how much information they like to receive. More information is sometimes preferred in specific traffic situations. E.g., it is known that drivers prefer more information during tunnel driving because it increases their sense-of-control and decreases anxiety (Vashitz, et al., 2008). It is unclear if this preference also corresponds to driver's compliance behaviour. Therefore, RQ2 is formulated:

RQ2: Does preference towards a level of information quantity affect compliance behaviour?

When there is an increase of complexity (NHTSA, 2016) or quantity and variety of information in messages, workload may also increase (Kroon et al, 2019; NHTSA, 2016; de Waard, 1996) and eventual overload may lead to dangerous situations (Strayer, et al., 2013). Also, adding more information units can lead to anxiety and negative feelings on the road, but still leads to safer driving performance (Koo et al., 2015).

Assessment of driver's mental workload is often based on self-report measures (e.g., Lyu, et al., 2017; da Silva, 2014; Wiese & Lee, 2004). Self-report subjective measures of workload give a valid indication of the experienced workload of the driver (de Waard, 1996). Besides, the following driving performance measures are also linked to how people experience mental workload. This gives a stronger indication of how the consequences of workload affect behaviour.

First, the frequency of steering wheel movements changes when adding another, visual, task to primary driving (Son & Park, 2011; Teh, et al, 2014; Verwey, 2000; Verwey, & Veltman, 1996). Results can show overcompensation during visual distraction and under-compensation in cognitive distraction (Aghaei, et al., 2016; Dong, Hu, Uchimura, & Murayama, 2011).

Second, studies (Harms, 1991; Son, Lee, & Kim, 2011; Son, et al. 2010) showed that drivers show compensating behaviour when their workload increases. Speed reduction is one of these compensating behaviours, which means that drivers create more time to focus on the secondary task. Furthermore, increased workload causes variation in speed or speed control (Aghaei, et al., 2016; Son, Lee, & Kim, 2011; Teh, et al., 2014). Linked to speed is the position of the accelerator pedal. The minimal movements of the accelerator pedal do not always influence the speed but can be an observation of increased task difficulty (Rakauskas, Gugerty, & Ward, 2004). The secondary task does influence the body movements (Hansen, Busso, Zheng, & Sathyanarayana, 2017), such as the accelerator pedal position. The frequency of these minimal accelerator pedal movements may be an indication of increased workload.

For measuring mental workload, two further research questions were formulated. First, the subjective mental workload was investigated to show the differences between the in-vehicle messages used in this study (RQ3). Next, the relation between mental workload and driving performance on steering, speed and accelerator position was investigated (RQ4). The last question is mainly to show the consequences of the level of information quantity on driver's performance on the road. Again, both questions were used for both traffic scenarios:

RQ3: Do different levels of information quantity result in different levels of mental workload?

RQ4: What is the relation between the driving performance measures steering, speed and accelerator position on the one hand and subjective mental workload on the other hand?

1.3. Pre-experimental phase

During the pre-experimental phase, conversations with national and local road managers from the Netherlands took place to determine which traffic scenarios are relevant to use in this study.

Furthermore, the in-vehicle messages were designed during this pre-experimental phase. The design choices are elaborated here.

1.3.1. Traffic scenarios and their target behaviour

Based on the conversations with road managers, two scenarios were chosen which differed in the level of urgency. Both scenarios represent a problematic road situation that Dutch road managers currently like to improve. From the basis, the urgency of these scenarios is different and these are chosen to investigate if different responses exist between traffic scenarios. The urgency of an in-vehicle message influences driver response (Wiese & Lee, 2004)

The in-vehicle message with a high urgency is a message about an approaching traffic jam (TJ scenario). When this message is presented to the driver, he or she should immediately respond to the information given in this message. Therefore, the defined target behaviour, after receiving the information, is:

Drivers watch the IVIS and subsequently reduce their speed by releasing the accelerator pedal or by pushing the brake pedal.

Second, the low urgency scenario contains a message with a recommended alternative destination (AD scenario). In this scenario, the driver plans to go to beach North that day. However, during the ride, it appears that this beach is very busy. Therefore, the in-vehicle message recommends another, comparable, beach South. It is the driver's choice to go to beach South, which makes it non-urgent. It does not cause dangerous traffic situations when the driver does not comply with the message. The defined target behaviour, after receiving the information, is:

Drivers watch the IVIS and subsequently change their lane from the lane from Beach North to the lane to Beach South.

1.3.2. Design in-vehicle messages

The design and modality of a notification are of importance to the level of mental workload a driver experiences. For example, the font, colour, structure of the visual message can greatly influence the ease with which the driver processes the content of the message. These design requirements are explained in the following section and applied to the in-vehicle message design. Designing is an iterative process, which means that feedback from different road users was necessary to improve the in-vehicle message. The pilot test results and feedback from different participants can be found in Appendix H.

1.3.2.1. Structure of the provided information

An in-vehicle message should be understandable for every driver. A clear and logical structure of the different pieces of information improves this comprehension. Gestalt principles such as *proximity*, objects physically close to each other are perceived as a group, and *similarity*, objects which look similar are perceived as a group, are used for the design of these messages (Johnson, 2014). Figure 4 shows that different visual groups are created by the looks of the colour and text frames. Figure 3 shows an example of proximity.

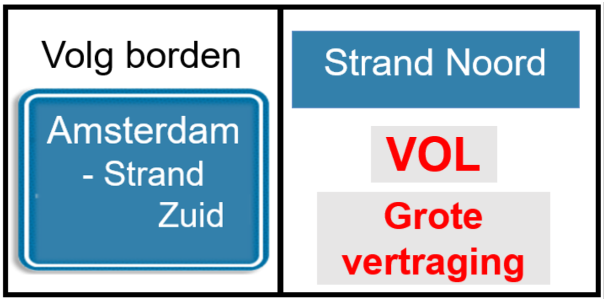


Figure 4. Example of similarity: view similar colours of the traffic signs.

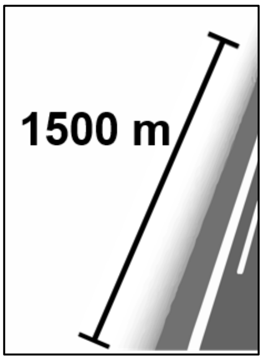


Figure 3. Example of proximity in the current design.

Furthermore, pieces of information need hierarchy, so readers can group more important and less important information. According to Johnson (2014), the most important information should be placed at the top. This rule is executed in the design of the alternative destination message. However, the pilot tests of the design of the traffic jam message show a contradictory result. Here, both participants explained that the most important information should be placed at the bottom because they see the message from the perspective of the driver (see Figure 5). In this case, the results from the pilot tests will be followed.

Figure 5. Visual hierarchy traffic jam message.

1.3.2.1. Written text

Reading is not a natural skill for a human being (Johnson, 2014). Quick legibility of text is important during a task in which reading does not have the highest priority (Wickens, Lee, Liu, & Gordon-Becker, 2004). Legibility can be improved by choosing the right font size and letter spacing (NHTSA, 2016). In the current study, the font type Arial was used, which is chosen as the clearest font type for these messages according to the sizes of the NHTSA (2016) and Mijksenaar (2009).

Figure 6. High information quantity message for the traffic scenario 'alternative destination' containing the maximum amount of information units.

Unfamiliar vocabulary should be avoided, which means that the language should be clean and simple (Johnson, 2014), and should be consistent with the knowledge of the driver (Wickens et al., 2004). Furthermore, some specific advice on the length of the text was given by Kroon et al. (2019). For all the messages, these pieces of advice were followed. One of the messages which reaches the limits of this advice is the message with high information quantity in the alternative destination scenario (Figure 6).

1.3.2.1 Background colour

Avoid noisy backgrounds which can disrupt recognition of the characters of the text (Johnson, 2014), so the accessibility of the information is high (Wickens et al. 2004). Furthermore, the mutual contrast value between the background and the text should have a minimum value of 70 (see Figure 7; Mijksenaar, 2009). This can be calculated by:

$$\frac{Highest\ colour\ value - lowest\ colour\ value}{highest\ colour\ value} \cdot 100$$

Hue	LR(%)
red	13
yellow	71
blue	15
orange	34
green	17
purple	18
pink	30
brown	14
black	8
grey	19
white	85
beige	61

Figure 7. Table of colours and their colour values, or the so-called light reflectance (LR).

All the stimuli in the present study were presented with a contrast between the colour of the letters versus the background higher than 70 (see e.g., Figure 4 and Figure 3) to enable equal accessibility of information to all the participants.

2. Method

2.1. Participants

In total 37 participants took part in the current study. Four of them did not finish because of simulation sickness. Table 1 shows an overview of demographics separated with and without these participants. The language used in this study was Dutch, so participants did not experience additional workload, because of speaking a foreign language. Participants possessed a valid driver's license. The quality of their eyes had to be sufficient, because of the use of the Virtual Reality (VR) glasses, which means that this study did not allow participants with glasses. Participants who were pregnant, suffered from epilepsy or used medication that would influence their driving skills were discouraged to participate, because of the possible health issues associated with VR glasses.

Table 1

Demographics for all participants who finished the online survey and for all participants who finished the simulator experiment.

	Finished online survey	**Finished simulator experiment**
Females	23	20
Males	14	13
Mean age	27.7	26.2
Mean driving hours per week	3.6	3.2

2.2. Experimental design

This study used a 2 (traffic scenario) x 3 (information complexity) within-subjects design (see Table 2): traffic jam (TJ) notifications with three levels (TJLow, TJIntermediate, TJHigh); alternative destination (AD) notifications with three levels (ADLow, ADIntermediate, ADHigh), and a control condition without notification. The order of presentation of the different messages and scenarios were counterbalanced to prevent learning effects (see Table 2). The message with the low information quantity was always placed at the start of the scenario. This choice was made to observe how people responded without advice for behaviour included in the message.

Table 2

The counterbalance of traffic scenarios (Traffic jam (TJ) and alternative destination (AD)) and the different information messages (Low (L) , Intermediate (I) and High (H) information quantity).

Participant number	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**	**Trial 6**	**Trial 7**
1, 5, 9, 17, 21, 25, 29, 33, 35 and 36	TJL	TJI	TJH	Control	ADL	ADI	ADH
2, 6, 10, 14, 18, 22, 26, 30 and 37	ADL	ADI	ADH	Control	TJL	TJI	TJH
3, 7, 11, 15, 19, 23, 27 and 31	TJL	TJH	TJI	Control	ADL	ADH	ADI
4, 8, 12, 16, 24, 28, 32 and 34	ADL	ADH	ADI	Control	TJL	TJH	TJI

2.3. Materials

As previously mentioned, three versions of messages were created for every traffic scenario. The first, low version, only consisted of the most basic information about the traffic situation, without advice for driving behaviour. The second, intermediate version, consisted of the basic information, but with advice for driving behaviour. The third, high version, consisted of more information about the traffic situation, e.g., the reason for the message and the specific location of the situation. The visual design of the messages used in the study may be found in Appendix A.

Before the sessions in the driving simulator, participants received a preview of the design of the in-vehicle information messages via a questionnaire in Qualtrics survey software. The design of these 'online survey'-messages were corresponding to the simulator messages, but with slightly different content. These content differences were comparable to regular differences in practice, e.g., the distance towards a traffic jam or the name of the destination. The messages used in this questionnaire may be found in Appendix B. This pre-exposure was done so participants could better comprehend the messages in the simulator experiment.

For the simulator experiment, the Logitech G920 Driving Force was used to simulate the driving experience (see Figure 8). This simulator is located in the Behavioural, Management and Social sciences (BMS) lab of the University of Twente. The steering wheel and pedals had been constructed to a frame with a car seat to get a realistic driving experience. The simulated environment had been developed in the cross-platform-game-engine Unity3D. This environment, in which participants drove, was a city with different buildings, roads and junctions. The traffic density on the streets was relatively high, so the driving task required some mental workload. The animation of the car was a sedan with an automatic gear shift. Within the animation interior of the car, an in-vehicle information display had been placed for the presentation of the traffic notifications. This display was located right to the steering wheel (see Figure 9), so the view on the road was not blocked by the display.

Additionally, the environment in which the participant drove was displayed with the Varjo VR-2 Virtual Reality glasses. These glasses have a human-eye resolution of over 60 PPD plus improved peripheral vision and colour consistency (Varjo, 2020a). Furthermore, they include the Varjo 20/20 Eye Tracker, which gives precise eye data by using high-speed video cameras (100 Hz stereo eye-tracking) and tracking both eyes. The sensors have a resolution of 1280 x 800 pixels and the recordings are all done in the infrared (IR) spectrum (Varjo, 2020b). iMotions software (version 8.2) was used for the acquisition and analysis of the data.

The Rating Scale Mental Effort (RSME; Zijlstra, 1993; see Appendix F) was used for measuring subjective workload. Furthermore, simulation sickness was monitored with the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993; see Appendix G).

Figure 8. Logitech G920 driving simulator.

Figure 9. Driver perspective with the in-vehicle display and two areas of interest (AOIs): green = Road, red = Display. Example of the ADHigh message.

2.4. Procedure

Participants received the instruction of the first part of the experiment (Appendix D) and filled out the informed consent (Appendix C), and a demographics section (Appendix E). After this, they filled out the online survey about their preference for in-vehicle messages and were instructed for the second part of the experiment.

Next, participants were seated in the driving simulator and the VR headset was prepared for use (e.g., multi-point calibration). First, they were given the opportunity to drive freely in the environment and practice in the simulated car. Participants were asked to drive in the simulator as if they controlled a real car. After that, one practice trial was executed so participants got used to the experimental sessions. In this way, participants experienced how the in-vehicle display was activated during the ride. In general, people needed 10 to 15 minutes to get used to the driving activity in a simulator (Ronen, & Yair, 2013), which was achieved with the 5 minutes free drive and one practice trial.

When participants finished the practice trial, the experiment proper started. They always started at the same location in the simulated city, just before a turn. After this turn, participants only needed to drive straight ahead. The number of cars in the surrounding traffic was always the same, but their location was varied in every session. After the turn, participants received a notification on the in-vehicle information display (see Figure 9). The moment of the presentation of the message was different for every traffic scenario. The information on the display was about one of the two previously mentioned traffic scenarios (‘Traffic Jam’ or ‘Alternative Destination’).

During the TJ scenario, participants were instructed to accelerate after the turn. At some point, they received a message to reduce speed, because of an approaching traffic jam. Participants could choose to respond to the message by reducing their speed (i.e., releasing the accelerator pedal and/or pushing the break).

During the AD scenario, participants were instructed to go to 'Beach North', but they received an in-vehicle message that they should go to 'Beach South' because it was too busy on the road to 'Beach North'. Participants saw traffic signs with 'Amsterdam, direction for Beach South' above the left lane, while the right lane was going to 'Beach North' (see Figure 9). At the start, they were instructed to drive on the right lane, so they could choose whether they complied with the message and make a lane change to the left lane or not. Furthermore, the traffic signs above the road instructed people to go the first left when they wanted to go to ‘Amsterdam, direction for Beach South’ (see Figure 9). This was to see if people understood the message and looked at the signs above the road.

After a while, the experimenter asked if the participant saw the message, if yes, the session finished. Of course, this was done after participants responded to the message in one way or the other. After the trial, they were asked to fill out the RSME scales (Appendix F). The message was shown to them again. Every 15 minutes, the Simulator Sickness Questionnaire (Kennedy, et al., 1993) was filled out (Appendix G). This procedure was repeated seven times, so every display was measured once, and one control session was executed. The timeline for one single session is visualized in Figure 10.

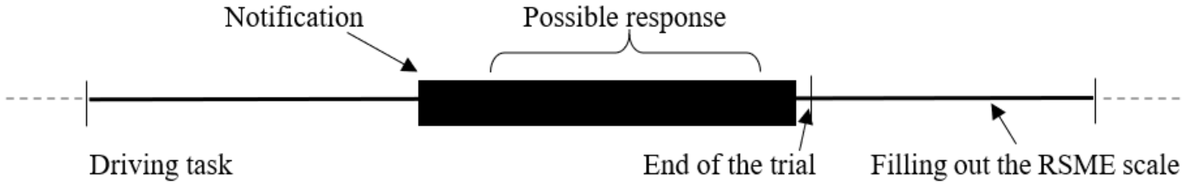


Figure 10. Timeline of one experimental session.

2.5. Measures

The used measures of the current study are elaborated below. An overview of all measures can be found in Table 3.

Table 3.

Overview of measures for the total study. Preference was measured during an online survey. The other measures were observed during the simulator experiment.

	Measurement	**Variable**
Preference		Preference
		Information density
		Content clarity
		Order clarity
Driver behaviour	Compliance behaviour	Speed reduction (TJ)
		RT for lane change (AD)
	Eye movements	Revisits
		Time spent
Mental workload	Subjective workload	RSME display interaction
		RSME total drive
Driving performance	Steering	Frequency
		Variance
	Speed	Reduction
		Variance
	Accelerator position	Frequency
		Variance

The present work accounted for four types of measurements of participants response to the different set of information:

- *Preference:* Preference for one of the three messages of both scenarios was asked in the online survey before the simulator experiment. Also, two other, corresponding, traffic scenarios were incorporated with the same urgency level as the already existing scenarios. One about road construction (RC) and one about an alternative route (AR). This was done to increase the data points of this online questionnaire and to be able to observe individual preference. Furthermore, three specific questions were asked per message. The clarity of the symbols, content and order of the information was measured by a five-point scale from ‘unclear’ to ‘clear’. Furthermore, participants needed to determine if the message contained enough information on a five-point scale from ‘too little information’ to ‘too much information’. The messages used in this questionnaire may be found in Appendix B. These contained slightly different information than the messages used in the simulator experiment.

- *Compliance and eye movement behaviour*: The compliance with the TJ scenario was measured when the driver reduced his/her speed. This means that speed was observed after 4 and 8 seconds to test if and how much participants reduced their speed. Furthermore, the eyes of the participants were tracked to test how much time they spent watching, and how often they revisited the display and the road. The duration of participants watching the in-vehicle display (i.e., time spent, or dwell time) and the number of times people watch the display (i.e., revisits) is a way of compliance with the in-vehicle message. Second, the compliance with the AD scenario was measured based on the response time (RT) of the lane change made. The RT for a lane change was based on the position of the vehicle on the road, which was measured by hitting a box collider in the virtual scene. Also, in this scenario, the eyes of the driver were tracked to observe the time spent watching the display and the road, and how often a revisit was necessary.
- *Subjective workload*: The Rating Scale Mental Effort (RSME; Zijlstra, 1993; see Appendix F) was used as a subjective mental workload scale. The RSME meets the requirements for the replacement of the National Aeronautics and Space Administration-Task Load Index (NASA-TLX; Hart, & Staveland, 1988). The NASA-TLX is a multidimensional scale, which would take several minutes to fill out. Since the participant gets presented with several different notifications, the RSME questionnaire was chosen as a subjective workload measurement. This questionnaire takes less than one minute, which would invoke little memory skills of the participants. Unidimensional scales are useful and as sensitive as multidimensional scales (e.g., Mulder, Dijksterhuis, Stuiver, and de Waard, 2008; Widyanti, Johnson, & de Waard, 2013; Verwey, & Veltman, 1996). Furthermore, a strong correlation has been found between the scores of the RSME and NASA-TLX (Ghanbary Sartang, Ashnagar, Habibi, & Sadeghi, 2017). The fact that participants should indicate their direct invested effort, instead of abstract mental workload, makes it a good scale for self-report subjective workload (de Waard, 1996). To separate the mental effort measured during different tasks, the RSME was filled out twice by the participant during this study. Once for the interaction with the IVIS and once for the driving task in total.
- *Driving performance*: Performance of the drivers after receiving instructions was measured in terms of steering frequency, steering variance, changes in speed, accelerator frequency, and accelerator variance. Frequency measures (steering and accelerator position) were defined as the time between two peaks of a steering wheel movement from left to right or the other way around. Also, the number of peaks was counted. Speed reduction was measured based on 4 and 8 seconds after the display activation. Finally, the variance of steering, speed and accelerator

position were calculated based on the root mean squared error (RMSE), or deviations from a predicted line. In the case of the steering and accelerator position, this line was 0. The predicted speed variance was based on the speed in the control condition.

2.6. Data analyses

All data analyses were performed in Rstudio (version 4.0.3). Subjective workload, eye-tracking and driving performance data were analysed separately for every scenario. Compliance behaviour data was different for both scenarios (i.e., Speed reduction (TJ) and RT on lane change (AD)).

In preparation for analysing the eye-tracking data, AOIs (Areas of Interest) were created. This was done in iMotions software (version 8.2). The two AOIs used in this study were the in-vehicle information display (Figure 9: red) and the road (Figure 9: green). For every AOI, the revisits, and time spent (% and ms) were analysed. Especially, the time spent on road AOI is interesting, because this may indicate how much time people attend on the road. Choice of measures was done partly based on the advice of Carter and Luke (2020).

Driving performance data was first transformed into usable data for analysis. In the case of the steering, speed and accelerator position, data was acquired 3 seconds before and 3 seconds after display activation, so only data was used when the car was on a straight road. Frequency calculation (for steering and accelerator positions) was based on the number of peaks, and the mean time between these peaks before and after display activation. A peak was defined as the greatest or the smallest value in correspondence with the surrounding values, so the change of direction of the steering wheel. The time between these peaks was the time from a peak to a valley or the other way around, given that the difference was greater than 1° (Verwey & Veltman, 1996). Only in the case of frequency, Python (version 3.8) was used for calculating the frequency data.

Furthermore, steering variance, speed variance and accelerator variance were calculated based on the Root Mean Squared Error (RMSE). In the case of the steering and accelerator variance, the predicted reference point was 0 (i.e., when the steering wheel was exactly in the middle, or when the accelerator pedal was not pushed). This indicates the spread of the residual scores are from the predicted line, as also used for steering variance in Rizzi, Jagacinski, & Bloom (2021). In the case of the speed variance, the control condition is taken as the predicted line.

Most of the analysed data were not normally distributed according to the Shapiro-Wilk test. Yet, Analysis of Variance (ANOVA) models were used for comparing group differences and interaction effects. According to Sawyer (2009), these are robust options for violations of normality when the group size is large enough. Also, Rash and Guiard (2004) confirmed that the ANOVA is quite robust against violations of normality. The same argument applies to the use of linear regression: when the sample size is large enough for every group, violations of normality do not noticeably impact the general results (Schmidt, & Finan, 2018).

3. Results

3.1. Compliance and eye-movement behaviour

Compliance behaviour was measured to investigate the response to the in-vehicle messages. Different compliance behaviour was defined for every traffic scenario. Furthermore, eye movement data were collected to observe drivers' eye behaviour during the tasks.

3.1.1. Traffic Jam scenario

Drivers decreasing their speed was the goal of the in-vehicle messages in the TJ scenario. Therefore, for every message, speed was observed at 0 seconds (i.e., the moment of display activation), at 4 seconds after display activation and at 8 seconds after display activation. The control session, i.e., the session without a presented message, is excluded from this analysis because no compliance data was collected here. A repeated-measures ANOVA showed that decrease of speed was significantly different during driving with different in-vehicle messages (TJLow, TJIntermediate, TJHigh), $F(2,68) = 11.00$, $p < .001$, $\eta^2 = .06$. Furthermore, participants drove significantly slower after 4 seconds (54.57 km/h) and 8 seconds (49.76 km/h) of display activation (67.09 km/h), $F(2,68) = 81.13$, $p < .001$, $\eta^2 = .41$. The Tukey HSD post hoc tests were conducted to show the differences between these three messages. Participants significantly decreased more speed during the intermediate message ($p < .001$) and the high message ($p = .015$) in comparison to the low message. This means that participants drove slower in the intermediate and high messages in comparison with the low message (see Figure 11).

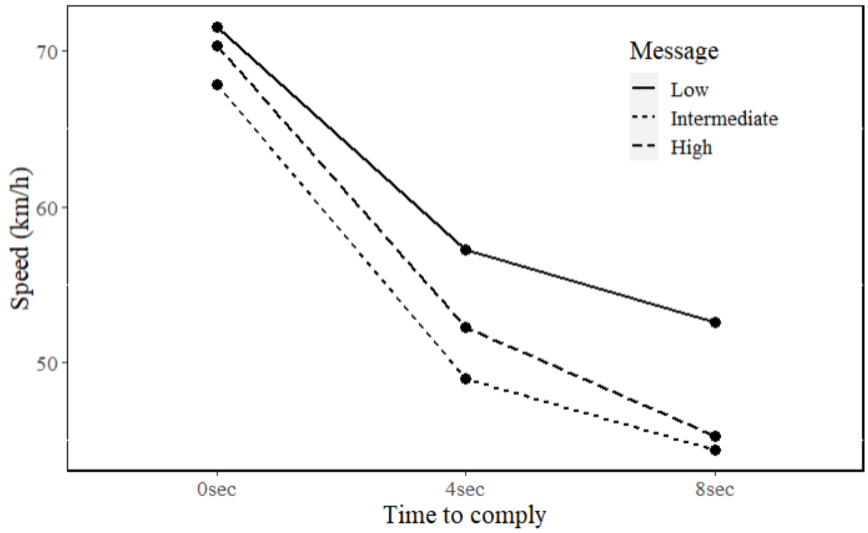


Figure 11. Compliance behaviour for different in-vehicle messages (Low = Message with low information quantity, Intermediate = Message with intermediate information quantity, High = Message with high information quantity). Time to comply was separated in three 'timestamps' (0, 4, and 8 seconds).

Next, eye-tracking data were collected during display activation. As previously explained, the in-vehicle display and the road were areas of interest (AOIs) during eye-tracking measurement. For these AOIs revisits (i.e., the number of returned gaze to the AOI) and the time spent (%) are measured. The descriptive eye-tracking results can be found in Table 4.

Table 4.

Averages for Revisits and Time spent in both AOIs Display and Road.

Type of message	**Display**		**Road**		**Total**
	Revisits	*Time spent (%)*	*Revisits*	*Time spent (%)*	*Time spent (%)*
Control	0.17	0.55	5.24	79.08	79.74
TJLow	1.62	11.79	4.49	69.83	81.74
TJIntermediate	1.00	9.85	4.73	66.57	76.47
TJHigh	2.44	13.68	4.24	62.43	76.18

According to the one-way ANOVA, participants revisited the display significantly different across all the messages, $F(3,132) = 20.47$, $p < .001$. Post hoc comparisons using the Tukey HSD test resulted in significant differences between almost all variables, $p < .05$. Except for the difference between the low and intermediate message, this was not significant. Furthermore, the time spent watching the display was also significantly different across the messages, $F(3,132) = 50.27$, $p < .001$. Participants watched the high message significantly longer than the intermediate message, $p = .006$. Also, the display was always watched shortest during the control session, $p < .001$. Lastly, the time spent watching the road was also significantly different, $F(3,132) = 10.16$, $p < .001$. However, only significant differences could be found between the type of messages and the control condition, $p < .05$. Differences in time spent between the messages and the Display and Road AOI are visualized in Figure 12.

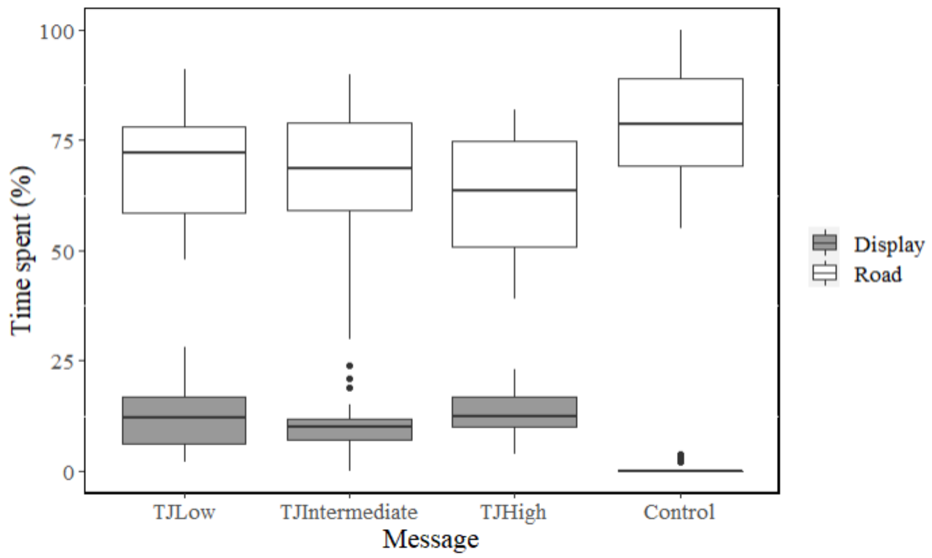


Figure 12. Eye-tracking results in the Traffic Jam (TJ) scenario. Time spent (%) watching the display and time spent watching the road.

3.1.2. Alternative Destination scenario

The goal of the in-vehicle message in the AD scenario was that participants changed lane to another destination, in this case, beach South. The response time (RT) of the lane change was taken as a measure for the AD scenario. Also, in this case, the control session, i.e., the session without a presented message, is excluded from this analysis because no compliance data was collected here. In this case, a one-way ANOVA was conducted to test group differences. This resulted in a significant difference between the messages, $F(2,93) = 167.34$, $p < .001$. According to the Tukey HSD post hoc test, participants were significantly slowest during the low message (14.8 sec), $p < .001$, in comparison to the intermediate and high message. Figure 13 shows this clear difference.

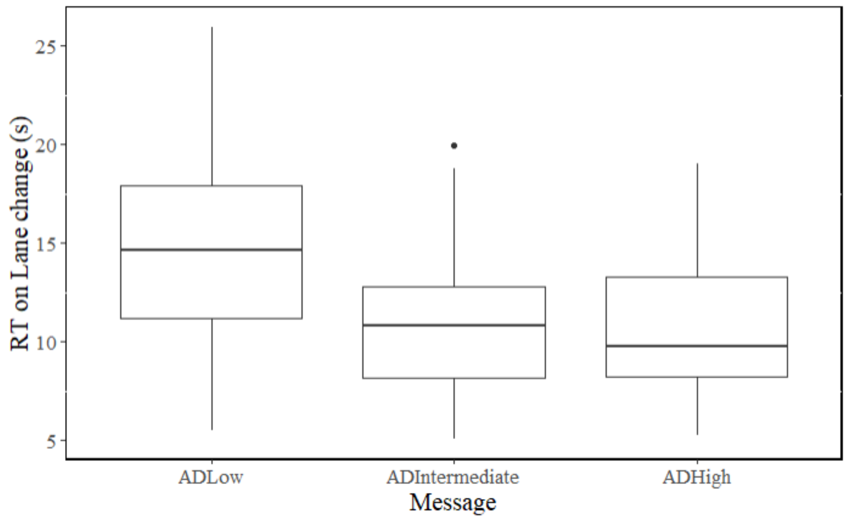


Figure 13. Compliance behaviour (i.e., RT on lane change) in the AD scenario.

Next, eye-tracking was analysed in the same manner as the TJ scenario. Table 5 presents all the averages for revisits and time spent watching the display and the road.

Table 5.

Averages for Revisits and Time spent in both AOIs Display and Road.

Type of message	**Display**		**Road**		**Total**
	Revisits	*Time spent (%)*	*Revisits*	*Time spent (%)*	*Time spent (%)*
Control	0.17	0.55	5.24	79.08	76.65
ADLow	2.82	9.14	7.85	73.58	82.76
ADIntermediate	2.88	9.31	7.30	70.13	75.34
ADHigh	3.59	13.54	7.56	66.14	77.65

A one-way ANOVA was conducted to test if group differences existed. Significant group differences existed between the number of revisits drivers needed on the display, $F(3,128) = 26.48$, $p < .001$. According to the Tukey HSD post hoc test, only a significant difference was found between the revisits during the control session on the one hand and sessions with a message on the other hand, $p < .001$. No significant effects between the revisits during other messages were found. Furthermore, the time spent watching the display was significantly different between the messages, $F(3,128) = 55.35$, $p < .001$ (Figure 14). The ADhigh message was watched significantly longer than the other two messages, $p < .001$. Lastly, significant differences were found between the time spent watching the road, $F(3,128) = 10.83$, $p < .001$. Also, in this case, most significant differences are based on time spent watching the road in the control session and the session with messages, $p < .01$. Furthermore, the ADLow and ADHigh messages differ significantly from each other, $p = .011$.

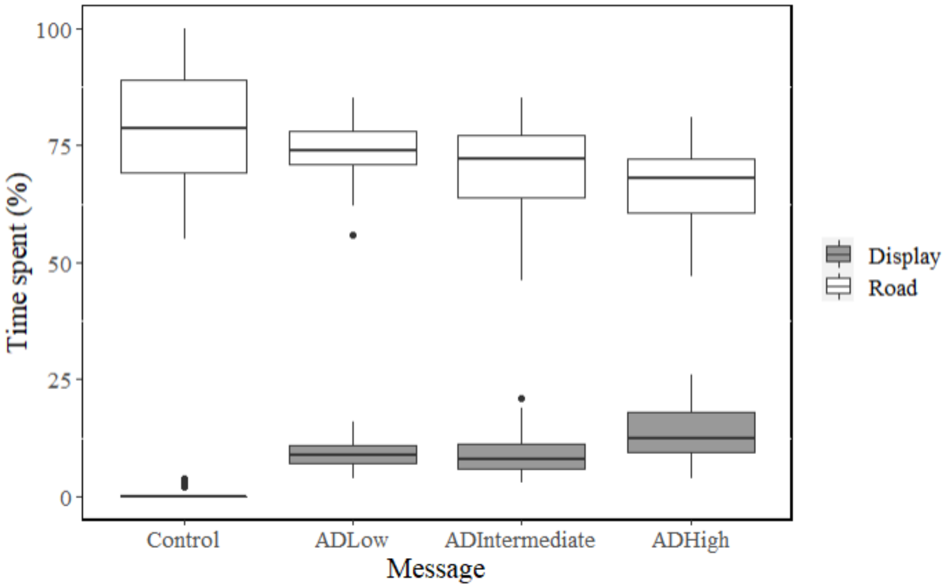


Figure 14. Eye-tracking results in the Alternative Destination (AD) scenario. Time spent (%) for the Display AOI (grey) and Road AOI (white).

3.1.3. Summary of the key findings: Compliance and eye movement behaviour

Table 6 shows an overview of all the key findings in the section on compliance behaviour and eye movements. Eye movements were measured in both scenarios, and compliance behaviour was measured differently per scenario.

Table 6.

Overview of the key findings in compliance and eye movements behaviour for both traffic scenarios. Compliance in the TJ scenario is the speed difference between display activation and 8 seconds after display activation. Compliance in the AD scenario is the RT on lane change.

Type of message	**Compliance**		**Eye-tracking**		
	TJ (km/h)	*AD (s)*	*Revisits Display*	*Time spent Display (%)*	*Time spent Road (%)*
Control (no message)	-	-	0.17	0.55	79.08
TJLow	-19.04	-	1.62	11.79	69.83
TJIntermediate	-23.49	-	1.00	9.85	66.57
TJHigh	-25.09	-	2.44	13.68	62.43
ADLow	-	14.8	2.82	9.14	73.58
ADIntermediate	-	11.1	2.88	9.31	70.13
ADHigh	-	10.6	3.59	13.54	66.14

Compliance results in the TJ scenario showed that drivers decreased significantly less speed when the TJLow message was shown. In the case of the AD scenario, the lane change time was significantly higher during the ADLow message. The eye movement data resulted in significantly more revisits during the TJHigh in comparison to the other two messages. Revisits of the messages in both scenarios were significantly different from the control session. Furthermore, drivers spent the most time watching the message with the high information quantity in both scenarios. While the significant difference in time spent watching the road is mostly based on the difference in the control and the messages in both scenarios. During the control session, drivers spent more time watching the road in comparison to driving with messages. Yet, during the ADLow message, drivers did significantly watch the road more in comparison to the ADHigh message.

3.2. Preference

Data for the preference regarding in-vehicle messages were collected during the online survey before the simulator experiment to check if the preference of receiving a certain type of information affected the compliance behaviour of participants.

3.2.1. Traffic Jam scenario

Before the experiment in the driving simulator, participants chose their preference for one of the three in-vehicle messages for the TJ scenario and RC scenario. Most of the participants preferred the intermediate message for both scenarios: 78.4% participants in the TJ scenario and 70.3% participants in the RC scenario. Furthermore, participants rated the information density, content clarity and order clarity for every message type (see Table 7). According to a one-way ANOVA, scores on information density, $F(2,108) = 77.2$, $p < .001$, and content clarity were significantly different, $F(2,108) = 13.19$, $p < .001$. Further analysis with a Tukey HSD post hoc test showed that information density in the TJIntermediate message was significantly found most appropriate, $p < .001$. The TJLow message scored significantly lowest on the content clarity item, $p < .001$.

Table 7.

Means and standard deviations (SD) of the online survey scores (1-5 Likert scale) on the information density(1 = too little information, 5 = too much information), content clarity and order clarity (1 = not clear, 5 = clear) of the TJ messages.

	Information Density		**Content Clarity**		**Order Clarity**	
	M	*SD*	*M*	*SD*	*M*	*SD*
TJLow	2.24	0.76	3.46	1.39	3.95	1.29
TJIntermediate	3.27	0.51	4.65	0.68	4.24	0.95
TJHigh	4.14	0.67	4.35	0.92	3.70	1.24

To test if preference is linked to the compliance behaviour measures (i.e., speed reduction), three groups were created based on the preferences expressed by participants. This means that the low group consisted of participants who overall prefer the message with the low information quantity displayed in the car. The intermediate group consisted of participants who overall prefer the message with the intermediate information quantity displayed in the car. The high group consisted of participants who overall prefer the message with the high information quantity displayed in the car. A three-way mixed ANOVA was conducted to test group difference of the within variables message (TJLow, TJIntermediate, TJHigh), and time to comply (0 sec, 4 sec, 8 sec) of speed between the preference groups. A significant main effect on ‘time to comply’ was found, $F(2,74) = 46.22$, $p < .001$, $\eta^2 = .17$. As well as the interaction effect of the messages and ‘time to comply’, $F(6,222) = 10.60$, $p < .001$, $\eta^2 = .07$. This significant effect was especially based on the interaction in the intermediate preference group. Just as with the previously reported compliance behaviour results, participants decreased less speed during the TJLow message ($p < .05$). However, the three-way interaction of preference group, message, and time to comply was not significant. This means no group differences were found between the different preference groups.

3.2.2. Alternative destination scenario

Participants chose, in the same way as in the TJ scenario, which type of display they preferred for the AD scenario and the AR scenario. 73% of the participants preferred the intermediate message in both the AD and AR scenario. According to a one-way ANOVA, a significant difference exists between the information density scores, $F(2,108) = 39.9$, $p < .001$. The post hoc Tukey HSD test shows that the

ADHigh message consisted of the highest information density ($p < .001$). For the content clarity scores, $F(2,108) = 7.09$, $p = .001$, only a significant difference existed between ADHigh and ADLow ($p < .001$). Scores on order clarity, $F(2,108) = 15.22$, $p < .001$, were significantly highest for the ADLow display ($p < .05$). See Table 8 for the descriptive scores.

Table 8.

Means and standard deviations (SD) of the online survey scores (1-5 Likert scale) on the information density(1 = too little information, 5 = too much information), content clarity and order clarity (1 = not clear, 5 = clear) of the AD messages.

Type of message	**Information Density**		**Content Clarity**		**Order Clarity**	
	M	*SD*	*M*	*SD*	*M*	*SD*
ADLow	2.46	0.65	4.68	0.67	4.84	0.44
ADIntermediate	2.81	0.46	4.35	1.06	4.11	1.26
ADHigh	3.78	0.82	3.78	1.27	3.43	1.34

To test if the preference for one of the messages is linked to compliance behaviour measure (i.e., the RT on lane change), three groups were created based on the preferences of the in-vehicle message. A two-way mixed ANOVA was conducted to observe the differences between the preference groups in compliance behaviour on the three messages. Only a significant main effect for the messages was found, $F(2,70) = 12.70$, $p < .001$, $\eta^2 = .15$. No interaction effect on the preference groups and the messages was found, which means that no group differences between the preference groups were observed.

3.2.3. Summary of the key findings: Preference

In all scenarios, most of the participants preferred the message with intermediate information. This also became clear in most of the items on information density, content clarity and order clarity. In the TJ scenario, the information density was most balanced in the intermediate message (i.e., closest to score 3). The content of the message was least clear in the TJLow message. In the AD scenario, the ADHigh scenario contained the highest information density and the lowest content clarity. Order clarity was significantly highest in the ADLow message. No effect of compliance behaviour was found between the preference groups.

3.3. Mental workload

To show the difference in mental workload between the three messages, subjective mental workload was measured based on the RSME scale. Furthermore, different driving performance measures (i.e., steering, speed, and accelerator position) were collected to investigate the relationship with mental workload.

3.3.1. Traffic Jam scenario

The Rating Scale Mental Effort (RSME) was used to test the subjective mental workload of the participants after driving with the in-vehicle messages. The RSME was tested in two ways, one time for the interaction with the in-vehicle display and one time for the total driving task. A one-way ANOVA was conducted to test the group differences in both RSMEs. Scores of the RSME on display interaction were significantly different for the three groups (TJLow = 27.51, TJIntermediate = 21.8, TJHigh = 48.5), $F(2,102) = 29.74$, $p < .001$. From a Tukey HSD post hoc analysis, this means that drivers experienced significantly the most workload during interaction with the TJHigh message ($p < .001$). Also, the scores of the RSME on total driving task were significantly different (TJLow = 27.74, TJIntermediate = 25.66, TJHigh = 33.23, Control = 21.35), $F(3,136) = 3.48$, $p = .018$. Only in this case, the control ride and the ride with the TJHigh message significantly differed, $p < .01$.

To predict the display scores based on the total task scores and the in-vehicle message a multiple regression for the RSME Display, RSME Total and messages was conducted. A significant regression equation was found, $R^2_{adj} = .6164$, $F(5,29) = 11.93$, $p < .001$. This means that 61.64% of the variation in display scores can be explained by the model containing total task scores. The total task scores significantly predicted the display interaction scores, $\beta = 0.70$, $t(29) = 3.49$, $p = .002$. The messages were not a significant predictor.

To give some insights on the driving performance measures in association with mental workload, Pearson correlations between the driving performance measures and the RSME scores were calculated. As shown in Table 9, a significant positive correlation exists between the Display scores and Total task scores of the RSME. Second, no correlation was found between one of the driving performance measures and the RSME scores. Therefore, no further analysis was done for the driving performance measures.

Table 9.

Correlation matrix of RSME scores and driving performance measures in the Traffic Jam scenario. Peaks count and mean peak time are two measures for steering frequency.

	1.	2.	3.	4.	5.	6.	7.
1.RSME Display	1						
2.RSME Total task	.58***	1					
3.Peaks count Steering	-.04	-.10	1				
4.Mean peak time Steering	.05	.23*	-.61***	1			
5.Steering variance	-.18	-.01	0.43***	-.05	1		
6.Speed variation	-.16	-.06	-.03	.13	-.06	1	
7.Speed decrease/increase	.02	-.03	.04	-.01	-.06	.03	1

Note: *$p < .05$, **$p < 0.1$, ***$p < .001$.

3.3.2. Alternative Destination scenario

The RSME was also tested for the display interaction and the total driving task in the AD scenario. The RSME display interaction scores were significantly different (ADLow = 31.23, ADIntermediate = 40.28, ADHigh = 55.46), $F(2,97) = 11.30$, $p < .001$. For which the ADHigh scores significantly highest in comparison to the other two messages, $p < .05$. Also, the scores on RSME total driving task are significantly different (ADLow = 38.23, ADIntermediate = 35.41, ADHigh = 44.51, Control = 21.35), $F(3, 131) = 10.43$, $p < .001$. This difference is mainly reflected in the differences between the control ride and the rides with a message, $p < .01$.

Also here, to predict the display interaction scores based on the total task scores and the in-vehicle messages, a multiple linear regression for the RSME Display, RSME Total and messages was conducted. was conducted. 59.87% of the variation in display scores can be explained by the model containing total task scores of the AD scenario, $F(5,27) = 10.55$, $p < .001$. The total task scores significantly predicted the display interaction scores, $\beta = 1.02$, $t(27) = 4.24$, $p < .001$. No significant prediction could be made from the messages.

Also, in the AD scenario, RSME scores on display and total task were placed in a correlation matrix with all the driving performance measures. The correlations between the two RSME also showed a positive correlation, just as in the TJ scenario (Table 10). Besides, the correlation between driving performance measures and RSME scores did not show any correlation. So, no further analysis is done on these measures.

Table 10.

Correlation matrix of RSME scores and driving performance measures in the Alternative Destination scenario. 'Peaks count' and 'Mean peak time' are two measures for steering frequency (3. & 4.) and accelerator position (8. & 9.).

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1.RSME Display	1								
2.RSME Total task	.65***	1							
3.Peaks count Steering	.09	.13	1						
4.Mean peak time Steering	-.07	-.05	-.70***	1					
5.Steering variance	-.05	.02	.12	-.03	1				
6.Speed variation	.00	-.06	.16	-.12	.00	1			
7.Speed decrease/increase	.02	-.10	.12	-.09	-.02	-.08	1		
8.Peaks count Accelerator position	-.12	-.04	.16	-.12	.08	-.15	.29***	1	
9.Mean peak time Accelerator position	.28*	.12	-.15	.10	.09	-.15	-.05	-.68***	1

Note: *$p < .05$, **$p < 0.1$, ***$p < .001$.

3.3.3. Summary of the key findings: Mental workload

Table 11 shows the overview of the subjective workload measures descriptive results. Based on the correlation matrices of both scenarios, no clear correlation was found between the RSME scores and the driving performance measures.

Table 11.

Overview of the key findings from the subjective mental workload measures. Scores on the Rating Scale Mental Effort (RSME) for the interaction with the display and the total driving task.

Type of message	**TJ – scenario**		**AD - scenario**	
	RSME Display	*RSME Total*	*RSME Display*	*RSME Total*
Low	27.51	27.74	31.23	38.23
Intermediate	21.8	25.66	40.28	35.41
High	48.5	33.23	55.46	44.51
Control (no message)	-	21.35	-	21.35

The interaction with the TJHigh message significantly caused the highest score in mental workload. For the total ride scores, the TJHigh was only significantly higher in comparison with the control ride. The same significant scores were also found in the AD scenario. The interaction with the display caused significant higher scores in the ADHigh scenario, while the total ride scores were only significantly different between the ADHigh message and the control ride.

4. Discussion

The current study investigated the differences in compliance behaviour and mental workload between different levels of information quantity presented in in-vehicle messages. The aim was to find the most appropriate level of information quantity in a message so that drivers comply correctly with the message without reducing road safety. Two traffic scenarios were used to investigate compliance in mental workload in different urgency levels. These scenarios were called “Traffic jam” and “Alternative destination”. For each of these scenarios, three messages were created and differences between compliance behaviour, preference and mental workload were investigated.

To answer the first research question – i.e., RQ1: *Do different levels of information quantity of in-vehicle messages result in different compliance behaviour?* – compliance behaviour was investigated differently for each scenario.

Regarding the TJ scenario, drivers should decrease speed to comply with the scenario. Also, eye movement behaviour was taken along in this observation. The decrease in speed was significantly less in the TJLow message in comparison to the other, TJIntermediate and TJHigh, messages. Regarding the eye movement behaviour, participants significantly revisited the display most often and they spent the most time watching the display during the TJHigh message. Lastly, participants spent the most time watching the road during the control session in comparison to the sessions with an in-vehicle message.

The compliance in the AD scenario was measured by the RT on the lane change. The ADLow messages resulted in the significantly slowest lane change. Regarding the eye movement behaviour, drivers watched significantly longest on the ADHigh message in comparison with the other messages. Most time was spent watching the road during the control session. However, drivers did watch the road for a significantly longer time in the ADLow message in comparison to the ADHigh message.

In both scenarios, the compliance behaviour was achieved during all messages, but the response differed in intensity and tempo. It is remarkable that during both scenarios, the response to the messages with the low information quantity was less intense. When watching the design of this message, the low version does not contain advice for behaviour that the driver needs to execute. The other two messages do contain this. It is known that a notification with higher comprehension results in a faster response (Hanowski, & Kantowitz, 1997), which also applies to risky situations (Williams, & Noyes, 2007). Furthermore, from research in semi-automated systems, Bhaskara, Skinner, and Loft (2020) reviewed that performance accuracy may increase during messages containing the information as reasoning and predicted outcomes, because of its transparency. This was, among other things, based on improved decision-making accuracy (Mercado, et al. 2015). Also, Kroon et al. (2019) advised adding the reason for behaviour in the message to increase comprehension and compliance behaviour. The low information quantity message in the current study only contained a reason for the message. Probably this message did not contain enough other information, which lowered its comprehension and transparency.

Continuing this, when something is less familiar to people, the viewing time to this object is longer (Carter & Luke, 2020). One can expect that the slow compliance behaviour of the message with low quantity would be partly caused by the long viewing time on the display, however, this was not found during this study. Drivers watched longest at the message with the high information quantity, which is more in line with Guo et al. (2016). They found that the amount of information on a traffic sign is highly correlated with visual behaviour, which means that the time spent on the object increases with adding information units.

To further investigate the compliance behaviour, the individual preference for one of the messages was also investigated – i.e., RQ2: *Does preference towards a level of information quantity affect compliance behaviour?* Despite most participants preferred to receive the message with intermediate information quantity over the other options, no significant difference was identified between preferences and compliance behaviour. So, preference did not affect compliance behaviour during this study.

To answer the third research question about mental workload – i.e., RQ3*: What is the difference in mental workload regarding the information complexity of the messages?* – a subjective mental workload scale (i.e., RSME) was investigated for the interaction with the display and the total driving task itself. A significant relationship was found between the display interaction and total task scores. Based on the total range of the RSME, no excessively high scores were found for both scenarios. This means that no mental overload was observed during all messages. Nevertheless, during both scenarios, significant differences were found between the messages. Regarding the scores on display interaction, the message with the high information quantity resulted in the highest mental workload in both scenarios. The total ride workload was especially different between the session without a message (i.e., control session) and all other sessions. The workload scores were significantly higher during the sessions with a message. In general, the workload scores in the AD scenario were higher than in the TJ scenario.

First, based on the total range of the RSME scale, no overload was identified during this experiment. The increase of mental workload is most of the time caused by a combination of different events, actions, and resources (Rosenberger, 2015; de Waard, 1996). No distracting, less relevant, factors such as talking to a passenger, or the presence of other task-irrelevant signs on the road were taken into consideration during this study. These factors would contribute to a higher workload in real-life (NHTSA, 2016; Strayer, et al., 2013; de Waard, 1996). Still, even though no excessively high mental workload was observed, this study confirms that information volume increases mental workload (Lyu, et al., 2017). It also corresponds to the increase in negative feelings in Koo et al. (2015) when adding more information. This increase in mental workload in the current study was only visible in the high information quantity messages. The differences between the low and intermediate messages were probably not big enough to show significant workload differences.

Lastly, the scores in the AD scenario resulted in a higher workload than the TJ scenario. De Waard (1996) explains that during complex actions, the task demand rises, and visual resources of the driver are increasingly used. For complying with the AD scenario, drivers needed to change a lane, which is, in general, more complex than decreasing speed.

The previous section showed that differences were found between subjective mental workload in the different messages. However, during the investigation of the fourth research question – i.e., RQ 4: *Do different levels of information quantity result in different levels of mental workload?*– no relationship was identified between the subjective mental workload scores and the driving performance measures steering, speed and accelerator position.

These results are in contrast with previous literature about the relation between the driving performance measures, steering frequency (e.g., Son & Park, 2011; Verwey & Veltman, 1996), speed reduction (e.g., Harms, 1991; Son, Lee, & Kim, 2011), and accelerator position (e.g., Hansen et al., 2017), and subjective mental workload. As already mentioned in the results of RQ3, no excessively high scores were found for subjective mental workload. This could explain the noncorrelation between subjective mental workload and driver performance. When the mental workload would increase more, one could expect that driving performance is affected (Mehler, Reimer, Coughlin, & Dusek, 2009).

4.1. Limitations

Some limitations should be considered regarding the experimental design and materials. First, this study was done in a driving simulator which was not yet validated, because it was fairly new. Participants were properly prepared to get used to the differences. Enough time was spent for drivers to try and practice driving in the simulator before starting the experiment. Furthermore, the virtual reality glasses improved the presence and immersion effect of the situation, which increased the ecological validity. Still, results should be interpreted cautiously and can only be interpreted as totally valid when this simulator is validated properly. Second, the message with the low information quantity of both scenarios was always tested as the first message of that scenario (see Table 2). This choice was made to investigate how drivers would respond without receiving advice for behaviour (see Appendix A). The newness of the message could have influenced how participants responded. In defence, participants were already prepared by corresponding messages in the prior online survey (Appendix B). This means that the design and content of the messages were not new anymore. Still, this might not completely abrogate the newness effect. Third, the traffic situation on the road was not always comparable. The number of cars was always the same, but their location differed. This means that traffic density was not always comparable during every session. Behaviour and workload could have been influenced by this. Last, the messages were created by the researcher and checked by an expert, but not validated. Of course, the two scenarios were intended to be different. However, other unintended design differences may have been noticed by the drivers that make the results of the scenarios unintentionally different.

4.2. Conclusion

In summary, drivers complied in almost all of the cases with the message, but the least intense response was measured during the low information quantity message. The intermediate information quantity message was preferred by most participants. The high information quantity message provoked the highest mental workload and the longest time spent watching the display. In this case, drivers were not overloaded, but a significant increase in mental workload was observed. Overload can be expected to occur when traffic situations are more complex.

The eventual goal of this research was to find the appropriate amount of information presented within in-vehicle messages, to make drivers comply correctly with the messages, but without reducing road safety. Based on this goal and the presented results, the message with intermediate information quantity resulted in the best option during this study. In general, it is advised to present at least a basic level of information quantity. The advice for behaviour and the reason for presenting the message are two elements that were added to the intermediate information quantity message. It is recommended that at least these two elements be presented in an in-vehicle message, causing safe and efficient compliance behaviour. This contributes to a better comprehension of the message. In addition, designers must be careful with extending this basic information quantity level. Drivers will significantly watch more time at the display and mental workload increases when more information elements are added. Therefore, it is advised to keep to this basic level of information quantity.

This study mainly focussed on three different versions of two traffic scenarios. However, no large differences existed between the two scenarios. Therefore, further research should focus more on the specific elements or information units within an in-vehicle message, so compliance and workload can be more specifically researched. By this, the information units on an in-vehicle message could be more specified in practice. The eventual target behaviour and the state of the environment of a message are of great importance for the incorporation of the information units. It is recommended to watch this target behaviour and critically evaluate which elements from the environment should form the basis of the message. This recommended basis should consist of the advice for behaviour and the reason for presentation. Deviations from this basis are not advised, because safety risks and mental workload will increase.

References

Aghaei, A. S., Donmez, B., Liu, C. C., He, D., Liu, G., Plantaniotis, K. N., … Sojoudi, Z. (2016). Smart driver monitoring: When signal processing meets human factors: in the driver's seat. *IEEE Signal Processing Magazine, 33*(6), 35-48. doi: 10.1109/MSP.2016.2602379

Bhaskara, A., Skinner, M., & Loft, S. (2020). Agent transparency: A review of current theory and evidence. *IEEE Transaction on Human-Machine Systems, 50*(3), 215-224. doi: 10.1109/THMS.2020.2965529

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology, 155*, 49-62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Diver inattention monitoring system for intelligent vehicles: A review. *IEEE Transaction on Intelligent Transportations Systems, 12*(2). doi: 10.1109/TITS.2010.2092770

Felici, E. (2019). Van data top tot digitalisering in mobiliteit [PowerPoint slides]. Retrieved from https://www.verkeerinbeeld.nl/static/files/tinymce/uploads/Presentatie%20Edoardo%20Felici.pdf

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics, 73*, 90-99. https://doi.org/10.1016/j.apergo.2018.06.006

Ghanbary Sartang, A., Ashnagar, M., Habibi, E., & Sadeghi, S. (2017). Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses. *Journal of Occupational Health and Epidemiology, 5*(4), 211-217. doi: 10.18869/acadpub.johe.5.4.211

Guo, Z., Wei., Z, & Wang, H. (2016). The expressway traffic sign information volume threshold and AGS position based on driving behaviour. *Transportation Research Procedia, 14,* 3801-3810. doi: 10.1016/j.trpro.2016.05.465

Hanowski, R. J., & Kantowitz, B. H. (1997). Driver memory retention of in-vehicle information system messages. *Transportation Research Record, 1573*(1), 8-16.

Hansen, J. H. L., Busso, C., Zheng, Y., & Sathyanarayana, A. (2017). Driver modelling for detection and assessment of distraction: Examples from the UTDrive Test Bed. *IEEE Signal Processing Magazine. 34*(4), 130-142. doi: 10.1109/MSP.2017.2699039

Harms, L. (1991). Variation in drivers' cognitive load. Effects of driving through village areas and rural junctions. *Ergonomics, 34*(2), 151-160.

Harms, I. M., Dijksterhuis, C., Jelijs, B., de Waard, D., & Brookhuis, K. (2018). Don't shoot the messenger: Traffic-irrelevant messages on variable message signs (VMSs) might not interfere with traffic management. *Transportation Research Part F, 65*, 564-575. https://doi.org/10.1016/j.trf.2018.09.011

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139-183.

International Organisation of Standardization (ISO) (2016). Detection-response task (DRT) for assessing attentional effects of cognitive load in driving. *Ergonomics*. Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:17488:ed-1:v1:en

Johnson, J. (2014). *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines*. https://doi.org/10.1016/B978-0-12-407914-4.12001-9

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing, 9*(4), 269-275. doi:10.1007/s12008-014-0227-2

Kroon, E. C. M., Martens, M. H., Brookhuis, K. A., de Waard, D., Stuiver, A., Westerhuis, F., … Hof, T. (2019). *Human factor guidelines for the design of safe in-car traffic information services.* Rijksuniversiteit Groningen.

Lansdown, T. C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics, 47*(1), 91-104. doi: 10.1080/00140130310001629775

Levy, J., & Pashler, H. (2008). Task prioritisation in multitasking during driving: Opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology, 22*(4), 507-525. doi: 10.1002/acp.1378

Lyu, N., Xie, L., Wu, C., Fu, Q., & Deng, C. (2017). Driver’s cognitive workload and driving performance under traffic sign information exposure in complex environments: A case study of the highways in China. *International Journal of Environmental Research and Public Health, 14*(2), 1-25. doi:10.3390/ijerph14020203

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board, 2138*(1), 6-12. doi: 10.3141/2138-02

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barber, D., Procci, K., & Barnes, M. (2015). *Effects of agent transparency on multi-robot management effectiveness.* (No. ARL-TR-7466). Retrieved from https://apps.dtic.mil/sti/citations/ADA622045

Mijksenaar (2009). *Dynamische Reizigers Informatie Systemen: Adviesrapport van bureau Mijksenaar over het weergeven van dynamische reizigersinformatie in opdracht van GOVI.* (Report No. 5471.01). Retrieved from https://www.crow.nl/downloads/documents/6657

Mulder, L. J. M., Dijksterhuis, C., Stuiver, A., & de Waard, D. (2009). Cardiovascular state changes during performance of a simulated ambulance dispatchers' task: Potential use for adaptive support. *Applied Ergonomics*, *40*(6), 965-977. https://doi.org/10.1016/j.apergo.2009.01.009

National Highway Traffic Safety Administration (2016). *Human factors design guidance for driver-vehicle interfaces.* (No. DTN22-11-00236/001). Retrieved from https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/812360_humanfactorsdesignguidance.pdf

Partnership Talking Traffic. (2019). De voordelen voor automobilisten. Retrieved from https://www.talking-traffic.com/nl/whats-in-it-for/automobilisten

Rakauskas, M. E., Gugerty, L. J., & Ward, N. J. (2004). Effects of naturalistic cell phone conversations on driving performance. *Journal of Safety Research, 35*(4), 453-464. doi:10.1016/j.jsr.2004.06.003

Rash, D., & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science, 46*, 175-208.

Rizzi, E., Jagacinski, R. J., & Bloom, B. J. (2021). Spatio-temporal flexibility of attention inferred from drivers’ steering movements. *Journal of Motor Behavior*, 1-13. doi: 10.1080/00222895.2020.1868968

Ronen, A., & Yair, N. (2013). The adaption period to a driving simulator. *Transportation Research Part F, 18*, 94-106. http://dx.doi.org/10.1016/j.trf.2012.12.007

Rosenberger, R. (2015). Driver distraction from dashboard and wearable interfaces. *IEEE Technology and Society Magazine*, *34*(4), 88-99. doi: 10.1109/MTS.2015.2494399

Sawyer, F. S. (2009). Analysis of variance: The fundamental concepts. *The Journal of Manual & Manipulative Therapy*, *17*(2), 27-38.

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology, 98*, 146-151. https://doi.org/10.1016/j.jclinepi.2017.12.006

Silva, F. P. da (2014). Mental workload, task demand and driving performance: What relation? *Social and Behavioral Sciences, 162*, 310-319. doi: 10.1016/j.sbspro.2014.12.212

Son, J., Lee, Y. & Kim, M. (2011). Impact of traffic environment and cognitive workload on older drivers’ behavior and simulated driving. *International Journal of Precision Engineering and Manufacturing*, *12*(1), 135-141. doi: 10.1007/s12541-011-0017-8

Son, J., & Park, M. (2011, November). Estimating cognitive load complexity using performance and physiological data in a driving simulator. In *Adjunct Proceedings of the Automotive User Interfaces and Interactive Vehicular Applications Conference.*

Son, J., Reimer, B., Mehler, B., Pohlmeyer, A. E., Godfrey, K. M., Orszulak, J., … Coughlin, J. F. (2010). Age and cross-cultural comparison of drivers’ cognitive workload and performance in simulated urban driving. *International Journal of Automotive Technology, 11*(4), 533-539. doi: 10.1007/s12239−010−0065−6

Stichting Digitale Infrastructuur Nederland. (2020). *Digitale infrastructuur gevisualiseerd*. Retrieved from https://www.dinl.nl/digitale-infrastructuur-gevisualiseerd/

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). *Measuring cognitive distraction in the automobile*. AAA Foundation for Traffic Safety.

Teh, E., Jamson, S., Carsten, O., & Jamson, H. (2014). Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance. *Transportation Research Part F, 22,* 207-217. https://doi.org/10.1016/j.trf.2013.12.005

Uus, O., Seitlinger, P. C., & Ley, T. T. (2020). Cognitive capacity in self-directed learning: Evidence of middle school students' executive attention to resist distraction. *Acta Psychologica, 209*. 1-11. https://doi.org/10.1016/j.actpsy.2020.103089

Varjo. (2020a). *Human-eye resolution with Bionic Display*. Retrieved from https://varjo.com/products/vr-2/

Varjo. (2020b). How the 20/20 Eye Tracker Works. Retrieved from https://varjo.com/blog/industrial-strength-eye-tracking-in-varjo/

Vashitz, G., Shinar, D., & Blum, Y. (2008). In-vehicle information systems to improve traffic safety in road tunnels. *Transportation Research Part F, 11*(1), 61-74. doi: 10.1016/j.trf.2007.07.001

Verwey, W. B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics, 43*(2), 187-209. https://doi.org/10.1080/001401300184558

Verwey, W. B., & Veltman, H. A. (1996). Detecting short period of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied, 2*(3), 270-285. doi: 10.1037//1076-898X.2.3.270

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 243-273). Hoboken, NJ: John Wiley & Sons.

Waard, D. de (1996). *The Measurement of Drivers' Mental Workload*. Haren, The Netherlands: University of Groningen.

Waard, D. de, Jessurun, M., Steyvers, F. J., Raggatt, P. T., Brookhuis, K. A., (2007). Effect of road layout and road environment on driving performance, drivers' physiology, and road appreciation. *Ergonomics 38*(7), 1395–1407. https://doi.org/10.1080/ 00140139508925197.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science., 3*(2), 159-177. doi: 10.1080/14639220210123806

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455. doi: 10.1518/001872008X288394

Wickens, C. D., Lee, J., Liu, Y., & Gordon-Becker, S. E. (2004). *An introduction to human factors engineering*. Upper Saddle River, NJ: Pearson Prentice Hall.

Widyanti, A., Johnson, A., & de Waard, D. (2013). Adaptation of the Rating Scale Mental Effort (RSME) for use in Indonesia. *International Journal of Industrial Ergonomics*, *43*(1), 70-76. doi: https://doi.org/10.1016/j.ergon.2012.11.003

Wiese, E. E., & Lee, J. D. (2004). Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performance. *Ergonomics*, *47*(9). https://doi.org/10.1080/00140130410001686294

Williams, D. J., & Noyes, J. M. (2007). How does our perception of risk influence decision-making? Implications for the design of risk information. *Theoretical Issues in Ergonomics Science*, *8*(1), 1-35. doi: 10.1080/14639220500484419

Zijlstra, F. R. H. (1993). Efficiency in work behaviour: A design approach for modern tools. Thesis (PhD), University of Delft.

Appendix A

In-vehicle information messages used in the driving simulator

Traffic jam	Alternative destination

Low

Intermediate

High

Appendix B

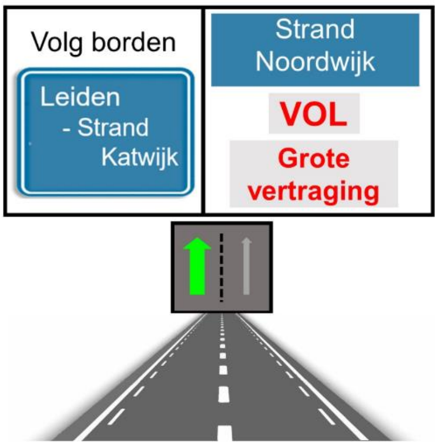
In-vehicle information messages used in the Qualtrics questionnaire

Traffic jam	Alternative destination

Low

Intermediate

High



Road Construction	Alternative route

Low

Intermediate

High

Appendix C

Informed Consent

Titel onderzoek: Compliance behaviour and cognitive workload during simulated driving with an in-vehicle information system.

Onderzoeker: Bente Rootmensen

Ik heb de informatie over het onderzoek gelezen en begrepen. Ik heb de mogelijkheid gekregen om vragen te stellen over het onderzoek en mijn vragen zijn naar mijn tevredenheid beantwoord.

Ja **Nee**

O O

Ik geef vrijwillig toestemming om deel te nemen aan dit onderzoek. Ik begrijp dat ik niet verplicht ben om vragen te beantwoorden en dat ik me op elk moment, zonder reden, kan terugtrekken.

Ja **Nee**

O O

Ik weet dat de data en de resultaten van dit onderzoek anoniem en in vertrouwen naar externe partijen gepubliceerd zullen worden. De eye-tracking opnames zullen worden verwijderd wanneer deze data geanalyseerd is, wat voltooid zal zijn binnen 2 maanden na het experiment.

Ja **Nee**

O O

Het onderzoeksproject is beoordeeld en goedgekeurd door de BMS Ethische Commissie. Ik verklaar, op een voor mij voor de hand liggende manier, dat ik ben geïnformeerd over de methode, het doel en [indien aanwezig] de risico’s en lading van het onderzoek.

Ja **Nee**

O O

Datum: …………………………….…… Handtekening participant: ……………….....………………...

In te vullen door de uitvoerende onderzoeker

Ik heb mondelinge en geschreven uitleg van het onderzoek gegeven. Ik zal overige vragen over het onderzoek beantwoorden. De deelnemer zal geen nadelige gevolgen ondervinden in geval van vroegtijdige beëindiging van deelname aan dit onderzoek.

Naam onderzoeker: Bente Rootmensen

Datum: …………………………….…… Handtekening onderzoeker: ……………….....………………

Appendix D

Participant instruction

Deelnemer instructie

Deel 1:

Welkom!

Je zal deelnemen in een experiment dat bedoeld is om beter te begrijpen wat het rijgedrag is van autobestuurder tijdens een rit met digitale verkeersinformatie. Ook wordt gemeten hoeveel werkbelasting autobestuurders hierbij nodig hebben. De onderzoeksvraag is hoeveel het gedrag en de werkbelasting verschilt per soort melding.

Het experiment bestaat uit twee delen. Het eerste deel is een online vragenlijst over het ontwerp van de verkeersberichten. Het tweede deel vindt plaats in de rijsimulator. Het ontwerp van de verkeersberichten die je in de online vragenlijst te zien krijgt zullen overeenkomen met de berichten die je later te zien krijgt in de rijsimulator. Echter, de inhoud van deze berichten zal wat verschillen. Na deze vragenlijst zal je meer instructies krijgen over het experiment in de rijsimulator. In totaal zal je ongeveer **60 minuten** bezig zijn met het onderzoek.

Stel gerust vragen over de aspecten die niet duidelijk zijn van het onderzoek.

De BMS studenten die zich hebben ingeschreven via het SONA-systeem voor dit onderzoek krijgen 2 SONA punten.

Veel succes met het experiment en alvast bedankt voor het deelnemen!

Bente Rootmensen
Afstudeerder voor de Master Human Factor and Engineering Psychology & Keypoint Consultancy Enschede

Deel 2:

Dankjewel voor dit eerste deel van het experiment.

Het tweede deel vindt plaats in een rijsimulator. Hier zul je dezelfde soort berichten te zijn krijgen als je in deze vragenlijst hebt gezien. Wees ervan bewust dat het rijden in een rijsimulator geen volledig realistische weergave is van een echte rit in een auto. Je krijgt daarom eerst de tijd om even te wennen, daarna zal er **1 oefensessie** zijn om te oefenen hoe het experiment zal verlopen. Deze oefensessie zal dezelfde opzet hebben als de andere sessies in het experiment. Het experiment zelf bestaat uit meerdere van dit soort sessies.

De simulatie zal worden gedaan met een Virtual Reality bril die ook je oogbewegingen bijhoudt. Deze bril zorgt voor een realistische weergave van de gesimuleerde omgeving. Het kan zijn dat de bril zorgt voor een ongemakkelijk of oncomfortabel gevoel. Je krijgt daarom zo veel pauze als u wilt tussen de sessies door. Wanneer je wilt stoppen met het onderzoek, mag je dit ten aller tijde aangeven. Je kunt altijd even wat eten of drinken tussen de sessies door.

Procedure van een sessie in de rijsimulator

De rijsimulator werkt hetzelfde als een automaat, dus je hoeft niet handmatig te schakelen. Je begint voor een rechte weg, zodat je tijd hebt om te accelereren. Zorg dat je op de rechterbaan blijft rijden, tenzij je anders wordt verteld.

Tijdens de rit zal er een informatiescherm worden geactiveerd, die dezelfde soort berichten presenteert als je hier gezien hebt. Zorg dat je doet alsof dit een echte situatie is die op de weg plaatsvindt.

De onderzoeker zal aangeven wanneer het experiment is afgelopen. Daarna stopt de simulator en word je gevraagd om een aantal vragen in te vullen.

Het is belangrijk om te weten dat het besturen van de auto altijd de hoogste prioriteit heeft, net zoals dit het geval is in een echte situatie op de weg.

Hygiënemaatregelen i.v.m. COVID-19.

De rijsimulator wordt volledig gedesinfecteerd na elk experiment. De onderzoeker houdt 1,5 meter afstand en zorgt dat alles volgens de corona-richtlijnen zal verlopen.

Stel gerust vragen over de aspecten die niet duidelijk zijn van het onderzoek. Als alles duidelijk is kun je deze vragenlijst afsluiten en zal de onderzoeker je helpen bij de opstart van de rijsimulator.

Veel succes met het experiment!

Bente Rootmensen
Afstudeerder voor de Master Human Factor and Engineering Psychology & Keypoint Consultancy
Enschede

Appendix E

Demographic questions participant

Demografische gegevens rijsimulator onderzoek

Hartelijk dank voor uw medewerking aan dit onderzoek. Ten eerste zou ik u willen vragen om wat demografische gegevens in te vullen.

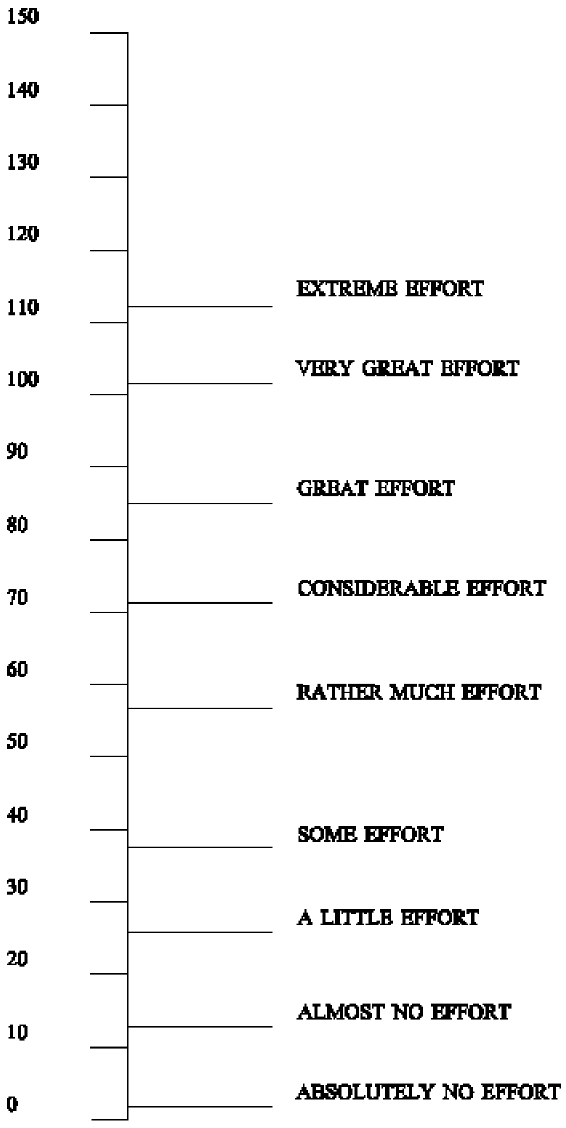
1. Wat is uw leeftijd?
2. Hoe lang heeft u uw rijbewijs?
    a. 1-5 jaar
    b. 6-10 jaar
    c. 11-20 jaar
    d. 21-30 jaar
    e. Langer dan 30 jaar
3. Hoeveel uur rijdt u per week?
4. Gebruik je een bril of lenzen om je oogkwaliteit te corrigeren? Ja/Nee
5. Opmerkingen of eventuele bijzonderheden die invloed zouden kunnen hebben op de metingen.

Appendix F

Rating Scale Mental Effort (Zijlstra, 1993)

Participant: ….   Trial: ….

Hoeveel moeite kostte de interactie met het digitale verkeersbericht?



Hoeveel moeite kostte de rijtaak in het algemeen?

Appendix G

Simulator Sickness Questionnaire (Kennedy, et al., 1993)

SIMULATOR SICKNESS QUESTIONNAIRE

Kennedy, Lane, Berbaum, & Lilienthal (1993)***

Instructions : Circle how much each symptom below is affecting you right now.

1. General discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Headache	None	Slight	Moderate	Severe
4. Eye strain	None	Slight	Moderate	Severe
5. Difficulty focusing	None	Slight	Moderate	Severe
6. Salivation increasing	None	Slight	Moderate	Severe
7. Sweating	None	Slight	Moderate	Severe
8. Nausea	None	Slight	Moderate	Severe
9. Difficulty concentrating	None	Slight	Moderate	Severe
10. « Fullness of the Head »	None	Slight	Moderate	Severe
11. Blurred vision	None	Slight	Moderate	Severe
12. Dizziness with eyes open	None	Slight	Moderate	Severe
13. Dizziness with eyes closed	None	Slight	Moderate	Severe
14. *Vertigo	None	Slight	Moderate	Severe
15. **Stomach awareness	None	Slight	Moderate	Severe
16. Burping	None	Slight	Moderate	Severe

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

Last version : March 2013

***Original version : Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, *3*(3), 203-220.

Appendix H

Documentation pilot test design traffic messages

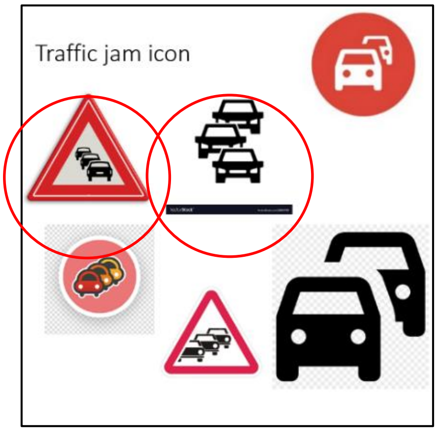
Iterative design process

The current information messages used in the driving situation of the experiment are designed based on the feedback of three different road users. The design process was informal for most of the time, but the essential basics of the user-centred design were applied in these tests. Especially, the iterative process of design was considered to design a message based on the needs and capabilities of the road user. This means that: a design was made, after that evaluated by a pilot test and then designed further. This constant design feedback loop improved the message based on users' needs. One pilot test, which is described here, was an evaluation for the choice of icons and pictograms used in the message. Two other pilot tests were executed for testing the understandability and recognizability of the concept messages.

Pilot test 1:

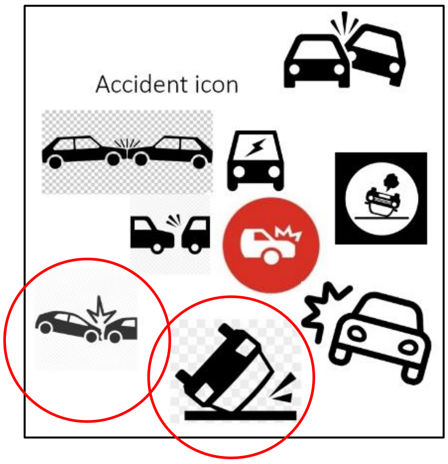
The pilot test aimed for the choice and evaluation of icons and pictograms.

A participant with no driver's licence (age 22), female.



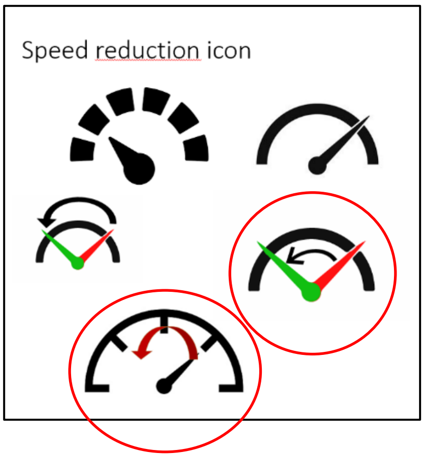
Traffic jam icon: The icons were collected from google. Some were already existing icons in traffic (physical traffics signs, Google Maps icon), some not.

Participant chose the first two icons as most representative as a traffic jam icon. One because this is a known traffic jam sign in the Netherlands and the other because the cars are in a queue (see red circles).



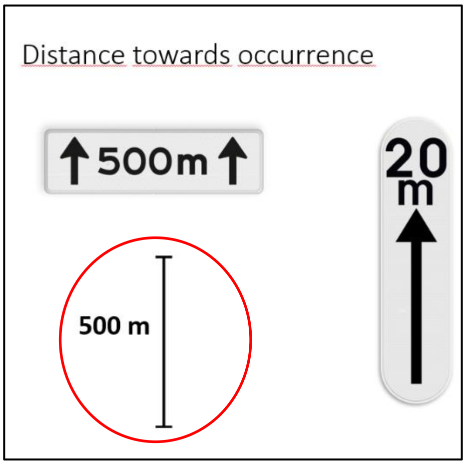
Accident icon: Also, these icons were collected from google. Some were already exiting icons in traffic (Google Maps), some not.

Participant chose circled icons as the most representative accident icon. Because it is clear an accident has taken place. The other icons can have an ambiguous meaning which can confuse the driver.



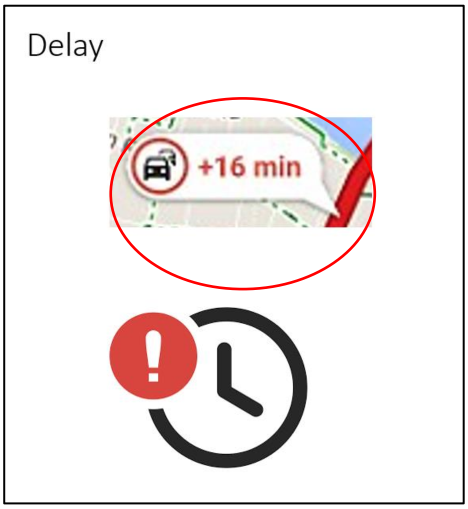
Speed reduction icon: These icons were collected from google. Some were adapted by Adobe Illustrator, to create more examples.

The participant chose the icons with the arrow as the most representative for speed reduction. Both, because they are presenting a need for change. Furthermore, some others are ambiguous for the driver.



Distance towards occurrence: These icons were collected from google or self-made from ideas.

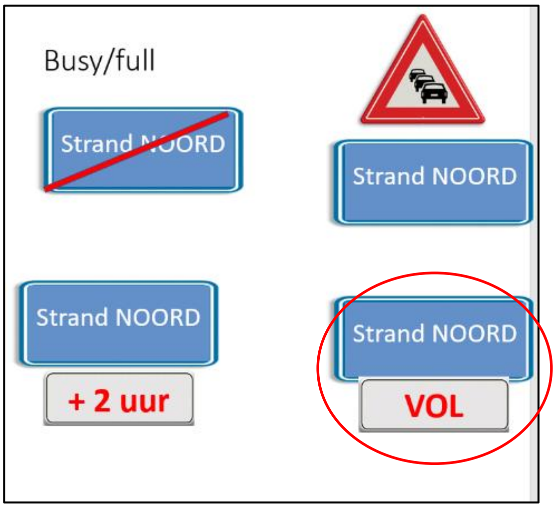
The participant chose the top one as most known, but afterwards, this seemed not the right sign for a distance towards an occurrence, but the length of an occurrence. Therefore, the lowest one was chosen as the clearest.



Delay: Not many icons were found for a representation of the information unit 'delay', but the ideas of these two icons are clear.

The participant chose the top one because she did not understand what the meaning was of the lower one.

Busy/full icon: These icons are self-made because the situation was very specific.



The participant commented that all icons had another meaning. So, it depended on which the message of this icon was to choose the right one. When the message was told to the participant, she chose the bottom right one.

Later, the message ‘great delay’ was added in the high message, which means the same as ‘VOL’, but is an addition to the message.

Pilot test 2:

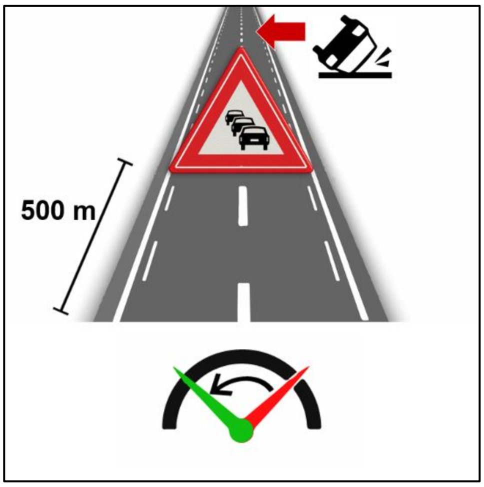
Pilot test for evaluation concept messages and their meaning:

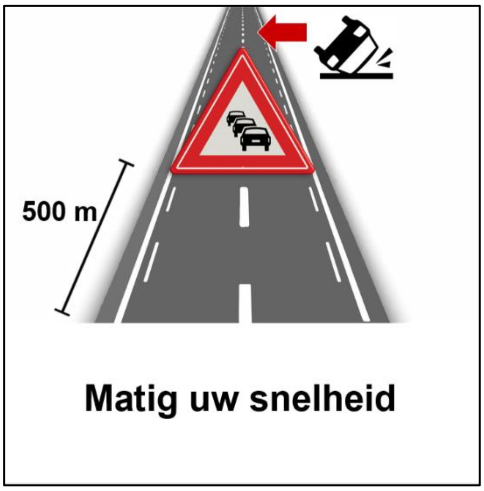
Participant: inexperienced driver (age 19), male

Traffic jam messages: Different versions of a message were generated, ranging from all icons to all text.

The distance was interpreted as length by the participant at first sight, but at second sight this was interpreted as distance.

Participant needed a second to understand the speed reduction sign. The speed reduction icon was the first thing the participant saw. This message was the second clearest according to the participant.

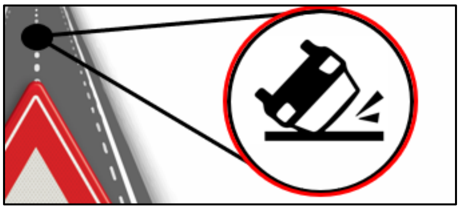




Participant chose this message as the clearest because there is some text to focus on.

Tip: it is not immediately clear that there has been occurred an accident. So, let the design be as a magnifier (as done in cartoon drawing).

Change: see below



Tip from the participant: switch the two written sentences, so the instruction on what to do is read first.

Same tip as the previous message. Switch the sentences.

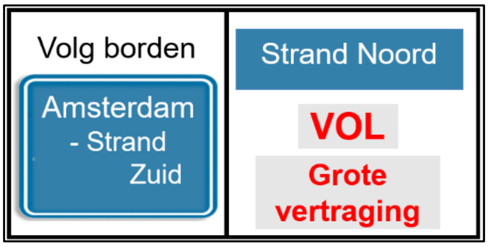
Not clear according to the participant, too much written text.

Too much text.

Alternative destination messages: One message was shown to the participant and the experimenter asked what the meaning was of this message.

The participant read the message from left to right, so: Volg borden → strand Noord → Amsterdam – strand zuid → vol → grote vertraging.

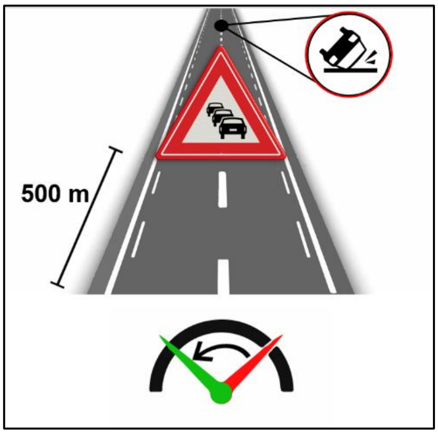
It was immediately clear that a vertical line should be placed between the two messages, like below.



Pilot test 3

This pilot aimed to test the last changes in the design. Several different versions of the design were shown to the participant.
Participant: experienced driver (age 28), male.

Traffic jam messages: Just as in the last pilot test, different versions of the messages were shown to the participant.

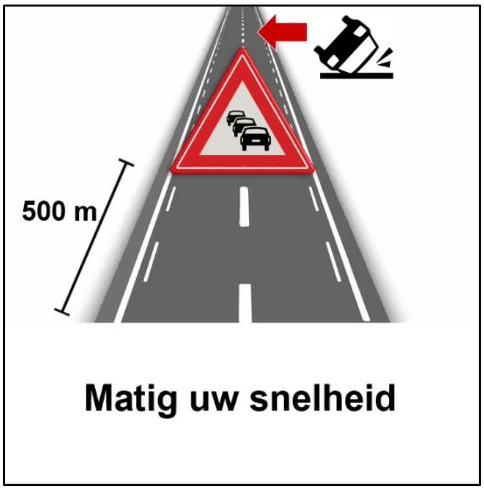


The participant described the meaning of the message: "Reduce speed because of an accident. This is in 500 meters…… possible causing a traffic jam."
Furthermore, there was some confusion about the 500 m icon: In 500 m or for the following 500 m? Also, in a real situation, this would be count down in correspondence with your speed.

Design changes: Traffic jam is made bigger. If possible, the distance towards the traffic jam will count down.

Participant: Better in comparison with the previous message, especially when seeing the message for the first time. Advice: possible to present the icon and 'matig uw snelheid' at the same time.
This is the clearest message according to the participant.

Design: It is chosen to only leave the text in. 'Matig uw snelheid' is clear enough for inexperienced and experienced drivers. The extra speed reduction icon will cause another information point to the process.



Participant: this accident sign is less clear than the previous one. The arrow is not connected to the accident icon, which makes it ambiguous.

Participant: This would mean: 'reduce your speed in 500 m'. Only after that, you go look at the meaning of the icons. This causes confusion.

Participant: This would mean: There is a traffic jam in 500 meters, after this, you look at the reason and advice behaviour. Because the road is prominently in the middle.

Sidenote: ongeluk = ongeval

Participant: Too much text. Show ‘matig uw snelheid’ at the top.

Too much text.

Participant: ‘VOL’ and ‘Grote vertraging’ may be an unnecessary repetition.

Participant doubts if the arrows are recognizable enough. He understands it and it is clear, but not in line with the already existing arrows. However, he concludes that this is the clearest way to present it.

No design changes are made here because 'Grote vertraging' is the reason that you should go another way.