TUDelft

Delft University of Technology

Take over! A video-clip study measuring attention, situation awareness, and decision-making in the face of an impending hazard

Lu, Zhenji; Happee, Riender; de Winter, Joost C.F.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Take over! A video-clip study measuring attention, situation awareness, and decision-making in the face of an impending hazard

Zhenji Lu, Riender Happee, Joost C.F. de Winter *

Department of Cognitive Robotics, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, the Netherlands

ABSTRACT

In highly automated driving, drivers occasionally need to take over control of the car due to limitations of the automated driving system. Research has shown that visually distracted drivers need about 7 s to regain situation awareness (SA). However, it is unknown whether the presence of a hazard affects SA. In the present experiment, 32 participants watched animated video clips from a driver's perspective while their eyes were recorded using eye-tracking equipment. The videos had lengths between 1 and 20 s and contained either no hazard or an impending crash in the form of a stationary car in the ego lane. After each video, participants had to (1) decide (no need to take over, evade left, evade right, brake only), (2) rate the danger of the situation, (3) rebuild the situation from a top-down perspective, and (4) rate the difficulty of the rebuilding task. The results showed that the hazard situations were experienced as more dangerous than the non-hazard situations, as inferred from self-reported danger and pupil diameter. However, there were no major differences in SA: hazard and non-hazard situations yielded equivalent speed and distance errors in the rebuilding task and equivalent self-reported difficulty scores. An exception occurred for the shortest time budget (1 s) videos, where participants showed impaired SA in the hazard condition, presumably because the threat inhibited participants from looking into the rear-view mirror. Correlations between measures of SA and decision-making accuracy were low to moderate. It is concluded that hazards do not substantially affect the global awareness of the traffic situation, except for short time budgets.
© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Highly automated driving and take-overs

It is likely that future cars will be equipped with highly automated driving systems. These systems permit the driver to engage in non-driving tasks, yet the driver would still need to drive manually when the road environment or traffic situation becomes too complicated for the automated driving system to function correctly. An important question, therefore, is whether human drivers are able to reclaim control when the automation fails or exceeds its operational design domain.

Well over one-hundred driving simulator studies (for a meta-analysis, see Zhang, De Winter, Varotto, Happee, & Martens, 2019) and a relatively small number of on-road and test-track studies (Eriksson, Banks, & Stanton, 2017; Frison, Wintersberger, Schartmüller, & Riener, 2019; Josten, Zlocki, & Eckstein, 2016; Naujoks, Purucker, Wiedemann, & Marberger, 2019; Pfromm, 2016; Weinbeer et al., 2017), have examined the ability of human drivers to take over control from the automated vehicle for various types of traffic situations. In their meta-analysis of 129 experimental studies, Zhang et al. (2019) examined the determinants of take-over time, that is, how quickly after a take-over request the driver grabs the steering wheel or presses the brakes. This meta-analysis showed that take-over times depend on, among others, the urgency of the situation and the type of take-over request. Several simulator studies have found that, in urgent conditions, drivers show mean take-over times as low as about 1 s (e.g., Cohen-Lazry, Katzman, Borowsky, & Oron-Gilad, 2019; De Winter, Stanton, Price, & Mistry, 2016; Politis, Brewster, & Pollick, 2017).

1.2. Situation awareness during take-overs

Situation awareness (SA) in the Endsley (1995) model comprises perception, comprehension, and prediction, or as Adam (1993) put it, SA is "knowing what is going on so you can figure out what to do". According to Endsley, SA is a prerequisite for effective decision making. Equivalently, for a driver to succeed in the take-over task, a high level of SA would be necessary. Take-over quality, that is, whether the driver executes the take-over decision safely is a less studied topic compared to take-over time (but see e.g., Köhn, Gottlieb, Schermann, & Krcmar, 2019; Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014; Wiedemann et al., 2018). Although drivers can grab the steering wheel or press the brakes 1 s after a take-over request, it is debatable whether such a short time interval is sufficient for becoming situationally aware and achieving high take-over quality. Zeeb, Buchner, and Schrauf (2016) found that drivers achieved motor readiness (i.e., eyes-on-road, hands-on-wheel) quickly, but the cognitive processing of the traffic situation (i.e., system deactivation) and take-over quality (i.e., deviations from lane center) were impaired if the driver was performing a secondary task while the automation was active. In summary, good SA is an essential requirement for a high-quality take-over.

Several studies have previously examined the effect of time budget on driver SA and take-over quality. Samuel, Borowsky, Zilberstein, and Fisher (2016) examined SA in situations where the driver received a take-over request 4 s, 6 s, 8 s, or 12 s before the appearance of a hazard. Eye-tracking results showed that drivers were less likely to detect the hazard for 4-s time budgets times as compared to 8- and 12-s time budgets. Similar results were obtained by Vlakveld, Van Nes, De Bruin, Vissers, and Van der Kroft (2018). They found that drivers were more likely to detect hazards for 6-s time budgets as compared to 4-s time budgets. A mediating role of SA was suggested by Köhn et al. (2019). These authors found improved self-reported SA and faster take-over times with temporary interruptions of the non-driving task compared to without. Yang, Karakaya, Dominioni, Kawabe, and Bengler (2018) found that their LED strip concept, which informed drivers about automation functionality and potential hazards, improved gaze behavior (participants were more likely to look at the road when the automation was active) and take-over quality (participants were more likely to steer and less likely to slam the brakes as their first response). Lu, Coster, and De Winter (2017), using a method inspired by Gugerty (1997), had participants observe 1-s to 20-s video clips of traffic situations from the driver's perspective. After each clip, participants had to rebuild the traffic situation in a top-down view. Lu et al. (2017) concluded that drivers needed about 7 s to regain SA, defined as having knowledge about the number of cars and the locations of those cars; up to 20 s were needed for also being aware of the relative speeds of the cars on the road.

A limitation of the study by Lu et al. (2017) is that the situations in their video clips were not hazardous and that there was no incentive or a possibility for participants to take over control. The presence of a hazard gives 'meaning' to the situation, and participants may achieve high levels of SA if they have to take over control in a meaningful context. This hypothesis ties into theories of perceptual chunking: humans are known to be better able to process and memorize situations if they recognize semantic patterns (Egan & Schwartz, 1979). For example, in the domain of chess, expert chess players are better able to recall actual game positions as compared to random game positions (Gobet & Simon, 1996). In a similar vein, Banks, Plant, and Stanton (2018) explained, using schema theory, that SA is construed via goal-directed interactions with the world. According to this viewpoint, SA is updated in a perception–action cycle, in which unexpected events (e.g., a hazard) guide attention (Clark, Stanton, & Revell, 2020; Stanton, Salmon, Walker, Salas, & Hancock, 2017). On the other hand, the presence of a hazard can be expected to diminish global SA, because the hazard (e.g., a car standing still on the ego-lane) attracts attention, as a consequence of which the driver may fail to perceive other vehicles on the road. This phenomenon is analogous to the 'weapon focus' effect (Loftus, Loftus, & Messo, 1987).

1.3. Study aims

This study aimed to examine participants' SA levels in non-hazard situations, in which there was no need to take over, versus hazard situations, in which there was a stationary car in the ego-lane. Participants watched animated video clips and after each clip, rebuilt the situation in a top-down view. We used non-interactive videos to ensure that each person experienced exactly the same visual stimuli. Furthermore, an eye-tracker was used to measure the extent to which the hazard in the ego-lane distracted from other cars on the road.

SA was assessed by comparing the rebuilt situation with the actual situation as it occurred at the end of the video clip. Furthermore, before asking drivers to rebuild the situation, we required drivers to make a decision (no need to take over, evade

left, evade right, or brake). It may be argued that the association between decision-making and performance in the rebuilding task is small because the rebuilding task takes into consideration all cars on the road (global SA) rather than the task-relevant cars (local SA). Accordingly, this study also used an index of local SA, which takes into account nearby cars only.

Finally, it may be argued that in actual highly automated driving, drivers will not passively watch a hazardous scene unfold. Instead, drivers will take over from the automation, evade the object, and accordingly mitigate the hazard. In order to account for this, we included a second part of the experiment, in which participants could intervene once they had decided about how to take over control from the automated car.

2. Methods

2.1. Participants

Forty people participated in the experiment as part of an MSc-level course at the Delft University of Technology. Students could also perform alternative assignments. The use of students is common in automated driving research (De Winter, Happee, Martens, & Stanton, 2014). In a meta-analysis of take-over experiments, Zhang et al. (2019) found that as much as 46% of 119 experiments in which age data were available featured a mean participant age equal to or below 30 years. Although students usually have a driving license, they tend not to drive daily because they favor public transport or cycling. For this reason, special attention must be paid to data quality control.

In our study, three participants who indicated not having a driving license were removed from the analysis. Furthermore, we decided to exclude three participants who did not adhere to the instructions (e.g., not following instructions, confusing 'take over' for 'overtake', assuming that the ego-car might brake automatically), one participant who performed poorly (sometimes placing zero cars in front of the car), and one participant who disclosed that her driving experience amounted to a 10-min driving exam only. Accordingly, 32 participants were retained in the analysis. Eye-tracking data from 1 out of the 32 participants were discarded due to a calibration problem.

The participants were 29 males and 3 females, aged between 22 and 29 years (M = 24.2, SD = 1.8). Participants held a driving license for an average of 5.5 years (SD = 2.3). In the last 12 months, three participants drove every day, five drove 4–6 days a week, ten drove 1–3 days per week, six drove between once a week and once a month, three drove less than once a month, and five did not drive. About half of the participants were Dutch and the other half were international. The international participants were mostly from China and India.

All participants provided written informed consent. The research was approved by the TU Delft Human Research Ethics Committee.

2.2. Apparatus

The videos were presented on a 24-inch monitor with a display area of 531 × 298 mm. An SR Research EyeLink 1000 Plus eye tracker was used to record monocular eye movements at 2000 Hz. The monitor was positioned 95 cm in front of the participant and 35 cm behind the eye-tracking camera and IR light source. The distance between the table surface and the lower edge of the display area was 20 cm. The horizontal and vertical viewing angles were approximately 31° and 18°, respectively. The room lights were turned on, and window blinders were put down when the experiment started.

Driving videos with animated traffic were generated with Prescan 8.0.0 (PreScan, 2017). The videos were shown for 1, 3, 6, 9, 12, or 20 s at a frame rate of 20 Hz. The videos had a resolution of 1920 × 1080 pixels. Rear-view, left, and right mirrors were integrated into the video (see Fig. 1).

2.3. Animated video clips

All videos showed a three-lane highway, with the ego-car always driving in the middle lane at 100 km/h. The videos were either 1-, 3-, 6-, 9-, 12-, or 20-s long. In the 1-, 3-, 6-, 9-, and 12-s videos, five surrounding cars were included. In a previous study (Lu et al., 2017), it was found that participants can accurately estimate the total number of cars with up to 6 cars present when the videos are 20-s long. In the present study, the 20-s videos contained six surrounding cars, to avoid creating the impression that all videos included the same number of cars.

The three lanes each contained one or two surrounding cars. Two to four cars drove in front of the ego-car. The farthest car was 50–80 m away from the ego-car. All surrounding cars drove at one of three constant speeds: 80, 100, or 120 km/h. The three lanes could host cars of all three speeds (80, 100, 120 km/h). Between 0 and 3 cars drove 80 km/h, between 1 and 3 cars drove 100 km/h, and between 1 and 3 cars drove 120 km/h. A 20 km/h speed difference with respect to the ego-car was regarded as representative of the speed variation one might encounter on real roads. None of the cars changed lanes. Car models and colors were randomly selected from 13 colors and 10 models.

Note that because cars driving 120 km/h could drive on all lanes, another car could start to overtake the participant on the left lane or the right lane. In this way, the design aimed to eliminate the participants' expectancies regarding overtaking country-specific regulations. For example, in the Netherlands, the overtaking lane is the left lane, whereas in India it is the right lane. Furthermore, in some countries, undertaking (i.e., overtaking on the inside) is prohibited whereas, in other countries, this is permitted.

Fig. 1. The driving environment. In this hazard video, a car blocks the right lane. There is a stationary car in the ego-lane. Hence, the correct decision is to evade to the left. In total, there are five surrounding cars.

Twelve non-hazard videos were created. In the non-hazard videos, all cars drove at a constant speed, and there was no need to take over. A mirroring technique was used: First, six situations were created, and these were left–right mirrored to create six more non-hazard situations. The non-hazard videos had lengths of 1, 3, 6, 9, 12, and 20 s.

In addition to the non-hazard situations, we created 16 videos in which there was a need to take over and avoid a collision. These hazard situations had lengths of 1, 3, 6, and 9 s. In each hazard video, the hazard was a car in front that decelerated at 5 m/s^2 from the start of the video. Note that this type of hazard would be impossible in 12-s and 20-s videos because of dynamic constraints. The hazardous car stood still during the last second of the video, to ensure the same visual looming effects for each video. At the end of the hazard video, the distance between the ego-car and the hazardous car was 19–22 m. At a speed of the ego-car of 100 km/h, this means that a collision was unavoidable by means of braking. Cars in the adjacent left or right lane were slowing down or speeding up. In this way, the cars in the left or right lane could block an evasive maneuver. The hazard situations and corresponding correct responses were as follows:

- evade (lane change) to the *right* to avoid a collision with the front and left car (4 videos),
- evade (lane change) to the *left* to avoid a collision with the front and right car (4 videos). These situations were left–right mirrored with respect to the 'evade right' situations,
- evade (lane change) to the *left* or *right* to avoid a collision with the front car (4 videos), or
- *braking* without steering to minimize the impact of an unavoidable collision with the front car, and avoid a side collision with the left and right car (4 videos).

Additionally, five training videos were created. The training videos had a length of 1, 3, 6, 9, or 12 s, and contained four to six surrounding cars with no hazard. In summary, a total of 33 videos were created: 12 non-hazard videos, 16 hazard videos, and 5 training videos.

2.4. Experimental design

The experiment was of a within-subject design in which all participants viewed the same video clips in a random order that was different for each participant. The experiment consisted of two parts. In Part 1, participants viewed the video clips from beginning to end. In Part 2, participants viewed the video clips and could press the spacebar to stop the video as soon as they had decided how to take over control. Part 2 aimed to examine whether the results obtained in Part 1 generalize towards more naturalistic conditions in which participants can decide to take over at any moment they wanted. In total, each participant completed 50 trials: 5 training trials before Part 1, 28 experimental trials in Part 1, 3 training trials before Part 2, and 14 experimental trials in Part 2. The duration of the experiment per participant was approximately 60 min.

2.4.1. Experimental procedure Part 1

The participants read and signed an informed consent form and completed a brief questionnaire about age, gender, and driving experience. They adjusted the height of their chair and head support to find a comfortable seating position. Participants had their head in the head support during the entire experiment, as required for the Eyelink eye tracker.

Participants received task instructions on both the consent form and the monitor. They were informed that they would be watching short driving videos (between 1 s and 20 s) from a driver's perspective and were asked to imagine that they were the driver of an automated car. It was stated that after each video, they had to indicate their control strategy to avoid a collision, to rate the danger of the situation, to rebuild the scene by placing cars based on what could be remembered from the last frame of the video, and to rate the difficulty of the car-placement task. It was also mentioned that all cars would continue to drive without changing speed or lane.

First, a standard nine-point calibration procedure of the Eyelink eye tracker was performed. The experiment started with five training videos. Next, participants completed the two parts of the Experiment. In Part 1, participants viewed the 28 video clips in random order. Details about the randomization process are provided in Table A1.

After each video, participants completed four tasks in the following order:

(1) *Decide:* participants had to select one of four response options for taking over control from the automated car. The options were 'Evade left', 'Evade right', 'Brake only', and 'No need to take over', provided in an isometric layout (top, left, right, bottom, respectively). Participants had 5 s to respond.

(2) *Rate Danger:* Participants rated the danger of the situation using a standard Likert item "The situation was dangerous" with anchors at Completely disagree (0) to Completely agree (10). Participants had 10 s to respond. Note that the word 'dangerous' instead of 'hazardous' was used in this question. The reason is that the word 'dangerous', which means 'power to cause harm', better reflects how the participants' themselves experienced the situation. That is, the other cars in the environment are hazards, which may or may not cause a dangerous situation for the participant.

(3) *Rebuild Situation:* Participants rebuilt the situation by placing cars around the ego-car based on the last moment of the previous video, see Fig. 2. Participants could pick and place cars, having different indications of speed (faster than the ego-car, equal speed as the ego-car, slower than the ego-car). In the training trials, participants were shown the correct answer after they had rebuilt the situation.

(4) *Rate Difficulty:* Participants responded to the Likert item "The rebuilding task was difficult", from Completely disagree (0) to Completely agree (10). The participants had 10 s to respond.

2.4.2. Experimental procedure Part 2

Participants were instructed on the monitor to imagine that they were the driver of an automated car and to stop the video by pressing the spacebar as soon as they had made a decision about how to take over control by evading left, evading
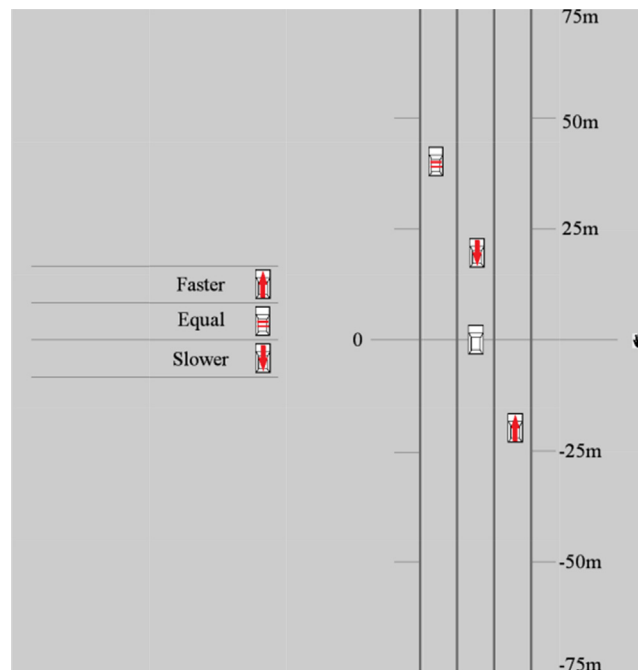

Fig. 2. Interface for rebuilding the situation (used in Part 1 of the Experiment).

right, or braking. In Part 2 of the experiment, participants viewed 14 videos they had also seen in Part 1, but in a new random order (see Appendix A for details). First, participants completed three training trials. Next, the participants viewed 14 videos in which they could stop the video by pressing the spacebar. If they pressed the spacebar, a decision-making interface would appear with the response options 'Evade left', 'Evade right', 'Brake only'. (The option 'No need to take over' was not offered because participants' had already pressed the spacebar to indicate they wanted to take over). If participants did not press the spacebar, they would finish watching the full video and the decision-making interface would appear with all four response options: 'Evade left', 'Evade right', 'Brake only', and 'No need to take over'. Participants had 5 s to select a response using the interface. After this, participants rated the danger of the situation. There was no rebuilding task in Part 2 of the experiment.

2.5. Dependent measures

The following dependent measures were calculated for each trial.

- *Decision accuracy* (0 or 1). Whether the decision was correct (1) or incorrect (0). For the non-hazard situations, the correct decision was 'No Need to Take Over'. For the hazard situations, the correct decision was 'Evade left', 'Evade right', or 'Brake only', as explained above. If the participant did not provide a response within the 5-s time budget, the response was regarded as incorrect. In Part 1 of the experiment, this occurred in 21 out of 896 trials (32 participants × 28 trials per participant). In Part 2 of the experiment, this occurred in 4 out of 448 trials (32 participants × 14 trials per participant).
- *Decision time* (s). How long participants took to select a decision using the interface. The maximum decision time was 5 s; if no decision was made, a decision time of 5 s was imputed.
- *Absolute error in the number of placed cars* (#). The absolute error of the number of cars placed with respect to the true number of surrounding cars. For example, if the participant had placed 3 cars while there were 5 cars in the video (which was the case for the 1-, 3-, 6-, 9-, and 12-s videos), the absolute error in the number of placed cars would be 2.
- *Total distance error* (%). The distance between the placed cars and true cars summed over all matched cars. The score was normalized to a scale from 0% (perfect placement) to 100% (very poor placement). More specifically, a score of 0% would correspond to placing all cars exactly on top of the true cars. A score of 100% would correspond to a distance error equal to the sum of distances between the true cars and the ego-car. In other words, a score of 100% would be obtained if the participant 'mindlessly' placed all cars on top of the ego-car.
- *Total speed error* (–). The absolute speed error between the speeds of the placed cars and the speeds of the true cars, summed over all matched cars. If the participant reported the same speed as the true speed of the car, the error for that matched pair would be 0. If the participant reported 'equal' while the car was driving faster or slower than the ego-car or if the participant reported 'slower' or 'faster' while the car was driving at a speed that was equal to the speed of the ego-car, then the error for that matched pair would be 1. Finally, if the participant reported 'faster' or 'slower' while the car was actually driving slower or faster respectively, then the error for that matched pair would be 2. Accordingly, a total speed error of 0 represents a perfect speed estimate, where the participant estimated the relative speed of all matched cars correctly. A score of 10 is the worst score that could be attained. Such a score would hypothetically occur if the participant reported an incorrect speed (i.e., reporting 'slower' where 'faster' would be correct or reporting 'faster' where 'slower' would be correct) for all 5 cars.
- *Self-reported difficulty* (%). The subjective rating of the difficulty of the rebuilding task. The response on the scale from 0 to 10 was converted to a percentage, where 0% represents 'completely disagree' and 100% represents 'completely agree'.
- *Self-reported danger* (%). The subjective rating of the danger of the situation. The response on the scale from 0 to 10 was converted to a percentage, where 0% represents 'completely disagree' and 100% represents 'completely agree'.
- *Braking* (%). The percentage of trials in which the participant decided to brake. This measure was included as braking when there was no need to brake was thought to reflect poor SA (Yang et al., 2018).
- *Take-over* (0 or 1). Whether the participant pressed the spacebar (1) or not (0) while viewing the video (only in Part 2 of the experiment). Note that in Part 2, participants were asked to press the spacebar as soon as they had made a decision about how to take over control by evading left, evading right, or braking.

Before computing the 'Total distance error' and 'Total speed error', it was necessary to match the placed cars with the true cars. For this purpose, we used a matching algorithm, previously described by Lu et al. (2017). The matching algorithm paired all possible combinations of placed cars and true cars. From all these combinations, the algorithm selected the combination with the lowest total distance error; this combination was the 'match'.

The above measures 'Absolute error in the number of placed cars', 'Total distance error', and 'Total speed error' can be seen as indexes of global SA because these measures consider all cars regardless of their relevance. In addition to global SA, we assessed local SA by examining whether the participant correctly placed a car in the left or right lane within an absolute longitudinal distance of 15 m from the ego-car. A high local SA is obtained if the decision of the participant (e.g., 'Evade left') corresponds with the placed car (e.g., car in the right lane).

Furthermore, as a validation check of whether the hazards were indeed experienced as hazardous, we assessed participants' pupil diameter. Pupil diameter is an often-used index of stress, emotional arousal, and cognitive load (e.g., Bradley, Miccoli, Escrig, & Lang, 2008; Marquart & De Winter, 2015; Pedrotti et al., 2014). Finally, we assessed participants' visual attention distribution as a function of the elapsed time of the video clip, to investigate the weapon focus effect. The weapon

focus effect is operationalized as a situation where participants look at the hazard (i.e., the stationary car in the ego-lane) while ignoring other parts of the scene.

2.6. Statistical analyses

We compared the mean scores on the dependent measures between the hazard and non-hazard situations. Paired-samples t-tests were performed between these two conditions, for 1-s, 3-s, 6-s, and 9-s videos separately. Note that for each video length, there were four hazard videos and two non-hazard videos. The scores were averaged per video length. By virtue of the central limit theorem, we felt justified in assuming a normal distribution of the difference scores between the non-hazard and hazard videos. p-values smaller than 0.005 were deemed statistically significant (Benjamin et al., 2018). The within-subjects Cohen's d_z was used as an effect size measure. At a sample size of 32 participants, an absolute d_z value of 0.535 or higher is statistically significant, $p < .005$.

3. Results

3.1. Effects of hazard presence (Part 1 of the experiment)

Participants' decisions were more accurate for non-hazard situations compared to hazard situations. In other words, participants were better at recognizing that there was no need to take-over than at indicating the correct take-over action (evade left, evade right, brake only) in case of a hazard. For 1-s videos, however, the difference between the two conditions was not statistically significant (Fig. 3A).

An often-made error was pressing the brakes when there was no need to. Pressing the brakes was the correct answer for 0% of the non-hazard situations, but was found in 35% of responses. Pressing the brakes was the correct answer for 25% of the hazard situations, but was found in up to 42% of the responses (Fig. 3H). Pressing the brakes may be a logical precautionary response for participants to minimize collision risk when being uncertain about the situation.

Participants' decision times for hazard and non-hazard situations were similar (Fig. 3B). The exception was 1-s videos, where participants took significantly more time in non-hazard situations compared to hazard situations (Fig. 3B).

Rebuilding performance improved with video length, which replicates the findings by Lu et al. (2017), see Fig. 3C–E. There were no statistically significant differences in rebuilding performance between hazard and non-hazard situations (Fig. 3C–E),
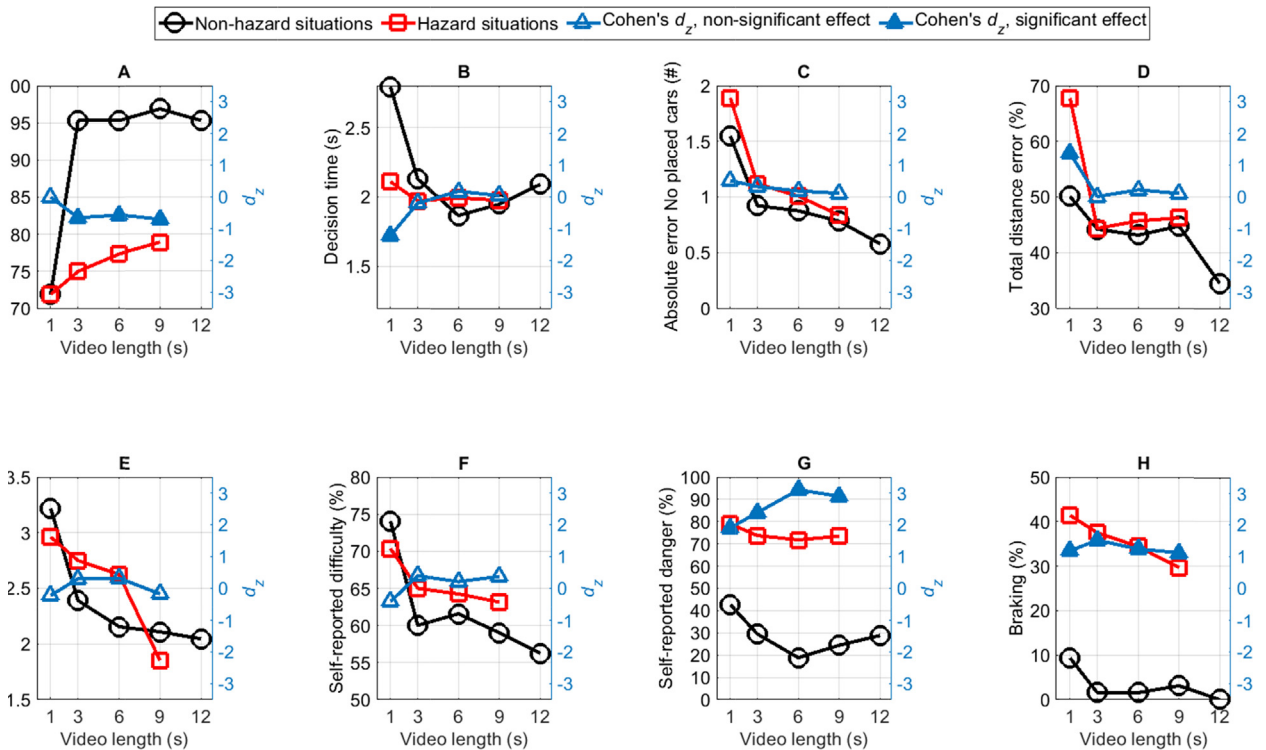

Fig. 3. Results for Part 1 of the experiment. Absolute d_z values of 0.54 and higher are statistically significant, $p < .005$, and are marked as solid instead of open triangular markers. A positive d_z means that the value is higher for hazard situations than for non-hazard situations.

except for the total distance error for 1-s videos, which was significantly higher for hazard than for non-hazard situations (Fig. 3D). Participants found the rebuilding of hazard and non-hazard situations equally difficult (Fig. 3F). However, the hazard situations were perceived as much more dangerous than non-hazard situations (Fig. 3G).

Appendix A provides the results of repeated-measures ANOVAs for testing the effect of time budget (i.e., video length) per dependent measure.

3.2. Pupil diameter (Part 1 of the experiment)

Pupil diameter was compared between hazard and non-hazard situations (Fig. 4). The higher pupil diameter with hazard near the end is consistent with the notion that the hazard induced stress or emotional arousal.

3.3. Attention distribution (Part 1 of the experiment)

We examined participants' attention distribution towards the rear-view mirror. The findings, shown in Fig. 5, are consistent with the weapon focus effect, especially for 1-s videos. That is, the hazard in front appeared to attract attention at the expense of attention towards the rear-view mirror. More specifically, at the end of the 1-s hazard videos, only 10% of the participants looked into the rear-view mirror, compared to 48% for non-hazard videos.

3.4. Predicting decision accuracy from global situation awareness (Part 1 of the experiment)

We examined Spearman rank-order correlations between SA and decision accuracy (Table 1). Participants with better decision accuracy (variable 3) did not have better SA regarding the number of placed cars (variable 5) and the total distance error (variable 6). However, there was a substantial correlation between decision accuracy and total speed error (ρ = -0.43).

The correlation matrix also showed that participants who drove more frequently (variable 2) had somewhat higher decision accuracy (variable 3) and better SA scores (variables 5–7). These correlations are consistent with Lu et al. (2017), but mostly too small to be statistically significant.

3.5. Predicting decision accuracy from local situation awareness (Part 1 of the experiment)

Table 2 shows all combinations of 'correct decision' and 'decision made' by the participants and the percentage of cases where the participant placed a car in the left or right lane. It can be seen that decision accuracy was high. To illustrate, in the
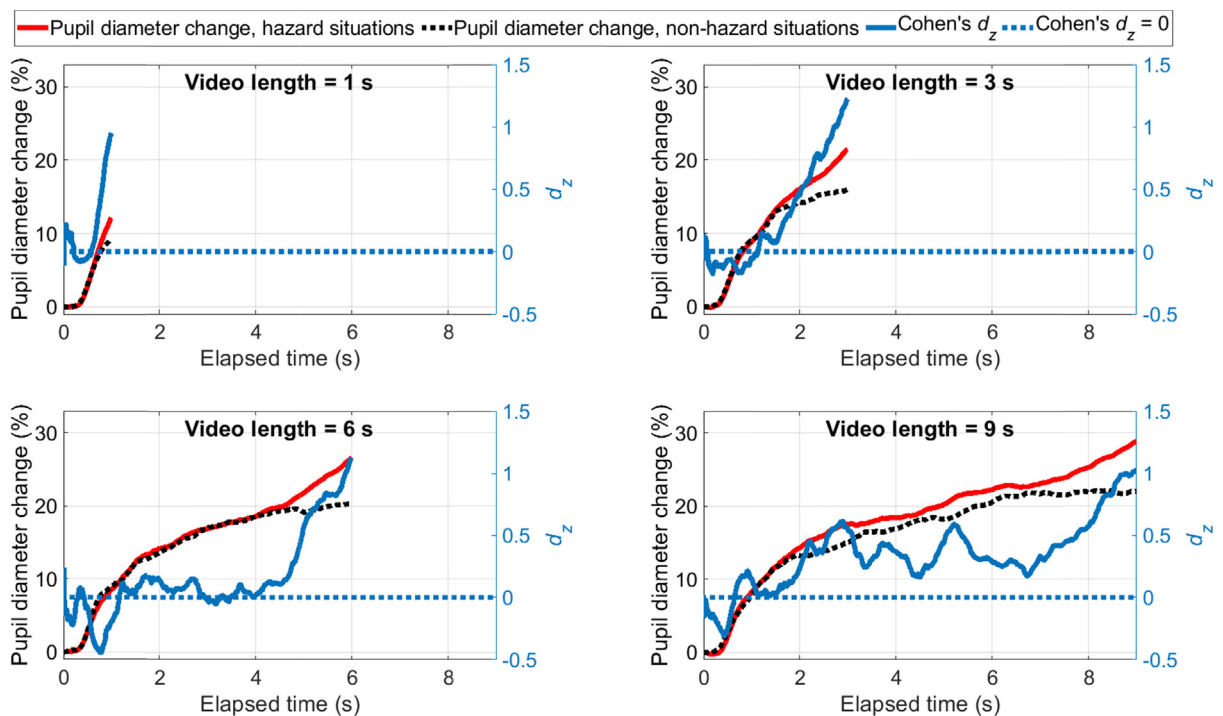

Fig. 4. Pupil diameter change for the 1-, 3-, 6-, and 9-s videos with respect the pupil diameter at an elapsed time of 0 s (Part 1 of the experiment). A positive d_z means that the value is higher for hazard situations than for non-hazard situations. At a sample size of 31 participants, an absolute d_z value of 0.545 or higher is statistically significant, $p < .005$.
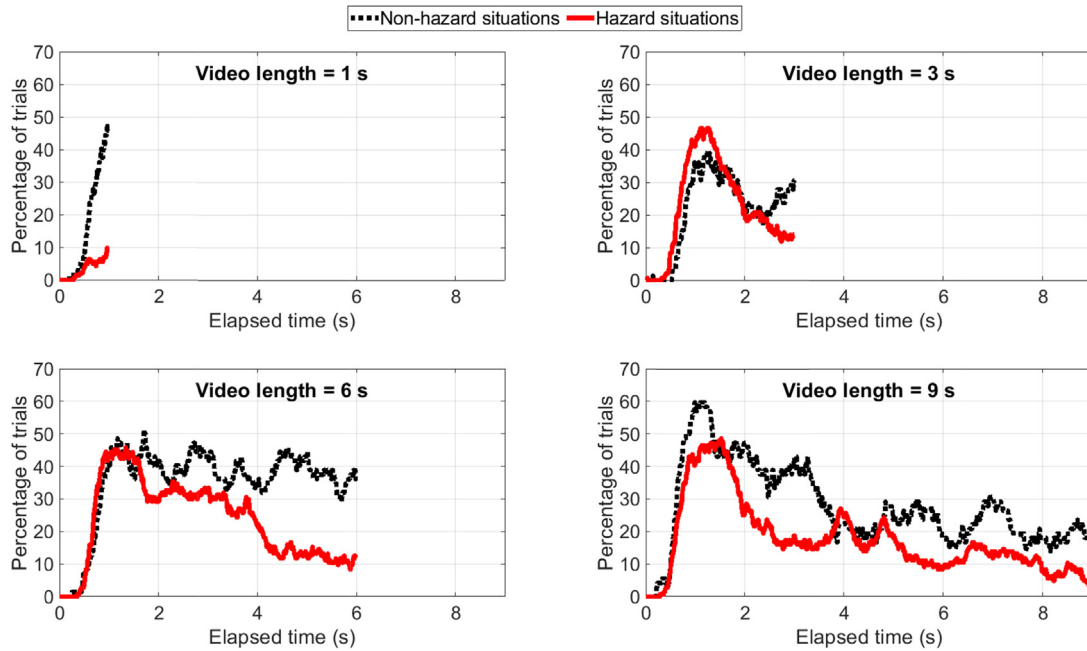
Fig. 5. Percentage of trials in which the participant looks into the rear-view mirror for 1-, 3-, 6-, and 9-s videos (Part 1 of the experiment).

Table 1
Spearman rank-order correlations matrix between personal characteristics and dependent measures (N = 32) for Part 1 of the experiment.

	Mean (SD)	1	2	3	4	5	6	7	8	9
1. Years of licensure	5.47 (2.29)									
2. Driving frequency (1 = never, 6 = every day)	3.50 (1.52)	0.38								
3. Decision accuracy (%)	82.14 (9.77)	−0.07	0.22							
4. Decision time (s)	2.06 (0.43)	0.07	−0.08	−0.34						
5. Absolute error in the number of placed cars (#)	1.10 (0.41)	0.10	−0.14	0.01	**0.47**					
6. Total distance error (%)	47.31 (9.31)	0.04	−0.31	−0.07	0.34	**0.75***				
7. Total speed error (-)	2.49 (0.61)	0.06	**−0.40**	**−0.43**	0.37	**0.58***	**0.54***			
8. Self-reported difficulty (%)	63.59 (15.45)	0.05	0.19	−0.19	0.16	0.04	−0.10	0.07		
9. Self-reported danger (%)	54.28 (10.75)	0.02	−0.08	−0.25	0.02	−0.12	−0.20	0.05	**0.43**	
10. Braking (%)	21.65 (10.22)	0.01	−0.19	**−0.62***	0.28	0.19	0.32	**0.39**	0.14	0.23

Note. Correlations that are significantly different from zero are marked in boldface for *p* < .05 and with an asterisk for *p* < .005. The scores for variables 3–10 were based on the average of 28 situations per participant.

'Evade left' scenario, participants correctly indicated 'Evade left' in 99 cases, while the incorrect answer 'Evade right' was selected 0 times. The imperfect accuracy in the 'Evade left' scenario was mostly caused by participants who selected 'Brake only' (23 cases); braking may be regarded as an ineffective yet reasonable response because braking reduces the impact of a collision.

Table 2 further shows that there was no straightforward association between local SA and decision accuracy. For example, for 'Evade right' situations, participants mostly made the correct decision (90 times 'Evade right' vs. 27 times 'Brake only', 5 times 'No need to take over', and 4 times 'Evade left'). Furthermore, participants who made the correct decision ('Evade right') afterward correctly recalled that a car was blocking the left lane (69% of 90 cases), indicating high local SA. However, this still this means that 31% of participants did not report a car in the left lane, despite making the correct decision. If participants made a wrong decision in the 'Evade right' scenario, they still tended to correctly report that there was a car in the left lane (80%, 75%, and 74% for 'No need to take over', 'Evade left', and 'Brake only' decisions, respectively). For the 'Brake only' scenarios, in which a car was present in both the left and right lanes, participants usually made the correct decision (107 of 125 cases). Even though participants made the correct decision, they did not place a car in the left and right lanes in 24% and 19% of the cases, respectively. These findings point to a disconnect between local SA and decision-making.

3.6. Effect of hazard presence (Part 2 of the experiment)

In Part 2 of the experiment, participants' decisions were generally more accurate for non-hazard situations compared to hazard situations (Fig. 6A). However, these effects were not statistically significant and smaller compared to Part 1 of the experiment (Figs. 6A vs. 3A, respectively).

Table 2
Overview of decision accuracy (correct decision vs. decision made by the participant) and measures of local situation awareness (whether the participant placed a car nearby in the left or right lane) for Part 1 of the experiment.

Correct decision in the situation	Decision made by the participant	#	Car placed nearby in the left lane (%)	Car placed nearby in the right lane (%)
No need to take over	*No need to take over*	348	39	31
	Evade left	10	10	30
	Evade right	8	50	0
	Brake only	11	45	36
Evade left	No need to take over	3	67	100
	Evade left	99	5	52
	Evade right	0	–	–
	Brake only	23	35	43
Evade right	No need to take over	5	80	0
	Evade left	4	75	0
	Evade right	90	69	6
	Brake only	27	74	19
Evade left or right	No need to take over	4	0	25
	Evade left	33	15	30
	Evade right	59	41	7
	Brake only	26	35	23
Brake only	No need to take over	6	83	50
	Evade left	6	50	83
	Evade right	6	33	17
	Brake only	107	76	81
Any	No decision provided	21		
	Total	**896**		

Note. Correct decisions are marked in Italics. The total number of situations equals 896 (32 participants × 28 videos).
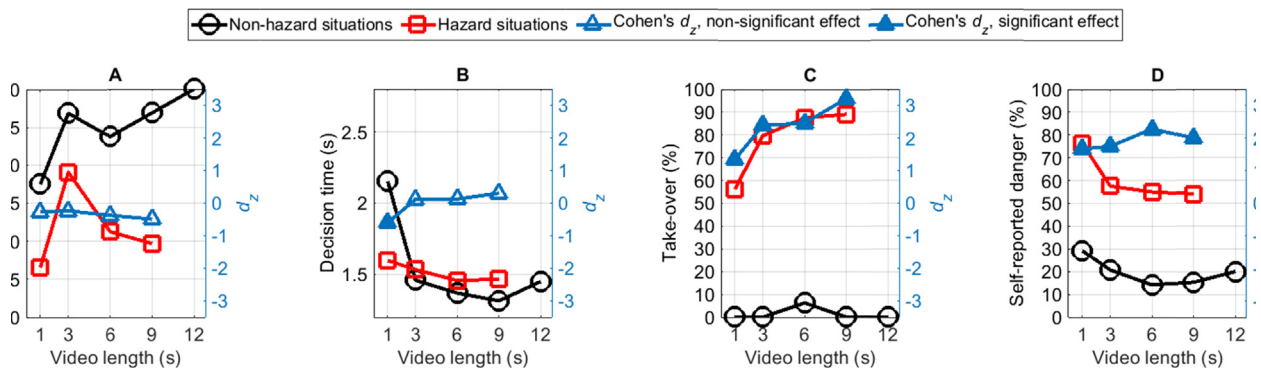

Fig. 6. Results for Part 2 of the experiment. Absolute d_z values of 0.54 and higher are statistically significant, $p < .005$, and are marked as solid instead of open triangular markers. A positive d_z means that the value is higher for hazard situations than for non-hazard situations.

Participants' decision time was faster in Part 2 of the experiment (Fig. 6B) compared to Part 1 (Fig. 3B), a difference that could be due to learning. The difference between the hazard situations and non-hazard situations followed the same pattern as in Part 1, with a faster decision time for hazard situations only for 1-s videos (Fig. 6B).

Again, the hazard situations were regarded as more dangerous than the non-hazard situations (Fig. 6D). More specifically, participants found the hazard situations moderately dangerous (50% to 60% for 3–9 s videos, Fig. 6D). In comparison, in Part 1 of the experiment, the situations were found to be highly dangerous (70% to 80% for 3–9 s videos, Fig. 3G). The difference in self-reported danger with Part 1 could be because in Part 2 participants often took over (i.e., pressed the spacebar) before the video ended, especially when there was more time to do so (Fig. 6C).

4. Discussion

4.1. Main results: Effect of the presence of a hazard on situation awareness

This study compared participants' SA levels between hazard and non-hazard take-over situations for different time budgets. Our results replicate earlier research by showing that SA, operationalized as situation-rebuilding performance,

improves with increasing time budget. We showed that the effect of the hazard compared to no hazard is large for experienced danger (d_z = 2 to 3 for self-reported danger, d_z = 1 for pupil diameter) but small for SA ($|d_z| < 0.5$). In other words, we observed no major differences in SA between hazard and non-hazard situations. An exception concerned situations of 1-s time budget, where participants showed a larger total distance error in hazard situations compared to non-hazard situations (Fig. 3D). The poor SA for 1-s hazard situations seems to be due to the weapon focus effect, where participants focused on the hazard while not looking at other cars such as cars in the rear-view mirror (Fig. 5, left top).

4.2. Effect of time budget on decision accuracy

For time budgets between 3 and 9 s, the decision accuracy was about 95% for non-hazard situations and only 75–80% for hazard situations. This difference can be explained by the fact that decision-making in hazard situations consists of two components: (1) recognizing that one has to take over (which is easy because of the salient stationary car), and (2) selecting the right decision among the three remaining options (Evade left, Evade right, Brake only). Selecting the correct option among three options is more difficult than selecting 'No need to take over' in non-hazard situations, which explains the low accuracy for hazard situations. For the 1-s time budget, decision accuracy was poor for both non-hazard and hazard situations. An increase in time budget did not offer much advantage for reaching a more correct decision in hazard situations, presumably because the hazards (i.e., cars in the left or right lane blocking a particular avoidance maneuver) became apparent only at the end of the video.

4.3. Validity of our method for measuring situation awareness

Our rebuilding task was inspired by early work of Gugerty (1997) and resembles the popular Situation Awareness Global Assessment Technique, SAGAT (Endsley, 1988). In short, in the SAGAT method, the simulation is suddenly interrupted and the participant has to answer queries about the state of the simulation, such as the positions and speeds of objects in the virtual world. The accuracy of the answers to these queries provides an indication of the participant's level of SA. The SAGAT is similar to our method, in which SA was derived from how accurately participants positioned the cars in the top-down view. The SAGAT has been criticized for measuring memory skills rather than SA (De Winter, Eisma, Cabrall, Hancock, & Stanton, 2019; Durso, Bleckley, & Dattel, 2006; Gugerty, 1998). It is also possible that the participants found our SAGAT-like method difficult because it required mental rotation; the participants had experienced the situation from an egocentric perspective, but had to position the cars from a top-down perspective.

It has been recommended that SA should not be measured via freeze-probe methods such as the SAGAT but by real-time probe or think-aloud methods instead (Jones & Endsley, 2004; Salmon, Stanton, & Young, 2012). Although this recommendation has merit, the advantage of our method is that we separated the effect of stimulus presentation (i.e., showing the video) from the measurement of decision-making and SA. Numerous take-over studies are already available in the literature (Zhang et al., 2019), and in each of these studies, participants took over control through steering or braking, which means that the situation changes accordingly, which in turn, inhibits a controlled comparison of how SA is affected by the presence of a hazard. This issue was corroborated in Part 2 of the experiment, where participants could take over when they felt this was needed. More specifically, the results for Part 2 showed that drivers adapted to the situation, in the sense that they anticipated the collision and moderated their task demands by preventing the hazard from coming close.

A limitation of our method is that the driving task was simple. Participants did not have to consider factors such as the heading of the cars (i.e., all cars drive in the center of their lane), interactions between vehicles, or distractions such as roadside objects or non-driving tasks. Furthermore, it may be argued that our study focused on level-1 SA (perception) and not as much on level-2 (comprehension) or level-3 SA (prediction). This concern, however, can be countered by the notion that participants were required to estimate the speed of other cars. By definition, the prediction of how a situation will evolve will have to take into consideration relative distance and relative speed (for a similar discussion, see Lu et al., 2017). Nonetheless, our study did not require participants to predict how covert hazards (also called latent or invisible hazards; cf. Vlakveld et al., 2018) or precursor events (cf. Crundall et al., 2012; Kahana-Levy, Shavitzky-Golkin, Borowsky, & Vakil, 2019) develop and materialize into an actual hazard. Instead, the hazard in our study was highly salient in the form of a stationary car in the ego-lane.

4.4. Eye-tracking results

We used eye-tracking to measure participants' attention distribution while performing the task. The obtained results are consistent with, and a refinement of, Lu et al. (2017), who used a relatively low-quality eye tracker that yielded relatively large amounts of missing data. The general pattern is that participants are initially likely to look into the rear-view mirror: around 1 s since the start of the video, participants looked into the rear-view mirror in about 50% of the cases. This value decreased with viewing time. These results can be interpreted as indicating that participants first try to obtain a general overview of the situation by locating where objects are, after which they allocate attention to how the situation unfolds in front of them. It is noted that in automated driving, drivers can relatively easily allocate their attention to other cars

and the mirrors. In comparison, in manual driving drivers have to monitor the position of their car in the lane in order to steer their car (e.g., Navarro, Osiurak, Ovigue, Charrier, & Reynaud, 2019).

Future research could attempt to measure SA in real-time using eye-tracking equipment (Moore & Gugerty, 2010). Having real-time knowledge of the driver's SA is attractive for creating automation systems that adapt to the driver's state. However, there are several caveats with such an approach. First, false positives and misses should be expected: a real-time prediction of SA will not have perfect predictive validity for task performance (De Winter et al., 2019). Second, driving situations are dynamic, and surrounding cars subtend different viewing angles: Cars driving close by will absorb a large part of the driver's field of view, whereas cars driving far ahead will take up a small region. Our preliminary attempts to make a real-time index of SA stranded because our algorithms were unable to classify whether the participant was looking at the car in the left, right, or middle lane in case multiple cars drove far ahead of the ego-car, even though we used an accurate eye tracker and head support. Accordingly, it may be more fruitful to examine the driver's attention distribution based on how frequently the driver glances to large areas of interest, such as whether or not the driver looks at the road ahead (Cabrall, Janssen, & De Winter, 2018).

In addition to measuring eye-gaze direction, we used the eye tracker to measure pupil diameter. Our findings showed an elevated pupil diameter for hazard situations compared to non-hazard situations, near the end of the video. These observations are consistent with the notion that the impending hazard causes a stress response among subjects. It is noted, however, that pupil diameter is sensitive to light (e.g., bright and dark regions on the monitor). Hence, we cannot conclusively state that arousal is the sole cause of the results shown in Fig. 4. Because of the sensitivity of pupil diameter to confounding variables such as the light reflex, we expect that pupillometry would not be fruitful for real-time applications in cars.

4.5. Association between situation awareness and decision accuracy

We observed weak-to-moderate associations between global (Table 1) or local (Table 2) SA and decision accuracy. For global SA (i.e., error in the number of placed cars, distance error, speed error), the weak-to-moderate association can be explained by the fact that global SA is unneeded for avoiding a collision: cars far behind or far ahead are irrelevant at safety–critical moments.

However, even for our measure of local SA, participants often made incongruent decisions. For example, a large portion of participants evaded to the left, but did not place a car close by in the right lane (Table 2). There are several explanations for this incongruence. First, participants had to decide under time pressure, which may have contributed to instinctive decision-making outside of conscious recall. They may have recognized the hazard in the adjacent lane but unsure about where to place it in the top-down view. Second, participants may have acted based on previously learned habits such as to overtake via the left in accordance with traffic laws.

In an early experiment using a similar method, Gugerty (1997) reported that "the direct and indirect measures were positively correlated (associated), suggesting that drivers' knowledge of nearby cars is largely explicit with little contribution of implicit knowledge." (p. 42). A closer inspection of Gugerty's results showed that correlations between explicit SA (rebuilding task performance) and implicit SA (hazard detection, blocking car detection, crash avoidance) were of only moderate strength (between −0.34 and −0.53, depending on whether all cars or only nearby cars were included), despite aggregating over 80 trials per participant. Hence, our results are actually in line with those of Gugerty. In summary, an important take-home message is that knowledge of where other cars are (explicit SA) only moderately related to what drivers do in a given emergency situation (implicit SA). Although explicit SA (obtained via a SAGAT method) is known to be a sensitive dependent variable in experimental studies (Endsley, 1988), it shows only moderate associations with task performance (for a survey see De Winter et al., 2019).

4.6. Conclusions and recommendations

We conclude that the answer to the question 'How much time do drivers need to obtain SA?' is largely independent of whether or not there is a hazard on the road, except when the time budget is only 1 s. Because decision-making accuracy in hazard situations was rather low (75–80%), the decision-making process in take-over scenarios should be automated where possible. We see potential in further developments of automated emergency braking systems, evasive maneuvering systems, or lane change assistance technology to ensure that drivers do not have to make time-critical decisions themselves. Future research could also aim to replicate our video-based experiment in a driving simulator with a 360-degree field of view, thus offering a higher degree of immersion. The present study was conducted with engineering students, who tend to have superior spatial ability (Lubinski, 2010). Future research should be conducted with a gender-balanced sample of drivers of different age groups. Finally, it is noted that the road environment in our study was symmetric, with the ego-car driving in the middle lane, being equally likely to overtake other cars (or be overtaken by other cars) driving in the left lane and the right lane. Participants were from countries that use left-hand traffic, such as the Netherlands, and countries that use right-hand traffic, such as India. Research shows that participants from Western Europe and India perceive different levels of risk when overtaking another car vehicle via the left lane or the right lane (Bazilinskyy, Eisma, Dodou, & De Winter, 2020). Future cross-national research could explore how country-specific traffic rules and associated expectations of drivers affect their SA.

CRediT authorship contribution statement

Zhenji Lu: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Riender Happee:** Supervision, Writing - review & editing, Funding acquisition. **Joost C. F. de Winter:** Conceptualization, Validation, Formal analysis, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Funding acquisition.

Acknowledgments

Data availability

The videos, raw data, and a MATLAB script that reproduces all results in this paper can be found here: https://doi.org/10.4121/uuid:a1f0ad0f-0033-431f-8a90-0efcae37dd6a.

Appendix A

In Part 1 of the experiment, half of the participants (those with an odd participant number) completed Set 1 first, followed by Set 2 (see Table A1 for a definition of the sets). The other half of the participants (those with an even participant number) completed Set 2 first, followed by Set 1. The videos were shown in a random order that was different for each participant.

In Part 2 of the experiment, participants viewed the same 14 videos they had first experienced in Part 1 (Set 1 for the odd-numbered participants, Set 2 for the even-numbered participants), but in a new random order.

We performed one-way repeated measures ANOVAs after rank-transformation (Conover & Iman, 1981), with the video length as the repeated-measures factor. Pairwise comparisons between video lengths were conducted using paired-samples t-tests of the ranked-transformed data. Corresponding effect sizes for the pairwise comparisons were expressed

Table A1
Situations presented in Part 1 of the experiment.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Set 1	N1	N3	N6	N9	N12	N20	R3	R9	L1	L6	B3	B9	LR1	LR6
Set 2	NM1	NM3	NM6	NM9	NM12	NM20	R1	R6	L3	L9	B1	B6	LR3	LR9

Note. N = No need to take over situations, R = Evade right situations, L = Evade left situations, LR = Evade left or right situations, B = Brake only situations, M = Mirrored. The numbers refer to the video length in seconds.

Table A2
Results of the repeated-measures ANOVA and effect sizes (d_z) of paired comparisons between video lengths (non-hazard situations of Part 1 of the experiment).

	df	F	p	η_p^2	1 vs. 3	1 vs. 6	1 vs. 9	3 vs. 6	3 vs. 9	6 vs. 9
Decision accuracy (%)	3,93	13.3	**<0.001**	0.30	−0.68	−0.78	−0.78	−0.08	−0.08	−0.02
Decision time (s)	3,93	27.6	**<0.001**	0.47	**1.00**	**1.39**	**1.20**	0.42	0.30	−0.15
Absolute error in the number of placed cars (#)	3,93	13.6	**<0.001**	0.31	**0.86**	**0.80**	**1.06**	0.09	0.24	0.12
Total distance error (%)	3,93	3.5	0.018	0.10	0.45	0.43	0.45	0.03	−0.03	−0.06
Total speed error (-)	3,93	10.1	**<0.001**	0.25	**0.86**	**0.76**	**0.85**	0.16	0.24	0.03
Self-reported difficulty (%)	3,93	17.6	**<0.001**	0.36	**0.96**	**0.79**	**1.06**	−0.15	0.09	0.22
Self-reported danger (%)	3,93	36.3	**<0.001**	0.54	**0.74**	**1.57**	**1.13**	**0.96**	**0.58**	**−0.61**
Braking (%)	3,93	2.5	0.061	0.08	0.35	0.35	0.25	0.00	−0.10	−0.18

Note. Significant effects ($p < .005$) are listed in boldface.

Table A3
Results of the repeated-measures ANOVA and effect sizes (d_z) of paired comparisons between video lengths (hazard situations of Part 1 of the experiment).

	df	F	p	η_p^2	1 vs. 3	1 vs. 6	1 vs. 9	3 vs. 6	3 vs. 9	6 vs. 9
Decision accuracy (%)	3,93	1.52	0.251	0.05	−0.18	−0.28	−0.36	−0.06	−0.18	−0.14
Decision time (s)	3,93	1.14	0.337	0.04	0.23	0.27	0.28	0.02	0.08	0.07
Absolute error in the number of placed cars (#)	3,93	46.9	**<0.001**	0.60	**1.28**	**1.52**	**1.75**	0.20	**0.67**	0.43
Total distance error (%)	3,93	42.1	**<0.001**	0.58	**1.76**	**1.60**	**1.65**	−0.11	−0.11	−0.01
Total speed error (-)	3,93	10.6	**<0.001**	0.25	0.15	0.29	**0.85**	0.17	**0.78**	**0.68**
Self-reported difficulty (%)	3,93	7.1	**<0.001**	0.19	0.45	**0.65**	**0.69**	0.16	0.16	0.02
Self-reported danger (%)	3,93	6.5	**<0.001**	0.17	0.45	**0.57**	0.51	0.20	0.06	−0.17
Braking (%)	3,93	2.9	0.038	0.09	0.12	0.33	0.46	0.19	0.33	0.19

Note. Significant effects ($p < .005$) are listed in boldface.

in terms of Cohen's d_z, a within-subjects effect size index (Faul, Erdfelder, Lang, & Buchner, 2007). Results are shown in Table A2 (non-hazard situations) and Table A3 (hazard situations).

It can be seen that the effect of time budget on the total distance error was statistically significant for hazard situations, but not for non-hazard situations; this can be explained by the high distance errors for 1-s hazard situations. The effects of time budget on decision accuracy and decision time were significant for non-hazard situations but not for hazard situations, which can be explained by the poor decision accuracy for 1-s situations compared to 3-, 9-, and 12-s situations in the non-hazard condition (Table A3).

References

Adam, E. V. (1993). Fighter cockpits of the future. In *Proceedings of the 12th IEEE/AIAA digital avionics systems conference, Fort Worth, TX* (pp. 318–323). https://doi.org/10.1109/dasc.1993.283529.

Banks, V. A., Plant, K. L., & Stanton, N. A. (2018). Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016. *Safety Science, 108*, 278–285. https://doi.org/10.1016/j.ssci.2017.12.023.

Bazilinskyy, P., Eisma, Y. B., Dodou, D., & De Winter, J. C. F. (2020). Risk perception: A study using dashcam videos and participants from different world regions. *Traffic Injury Prevention.* https://doi.org/10.1080/15389588.2020.1762871.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., … Cesarini, D. (2018). Redefine statistical significance. *Nature Human Behaviour, 2*, 6–10. https://doi.org/10.1038/s41562-017-0189-z.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*, 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x.

Cabrall, C. D. D., Janssen, N., & De Winter, J. C. F. (2018). Adaptive automation: Automatically (dis)engaging automation during visually distracted driving. *PeerJ Computer Science, 4*, e166. https://doi.org/10.7717/peerj-cs.166.

Clark, H., Stanton, N. A., & Revell, K. M. (2020). Automated vehicle handover interface design: Focus groups with learner, intermediate and advanced drivers. *Automotive Innovation, 3*, 14–29. https://doi.org/10.1007/s42154-019-00085-x.

Cohen-Lazry, G., Katzman, N., Borowsky, A., & Oron-Gilad, T. (2019). Directional tactile alerts for take-over requests in highly- automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 65*, 217–226. https://doi.org/10.1016/j.trf.2019.07.025.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*, 124–129. https://doi.org/10.1080/00031305.1981.10479327.

Crundall, D., Chapman, P., Trawley, S., Collins, L., Van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis and Prevention, 45*, 600–609. https://doi.org/10.1016/j.aap.2011.09.049.

De Winter, J. C. F., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology and Work, 21*, 99–111. https://doi.org/10.1007/s10111-018-0527-6.

De Winter, J. C. F., Happee, R., Martens, M., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 196–217. https://doi.org/10.1016/j.trf.2014.06.016.

De Winter, J. C. F., Stanton, N., Price, J. S., & Mistry, H. (2016). The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *International Journal of Vehicle Design, 70*, 297–324. https://doi.org/10.1504/IJVD.2016.076736.

Durso, F. T., Bleckley, M. K., & Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors, 48*, 721–733. https://doi.org/10.1518/001872006779166316.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition, 7*, 149–158. https://doi.org/10.3758/BF03197595.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). In *Proceedings of the IEEE 1988 national aerospace and electronics conference* (pp. 789–795). https://doi.org/10.1109/NAECON.1988.195097.

Endsley, M. R. (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64. https://doi.org/10.4324/9781315087924-3.

Eriksson, A., Banks, V. A., & Stanton, N. A. (2017). Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention, 102*, 227–234. https://doi.org/10.1016/j.aap.2017.03.011.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146.

Frison, A. K., Wintersberger, P., Schartmüller, C., & Riener, A. (2019). The real T(h)OR: Evaluation of emergency take-over on a test track. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications: Adjunct proceedings* (pp. 478–482). https://doi.org/10.1145/3349263.3349602.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31*, 1–40. https://doi.org/10.1006/cogp.1996.0011.

Gugerty, L. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied, 3*, 42–66. https://doi.org/10.1037/1076-898X.3.1.42.

Gugerty, L. (1998). Evidence from a partial report task for forgetting in dynamic spatial memory. *Human Factors, 40*, 498–508. https://doi.org/10.1518/001872098779591412.

Jones, D. G., & Endsley, M. R. (2004). Use of real-time probes for measuring situation awareness. *The International Journal of Aviation Psychology, 14*, 343–367. https://doi.org/10.1207/s15327108ijap1404_2.

Josten, J., Zlocki, A., & Eckstein, L. (2016). Untersuchung der Bewältigungsleistung des Fahrers von kurzfristig auftretenden Wiederübernahmesituationen nach teilautomatischem, freihändigen Fahren [Investigation of drivers' handling performance in sudden take-over situations after semi-automated, hands-off driving]. *FAT Schriftenreihe, 289*.

Kahana-Levy, N., Shavitzky-Golkin, S., Borowsky, A., & Vakil, E. (2019). Facilitating hazard awareness skills among drivers regardless of age and experience through repetitive exposure to real-life short movies of hazardous driving situations. *Transportation Research Part F: Traffic Psychology and Behaviour, 60*, 353–365. https://doi.org/10.1016/j.trf.2018.09.021.

Köhn, T., Gottlieb, M., Schermann, M., & Krcmar, H. (2019). Improving take-over quality in automated driving by interrupting non-driving tasks. In *Proceedings of the 24th international conference on intelligent user interfaces* (pp. 510–517). https://doi.org/10.1145/3301275.3302323.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". *Law and Human Behavior, 11*, 55–62. https://doi.org/10.1007/BF01044839.

Lu, Z., Coster, X., & De Winter, J. C. F. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics, 60*, 293–304. https://doi.org/10.1016/j.apergo.2016.12.003.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences, 49*, 344–351. https://doi.org/10.1016/j.paid.2010.03.022.

Marquart, G., & De Winter, J. C. F. (2015). Workload assessment for mental arithmetic tasks using the task-evoked pupillary response. *PeerJ Computer Science, 1*, e16. https://doi.org/10.7717/peerj-cs.16.

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

Naujoks, F., Purucker, C., Wiedemann, K., & Marberger, C. (2019). Noncritical state transitions during conditionally automated driving on German freeways: Effects of non-driving related tasks on takeover time and takeover quality. *Human Factors, 61*, 596–613. https://doi.org/10.1177/0018720818824002.

Navarro, J., Osiurak, F., Ovigue, M., Charrier, L., & Reynaud, E. (2019). Highly automated driving impact on drivers' gaze behaviors during a car-following task. *International Journal of Human-Computer Interaction, 35*, 1008–1017. https://doi.org/10.1080/10447318.2018.1561788.

Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J. R., Mérienne, F., Benedetto, S., & Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction, 30*, 220–236. https://doi.org/10.1080/10447318.2013.848320.

Pfromm, M. (2016). Menschbezogener Umgang mit Systemstörungen bei teilautomatisierter Fahrzeugführung [Human-related handling of system errors in semi-automated vehicle guidance] (Doctoral dissertation). Technische Universität Darmstadt, Darmstadt, Germany.

Politis, I., Brewster, S., & Pollick, F. (2017). Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers. *International Journal of Mobile Human Computer Interaction, 9*, 1–16. https://doi.org/10.4018/ijmhci.2017070101.

PreScan (2017). A physics-based simulation platform. https://tass.plm.automation.siemens.com/prescan.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*, 2063–2067. https://doi.org/10.1177/1541931214581434.

Salmon, P. M., Stanton, N. A., & Young, K. L. (2012). Situation awareness on the road: Review, theoretical and methodological issues, and future directions. *Theoretical Issues in Ergonomics Science, 13*, 472–492. https://doi.org/10.1080/1463922X.2010.539289.

Samuel, S., Borowsky, A., Zilberstein, S., & Fisher, D. L. (2016). Minimum time to situation awareness in scenarios involving transfer of control from an automated driving suite. *Transportation Research Record, 2602*, 115–120. https://doi.org/10.3141/2602-14.

Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science: Situation awareness in individuals, teams and systems. *Ergonomics, 60*(4), 449–466. https://doi.org/10.1080/00140139.2017.1278796.

Vlakveld, W., Van Nes, N., De Bruin, J., Vissers, L., & Van der Kroft, M. (2018). Situation awareness increases when drivers have more time to take over the wheel in a Level 3 automated car: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*, 917–929. https://doi.org/10.1016/j.trf.2018.07.025.

Weinbeer, V., Baur, C., Radlmayr, J., Bill, J., Muhr, T., & Bengler, K. (2017). Highly automated driving: How to get the driver drowsy and how does drowsiness influence various take-over-aspects? *Paper presented at the 8th Tagung Fahrerassistenz, Einführung hochautomatisiertes Fahren, München, Germany*.

Wiedemann, K., Naujoks, F., Wörle, J., Kenntner-Mabiala, R., Kaussner, Y., & Neukum, A. (2018). Effect of different alcohol levels on take-over performance in conditionally automated driving. *Accident Analysis & Prevention, 115*, 89–97. https://doi.org/10.1016/j.aap.2018.03.001.

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 (ITSC)* (pp. 650–655). https://doi.org/10.1109/ITSC.2018.8569986.

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis & Prevention, 92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002.

Zhang, B., De Winter, J. C. F., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F, 64*, 285–307. https://doi.org/10.1016/j.trf.2019.04.020.