Human machine interface design for continuous support of mode awareness during automated driving
An online simulation

Tinga, Angelica M.; Cleij, Diane; Jansen, Reinier J.; van der Kint, Sander; van Nes, Nicole

DOI
10.1016/j.trf.2022.03.020

Publication date
2022

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Transportation Research Part F: Psychology and Behaviour 87 (2022) 102–119

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

Human machine interface design for continuous support of mode awareness during automated driving: An online simulation

Angelica M. Tinga [a,*], Diane Cleij [a], Reinier J. Jansen [a], Sander van der Kint [a], Nicole van Nes [a,b]

[a] *Department of Human Factors and Vehicle Automation, Dutch Institute for Road Safety Research (SWOV), Bezuidenhoutseweg 62, 2594 AW The Hague, The Netherlands*

[b] *Department of Human Centered Design, Faculty of Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands*

ARTICLE INFO

Keywords:
HMI
Automated driving
Mode awareness
Human factors in vehicle automation
Think-aloud
NDRA

ABSTRACT

In the transition towards higher levels of vehicle automation, one of the key concerns with regards to human factors is to avoid mode confusion, when drivers misinterpret the driving mode and therewith misjudge their own tasks and responsibility. To enhance mode awareness, a clear human centered Human Machine Interface (HMI) is essential. The HMI should support the driver tasks of both supervising the driving environment when needed and self-regulating their non-driving related activities (NDRAs). Such support may be provided by either presenting continuous information on automation reliability, from which the driver needs to infer what task is required, or by presenting continuous information on the currently required driving task and allowed NDRA directly. Additionally, it can be valuable to provide continuous information to support anticipation of upcoming changes in the automation mode and its associated reliability or required and allowed driver task(s). Information that could support anticipation includes the available time until a change in mode (i.e. time budget), information on the upcoming mode, and reasons for changing to the upcoming mode. The current work investigates the effects of communicating this potentially valuable information through HMI design. Participants received information from an HMI during simulated drives in a simulated car presented online (using Microsoft Teams) with an experimenter virtually accompanying and guiding each session. The HMI either communicated on automation reliability or on the driver task, and either included information supporting anticipation or did not include such information. Participants were thinking aloud during the simulated drives and reported on their experience and preferences afterwards. Anticipatory information supported understanding about upcoming changes without causing information overload or overreliance. Moreover, anticipatory information and information on automation reliability, and especially a combination of the two, best supported understandability and usability. Recommendations are provided for future work on facilitating supervision and NDRA self-regulation during automated driving through HMI design.

* Corresponding author.
E-mail address: angelique.tinga@swov.nl (A.M. Tinga).

https://doi.org/10.1016/j.trf.2022.03.020
Received 27 August 2021; Received in revised form 24 March 2022; Accepted 31 March 2022
Available online 9 April 2022

1. Introduction

Vehicle automation has the potential to improve traffic safety by handling (parts of) the driving task. In the transition to fully automated vehicles in which the complete driver task is taken care of by the automation (SAE Level 5, Society of Automotive Engineers International J3016, 2021) lower level automated vehicles are being introduced. These lower level automated vehicles still impose important requirements for the human driver, depending on the level of automation. When driving with partial automation (SAE Level 2) the driver is required to constantly monitor the driving situation. With conditional automation (SAE Level 3) the driver does not need to monitor the driving situation continuously, but has to remain fit enough to take over the driving task when requested. With highly automated driving (SAE Level 4) the driver can be completely 'out of the loop' as the automation is able to bring the car to a safe stop if needed within a designated operational design domain (ODD; the specific conditions under which an automated system or feature thereof is designed to function). From a technical perspective, automated driving starting from SAE Level 3 no longer requires constant supervision by the driver. The possibility to then engage in non-driving related activities (NDRAs) is one of people's most valued expected benefits of automated driving (König & Neumayr, 2017). With SAE Level 2 and 3 automated driving, however, the driver needs to be a back-up for the automation when the automation is unable to handle the driving task. This requires the driver to disengage from an NDRA and monitor the driving situation to safely take over (parts of) the driving task when needed. A Human Machine Interface (HMI) has the potential to support the driver to optimally fulfil this requirement. When a take-over is imminent, an HMI should support the take-over process or, as described by van den Beukel et al. (2016), provide support for intervention. Both when a take-over is imminent and *during* automated driving (especially SAE Level 2 and 3), an HMI should stimulate mode awareness, or understanding of the driving mode and the driver's responsibilities, and prevent mode confusion. This is one of the key requirements for supporting both supervision of the driving environment as well as self-regulation of NDRA uptake and disengagement. Most research on the role of HMI design in the interaction between driver and automation focuses on support for intervention (e.g. Melcher et al., 2015; Mirnig et al., 2017; Naujoks et al., 2014; Petermeijer et al., 2017). Less attention, however, has been given to HMI design focusing on support for supervision and NDRA self-regulation. The current research addresses this knowledge gap.

1.1. Background

Self-regulation of NDRA engagement is necessary when the driver is expected to perform a supervisory task and maintain take-over readiness. Schömig and Metz (2013) proposed a model of NDRA self-regulation during manual driving consisting of three levels: 1) Planning; 2) decision; and 3) control. Wandtner et al. (2018) adapted this model for automated driving. On the planning level, proper self-regulation consists of performing NDRAs during parts of the trip only with automated driving. On a decision level, the driver should assess the current driving situation and should only decide to engage in an NDRA if appropriate, considering estimated NDRA duration and predicted system availability. Finally, on a control level, drivers should maintain take-over readiness, such that they can take over when requested by the automation. Several studies have shown that drivers self-regulate their NDRA behavior and adjust it to the driving situation complexity both during manual driving (Schömig, Metz, & Krüger, 2011; Tivesten & Dozza, 2015) as well as during automated driving (Lin et al., 2019; Wandtner et al., 2018).

There is general consensus that HMI design should communicate transparently to the driver to support the driver with supervision and self-regulation. Such communication can be used to avoid mode confusion and automation surprises and facilitate appropriate trust and reliance (Carsten & Martens, 2019; Kyriakidis et al., 2019; Martens & van den Beukel, 2013). Several studies have examined different types of information presented by HMI design to mitigate these human factors issues. We identified two promising approaches in the literature differing in the focus of the communicated information: 1) Communicating automation reliability, and 2) communicating the required or allowed task. These approaches will be discussed next, followed by the potential benefit of including anticipatory information well ahead of an upcoming change in an upcoming transition.

1.1.1. Communicating automation reliability

Driving simulator experiments (Beller et al., 2013; Helldin et al., 2013; Stockert et al., 2015) showed beneficial effects of providing continuous information on automation reliability while driving with partial automation. This information was either communicated on a continuous scale via a bar display (Helldin et al., 2013; Stockert et al., 2015) or dichotomously through an anthropomorphic icon (i.e., a 'confused emoticon' in Beller et al., 2013). These studies indicate that continuously providing information on reliability of the automation reduces workload (Stockert et al., 2015), facilitates calibrated trust and enhances acceptance (Beller et al., 2013; Helldin et al., 2013; Stockert et al., 2015). Moreover, studies investigating information needs demonstrated that drivers appreciate information on automation reliability (Beggiato et al., 2015; Feierle et al., 2020; Hecht et al., 2019).

The effect of continuously communicating automation reliability during NDRA engagement was examined in a study by Large et al. (2017) in which participants drove with conditional automation in a simulator. During the drive, information on automation reliability was provided through a 'health-bar'; a colored bar icon. In general, participants attended the health-bar $<3\%$ of the time. The health-bar was potentially not attended often because the relatively small health bar icon presented behind the steering wheel did not attract a lot of attention. Yet, when participants did attend the health-bar, positive effects were observed. One participant adjusted his/her posture when reliability decreased, in such a way that a better take-over readiness was achieved. Another participant who regularly attended the health bar reported enhanced situation awareness.

In Yang et al. (2017) participants were provided with continuous information on the automation reliability though an LED strip during simulated driving with conditional automation and while engaging in NDRAs. This LED strip was positioned at the bottom of the windshield ranging from the outer left to the outer right, ensuring the strip could also be perceived peripherally. The automation reliability was communicated in three levels indicated by both the color and light pulse frequency. Findings of the study showed that especially the worst performers in terms of take-over quality (including minimal time to collision and standard deviation of the lateral position) were supported by the HMI. However, the HMI decreased the take-over quality of the best performers. Yang et al. (2017) hypothesized that the latter result was caused by participants confusing the reliability information for warning signals. Thus, it could be valuable to provide the driver with information on automation reliability when it is ensured that this information is perceived when the driver is engaged in an NDRA and that the information is understood correctly.

1.1.2. Communicating the required or allowed task

As an alternative HMI approach, information can be provided on tasks that are required of or allowed for the driver, both of which are related to the momentary reliability level of the automation. Several studies have investigated HMI designs that inform the driver on the need to monitor the automation, in addition to providing take-over requests, while driving with conditional automation. Simulator studies described in Yang et al. (2018) and Lu et al. (2019) compared take-over performance when providing the driver only with take-over requests to providing the driver with monitoring requests that occasionally resulted in a take-over request and found that the latter resulted in better take-over performance. Increased situation awareness and hazard detection were found when continuously informing drivers on their monitoring task, rather than not informing (Yang et al., 2018), or when warning the drivers only once (van den Beukel et al., 2016). Similar to the study by Large et al. (2017), the study described by Yang et al. (2018) found that in response to low automation reliability drivers adjusted their posture to a more upright position in which they could better monitor the road. In addition, van den Beukel et al. (2016) demonstrated that information on the driver task provided through colored LEDs at the bottom of the windshield better supported the supervisory task than simple auditory tones or display icons. Lu et al. (2019), however, also demonstrated that drivers developed overreliance on the request to intervene, even if they were monitoring the road in response to the monitoring request. This is an issue that needs to be prevented, as overreliance can degrade take-over performance (Feldhütter et al., 2019).

1.1.3. Communicating anticipatory information

The above studies investigated the beneficial effect of communicating the momentary status of automation reliability or the corresponding task requirements. Aside from communicating momentary states, further support might be provided by communicating state changes in advance. Such communication likely supports anticipation and therefore reduces automation surprises (Carsten & Martens, 2019; Martens & van den Beukel, 2013). Moreover, it can be valuable to provide information on the reasons for changes. Such information, at least when it comes to reasons for changes in automation reliability, can improve system understanding, which can in turn support the driver in anticipating when state changes are likely to occur and improve take-over performance (Körber et al., 2018; Naujoks et al., 2017; Richardson et al., 2018). Moreover, both Thill et al. (2018) and Ruijten et al. (2018) concluded that providing reasons for recommendations provided by a support system improves adherence to the recommendations compared to only presenting the recommendation.

Anticipation of changes can also be supported by communicating the time remaining in the current reliability level or required/allowed driver task. We refer to this time window as the 'time budget'. When communicating the time budget it is additionally expected that the HMI should communicate on the expected next state in order to support the driver's understanding about changes that are coming up. The potential value of information on the time budget is first of all raised by drivers themselves, who indicated in several user studies (Beggiato et al., 2015; Hecht, Kratzert, et al., 2020) that they would appreciate information on remaining time in the current automation level. Communicating the time budget and the upcoming state also has the potential to support supervision and NDRA self-regulation, especially on a planning and a decision level. For instance, Wandtner (2018) examined the effect of communicating on automation availability and upcoming take-over situations in the next 5 km. Participants were informed on the remaining distance in which the automation would be engaged, which also provides an indication of the available time budget. The remaining distance was communicated by a visual representation including a horizontal bar in which part of the bar was shaded in a color. This shaded area depleted with decreasing distance until a change in automation mode. Findings demonstrated the HMI enhanced NDRA (dis)engagement and monitoring behavior, especially when it additionally included an explicit pre-alert. A study by Hecht, Kratzert, et al. (2020) examined the effect of communicating about the remaining time to an imminent take-over from automated to manual driving only seconds before or minutes in advance. The time budget to the take-over was communicated through a text including a countdown until the take-over. Results indicated that the HMI supported planning of NDRA engagement at the beginning of the automated drive and in some instances between activation and deactivation of the automation, but the HMI did not support preparation for the take-over. In addition to these findings, it has even been acknowledged that HMI design could aid in planning by also providing information over the complete trip, rather than just providing information on the current or near future automation status (Hecht, Sievers, et al., 2020).

1.2. The current study

Considering the above, it appears that supervision and NDRA self-regulation can be supported by providing 1) continuous information on the current mode in terms of automation reliability or the required/allowed driver task and 2) continuous information supporting anticipation of upcoming changes in the current mode. This anticipatory information includes information on the available time budget, information on the upcoming mode, and reasons for changing to the upcoming mode. Combining the two in a single HMI design potentially results in an additive effect in supporting supervision and NDRA self-regulation. The current work explores the effects of communicating this potentially valuable information through HMI design.

Four different HMI designs are explored by presenting participants with simulated drives online (using Microsoft Teams) with an experimenter virtually accompanying and guiding each session. In the simulated drives participants received continuous information from one of the four HMI designs (following a within-subjects 2 × 82 design). Two aspects were manipulated in the HMI designs. The first aspect concerned the focus of the communicated information: 1) Information on automation reliability, from which the driver can potentially infer what task is required (e.g. monitoring) or whether and what type of NDRA is allowed; or 2) direct information on the required or allowed task. The second aspect concerned the inclusion or exclusion of information that supports anticipation (i.e., time budget, upcoming mode, reasons for changing to the upcoming mode). Manipulating these two aspects allows for exploring the effect of information focus and anticipatory information and whether the two aspects interact.

To be able to support supervision and NDRA self-regulation, it is first of all important to establish whether the information communicated by the HMI is understood, usable and accepted. The current study is therefore dedicated to evaluating effects of the manipulations to the HMI design on understandability, usability, and user preferences. Understandability refers mainly to whether the presented information by the HMI design is correctly understood. Usability includes the perceived usefulness of the provided information for both NDRA self-regulation and supervision and includes anticipation of future changes and inferences based on the provided information (e.g., inferring driver task from automation reliability), as well as the participant's perceived information load and potential overreliance. Evaluation of user preferences focusses mainly on reasons for preferring one HMI concept over another and how preferences relate to individual differences, thereby gaining insight into how acceptance in HMI design could be supported. Measures of understandability, usability and user preferences were collected by using the concurrent think-aloud procedure (Eccles & Arsal, 2017), which entailed participants thinking aloud during the simulated drives, and through collecting questionnaire data after presentation of each HMI design and at the end of the experiment. The exploration of the current study aims to support the next steps in HMI design to facilitate supervision and NDRA self-regulation.

2. Methods

2.1. Participants

Sixteen participants (3 female, 13 male, mean age = 41.12 years, *SD* = 10.51, min = 24, max = 57) were included in the current study. The number of participants was based on previous research that used inferential statistics on think-aloud data of about 10 participants (range 6–17) in total in a within-design or per group in a between-design (Key et al., 2016; Kircher & Ahlstrom, 2018; Rose et al., 2019; Walker et al., 2011). Requirements for participation were a valid driving license and experience with assisted driving technology (i.e., a lane keeping system and/or adaptive cruise control). To be able to recruit such participants, calls for participants were placed on online discussion platforms with a focus on automated vehicles (e.g., fora about Tesla). Participants owned a driving license for an average of 21.81 years (*SD* = 10.19, min = 6, max = 39). The median number of km driven per year was 20,000–50,000 km (min = <5000 km, max = 50,000–100,000 km). All but one participant owned a car with SAE L2 automation options, with an average of 2.53 years (*SD* = 1.72, min = 1, max = 5). The participant who did not own such a vehicle, did so in the recent past. Two additional participants participated, but these participants were excluded from analyses due to insufficient data quality. The study was approved by the Research Ethics Committee of the Dutch Institute for Road Safety Research (SWOV).

2.2. Apparatus

Simulated driving scenarios were made in Unity 2019.3.13f1 with the virtual environment being built on the openly available assets Windridge City (Nature Manufacture) and AirSim (Shah et al., 2018). The viewer's perspective in the car (see also Fig. 1) was on eye height (vertical position) of someone sitting in the middle of the vehicle in between the driver's seat and the front passenger's seat (lateral position), moved slightly towards the back of the car from the driver's seat (longitudinal position), looking outside through the front windshield (direction). This specific viewer's perspective ensured a good view on the HMI designs. HMI designs were superimposed on the drives and were developed in Max (2020). The experiment took place online (with an experimenter virtually accompanying and guiding each session) using Microsoft Teams version 1.3.00.21759. [1] The drives with the superimposed HMI designs were shown and questionnaires were presented to the participants using LimeSurvey Professional Version 3.23.1. The participant shared his/her screen through the 'share screen' functionality in Microsoft Teams while going through the experiment in LimeSurvey. Verbal utterances and the video image showing the participant and his/her screen were recorded using the recording function in Microsoft Teams for subsequent transcription.

2.3. Driving scenarios

Four driving scenarios lasting five minutes each were developed, following the same route between two points, points A and B, in the virtual driving environment. Of the four scenarios, two followed the route from point A to B and two followed the route from point B to A. These four driving scenarios were presented in a counterbalanced order in which the route from A to B was always followed by the route from B to A and vice versa. During each driving scenario seven changes in modes of automated driving occurred due to an

[1] . At the time of conducting the current study it was not possible to perform the experiment in the laboratory due to the COVID-19 pandemic. However, all experimental sessions were virtually accompanied and guided by an experimenter to ensure the validity of the current study's results.

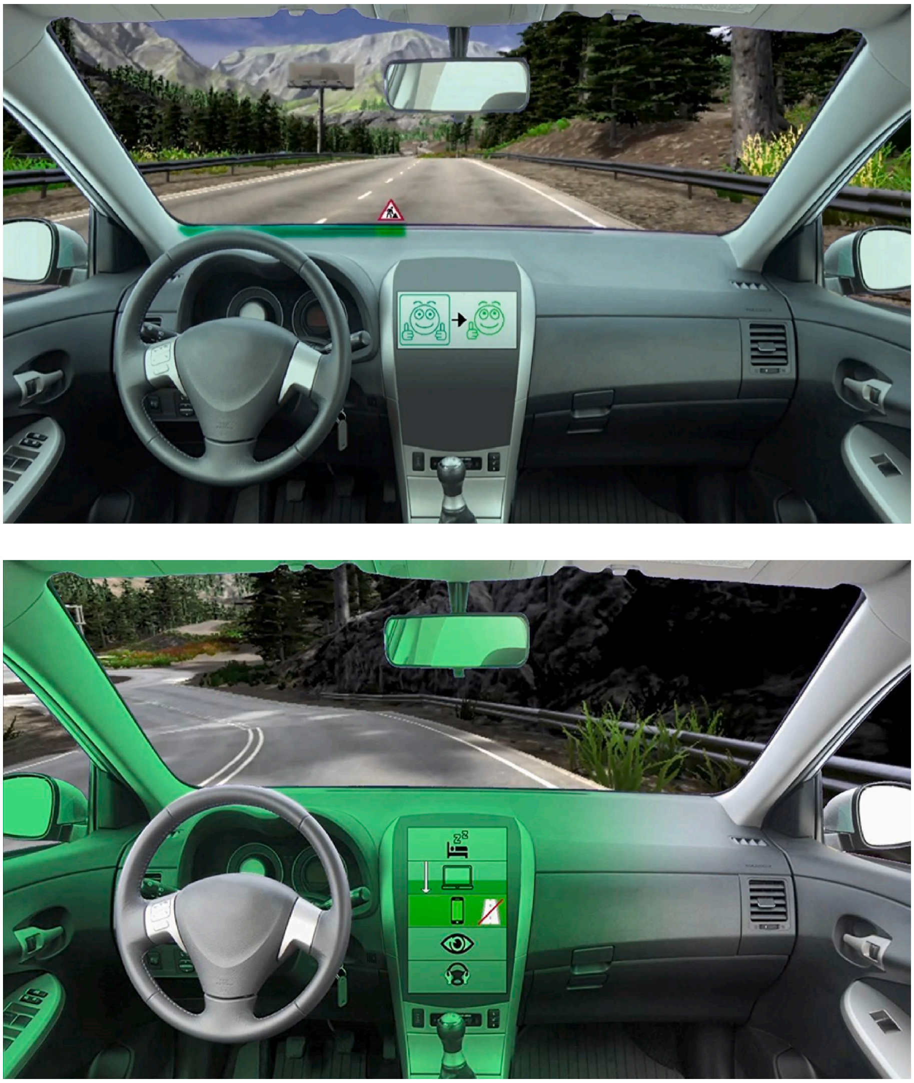

Fig. 1. Depiction of the Anticip conditions. Above: AR_Anticip, with a LED strip below the window and emoticons communicating on automation reliability presented in the center console. Below: DT_Anticip, with icons in the center console communicating the desired driving task associated with a continuous transition from a high mode of automated driving (top panel) to a low mode of automated driving (bottom panel). Note that the icon on the lowest bottom of the panel was never selected in the scenarios in the current study.

event along the route. These events included transitions between different road types (urban, rural, highway), roadworks, no road markings, and fog. Except for the transitions between different types of roads, each specific event occurred across the different scenarios at two different places along the route. For the two drives from point A to B and the two drives from point B to A, the events occurred at a different location along the route. The automation was able to execute the driving task at all times and therefore no take-over (request) took place. The routes and the events occurring along the routes are depicted in Figure S1 in the supplementary materials. Table S1 provides an overview of the modes of automated driving that occurred during the driving scenarios and their relationship to the events. For the purpose of the current study four different modes were included reflecting both automation reliability and required and allowed driver tasks. In Mode 1 the driver should monitor the automation and the on-road conditions as a take-over request can be prompted instantly. Therefore, the driver should not perform any NDRA. In Mode 2 the driver should supervise the automation and should be aware of short take-over times. Therefore, during this mode the driver would be able to perform an NDRA which can easily be disengaged from. An example of such an NDRA is checking messages. In Mode 3 the driver should supervise the automation sporadically and there is a sufficient time for take-overs. In this mode the driver would therefore be able to perform an NDRA that would take somewhat longer to disengage from, such as watching a movie or working on a laptop. In Mode 4, there is no need for the driver to supervise the automation as any take-over request would be optional as the automation can always bring the vehicle to a safe state. Therefore, during this last mode any NDRA can be performed and even sleeping would be an option (when this would be legally permitted). The four modes were presented an equal number of times across scenarios. Note that Mode 1 and Mode 2 occurred in an urban environment, which is currently a challenge for vehicle automation, but not necessarily an unrealistic scenario. For example, Andreone et al. (2021) and Tinga et al. (2021) describe pilots with conditionally automated cars and highly automated shuttle busses in urban environments respectively. Information on modes and events was fed into the HMI concept in order to use this information for communication to the participant.

2.4. Experimental conditions

Four experimental conditions of HMI designs were applied in the current experiment: 1) AR_Base; 2) AR_Anticip; 3) DT_Base; and 4) DT_Anticip. In the AR conditions (1 and 2) the type of communicated information focused on the *automation reliability* while in the DT conditions (3 and 4) this information focused on the *driver task*. In the Anticip conditions (2 and 4) information supporting *anticipation* was provided, including information on the available time budget, the upcoming mode, and reasons for changing to the upcoming mode. The Base conditions (1 and 3) served as a *baseline* in which no such information supporting anticipation was provided, only the reasons for a change in mode were communicated at the time that a change took place. More details on the information provided in each condition and the way in which the information was communicated will be discussed next. For an overview, see Table S2.

The Anticip conditions are depicted in Fig. 1. The first Anticip condition, AR_Anticip, included an HMI that communicated on the reliability of the automation directly through four emoticons (based on Beller et al., 2013, in which a single emoticon was effective in communicating certainty of the automation). A pilot study asking participants ($N = 10$) to rate and rank-order the emoticons on (un) certainty demonstrated that these emoticons were effective in communicating (un)certainty. The emoticons are depicted in Figure S2 in the supplementary materials. These emoticons were additionally coupled to colors that were demonstrated in another pilot study ($N = 3$) to facilitate understanding of the meaning of the emoticons. The color range was as follows with increasing reliability: orange – (bright) green – green – (dark) green and is depicted in Figure S3 in the supplementary materials. The emoticons, in their respective color, were presented in the center console. The emoticon on the left (being outlined) conveyed information about the current automation reliability and the emoticon on the right (without being outlined) conveyed information on the upcoming automation reliability. To make the direction of change clearer, an arrow pointing from the emoticon on the left to the emoticon on the right was added. A LED strip at the bottom of the windshield was used to communicate the time budget in addition to communicating the current and upcoming mode. The LED strip would 'deplete' from right to left, where the size of the strip was linearly proportional with the remaining time budget (depletion of the LED strip to indicate time in current/time to next level somewhat followed Wandtner, 2018). The illuminated portion of the LED strip consisted of two parts. The left part of the illuminated portion indicated the current mode and had the same color as the color of the emoticon of the current level. The right part of the illuminated portion indicated the next mode and had the same color of the emoticon of the next mode. Additionally, a transition icon was included above the illuminated portion on the right to indicate the reason for an upcoming change in mode (e.g., construction work ahead). In addition, ambient light in the car was simulated by overlaying the interior of the car with a transparent layer in the color corresponding to the current mode. The intensity of the ambient light effect, here simulated with a changing transparency, also communicated the time budget: The intensity was high when there was still a lot of time left in the current mode and the intensity decreased with decreasing time left.

The second Anticip condition, DT_Anticip, included an HMI that communicated the desired (or allowed non–)driving task associated with the automation mode through icons in the center console, which were highlighted in the same color range as used in AR_Anticip. Three icons indicated an allowed task (with decreasing mode of automated driving: sleeping, using a laptop, using a telephone), and two icons indicated a required task (with decreasing mode of automated driving: paying attention to the road, taking over the driving task). These icons were highlighted when they were applicable, of which the icon indicating the lowest mode (taking over the driving task) was never selected in the driving scenarios of the current study. The allowance icons were additive: when sleeping was allowed, drivers were also allowed to use a laptop and/or to use a telephone, in which case all allowance icons were highlighted. The applied icons were deemed suitable in communicating a desired task based on a pilot study asking about the meaning of each icon through questionnaires ($N = 10$), and especially when being highlighted in in their respective color as demonstrated by another pilot study ($N = 6$). The time budget was communicated by the color in which the icon was highlighted. The highlighted area 'depleted' with less time remaining in the current mode. When the next mode was lower, the highlighted area depleted downwards. When the current mode would be associated with a required action with the next mode being associated with an allowed action, the highlighted area depleted upwards. To make the direction of change clearer, an arrow pointing upwards or downwards was added for an upcoming change to a higher or lower mode respectively. In addition, a transition icon indicating the reason for an upcoming change in mode was presented next to the icon communicating a desired driver task. Ambient light was used to communicate the time budget across all modes. The ambient light was simulated by overlaying the interior of the car with a layer in the highlighted color of the icon that was currently applicable. The ambient light effect was centered around the steering wheel with a radius linearly decreasing with a smaller time budget.

Baseline conditions were developed for each of the two Anticip conditions described above: AR_Base and DT_Base. The baseline conditions did not feature information that allows for anticipating an upcoming change in mode. In AR_Base only one emoticon was presented at a time that communicated the current level of reliability of the automation. In DT_Base only one icon was presented at a time that communicated the current desired (or allowed non–)driving task. The baseline versions included the same transition icons presenting the reason for a change in the reliability level. In AR_Base, the icon was presented for 5 s next to the emotion on the mid console at the moment a transition towards another mode occurred. In DT_Base, the icon was presented at the same location as in the DT_Anticip condition, but rather than presenting the icon in advance of a change in mode, the icon was only present for 5 s at the onset of such a change.

The four experimental conditions were combined with the four different driving scenarios, in such a way that each condition was coupled to a drive from A to B and a drive from B to A and that each condition was coupled to events occurring at a different place.

2.5. Measures

2.5.1. Think-aloud

The concurrent think-aloud method (Eccles & Arsal, 2017) was applied in the current experiment to gain insight into the understandability and usability of the information communicated by the HMI design in the different conditions. During this method participants speak aloud anything that comes to their mind during a certain task. The concurrent think-aloud method has been demonstrated to validly provide insight into participants' thinking (Charters, 2003). Moreover, this procedure has proved to be effective in gaining insight into situation awareness during processing of provided information in a study by Rose et al. (2019) examining situation awareness during operation of a train simulator.

2.5.2. Questionnaires

The ITC-SOPI spatial presence scale (Lessiter et al., 2001) was used to evaluate whether the set-up of the current experiment induced a sufficient feeling of presence (on a scale from 1 to 5). This evaluation was considered to be of importance, as a sufficient feeling of presence would support the validity of the findings on the other outcome measures.

Several questionnaires were used in order to gain insight into understandability and usability. Participants were asked to rate on a scale from 1−5 to what degree each HMI concept helped the participant to understand where the attention needed to be focused, what s/he was allowed to do at what moment and to what degree it supported understanding of future events. The System Usability Scale (Brooke, 1986) was used to gain insight into how participants rated the usability of the HMI concepts in the different conditions. This scale measures system usability from 0 to 100. The subset of items providing insight into monitoring (i.e., checking in during performance) of the automation-induced complacency scale (Merritt et al., 2019; items 7–10) was used to gain insight into whether monitoring performance will be adequate, in order to provide insight into potential overreliance for each condition (on a scale from 1−5). The shortened NASA-TLX (Hart & Staveland, 1988) was used to gain insight into task load induced by processing the presented information for each condition, with each of the items rated on a scale from 1−7. In this way insight can be gained into potential information overload.

Regarding user preferences, participants were asked to indicate their preferred HMI concept out of the presented concepts and to order the concepts from most preferred to least preferred.

The NASA-TLX, the ITC-SOPI spatial presence scale and the automation-induced complacency scale were shortened for the purpose of the current experiment. In these shortened versions a few items were discarded that were not applicable to the current experiment. In this way, only applicable items were answered by the participants. This approach is applied more often in research (e.g. Hart, 2006; Tjon et al., 2019). Regarding the NASA-TLX, three items on mental demand, temporal demand and frustration level were selected. Regarding the ITC-SOPI, five items were selected, omitting the items of the spatial presence scale focusing on physical interaction, following Tjon et al. (2019). For the subset of items of the automation-induced complacency scale one item was omitted. In this way, items were excluded which do not apply to a passive task such as watching a simulated drive. All items were translated to Dutch in order to ensure the items were understood by participants.

Additionally, in order to gain insight into characteristics of participants and potential individual differences in relation to preferences, questions on demographics (i.e., age, sex) and driving experience (i.e., years of having a driver's license, kilometers driven each year, whether someone has a car with automated functionalities such as LKS, ACC, and years of owning an vehicle with automated functionalities) were employed. In addition, questions on trust in automation (Payre et al., 2016), trust in technology (Merritt et al., 2013) and driving enjoyment (Ernst & Reinelt, 2017; including the constructs behavioral intention to use an autonomous car, personal driving enjoyment, perceived enjoyment of driving an autonomous car, perceived traffic safety of an autonomous car, and perceived usefulness of an autonomous car) were used, with each of the items having a scale from 1−5.

2.6. Procedure

Participants received an instruction sheet on the think-aloud protocol when an appointment for participation in the experiment was made. The instructions included 1) informing the participant that s/he will watch a simulated drive in a self-driving car while receiving information from a system in the self-driving car, 2) asking the participants to verbalize everything they are seeing and thinking, acting as if they are alone in the room speaking to themselves (following Eccles & Arsal, 2017; Jaspers et al., 2004), and 3) informing the participant that the experimenter would remain silent while watching the simulated drives only to remind the participant to keep talking if s/he falls silent (following Jaspers et al., 2004). To prepare participants for thinking aloud during the experiment, the instruction sheet also included an example of how one would think aloud when searching for what one would like on his/her sandwich in the kitchen. A different setting was used in the example to avoid priming participants on content of a similar context as provided in the experiment. A similar approach to train the think-aloud protocol has been used by Key et al. (2016). In addition to these instructions, participants were asked to facilitate an uninterrupted session, by ensuring that no other people or animals would enter the same room, and by reducing environmental noise to the best extent possible.

On the day of the experiment, participants first provided informed consent. An opportunity was provided to read the instruction sheet on the think-aloud protocol again and to ask questions (following Key et al., 2016). Participants were instructed that they will be taking a ride in a self-driving car while they would receive information from a system in the self-driving car. In this short instruction no information was provided on the meaning of automation modes or HMI elements.

The four different conditions (AR_Anticip, DT_Anticip, AR_Base, and DT_Base) were presented during four different simulated driving scenarios. Experimental conditions were tested within participants, and therefore it was important to ensure that the experience during the AR and DT Base conditions would not be affected by the participants already having received the associated Anticip condition with additional information. Therefore, AR_Base was always shown right before AR_Anticip and DT_Base was always shown right before DT_Anticip. This resulted in a selection of four different orders of the four different conditions with the four different driving scenarios. The order of the combinations that was presented to participants was determined using block randomization, in this way it was ensured that an equal number of participants was included for each of the four different orders (Goodwin, 2009). Before starting the first condition's simulated drive, the experimenter turned off his/her video image. While the participant watched the scenario the participant was thinking aloud. The experimenter prompted the participant to try to keep talking after the participant fell silent for a fixed interval of 15 s (following Jaspers et al., 2004). When the scenario ended the experimenter asked the participant whether s/he wanted to mention anything else. In order to minimize practice effects, participants were not provided with feedback on their thinking aloud and no further instruction was given (following Rose et al., 2019). Each time one condition was presented, participants answered to the items of the NASA-TLX, the System Usability Scale and the automation-induced complacency scale. After being presented with all four conditions, participants answered to the items of the ITC-SOPI spatial presence scale, the items on subjective understanding, and participants indicated their preferences. Finally, participants answered to the questions providing insight into participants' characteristics.

2.7. Data processing and analyses

The think-aloud recordings were transcribed verbatim and grouped into a series of statements, following Eccles and Arsal (2017). Grouping was performed based on temporal adjacency (e.g., a silence of one or more seconds is typically an indicator of a new statement) and content (i.e., a change of focus initiates a new statement). After transcribing the recordings completely, the recordings were distributed across two experts in human factors in vehicle automation (hereafter: coders) and each transcribed statement was coded using a priori coding scheme. For each statement the coders assessed whether its contents expressed awareness in relation to the vehicle's HMI, the automation and/or the driver's task. If this was the case, the statement was flagged as 'relevant'. The coders categorized relevant statements on the three levels of situation awareness following Rose et al. (2019) and Endsley (1988): 1) Perception (i.e., statement referring to a signal), 2) comprehension (i.e., statement about the meaning of the communicated information), and 3) projection of future states (i.e., statement referring to a future event or action). For each relevant statement it was also determined whether the statement reflected a correct or incorrect understanding about the HMI, automation and/or the driver's task. The detailed coding scheme is presented in Table S3.

The reliability of the coded data was assessed by calculating inter-rater reliability (hereafter: IRR) with 'irr' in *R* (R Core Team, 2017) using Cohen's Kappa (Cohen, 1960). IRR was assessed three times in total. A random transcription was selected for the first IRR-assessment. IRR was first calculated for the selection of relevant statements. Subsequently, IRR on the level of situation awareness at which the statement was to be judged was calculated only for those statements on which agreement on relevance was found. Likewise, IRR on correctness of the statement was calculated only for statements on which both relevance and level of situation awareness were in agreement. As a moderate inter-rater reliability of 0.59–0.61 was obtained in the study of Rose et al. (2019), the minimum required inter-rater reliability for each variable under review was set at 0.59. This might seem relatively low compared to other types of coded data, but note that inter-rater reliability for this coding procedure is not only about *how* to code the statements but also *what* should be coded (Rose et al., 2019), which makes this coding procedure inherently different from most other coding procedures. In the first two IRR-assessments, coders discussed and refined their interpretation of the variables. This resulted in the IRR-assessment meeting the aforementioned performance criteria in the first and/or second assessment. Subsequently, the remaining transcriptions were distributed across the coders. The last transcription was coded by both coders in the third and last IRR-assessment. The performance criteria were met again, suggesting that the IRR was consistently sufficient throughout the coding process.

The coded think-aloud data were analyzed using zero-inflated regression models, because count data, such as the coded data in the current study, exhibit relatively many zero observations and often have a non-normal distribution (Greene et al., 2011; Zeileis et al., 2008). Thus, generalized linear models (GLMs) with zero-inflation implemented were applied using 'glmmADMB' in *R* to analyze the coded think-aloud data. This specific implementation was chosen as it is suitable for mixed models (which is not the case for 'standard' zero-inflated models in *R* such as 'zeroinfl'). These mixed zero-inflated regression models were applied to test the effect of communicating information supporting anticipation (Base conditions versus Anticip conditions; Anticipatory information), the effect of the focus of the communicated information which was either on automation reliability or the driver task (AR conditions versus DT condition; Information focus), and the interaction between these two factors. This was tested for statements per level of situation awareness and on all levels combined, both for correct and incorrect statements. Participant was added as a random factor to all models. Distributions that are theoretically possible for zero-inflated data were fitted to the data on correct statements and to the data on incorrect statements. The best fits were applied to the models, resulting in applying models with a poisson distribution for data on correct statements and with a negative binomial distribution for data on incorrect statements. The resulting models were evaluated with Wald Chi-Squared (W) tests and p-values to assess the statistical significance of the effects. Marginal and conditional R^2 were calculated using the function 'r2′ from the package 'performance' in *R*. Marginal R^2 expresses the proportion of variance explained by the fixed factors and conditional R^2 expresses the proportion of variance explained by both the fixed and random factors (Nakagawa & Schielzeth, 2012).

Regarding questionnaire data, items on validated questionnaires (i.e., NASA-TLX, System Usability Scale, automation-induced complacency scale, ITC-SOPI, trust in automation, trust in technology) were scored as recommended in the literature. Yet, the items of the NASA-TLX and the automation-induced complacency scale were separately analyzed, as the scales were shortened for the purpose of the current experiment. For analyses on items with a rating scale testing the effect of Anticipatory information, Information focus, and the interaction between the two, data was analyzed using linear mixed-effects models using 'LME4′ in *R* with participant as random factor. Regarding analyses on preferences, effects on rankings were tested by fitting a cumulative link mixed model using the *R*-package 'ordinal', again Anticipatory information and Information focus were added as fixed factors to the model and participant was added as a random factor. To explore what aspects affects preferences, scores on subjective understandability, usability, complacency and workload for each participant's most preferred concept were compared to the scores for each participant's least preferred concept using linear mixed-effects models with participant as random factor. Additionally, individual differences related to the preferences were explored by comparing age, years of owning an vehicle with automated functionalities, years of having a driver's license, trust in automation, trust in technology and driving enjoyment in participants preferring a Base concept with participants preferring a Anticip concept and in participants preferring an AF concept with participants preferring a DT concept. These comparisons were executed using a linear model. All resulting models for the questionnaire data were evaluated with *F* tests or Chi-Squared (G^2) tests (cumulative link mixed models), *p*-values and (marginal and conditional) R^2 to assess the statistical significance of the effects and proportion of variance explained by the effects. All effects were tested against a significance level of $p < .050$.

3. Results and discussion

3.1. Think-aloud

Transcription of all participant responses resulted in a total of 2146 statements. Transcriptions indicated that participants focused on the presented simulations and did not require reminders to think-aloud. All statements were related to the experimental materials, except for two statements that lasted no longer than a few seconds. One participant noted there was a mosquito in the vehicle (which in reality was situated in the room in which the participant was located). Another participant mentioned children returning to school (who were not present in the room) at the start of the first condition. This participant complied with the think-aloud protocol after having been reminded once to focus on what could be seen in the simulation.

Out of the total of 2146 statements 947 statements were coded as relevant statements, which in turn proved to be correct in 731 (77.19%) cases and incorrect in 216 (22.81%) cases. Of the 947 relevant statements, 161 were coded as 'perception' in terms of situation awareness level, of which 99.38% were correct. Another 596 were coded as 'comprehension' (67.62% correct), and 190 were coded as 'projection' (88.42% correct).

Regarding the effects of the different conditions, results of the statistical analyses testing the effect of Information focus (AR versus DT) and Anticipatory information (Base versus Anticip), and the interaction between the two on correct and incorrect statements in total and on the three levels of situation awareness are presented in Table 1. The average number of statements on each level of situation awareness per condition are presented in Fig. 2 (correct statements on the left, incorrect statements on the right).

The results of the statistical tests presented in Table 1 demonstrate that, depending on the level of situation awareness, correct statements were affected by Information focus, Anticipatory information and/or an interaction between the two. Considering the effect of Information focus, more correct statements on a projection level occurred in the AR conditions compared to the DT conditions. Considering the effect of Anticipatory information, the Anticip conditions, compared to the Base conditions, led to a smaller number of correct statements on a comprehension level, but to a larger number of correct statements on a projection level and in total on all levels combined. The interactions between Information focus and Anticipatory information indicate a larger increase in number of correct perception statements from DT_Base to DT_Anticip than from AR_Base to AR_Anticip. Yet, the increase in number of correct projection statements from the Base to the Anticip conditions was larger for AR than for DT. Regarding the effect of anticipatory information, it appears that anticipatory information overall results in more correct statements, which is especially driven by the increase in correct projection statements. Although anticipatory information resulted in less correct comprehension statements, the effect size mR^2 indicates the increase in correct projection statements is considerably larger ($mR^2 = 0.84$) than the decrease in correct comprehension

Table 1
Outcomes of the mixed effects zero-inflated regression models for correct and incorrect statements for a perception, comprehension and projection level individually and for all levels combined (mR^2 is marginal R^2, cR^2 is conditional R^2).

	Information focus			Anticipatory information			Information focus * Anticipatory information		
CORRECT	W	p	mR^2/cR^2	W	p	mR^2/cR^2	W	p	mR^2/cR^2
Total	0.00	0.971	0.00/0.51	16.72	<0.001	0.12/0.57	0.75	0.386	0.12/0.57
Perception	2.59	0.108	0.02/0.55	0.68	0.410	0.01/0.55	5.25	0.022	0.06/0.58
Comprehension	1.09	0.296	0.01/0.34	8.01	0.005	0.08/0.39	3.27	0.070	0.13/0.42
Projection	15.80	<0.001	0.33/0.52	59.91	<0.001	0.84/0.88	9.31	0.002	0.86/0.91
INCORRECT	*W*	*p*	mR^2/cR^2	*W*	*p*		*W*	*p*	mR^2/cR^2
Total	16.60	<0.001	–	0.52	0.470	–	1.11	0.293	–
Perception	*Models failed to converge due to the low occurrence of incorrect statements on a perception level (total number of incorrect statements was 2)*								
Comprehension	20.05	<0.001	–	2.74	0.098	–	0.28	0.597	–
Projection	0.00	0.998	–	7.21	0.007	–	0.30	0.581	–

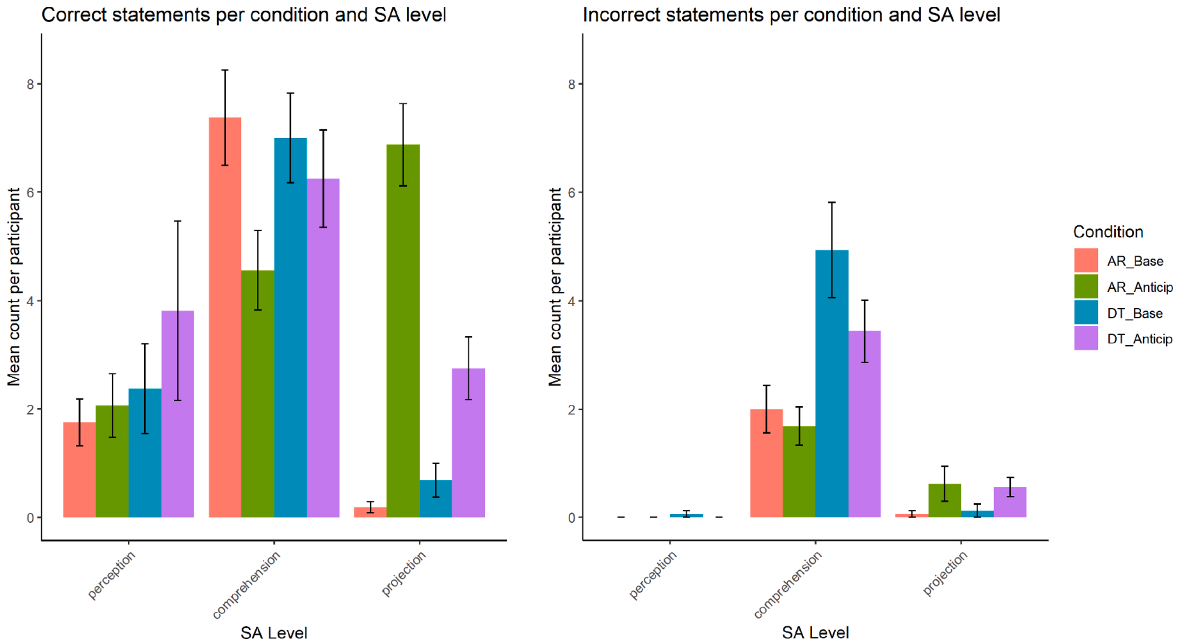

Fig. 2. Average number of statements on each situation awareness (SA) level per condition for correct statements (left) and incorrect statements (right). Error bars represent ± 1 standard error of the mean.

statements ($mR^2 = 0.08$). The decrease in correct comprehension statements with anticipatory information can also be driven by the fact that grouping of statements was performed based on temporal adjacency and that the highest level of situational awareness was coded. For example, a phrase uttered at a comprehension level followed directly by a phrase at a projection level would be considered as a single statement on a projection level. All in all, we view the large increase in correct projection statements with anticipatory information as more telling for the overall degree of situation awareness than the small decrease in correct comprehension statements. Additionally, projection statements are more often correct when information on automation reliability is presented compared to information on the driver task. Moreover, anticipatory information combined with automation reliability appears to support correct projection the most, highlighting the potential value of combining the two.

Regarding the statistical tests for the incorrect statements, models failed to converge for statements on a perception level as the number of incorrect statements on a perception level were averaging around 0. Moreover, for all models on incorrect statements R^2 values were unreliable and therefore these values are omitted. The results presented in Table 1 demonstrate that more incorrect statements were uttered on a comprehension level and in total by participants in the DT conditions compared to the AR conditions. Additionally, more incorrect projection statements occurred during the Anticip conditions compared to the Base conditions. These findings indicate that information on automation reliability was better understood than information on the driver task. The finding that more incorrect projection statements occurred with anticipatory information, probably reflects that in general more projection statements were made with anticipatory information ($N = 173$) than without such information ($N = 57$) regardless of correctness. Note also that there are considerable more correct projection statements ($N = 168$) than incorrect ones ($N = 22$), and that the absolute increase in number of correct statements from the baseline to the conditions with anticipatory information (increase of 140) was considerably higher than the increase in number of incorrect statements (increase of 16). The results therefore point towards a stronger effect of anticipatory information in supporting projection.

Yet, incorrect statements are of importance to consider, as a misunderstanding of the communicated information can potentially have detrimental effects. There were two frequently recurring themes in the incorrect projection statements when receiving anticipatory information. The first theme regarded participants relating an emoticon (AR condition) or an icon (DT condition) to an incorrect action. Typically, the participants thought they had to be more involved with the driving task than intended with the communicated information. For example, when the automation reliability is communicated indicating more attention to the driving environment is required, but that no take-over is (yet) required: “And now I see that construction work is approaching and then probably I need to take over the steering wheel.” The second theme regarded participants not understanding the meaning of a transition icon when they were anticipating an upcoming event, for example: “I am seeing a green bar on my windshield, with - I guess - something like fog that I am approaching”. In this case the participant was already driving through the fog; the transition icon announced the end of the fog approaching. These misinterpretations might be resolved when a driver would have received more information about the HMI or would have had more experience with the HMI. Yet, it might be of value to resolve this through enhancing the clarity of this information in future (studies on) HMI design.

3.2. Questionnaires

3.2.1. Spatial presence

Spatial presence experienced throughout the complete experiment was rated on average with 3.31 ($SD = 0.78$) on a scale from 1 to 5. These obtained ratings indicate a similar feeling of presence as in earlier experiments in which participants passively viewed a 360° scenario in the laboratory in virtual reality (Tjon et al., 2019).

3.2.2. Subjective comprehension and usability

Results on subjective comprehension and usability of the presented information are depicted on the left in Fig. 3. Participants indicated to have a better understanding about when they needed to pay attention to the road in the Anticip conditions ($M = 3.75$, $SD = 1.16$) compared to the Base conditions ($M = 3.28$, $SD = 1.14$), $F(1, 47) = 5.75$, $p = .021$, marginal $R^2 = 0.04$, and conditional $R^2 = 0.57$. Additionally, participants indicated to have a clearer understanding about what they were allowed to do when they were not expected to pay attention to the road in the DT conditions ($M = 3.19$, $SD = 1.40$) compared to the AR conditions ($M = 1.88$, $SD = 1.04$), $F(1, 47) = 30.53$, $p < .001$, marginal $R^2 = 0.22$, and conditional $R^2 = 0.54$. It was also indicated that the Anticip conditions ($M = 4.19$, $SD = 0.74$) better supported understanding about future events (like a road block) compared to the Base conditions ($M = 3.72$, $SD = 1.17$), $F(1, 47) = 5.47$, $p = .024$, marginal $R^2 = 0.05$, and conditional $R^2 = 0.37$. No other effects of Anticipatory information and Information focus, and the interaction between the two are were found on the items measuring subjective comprehension, all $F \leq 3.07$, $p \geq 0.086$, marginal $R^2 \leq 0.22$, and conditional $R^2 \leq 0.66$.

Regarding data on the System Usability Scale, Fig. 3 on the right, a significant main effect for Information focus was found, $F(1, 47) = 6.56$, $p = .014$, marginal $R^2 = 0.09$, and conditional $R^2 = 0.10$. This effect demonstrated higher average system usability scores for the AR conditions ($M = 60.31$, $SD = 19.89$) compared to the DT conditions ($M = 47.11$, $SD = 21.56$). No effect of Anticipatory information or the interaction between the two factors on the System Usability Scale data was found, both $F \leq 0.10$, $p \geq 0.757$, marginal $R^2 \leq 0.09$, and conditional $R^2 \leq 0.10$.

Considering the findings on subjective comprehension and usability together, the findings indicate that anticipatory information supported subjective comprehension about when to pay attention to the road and about whether an event would take place. Additionally, information on the driver task supported subjective comprehension about what tasks were allowed when there was no need to pay attention to the road. Yet, participants subjectively considered information on the automation reliability to be overall more usable than information on the driver task.

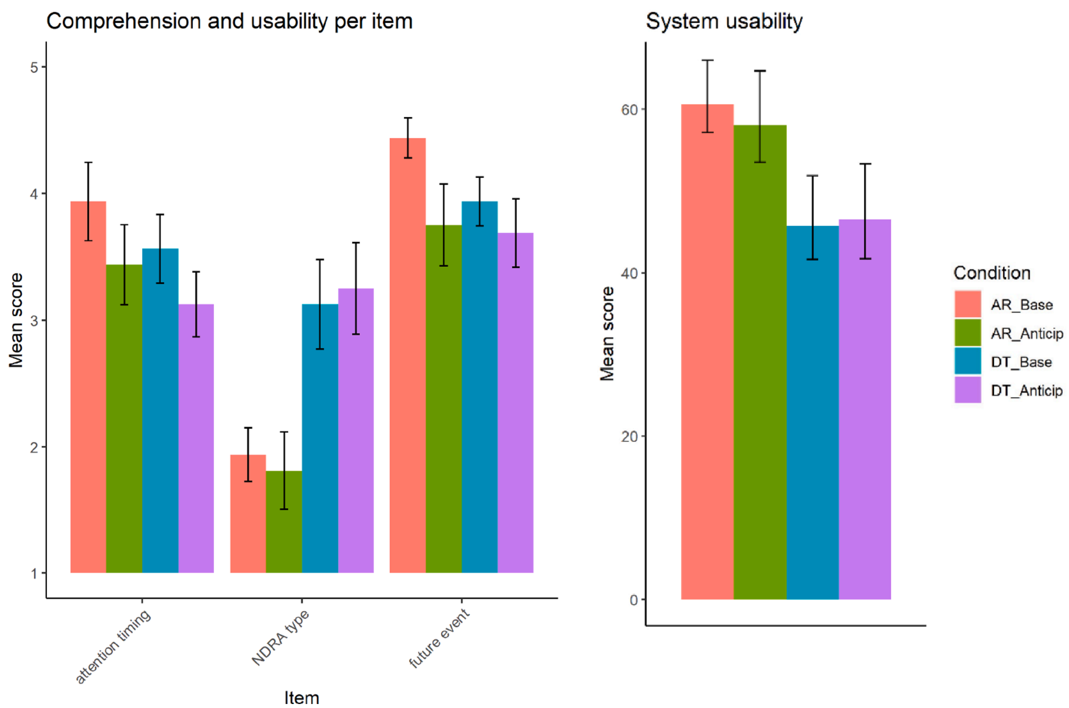

Fig. 3. Left: Average scores for each item measuring subjective comprehension and usability per condition. 'Attention timing' = 'The information system supported my understanding about when I needed to pay attention to the road.', 'NDRA type' = 'The information system supported my understanding about what I was allowed to do at moments when there was no need to pay attention to the road', 'Future event' = 'The information system supported my understanding of future events (like a road block).' Right: Average system usability as measured by the System Usability Scale. Error bars represent ± 1 standard error of the mean.

3.2.3. Complacency

Fig. 4 presents the average scores on each of the items of the complacency scale, thereby providing insight into potential over-reliance. One out of four items was significantly affected by the experimental conditions. Namely, carefully watching the HMI concepts presented in the DT conditions was considered as taking more time away from more important or interesting things ($M = 3.78$, $SD = 0.91$) than carefully watching the HMI concepts presented in the AR conditions ($M = 3.09$, $SD = 1.03$), $F(1, 47) = 8.48$, $p = .005$, marginal $R^2 = 0.11$, and conditional $R^2 = 0.16$. No other effects of Anticipatory information and Information focus, and the interaction between the two are were found on the items measuring complacency, all $F \leq 1.83$, $p \geq 0.183$, marginal $R^2 \leq 0.12$ and conditional $R^2 \leq 0.42$. The significant effect found for complacency suggests that when communicating on the driver task compared to when communicating on automation reliability other things are considered to be more important or interesting.

3.2.4. Information load

No significant main nor interaction effects were found on information load as measured by the NASA-TLX, all $F \leq 3.82$, $p \geq 0.057$, marginal $R^2 \leq 0.06$, and conditional $R^2 \leq 0.21$. This indicates that experienced information load was unaffected by the different conditions, even though more information was presented to the driver in the Anticip compared to the Base conditions.

3.2.5. Preferences

No effects of Anticipatory information and Information focus were found on the rankings of the HMI concepts presented in the conditions, all $G^2 \leq 0.20$, $p \geq 0.654$, marginal $R^2 = 0.00$, and conditional $R^2 = 0.00$. This finding indicates that generally participants did not have a clear preference for specific information.

Although no clear preference for a specific concept is demonstrated, the question arises whether a preferred concept is related to a difference experience of the concept than a non-preferred concept. We compared subjective experiences between participants' most preferred and least preferred concepts, yielding several significant effects. First of all, for the most preferred HMI concept participants indicated that it better supported understanding about future events (like a road block) ($M = 4.31$, $SD = 0.60$) compared to the least preferred concept ($M = 3.63$, $SD = 1.15$), $F(1, 15) = 7.35$, $p = .016$, marginal $R^2 = 0.13$, and conditional $R^2 = 0.47$. Participants' most preferred concept was also rated higher on the System Usability Scale ($M = 63.75$, $SD = 25.25$) compared to the least preferred system ($M = 39.84$, $SD = 17.90$), $F(1, 15) = 9.71$, $p = .007$, marginal $R^2 = 0.24$, and conditional $R^2 = 0.24$.

Regarding complacency, the most preferred concept took less time away from more important or interesting things ($M = 2.94$, $SD = 1.29$) than the least preferred concept ($M = 4.00$, $SD = 0.63$), $F(1, 15) = 8.76$, $p = .010$, marginal $R^2 = 0.22$, and conditional $R^2 = 0.22$.

Moreover, an exploration of the NASA-TLX subscales revealed that HMI designs that induced the least information load were preferred. First, mental workload was rated lower for the most preferred design ($M = 2.75$, $SD = 1.77$) compared to the least preferred design ($M = 4.13$, $SD = 1.71$), $F(1, 15) = 6.32$, $p = .024$, marginal $R^2 = 0.14$, conditional $R^2 = 0.32$. Second, temporal demand was lower for the most preferred ($M = 1.88$, $SD = 0.89$) compared to the least preferred design ($M = 2.94$, $SD = 1.39$), $F(1, 15) = 6.66$, $p = .021$, marginal $R^2 = 0.18$, conditional $R^2 = 0.18$.

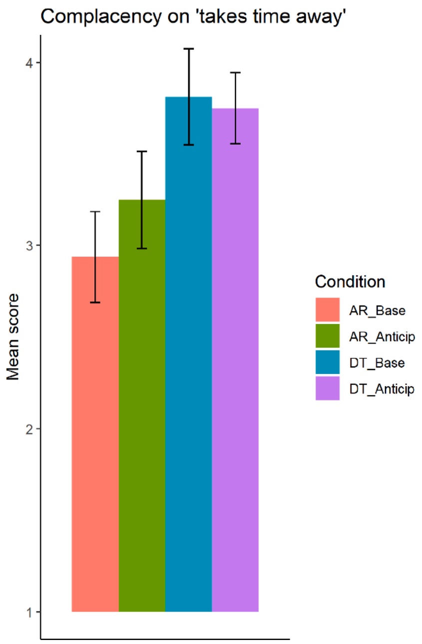

Fig. 4. Average score for the item 'Takes time away' (i.e., 'Carefully watching this information system takes time away from more important or interesting things.') of the complacency scale. Error bars represent ± 1 standard error of the mean.

No other effects were found for the relationship on experiences and preferences, all $F \leq 3.95$, $p \geq 0.066$, marginal $R^2 \leq 0.09$, and conditional $R^2 \leq 0.64$.

In addition to the relation between subjective experiences based on the HMI concepts and preferences for those concepts, it was explored whether individual differences might play a role in preferring one concept over another. A total of 7 participants preferred a concept which communicated information in advance (i.e., AR_Anticip and DT_Anticip), whereas 9 participants expressed a preference for a concept which did not communicate such information (i.e., AR_Base and DT_Base). The groups were compared on the following variables: age, years of owning an vehicle with automated functionalities, years of having a driver's license, trust in automation, trust in technology, driving enjoyment (including the constructs behavioral intention to use an autonomous car, personal driving enjoyment, perceived enjoyment of driving an autonomous car, perceived traffic safety of an autonomous car, and perceived usefulness of an autonomous car. Participants that preferred a Anticip concept had a higher trust in technology ($M = 4.42$, $SD = 0.55$) than participants that preferred a Base concept ($M = 3.96$, $SD = 0.09$), $F(1, 14) = 4.65$, $p = .049$, $R^2 = 0.25$, adjusted $R^2 = 0.20$. Differences were also demonstrated on driving enjoyment, with participants preferring a Anticip concept having a higher intention to use an autonomous car when it's available (mean intention to use = 4.10, $SD = 0.92$), expecting a higher traffic safety of autonomous cars (mean perceived safety = 4.37, $SD = 0.72$) and expecting a higher perceived usefulness of autonomous cars (mean perceived usefulness = 4.59, $SD = 0.59$) than participants preferring a Base concept (mean intention to use = 2.88, $SD = 0.83$; mean perceived safety = 3.43, $SD = 0.83$; mean perceived usefulness = 3.48, $SD = 1.29$), $F(1, 14) = 7.65$, $p = .015$, $R^2 = 0.35$, adjusted $R^2 = 0.31$ and $F(1, 14) = 5.82$, $p = .030$, $R^2 = 0.29$, adjusted $R^2 = 0.24$ and $F(1, 14) = 5.35$, $p = .036$, $R^2 = 0.28$, adjusted $R^2 = 0.23$, respectively. No other significant differences were found, all $F \leq 3.21$, $p \geq 0.095$, $R^2 \leq 0.19$, adjusted $R^2 \leq 0.13$.

In contrast, no significant differences were found for these variables related to individual differences for participants preferring a concept communicating on automation fitness (i.e., AR_Base and AR_Anticip, $N = 8$) compared to participants preferring a concept communicating on driver task (i.e., DT_Base and DT_Anticip, $N = 8$), all $F \leq 2.23$, $p \geq 0.158$, $R^2 \leq 0.14$, adjusted $R^2 \leq 0.08$.

The results on the preference data suggest that although participants overall did not have a clear preference for a specific concept, preferences appear to be related to whether the concept supported subjective understanding about event occurrences, subjective system usability, whether the concept is thought to take time away from more important or interesting things, and the subjective information load induced by the concept. Additionally, anticipatory information appears to be preferred by participants that have a higher trust in technology, a higher intention to use an autonomous car and that expect a higher traffic safety and usefulness of autonomous cars.

4. General discussion

The current study examined HMI design to continuously support supervision and NDRA self-regulation during automated driving. Specifically, effects were examined of HMI design 1) communication on the automation reliability or on the required/allowed driver task, and 2) presenting information supporting anticipation (including information on the available time budget, the upcoming mode, and reasons for changing to the upcoming mode). In order to assess these effects, participants were presented with simulated drives in which they received different types of information from an HMI in the simulated car. Participants were thinking aloud during the simulated drives and answered questions about their experience and preferences after the simulated drives. The subsequent analysis focused on understandability, usability and user preferences as a result of the manipulations of information focus and (presence of) anticipatory information. In the next sections a general discussion on the effects as described in the results and discussion section will be provided.

4.1. Information focus

The think-aloud data demonstrated that information on automation reliability was better understood and better supported anticipation of upcoming changes (as demonstrated by an interaction effect for correct projection statements) than information on the driver task. Although communicating on the driver task facilitated participants' subjective understanding about what they were allowed to do when they were not expected to pay attention to the road, it was found that information on automation reliability was rated higher by participants on subjective system usability.. Moreover, information on the driver task is potentially more likely to induce overreliance as participants considered other things to be more important or interesting than this type of information. Information focus did not impact information load, or preference.

The demonstrated beneficial effects of providing continuous information on automation reliability are in line with findings of previous user studies (Beggiato et al., 2015; Feierle et al., 2020; Hecht et al., 2019) and previous driving simulator studies (Beller et al., 2013; Helldin et al., 2013; Large et al., 2017; Stockert et al., 2015). There are also previous studies (Lu et al., 2019; van den Beukel et al., 2016; Yang et al., 2018) that suggest providing information on the desired driver task can be beneficial for increasing situational awareness. The current study suggests, however, that it is better to inform on automation reliability than on the desired driver task. Yet, these previous studies are different from the current study in an important respect. The current study provided specific information about the driver which not only indicated when the driver needed to pay attention but also when NDRA engagement was possible including which specific NDRA would be appropriate. Previous studies communicated when supervision was needed, but not which specific NDRA was allowed when there was no need for supervision. Therefore, providing information about when the driver should take on their supervising task can still be beneficial, especially to prepare the driver for a (potential) take-over. Yet, the current study suggests that informing on a specific NDRA might not be beneficial.

4.2. Anticipatory information

The think-aloud data demonstrated that anticipatory information was associated with an enhanced general situational awareness (i.e., total correct statements), and specifically supported anticipation of upcoming changes (i.e., correct projection statements). Participants also indicated subjectively that anticipatory information supported their understanding about when they needed to pay attention to the road and that it supported their understanding about future events (such as a road block). The beneficial effects of communicating anticipatory information were attained without impacting information load and complacency, suggesting that even though more information is presented to the driver, this is not causing any information overload or overreliance. Although anticipatory information had beneficial effects, this information was not subjectively preferred over not receiving such information.

It is important to realize that anticipatory information also led to more misinterpretations in anticipating upcoming changes (i.e., incorrect projection statements). Yet, the beneficial effects of anticipatory information were considerably larger. It is, however, important to enhance the clarity of the anticipatory information in future research to prevent such misinterpretations.

The demonstrated beneficial effects of anticipatory information are in line with research suggesting the value of communicating time budgets (Beggiato et al., 2015; Hecht, Kratzert, et al., 2020; Wandtner, 2018) and (reasons for) changes in automation reliability (Carsten & Martens, 2019; Körber et al., 2018; Martens & van den Beukel, 2013; Richardson et al., 2018). Although we included information on the *time* left in a current mode as anticipatory information, it could alternatively be valuable to provide information on the remaining *distance* in a current mode. For example, Holländer and Pfleging (2018) and Richardson et al. (2018) found that information on distance until next take-over was preferred to and considered to be more useful than information on time until next take-over. In Holländer and Pfleging (2018), however, time information was provided by using a digital countdown indicating the time budget very specifically in minutes and seconds while information on distance was provided by a bar that depleted with less distance until a take-over. Therefore, in the study by Holländer and Pfleging (2018) the demonstrated effect could also represent that more abstract information on remaining time and/or distance is preferred to more precise information. In Richardson et al. (2018) both time and distance information was communicated specifically in seconds or meters respectively. Future studies could examine how to best communicate the time budget, including whether it might be better to replace time with distance information. Additionally, it has even been suggested that HMI design could aid in self-regulation by also providing information about automation reliability over the complete trip, rather than just providing information on the current or near future automation status (Hecht, Sievers, et al., 2020). The effect of such information could be examined in future studies.

It is important to also consider potential implementation of the current findings in real-world settings. For example, the take-over paradigm employed in many empirical studies has been criticized as not being sufficiently grounded in real-life issues and hindering out-of-the-box research (de Winter et al., 2021). Providing a driver with anticipatory information on automation reliability actually has the potential to change the standard studied take-over situation in which a take-over request is only provided maximally 15 s in advance. Such small take-over times are thought to be unrealistic in practice and informing a driver a longer time in advance for such a change is probably required (de Winter et al., 2021). When implementing communication of anticipatory information the question arises whether such information is available to be communicated well in advance. This depends largely on how well the automation is able to predict upcoming changes. Initially, predictions might be based on comparing the automation's ODD with the expected circumstances on the selected route. These expected circumstances can be derived from, for example, HD maps and real-time weather and traffic information. In the European Union's Horizon 2020 project 'MEDIATOR' (No 814735), involving both researchers and experts from industry, first steps in deriving such predictions in such a way have already been taken (Cleij et al., 2020; Mano et al., 2022). In this project a system is being developed that mediates between driver and automation, taking into account who is fittest to drive. Algorithms that derive the required automation time budgets in real-time will be implemented in prototypes and tested in both simulator and on-road studies. Such studies will allow for expanding and validating the current study's findings. The accuracy of time budgets can potentially be further improved in the future by cooperative perception technology that extends sensor ranges by sharing and combining sensor data of multiple vehicles using vehicle to vehicle communication (Kim et al., 2013; Naujoks et al., 2015, 2017; Rauch et al., 2012; Yoon et al., 2021). Future work should examine how predictions can best be made to provide the most reliable and useful anticipatory information. Moreover, it should be studied how inaccuracies in the predictions affect the effectiveness of providing this information. For example, in the MEDIATOR project the time left in a current mode is foreseen as a likely, but not a certain, time window (Cleij et al., 2020).

4.3. Individual differences and preferences

When considering the proportion of variance that is explained by the effects examined in the current study it becomes apparent that in many instances individual differences also play a substantial role. To be precise, the explained variance increases on average when the variance of participant as random factor was considered in addition to the main variables as fixed factors. In line with Cohen (1988) that defines an R^2 value equal and above 0.26 as reflecting a large effect, 5 effects in the current study (all related to think-aloud data) would be considered as explaining a large proportion of variance without accounting for individual differences, whereas more than 20 effects would be considered as explaining a large proportion of variance when individual differences are accounted for. Even though individual differences appear to be of importance, the effect of anticipatory information on anticipation of upcoming changes as demonstrated for the think-aloud data was so strong (*marginal* $R^2 = 0.84$) that additionally considering individual differences did not account for much extra variance (*conditional* $R^2 = 0.88$). This finding highlights the importance of communicating information supporting predictability.

The exploratory analyses on preferences and individual differences also demonstrated that individual differences play a role. No clear preference for a specific HMI concept was found. Yet, overall concepts were preferred that were subjectively experienced as supporting understanding about event occurrences, as having a better system usability, as taking less time away from more important or interesting things, and as inducing less subjective information load. Moreover, anticipatory information appears to be preferred by people that have a higher trust in technology, a higher intention to use an autonomous car and that expect a higher traffic safety and usefulness of autonomous cars. Potentially, the subjective appreciation of anticipatory information might be supported by making people aware about the potential safety benefits and usefulness of autonomous cars and the trustworthiness of technology. Yet, as the role of individual differences in preferences was examined in the current study in an exploratory fashion, these effects of individual differences (in subjective experiences) and what the impact would be of intervening on this should be examined in more detail in future work.

4.4. Limitations and recommendations for future research

In the interpretation of the current findings it is important to take into account that the current study concerned an exploratory online experiment with a focus on examining understandability, usability and user preferences. Although the experiment took place online, the methodology of the current work was matched to previous think-aloud studies as much as possible, including an experimenter virtually accompanying and guiding each session. No interruptions during the session were observed (such as people entering the room), suggesting that the instructions sent prior to the experiment to minimize potential distractions were followed. A total of 99.9% of the think-aloud transcriptions concerned statements about the experimental materials, suggesting that participants focused on the experiment during the online sessions. Moreover, the current study suggests a similar feeling of presence as in earlier experiments in which participants passively viewed a 360° scenario in the laboratory in virtual reality (Tjon et al., 2019). These findings support the validity of the current study's methodology. While the authors believe that the validity will increase with more realistic settings, these findings suggest the current study's methodology could be employed in future online studies for exploring the effectiveness of HMI designs. We would recommend the experimenter to be virtually present at all times, to send instructions beforehand to minimize potential distractions, and to ensure a stable internet connection with high sound quality, such that experimental materials are presented without interference and to facilitate subsequent transcriptions. Yet, further validation of this online methodology would be valuable, by for example comparing data collected online to data collected in the lab.

Even though it was demonstrated that the presented scenarios induced a feeling of presence comparable to earlier work in virtual reality (Tjon et al., 2019), examining the effectiveness of communicating anticipatory information and information on automation reliability in a more immersive setting, such as a simulator or virtual reality experiment, would be an important next step. Such an immersive setting would allow for additionally examining the effect of communicating this information on actual driver behavior indicative of supervision and NDRA self-regulation, including take-overs. Moreover, the current study communicated four reliability modes of automated driving, reflecting both automation reliability and required and allowed driver tasks, instead of just two (on versus off) or three modes (on versus off versus monitoring). Communicating effectively on multiple reliability modes could be an enabler for introducing SAE Level 3. Approval was recently obtained for an SAE Level 3 system in Germany, the Mercedes traffic jam assist, which takes over the driving task in highway traffic jam situations (Templeton, 2021). With such a system, the driver will have a supervisory/monitoring role and should remain on standby, ready to take over at any time. The driver will, however, also have the opportunity to engage in NDRAs, which could potentially hinder the driver's supervisory responsibilities in the driving task. It is currently unclear whether SAE Level 3 systems will ever be suitable for urban settings, where the highly dynamic environment requires ample time for the driver to regain situation awareness before safely taking over control. To avoid risks, some vehicle manufactures have proposed to skip SAE Level 3 altogether, and to jump from SAE Level 2 to SAE Level 4 (Schartmüller et al., 2019). Yet, technologically it is challenging to make such a big step at once. An effective distinction in and communication of reliability modes could enable drivers to better understand and take on their role in SAE Level 3 and therewith it could potentially support implementation of this level of vehicle automation well before fully automated vehicles will be available. This would also mean that the corresponding safety benefits and opportunities to collect data on the road to aid the development of automated vehicles are utilized at an earlier stage. The current findings suggest the four modes communicated in the current study were understood well, as evidenced by the relatively low number of incorrect statements in the think-aloud data overall even though no instructions on the meaning of these modes were provided and a lot of changes in modes were occurring in a relatively short period of time. However, a study by Seppelt et al. (2019) demonstrated that, when it comes to labels used to communicate different automation modes to the driver, extremes of automation driving modes (e.g., 'no driving automation' and 'driverless') were better understood than levels falling in between those extremes (e.g., 'conditional self-driving'). Therefore, the authors raise the question if the driver's understanding in this study of (communication on) different levels in between the extremes is high enough for safe operation of automated vehicles. Further research, using for example a driving simulator or virtual reality, is needed to examine the effects of communicating four different modes on actual driving behavior and NDRA self-regulation.

Future work could also consider examining effects in a larger sample of participants. Although the number of participants in the current study is in line with previous research using inferential statistics on think-aloud data, potentially meaningful effects might have remained undetected. For example, no effects were found on information load, with all $F \leq 3.82$, $p \geq 0.057$, marginal $R^2 \leq 0.06$, and conditional $R^2 \leq 0.21$. This indicates that for information load there was one statistically insignificant effect that could explain 6% of the variance by the fixed factors and 21% of the variance by both the fixed and random factors, which could have potentially turned out statistically significant with a higher number of participants.

It might not only be valuable to examine effects in a larger groups of participants, but also in a more varied group of participants, especially since the current findings suggest individual differences play a role. Moreover, the current study only included participants that had experience with assisted driving technology. Yet, HMI design should not only be effective for experienced users, but also for first-time users. Therefore, future work should examine the generalizability of the reported effects towards novice users.

Another important aspect to take into account in interpreting the current findings is that HMI implementations different from the ones employed in the current study could be applied to provide anticipatory information and/or information on automation reliability. It can only be concluded what exact implementation would be most valuable when the effects of implementations covering the potential design space more broadly have been examined. Moreover, including information in addition to anticipatory information and/or information on automation reliability can also be beneficial. For example, providing information on automation behavior and automation perception can also support the supervisory role by improving the driver's understanding of and trust in the automation. Reasons for ongoing maneuvers and previews of next maneuvers of the automation and information on detected surrounding vehicles were considered to be valuable in Beggiato et al. (2015). In a study by Diels and Thompson (2017) participants with no experience with automated driving functionalities also indicated to prefer a visualization of detection and identification of hazards by the automation. Diels and Thompson (2017) and Beggiato et al. (2015) both concluded, however, that there is great variance in the need for this type of information depending on the level of automation and depending on the amount of experience with and trust in automation. In situations where this information on automation behavior and automation perception can be valuable it should be examined how this can best be combined with information on automation reliability and anticipatory information without inducing information overload.

5. Conclusion

All in all, the findings of the current study highlight the potential of providing the driver with anticipatory information and information on automation reliability, and especially a combination between the two, to support supervision and NDRA self-regulation through HMI design as long as automated vehicles are not yet fully automated.

Funding

The current research was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 814735 (Project MEDIATOR).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The funder played no role in study design, collection, analysis, interpretation of data, writing of the report, or in the decision to submit the paper for publication. The funder accepts no responsibility for the contents.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trf.2022.03.020.

References

Andreone, L., Borodani, P., Pallaro, N., Tango, F., Bellotti, F., Weber, H., Altpeter, B., Reimer, F., Griffon, T., Sauvaget, J. L., Geronimi, S., Page, Y., Guerineau, T., Willemotte, G., Mäkinen, T. (2021). Pilot Reporting Outcomes, Deliverable D6.5 of the H2020 project L3Pilot.

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J., Othersen, I., & Petermann-Stock, I. (2015). *What would drivers like to know during automated driving?* Tagung Fahrerassistenzsysteme: Information needs at different levels of automation.

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver-automation interaction: An approach using automation uncertainty. *Human Factors, 55*(6), 1130–1141.

Brooke, J. (1986). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. London: Taylor and Francis.

Carsten, O., & Martens, M. H. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. *Cognition, Technology & Work, 21*(1), 3–20.

Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education Journal, 12*(2), 68–82.

Cleij, D., Bakker, B., Borowsky, A., Christoph, M., Fiorentino, A., van Grondelle, E., Mano, D., van Nes, N. (2020). Mediator system and functional requirements, deliverable D1.4 of the H2020 project MEDIATOR.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Cohen, J. (1988). *Statistical power analysis for the behaviour science*. New York: Lawrence Erlbaum Associated.

Cycling '74 (2020). Max 8 [computer software]. Retrieved from: https://cycling74.com/products/max.

Diels C., Thompson S. (2017) Information expectations in highly and fully automated vehicles. In: N. Stanton (Eds.), Advances in human aspects of transportation. AHFE 2017. *Advances in Intelligent Systems and Computing* (Vol. 597).

de Winter, J., Stanton, N., & Eisma, Y. B. (2021). Is the take-over paradigm a mere convenience? *Transportation Research Interdisciplinary Perspectives, 10*, 100370. https://doi.org/10.1016/j.trip.2021.100370

Eccles, D. W., & Arsal, G. (2017). The think aloud method: What is it and how do I use it? *Qualitative Research in Sport, Exercise and Health, 9*(4), 514–531.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). In *Proceedings of the national aerospace and electronics conference* (pp. 789–795).

Ernst, C. P. H., & Reinelt, P. (2017). Autonomous Car Acceptance: Safety vs. Personal Driving Enjoyment. *Twenty-third Americas Conference on Information Systems* (pp. 1–8).

Feierle, A., Danner, S., Steininger, S., & Bengler, K. (2020). Information needs and visual attention during urban, highly automated driving—An investigation of potential influencing factors. *Information, 11*(2), 62.
Feldhütter, A., Härtwig, N., Kurpiers, C., Hernandez, J. M., & Bengler, K. (2019). Effect on mode awareness when changing from conditionally to partially automated driving. In *Advances in intelligent systems and computing* (p. 823).
Goodwin, C. J. (2009). *Research in psychology: Methods and design* (5th ed.). Hoboken: Wiley.
Greene, J. A., Costa, L. J., & Dellinger, K. (2011). Analysis of self-regulated learning processing using statistical models for count data. *Metacognition and Learning, 6*(3), 275–301.
Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904–908.
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology. Human Mental Workload* (p. 139–183).
Hecht, T., Darlagiannis, E., & Bengler, K. (2019). Non-driving related activities in automated driving–An online survey investigating user needs. In *International conference on human systems engineering and design: Future trends and applications* (pp. 182–188).
Hecht, T., Kratzert, S., & Bengler, K. (2020). The effects of a predictive HMI and different transition frequencies on acceptance, workload, usability, and gaze behavior during urban automated driving. *Information, 11*(2), 73.
Hecht, T., Sievers, M., & Bengler, K. (2020). Investigating user needs for trip planning with limited availability of automated driving functions. *International Conference on Human-Computer Interaction*, 359–366.
Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications* (pp. 210–217).
Holländer, K., & Pfleging, B. (2018). Preparing drivers for planned control transitions in automated cars. In *The 17th international conference on mobile and ubiquitous multimedia* (pp. 83–92).
Jaspers, M. W., Steen, T., Van Den Bos, C., & Geenen, M. (2004). The think aloud method: A guide to user interface design. *International Journal of Medical Informatics, 73*(11–12), 781–795.
Key, C. E. J., Morris, A. P., & Mansfield, N. J. (2016). Situation awareness: Its proficiency amongst older and younger drivers, and its usefulness for perceiving hazards. *Transportation Research Part F: Traffic Psychology and Behaviour, 40*, 156–168.
Kim, S. W., Chong, Z. J., Qin, B., Shen, X., Cheng, Z., Liu, W., & Ang, M. H. (2013). Cooperative perception for autonomous vehicle control on the road: Motivation and experimental results. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5059–5066.
Kircher, K., & Ahlstrom, C. (2018). Evaluation of methods for the assessment of attention while driving. *Accident Analysis & Prevention, 114*, 40–47.
König, M., & Neumayr, L. (2017). Users' resistance towards radical innovations: The case of the self-driving car. *Transportation Research Part F: Traffic Psychology and Behaviour, 44*, 42–52.
Körber, M., Prasch, L., & Bengler, K. (2018). Why do I have to drive now? Post hoc explanations of takeover requests. *Human Factors, 60*(3), 305–323.
Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M. H., Bengler, K., Andersson, J., Merat, N., Reed, N., Flament, M., Hagenzieker, M., & Happee, R. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science, 20*(3), 223–249.
Large, D. R., Burnett, G., Morris, A., Muthumani, A., & Matthias, R. (2017). A longitudinal simulator study to explore drivers' behaviour during highly-automated driving. *International Conference on Applied Human Factors and Ergonomics*, 583–594.
Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments, 10*(3), 282–297.
Lin, R., Liu, N., Ma, L., Zhang, T., & Zhang, W. (2019). Exploring the self-regulation of secondary task engagement in the context of partially automated driving: A pilot study. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, 147–160.
Lu, Z., Zhang, B., Feldhütter, A., Happee, R., Martens, M., & De Winter, J. C. F. (2019). Beyond mere take-over requests: The effects of monitoring requests on driver attention, take-over performance, and acceptance. *Transportation Research Part F: Traffic Psychology and Behaviour, 63*, 22–37.
Mano, D., Kotelnikov, E., Van Gent, P., Knauss, A., Baker, B., Enhuber, S. (2022). Prototype Component for Estimation of Automation State in Driving Context, Deliverable D2.3 of the H2020 project MEDIATOR. [Submitted for publication].
Martens, M. H., & van den Beukel, A. P. (2013). The road to automated driving: Dual mode and human factors considerations. In *16th international IEEE conference on intelligent transportation systems* (pp. 2262–2267).
Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-over requests for automated driving. *Procedia Manufacturing, 3*, 2867–2873.
Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I Trust It, but I Don't Know Why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors, 55*(3), 520–534.
Merritt, S. M., Ako-Brew, A., Bryant, W. J., Staley, A., McKenna, M., Leone, A., & Shirase, L. (2019). Automation-induced complacency potential: Development and validation of a new scale. *Frontiers in Psychology, 10*, 225.
Mirnig, A. G., Gärtner, M., Laminger, A., Meschtscherjakov, A., Trösterer, S., Tscheligi, M., McCall, R., & McGee, F. (2017). Control transition interfaces in semiautonomous vehicles: A categorization framework and literature analysis. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications* (pp. 209–220).
Nakagawa, S., & Schielzeth, H. (2012). A general and simple method for obtaining R2 from generalized linear mixed- effects models. *Methods in Ecology and Evolution, 4*, 133–142.
Naujoks, F., Forster, Y., Wiedemann, K., & Neukum, A. (2017). A human-machine interface for cooperative highly automated driving. *Advances in Human Aspects of Transportation. Advances in Intelligent Systems and Computing* (vol 484).
Naujoks, F., Grattenthaler, H., Neukum, A., Weidl, G., & Petrich, D. (2015). Effectiveness of advisory warnings based on cooperative perception. *IET Intelligent Transport Systems, 9*(6), 606–617.
Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. Advances in Human Aspects of. *Transportation, 7*(Part I), 1–431.
Payre, W., Cestac, J., & Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 58*(2), 229–241.
Petermeijer, S., Bazilinskyy, P., Bengler, K., & De Winter, J. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics, 62*, 204–215.
R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.
Rauch, A., Klanner, F., Rasshofer, R., & Dietmayer, K. (2012). Car2x-based perception in a high-level fusion architecture for cooperative perception systems. *IEEE Intelligent Vehicles Symposium*, 270–275.
Richardson, N. T., Flohr, L., & Michel, B. (2018). Takeover requests in highly automated truck driving: How do the amount and type of additional information influence the driver-automation interaction? *Multimodal Technologies and Interaction, 2*(4), 68.
Rose, J., Bearman, C., Naweed, A., & Dorrian, J. (2019). Proceed with caution: Using verbal protocol analysis to measure situation awareness. *Ergonomics, 62*(1), 115–127.
Ruijten, P. A. M., Terken, J. M. B., & Chandramouli, S. N. (2018). Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies Interact, 2*(4), 62.
Schartmüller, C., Weigl, K., Wintersberger, P., Riener, A., & Steinhauser, M. (2019). Text comprehension: Heads-up vs. Auditory displays: Implications for a productive work environment in SAE Level 3 automated vehicles. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications* (pp. 342–354).
Schömig, N., & Metz, B. (2013). Three levels of situation awareness in driving with secondary tasks. *Safety Science, 56*, 44–51.
Schömig, N., Metz, B., & Krüger, H. P. (2011). Anticipatory and control processes in the interaction with secondary tasks while driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(6), 525–538.

Seppelt, B., Reimer, B., Russo, L., Mehler, B., Fisher, J., & Friedman, D. (2019). Consumer confusion with levels of vehicle automation. In *Proceedings of the 10th international driving symposium on human factors in driver assessment, training and vehicle design* (pp. 391–397).

Shah, S., Dey, D., Lovett, C., & Kapoor, A. (2018). Airsim: High-fidelity visual and physical simulation for autonomous vehicles. *Field and Service Robotics*, 621–635.

Society of Automotive Engineers International J3016 (2021). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems.* Warrendale, PA: SAE International.

Stockert, S., Richardson, N. T., & Lienkamp, M. (2015). Driving in an increasingly automated world–approaches to improve the driver-automation interaction. *Procedia Manufacturing, 3*, 2889–2896.

Thill, S., Riveiro, M., Lagerstedt, E., Lebram, M., Hemeren, P., Habibovic, A., & Klingegård, M. (2018). Driver adherence to recommendations from support systems improves if the systems explain why they are given: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 56*, 420–435.

Tinga, A. M., De Zwart, R., Jansen, R. J., & De Goede, M. (2021). Automatic for the people-how prior encounters shape prospected interactions with automated shuttles. In *2021 IEEE intelligent vehicles symposium (IV)* (pp. 1485–1492).

Tivesten, E., & Dozza, M. (2015). Driving context influences drivers' decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study. *Journal of Safety Research, 53*, 87–96.

Tjon, D. M., Tinga, A. M., Alimardani, M., & Louwerse, M. M. (2019). Brain activity reflects sense of presence in 360 video for virtual reality. *Information Systems Development: Information Systems Beyond 2020*.

van den Beukel, A. P., van der Voort, M. C., & Eger, A. O. (2016). Supporting the changing driver's task: Exploration of interface designs for supervision and intervention in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 279–301.

Walker, G. H., Stanton, N. A., & Salmon, P. M. (2011). Cognitive compatibility of motorcyclists and car drivers. *Accident Analysis & Prevention, 43*(3), 878–888.

Wandtner, B. (2018). Non-driving related tasks in highly automated driving – Effects of task characteristics and drivers' self-regulation on take-over performance. *Inaugural-Dissertation, Julius-Maximilians-Universität Würzburg*.

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Secondary task engagement and disengagement in the context of highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*, 253–263.

Yang, Y., Götze, M., Laqua, A., Dominioni, G. C., Kawabe, K., & Bengler, K. (2017). A method to improve driver's situation awareness in automated driving. In *Proceedings of the human factors and ergonomics society* (pp. 29–47).

Yang, Y., Karakaya, B., Dominioni, G. C., Kawabe, K., & Bengler, K. (2018). An HMI concept to improve driver's visual behavior and situation awareness in automated vehicle. In *21st International conference on intelligent transportation systems* (pp. 650–655).

Yoon, D. D., Ayalew, B., & Ali, G. M. N. (2021). Performance of decentralized cooperative perception in v2v connected traffic. *IEEE Transactions on Intelligent Transportation Systems*.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software, 27*(8), 1–25.

Templeton, B. (2021, December 13). *Mercedes gets approval for traffic jam pilot, where is tesla?* Forbes, https://www.forbes.com/sites/bradtempleton/2021/12/13/mercedes-gets-approval-for-traffic-jam-pilot-where-is-tesla/?sh=8edcf955fc8e.