

How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving

Lu, Zhenji; Coster, Xander; de Winter, Joost

DOI

[10.1016/j.apergo.2016.12.003](https://doi.org/10.1016/j.apergo.2016.12.003)

Publication date

2017

Document Version

Final published version

Published in

Applied Ergonomics: human factors in technology and society

Citation (APA)

Lu, Z., Coster, X., & de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics: human factors in technology and society*, 60, 293-304. <https://doi.org/10.1016/j.apergo.2016.12.003>

Important note

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

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

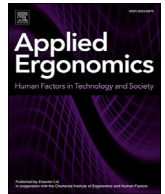
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' – Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.



How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving



Zhenji Lu^{*},¹, Xander Coster¹, Joost de Winter

Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands

ARTICLE INFO

Article history:

Received 6 June 2016
Received in revised form
28 November 2016
Accepted 4 December 2016
Available online 27 December 2016

Keywords:

Scene perception
Spatial memory
Eye tracking
Hazard perception

ABSTRACT

Drivers of automated cars may occasionally need to take back manual control after a period of inattentiveness. At present, it is unknown how long it takes to build up situation awareness of a traffic situation. In this study, 34 participants were presented with animated video clips of traffic situations on a three-lane road, from an egocentric viewpoint on a monitor equipped with eye tracker. Each participant viewed 24 videos of different durations (1, 3, 7, 9, 12, or 20 s). After each video, participants reproduced the end of the video by placing cars in a top-down view, and indicated the relative speeds of the placed cars with respect to the ego-vehicle. Results showed that the longer the video length, the lower the absolute error of the number of placed cars, the lower the total distance error between the placed cars and actual cars, and the lower the geometric difference between the placed cars and the actual cars. These effects appeared to be saturated at video lengths of 7–12 s. The total speed error between placed and actual cars also reduced with video length, but showed no saturation up to 20 s. Glance frequencies to the mirrors decreased with observation time, which is consistent with the notion that participants first estimated the spatial pattern of cars after which they directed their attention to individual cars. In conclusion, observers are able to reproduce the layout of a situation quickly, but the assessment of relative speeds takes 20 s or more.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past few decades, an increasing number of automated driving systems have become available, both for research and consumer purposes. Automated driving systems in which the driver can opt to be ‘out-of-the-loop’ by engaging in non-driving tasks such as reading or resting are expected within a decade (Kyriakidis et al., 2015; Underwood, 2014). When the automation malfunctions or reaches its functional limits, control has to be handed back to the driver. In such cases, the automation typically issues a warning signal (also called a take-over request, see Lu et al., 2016 for a review) after which the driver has to resume the driving task (SAE International, 2014).

A critical design parameter in the development of automated driving system is the available time for taking over control, sometimes referred to as ‘lead time’, ‘time buffer’, or ‘time

budget’ (Gasser and Westhoff, 2012; SAE International, 2014; Zeeb et al., 2016). If drivers do not have sufficient time to assess the situation prior to taking control, an accident may result (Mok et al., 2015). Drivers may prefer long lead times to prepare for the upcoming transition of control, but in reality, this is not always technologically feasible. For example, limitations in sensors (e.g., the limited range of a forward-facing radar) pose barriers regarding the maximum lead time that is feasible. In summary, it is important to understand how much time drivers need for gaining situation awareness (SA), because this sets demands on the automated driving technology.

Various studies have previously examined the effects of lead time on drivers’ behaviour after resuming control (Clark and Feng, 2015; Gold et al., 2013; Mok et al., 2015). For example, a driving simulator study by Gold et al. (2013) found that the less time is available until colliding with a stationary object (5 s vs. 7 s), the more abrupt are the drivers’ braking and steering inputs after receiving a take-over request. This study reported an average gaze reaction time (i.e., the time between the take-over request and the eye-gaze moving away from the non-driving task) of 0.5 s, an

^{*} Corresponding author.

E-mail address: z.lu@tudelft.nl (Z. Lu).

¹ Joint first authors.

average hands-on-steering-wheel time of 1.5 s, and an average mirror-scan time of about 3 s (for similar results see [Kerschbaum et al., 2015](#)). [Van den Beukel and Van der Voort \(2013\)](#) found a decrease in the number of accidents and higher self-reported SA scores when more time was available. [Mok et al. \(2015\)](#) found that only few participants in a 2 s lead-time condition safely negotiated a hazardous situation, while the 5 s and 8 s conditions yielded considerably safer driver behaviours. A driving simulator study by [Samuel et al. \(2016\)](#) compared 4, 6, 8, and 12 s lead times, and found that participants needed a lead time of at least 8 s in order to detect a latent pedestrian hazard with the same accuracy as they did when being in control of the vehicle. Driving simulator research by [Merat et al. \(2014\)](#) and by [Desmond et al. \(1998\)](#) suggests that it may take up to 20 s or 40 s before vehicle control is fully stabilised after reclaiming control. Although the above studies provide valuable knowledge, they do not offer much insight into the cognitive processes of how drivers are able to build SA of a traffic situation as a function of the available time.

Over the last 25 years, the topic of SA has been extensively investigated ([Endsley, 2015](#)). The Situation Awareness Global Assessment Technique (SAGAT) is one of the standard instruments for measuring SA ([Garland and Endsley, 1995](#)). In this method, the screens of a simulator are temporarily blanked, and participants subsequently have to answer queries about objects and unfolding events in the simulation. Although SAGAT has been criticized for the fact that it measures SA intermittently rather than continuously, and for relying heavily on memory skills (for discussion see [Durso et al., 2006](#); [Gutzwiler and Clegg, 2013](#); [Stanton et al., 2015](#)), there is now a sound body of literature showing that SAGAT scores exhibit criterion validity with respect to task performance ([Durso et al., 1995](#); [Gardner et al., 2015](#); [Loft et al., 2015](#); [Salmon et al., 2009](#)). Various promising alternative methods have been proposed for measuring SA, such as real-time probing (e.g., [Loft et al., 2013](#); [Martelaro et al., 2015](#); [Pierce, 2012](#)) and physiological techniques (e.g., [Crundall et al., 2003](#); [Gugerty, 2011](#)), but at present SAGAT still appears to be the most widely applied and validated SA assessment tool (see also [Endsley, 2000, 2015](#)).

The answer categories in SAGAT are usually discrete or discretised values of the state of the virtual environment (e.g., [Salmon et al., 2006](#); [Loft et al., 2015](#)). [Gugerty \(1997\)](#) used a similar technique as SAGAT for measuring participants' dynamic spatial memory by means of continuous values. Specifically, participants watched animations of traffic situations that lasted 18–35 s, and after each video, they indicated the positions of surrounding cars from a top-down view. Participants' level of SA was operationalized by comparing the positions of the placed cars with the actual positions of the cars in the animation. [Gugerty](#) found that the more cars are to be recalled, the poorer the performance on the SA task. Furthermore, he found that when the number of cars was larger, participants showed a prioritization effect whereby the most hazardous cars were remembered best.

In the present research, we refined the method used by [Gugerty \(1997\)](#) for determining the effect of time on SA scores. Specifically, we investigated the effect of viewing time (i.e., video length) with two levels of traffic density, namely 4 or 6 cars in surrounding traffic. The use of 4 and 6 cars is in approximate agreement with [Pylyshyn and Storm \(1988\)](#), who found that people can track up to five moving objects in a perceptual task, and with [Gugerty \(1997\)](#) who used 3 to 8 cars in his research. In our study, six different video lengths were adopted, ranging between 1 s and 20 s. The video lengths were based partly on a pilot study conducted prior to the present study ([Coster, 2015](#)). In this pilot, seven participants watched videos of animated traffic scenes and pressed the spacebar

when they had assessed the situation to such an extent that they would feel safe to take over control. The results showed that a viewing time of 12 s was generally deemed sufficient, with an overall minimum of 3 s. In visual processing research, it has been found that participants can recognize the gist of a scene when having viewed it for only 20 ms ([Thorpe et al., 1996](#)). [Oestmann et al. \(1988\)](#) found that radiologists were able to detect 'subtle cancers' and 'obvious cancers' from a radiograph in 0.25 s with true positive rates of 30% and 70%, respectively (cf. 74% and 98%, respectively, for unlimited viewing times). However, sub-second viewing times are probably too short for processing dynamic traffic scenes that require visual search by means of multiple fixations and saccades (see [Rayner, 2009](#) for a review on eye movements and visual search). Lead times that are typically used in driving simulator research range between 2 s and 12 s ([Gold et al., 2013](#); [Körber et al., 2015, 2016](#); [Melcher et al., 2015](#); [Mok et al., 2015](#); [Samuel et al., 2016](#)). In summary, our range of video lengths encompasses lead times that are commonly used, and range from extremely short (1 s) to longer than has been studied before (20 s).

The dependent measures in this study were: (1) self-reported task difficulty and time sufficiency, (2) the absolute error between the number of placed cars and the actual number of cars, (3) the error between the positions and indicated speeds of the placed cars relative to the actual positions/speeds of the cars, making use of an algorithm that globally selects a match between placed and actual cars by minimizing the positional difference, (4) the geometric difference between the positions of the placed and actual cars, and (5) eye-gaze activity. We expected that when the viewing time is longer, participants would find the task easier and have a better reproduction performance. Our corresponding aim was to explore at which video length these effects would saturate.

The geometric difference method is an innovation in SA assessment. It is based on a method for comparing polygons previously introduced by [Arkin et al. \(1991\)](#), which was said to be "invariant under translation, rotation, and change of scale, reasonably easy to compute, and intuitive" (p. 209). We applied this technique to obtain a generic index of difference that avoids the use of arbitrary parameters, such as correction factors related to the fact that people have a tendency to underestimate the distance to objects in virtual environments.

Eye tracking is widely used in studies of hazard perception, a term often equated with SA ([Horswill and McKenna, 2004](#); [Hosking et al., 2010](#); [Underwood et al., 2002, 2013](#)). We used eye tracking to gain a deeper understanding of how participants build SA as a function of time. It is well known that eye movements are correlated with bottom-up and top-down attention ([Borji and Itti, 2013](#); [Henderson, 2003](#); [Itti and Koch, 2001](#)) and memory of visual objects ([Irwin and Zelinsky, 2002](#); [Moore and Gugerty, 2010](#)). For example, using a SAGAT method, [Moore and Gugerty \(2010\)](#) found that the more participants fixated on an aircraft in an air traffic control task, the higher their SA (i.e., responses to state queries) for that aircraft. [Unema et al. \(2005\)](#) and [Over et al. \(2007\)](#) found that in visual search tasks, participants exhibit a course-to-fine eye-movement strategy, whereby the first fixations and saccades had a short duration and large amplitude, respectively, and later fixations became longer-lasting with smaller-amplitude saccades in between the fixations. In this paper, we measured whether participants glanced at the road or at the mirrors, and how frequently they glanced at the mirrors, as a function of observation time. We explored whether these measures of attention distribution and glance frequency exhibit similar saturation profiles as the objective task scores.

2. Method

2.1. Hardware

The videos and graphical user interface (GUI) were presented on a 24-inch widescreen HD monitor of the Smart Eye DR120 remote eye tracking system. The videos were programmed using PreScan 7.0 (Tass International, 2015) and had a resolution of 1920×1080 pixels. The participants reproduced the traffic situations using a standard Dell mouse.

2.2. Videos

Each video began with a 5 s crosshair display for participants to focus on, after which the traffic situation from an ego-centric viewpoint was presented for 1, 3, 7, 9, 12, or 20 s. At the end of the video, a black screen was displayed for 0.5 s. The rear view mirror and left wing mirror were positioned in such a way as to resemble real positions. Due to geometric constraints, this was not possible for the right wing mirror, which was therefore placed at the right edge of the video (Fig. 1).

In real driving, sound cues may aid in the formation of SA. For example, a driver may infer the location and speed of a nearby car through the sounds of that car's engine and tires. In our study, we decided to eliminate sound cues and make the SA task visual-only. Therefore, during each video, a standard sound of driving on a highway was played that was unrelated to the traffic in the video.

2.3. Participants

Thirty-four participants (5 female, 29 male) with a valid driver's license, aged between 20 and 31 years ($M = 24.6$, $SD = 2.6$ years) participated in this study. The mean participants' age when obtaining the first driver's license was 19.2 years ($SD = 2.4$), and the mean number of years of licensure (i.e., current age minus the age when obtaining the first driver's license) was 5.41 years ($SD = 2.92$). On a scale of 0 (every day), 1 (4–6 days a week), 2 (1–3 days a week), 3 (once a month to once a week), 4 (less than once a month), and 5 (never), the mean answer to "On average, how often did you drive a car or motorcycle during the last 12 months?" was 2.62 ($SD = 1.04$, min = 0, max = 4). Furthermore, on a scale from 0 (0 km), 1 (1–1000 km), 2 (1001–5000 km), 3 (5001–10,000 km), 4 (10,000–15,000 km) to 10 (more than 100,000 km), the answer to "About how many kilometres did you drive during the last 12 months?" was on average 2.59 ($SD = 1.79$, min = 1, max = 9).



Fig. 1. Screenshot of a video that includes six surrounding cars.

All participants read and signed a consent form, explaining the purpose and procedures of the experiment. Participants received €5 for their participation. They were split into two groups based on their recruitment number (i.e., group A if the participant number was odd, and group B if it was even). These two groups viewed 24 test videos, whereby each video in one group maps to a video in the other group, with these videos featuring the same traffic and the same ending moment, but a different starting moment. Both groups consisted of 17 participants: 3 females, 14 males, aged between 21 and 31 ($M = 25.0$, $SD = 2.8$ years) for group A; 2 females, 15 males, aged between 20 and 29 ($M = 24.2$, $SD = 2.4$ years) for group B. The mean number of years of licensure was 5.65 ($SD = 3.32$) and 5.18 ($SD = 2.56$) for groups A and B, respectively. For groups A and B, the mean driving frequency on the scale from 0 to 5 was 2.88 ($SD = 0.93$) and 2.35 ($SD = 1.69$), respectively, and the mean mileage on the scale from 0 to 10 was 2.35 ($SD = 1.14$) and 2.82 ($SD = 1.91$), respectively. According to an independent-samples *t*-test, groups A and B were not statistically significantly different with regard to age, gender, license age, years of licensure, driving frequency, and mileage ($p = 0.361, 0.641, 0.672, 0.646, 0.142, \& 0.453$, respectively).

2.4. Situations

Participants viewed videos of traffic situations on a three-lane highway, on which the ego-vehicle was driving in the middle lane with a constant speed of 28 m/s. The training videos lasted 12 s and featured three cars of surrounding traffic, whereas the test videos incorporated four or six surrounding cars. Both traffic densities were used 12 times. The test videos lasted 1, 3, 7, 9, 12 or 20 s.

Each video length occurred twice per traffic density and four times in 24 test videos. The test videos were shown to each participant in randomized order.

All traffic in each video met the following criteria. Each car was

- within a range of 80 m behind to 120 m ahead of the ego-vehicle during the full length of the video.
- visible during the full length of the video, except when driving through the blind spot. This visibility criterion had the effect that there could be only one car directly in front of and/or behind the ego-vehicle on the middle lane.
- starting and ending outside of the blind spot.
- driving at one of five constant speeds: 25.5, 26.75, 28, 29.25 or 30.5 m/s.
- staying in its own lane during the full length of the video.
- at least 5 m in front of or behind other cars at all times.

The model and colour of each car were randomly assigned out of 13 possible colours and 10 possible models. Averaged across the 24 scenarios, 57% of the cars were in front of the ego-vehicle at the end of the scenarios (58% at the beginning of the scenarios). In groups A and B, an overtaking event of the ego-vehicle occurred in 6 and 8 of the 24 scenarios, respectively.

2.5. Procedure

Prior to the test, the participants filled out a questionnaire about their driving experience. Participants were asked to adjust the chair in order to face the monitor mid-front, with the hind legs of the chair within a demarcation on the floor, approximately 65 cm away from the monitor. The height of the monitor was adjusted to the participant's height, after which the eye tracker was calibrated. Participants were given 2 training trials, followed by 24 test trials, viewing videos of traffic situations on a three-lane highway. After

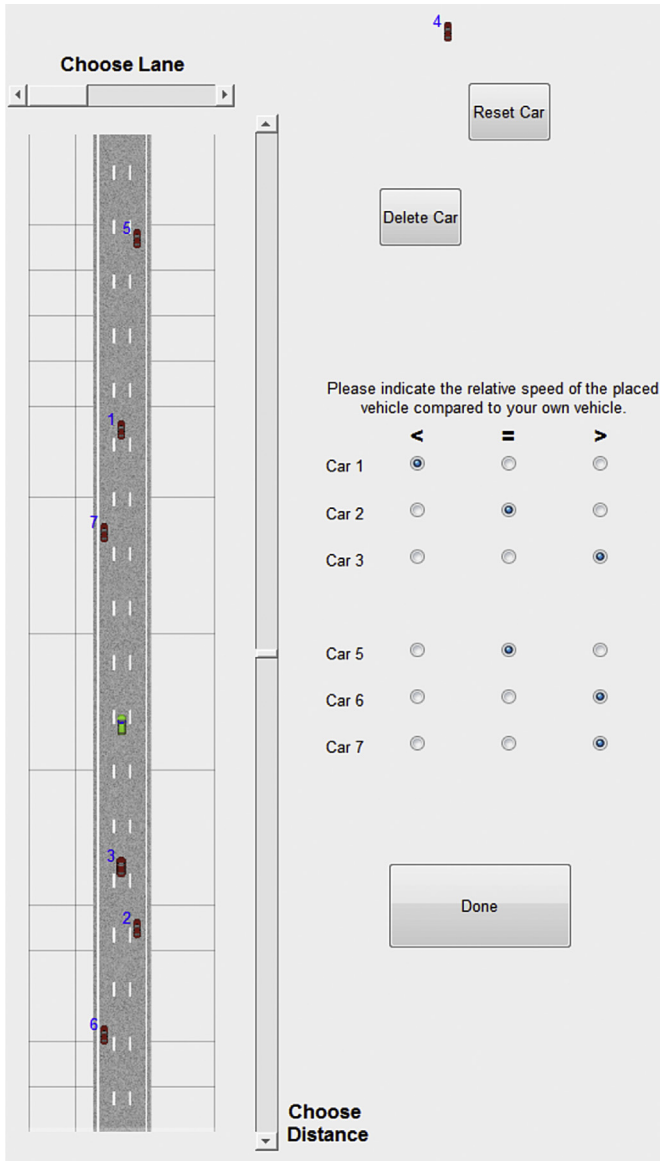


Fig. 2. Reproduction of a traffic situation. Red cars represent the surrounding cars. The green car represents the ego-vehicle. The slider bars were used to set the positions of surrounding vehicles. The radio buttons were used to indicate the relative speed. Placed cars could be deleted by pressing the 'Delete Car' button, and deleted cars could be placed again by pressing the 'Reset Car' button. Additional cars could be placed by pressing the 'New Car' button (only visible when no cars were deleted). The 'Done' button was pressed when a participant had finished the reproduction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

each video, participants reproduced the final positions of the surrounding cars by placing a minimum of 1 and a maximum of 8 vehicles without time restriction. Participants also indicated the speeds of the placed cars in relation to the ego-vehicle (Fig. 2).

After reproducing a situation, a two-item questionnaire measured the subjective task difficulty ('The task was difficult') and time sufficiency ('I had sufficient time to perform the task') on a scale from 0 (completely disagree) to 100 (completely agree). Participants received oral instructions during the training trials about (1) how to place the cars and use the slider bars, (2) how to interpret the time sufficiency question as 'the video was long enough for me to perform the task', and (3) that surrounding traffic would not necessarily follow traffic rules (e.g., cars could overtake

on the right). The traffic was not constrained by traffic rules, because we wanted to retain symmetry in the videos by letting all cars drive with constant speeds without changing lanes. For example, when preventing the traffic from overtaking the participant on the right, the mean speed on the left lane would be higher than the mean speed on the right lane. Such asymmetric videos would have complicated the analyses and interpretation of how well participants were able to place cars.

After each trial, the placed cars and the actual cars were shown side by side in two top-down views, providing feedback to the participant. This feedback was provided to enhance participants' engagement in the task, and to prevent misunderstandings and biases that may occur when participants have to map the three-dimensional video to a two-dimensional representation. The duration of the experiment was approximately 60 min including preparation time.

2.6. Dependent measures

The following dependent measures were used:

1. Self-reported time sufficiency (%)
2. Self-reported task difficulty (%)
3. Absolute error of the number of placed cars, defined as the absolute error of the number of placed cars, calculated according to Eq. (1), where n_p is the number of cars that were placed and n_a is the actual number of cars in the video.

$$S_1 = |n_p - n_a| \quad (1)$$

4. Total distance error between the placed cars and actual cars. First, the placed and actual cars were matched to each other. A particular combination of matches $c_\alpha(m_1, m_2, \dots, m_{\max(n_p, n_a)})$ connects placed and actual cars, where a match m is between a placed car and an actual car. The number of possible combinations $N(c)$ of matches between placed and actual cars is given according to Eq. (2).

$$N(c) = S(\max(n_p, n_a), \min(n_p, n_a)) \min(n_p, n_a)! \quad (2)$$

where $S(n, k)$ is the Stirling number of the second kind. For example, if $n_p = 6$ and $n_a = 6$, then $N(c) = 720$, if $n_p = 5$ and $n_a = 6$, then $N(c) = 1,800$, and if $n_p = 6$ and $n_a = 4$, then $N(c) = 1560$. When one placed car $p(i) = (p_{i,x}, p_{i,y}, p_{i,v})$ and one actual car $a(j) = (a_{j,x}, a_{j,y}, a_{j,v})$ are matched, the distance error and speed error between these two cars are given by Eqs. (3) and (4), respectively

$$dE(m(p(i), a(j))) = \sqrt{(p_{i,x} - a_{j,x})^2 + (p_{i,y} - a_{j,y})^2} \quad (3)$$

$$sE(m(p(i), a(j))) = |p_{i,v} - a_{j,v}| \quad (4)$$

Here, $(p_{i,x}, p_{i,y})$ and $(a_{j,x}, a_{j,y})$ are the lateral and longitudinal positions of the placed car p_i and actual car a_j , where the centre of the ego-vehicle is the coordinate origin. $p_{i,v}$ and $a_{j,v}$ are the speeds of the placed car p_i and actual car a_j , respectively. For both the placed car p_i and actual car a_j , $p_{i,v}$ and $a_{j,v}$ are equal to -1 , 0 , or $+1$, when their speeds are slower than, equal to, or faster than the ego-vehicle, respectively.

The total distance error of one combination of matches is

$$DE(c_\alpha) = \sum_{k=1}^{\max(N_p, N_a)} (dE(m_k))_{|m_k \in c_\alpha} \quad (5)$$

The total distance error is $DE(c_{\min})$, where $c_{\min}$ is the combination that gives the minimal distance among all $N(c)$ combinations (Eq. (6))

$$DE(c_{\min}) = \min\{DE(c_1), DE(c_2), \dots, DE(c_{N(c)})\} \quad (6)$$

5. *Total speed error.* This measure is calculated based on the minimal distance for combination match $c_{\min}(m_1, m_2, \dots, m_{\max(n_p, n_a)})$ (Eq. (7))

$$SE(c_{\min}) = \sum_{k=1}^{\max(N_p, N_a)} (sE(m_k)|_{m_k \in c_{\min}}) \quad (7)$$

6. *Geometric difference.* The turning function algorithm (Arkin et al., 1991) is widely used in computer vision to calculate the difference between two polygons. It maps a two-dimensional shape to a one-dimensional function. The first step is to construct polygons that represent the positions of placed and actual cars. Fig. 3a provides an example of P with all its nodes, where each node represents the centre of the placed car p_i . The following 10 possible nodes can be used to construct P :

1) $(-3.5, \min(p_{i,y} _{p_{i,y} > 0 \cap p_{i,x} = -3.5}))$	2) $(-3.5, \max(p_{i,y} _{p_{i,y} > 0 \cap p_{i,x} = -3.5}))$
3) $(0, \max(p_{i,y} _{p_{i,y} > 0 \cap p_{i,x} = 0}))$	4) $(0, \min(p_{i,y} _{p_{i,y} > 0 \cap p_{i,x} = 0}))$
5) $(3.5, \max(p_{i,y} _{p_{i,x} = 3.5}))$	6) $(3.5, \min(p_{i,y} _{p_{i,x} = 3.5}))$
7) $(0, \min(p_{i,y} _{p_{i,y} < 0 \cap p_{i,x} = 0}))$	8) $(0, \max(p_{i,y} _{p_{i,y} < 0 \cap p_{i,x} = 0}))$
9) $(-3.5, \min(p_{i,y} _{p_{i,y} < 0 \cap p_{i,x} = -3.5}))$	10) $(-3.5, \max(p_{i,y} _{p_{i,y} < 0 \cap p_{i,x} = -3.5}))$

The polygon is constructed with a sequence from node 1 to 10. A node can be skipped if it does not exist, or if it is the same node as a neighbouring node in front. The same process is applied to the actual cars a_i that compose polygon A (Fig. 3b). The second step is to represent the polygon as a turning function. As illustrated in Fig. 3c, the turning function measures the external angle of the clockwise tangent as a function of the arc-length s from the starting node of P . The function decreases with right-hand turns and increases with left-hand turns. The perimeter length is scaled to 1 to make a comparison possible. The function always starts at $(0, 2\pi)$ and always ends at $(1, 0)$. Fig. 3d shows the turning function for A. The third and final step is to calculate the distance between the two turning functions (Eq. (8), Fig. 3e).

$$L(P, A) = \int_0^1 |T_P(s) - T_A(s)| ds \quad (8)$$

2.7. Eye-tracking analysis

Participants' eye movements were analysed in order to understand how participants distributed their attention while watching the video. Four equally sized rectangular areas were defined, namely (1) road centre, (2) rear view mirror, (3) left wing mirror, and (4) right wing mirror. We defined the net dwell time percentage across all situations of all participants, as a

function of observation time in bins of 1 s. Furthermore, in order to obtain a measure of visual search and gaze activity, we measured how often per second participants glanced at one of the three mirrors. For this latter analysis, a glance was counted if the dwell time to the mirror was at least 200 ms. This 200 ms threshold corresponds to typical measures of fixation duration (e.g., Velichkovsky et al., 2002). Note that our Smart Eye eye-tracker did not have the precision and accuracy to be able to distinguish between cars within an area of interest (see Funke et al., 2016 for an assessment of eye trackers).

2.8. Statistical analyses

Three types of statistical analyses were conducted. First, a one-way repeated measures ANOVA after rank-transformation was performed (Conover and Iman, 1981), with the video length as the repeated-measures factor. Pairwise comparisons between video lengths were conducted by means of paired-samples t tests followed by a Bonferroni correction, which effectively means that the significance level (α) was set to 0.00333 ($= 0.05/15$). Effect sizes for the pairwise comparisons were expressed in terms of Cohen's d_z for assessing within-subject effects (Faul et al., 2007).

Second, we performed a between-subjects comparison between the participants in group A versus the participants in group B. These two groups had viewed the same videos with the same endpoint; only the video length was different. We assessed in how many of the 24 situations the longer videos yielded higher scores on the dependent measures than the shorter videos.

Third, at the level of the 24 situations, we assessed Spearman rank-order correlations between video length and the scores on the dependent measures. An objective measure of task difficulty, defined as the distance from the ego-vehicle to the actual car summed across all 4 or 6 cars, was partialled out in order to determine whether the effects between video length and task performance were robust to situation-specific effects.

3. Results

Some data had to be excluded due to quality issues, see Table 1 for an overview. A small number of trials were excluded due to technical malfunctions (e.g., video not shown, data storage error) or because the participant initially misunderstood the task. The data for two participants were excluded entirely because the experimenter's log files revealed that they had used mnemonics (hands, fingers) to enhance their task performance. Moreover, for the eye-tracking analysis, data could be used from 21 out of 34 participants.

Figs. 4–9 show the results for the self-reported time sufficiency, self-reported difficulty, mean absolute error of the number of placed cars, mean total distance error, mean total speed error, and the mean geometric difference score, respectively, for each of the 24 videos. The means of group A and group B are connected by a line.

3.1. Subjective measures

Self-reported time sufficiency and self-reported difficulty showed consistent effects of video length (Figs. 4 and 5). The pairwise comparisons between video lengths indicate that self-reported time sufficiency has statistically significant effects for almost all pairs of video lengths (Table 2), with an improvement even from 12 s to 20 s videos. The effects were less strong for task difficulty than for time sufficiency (Table 2).

Table 1

Overview of excluded trials, number of participants for which trials were excluded, and types of data that were excluded.

Malfunction/behaviour/limited conditions	Number of participants for which data were excluded	Total number of erroneous trials	Data not included for erroneous trials
Used hands/fingers as memory support	2	2*24	Reconstruction data, and self-report data
Video malfunction (video did not show)	1	1	Reconstruction data
Data storage error	6	6	Reconstruction data and self-report data
Misunderstanding of GUI controls	1	1	Reconstruction data and self-report data
Reconstructed beginning (instead of ending) of the video	2	5	Reconstruction data and self-report data
Accidentally pressed 'Done' button	1	1	Reconstruction data and self-report data
Misunderstood the meaning of 'time sufficiency'	1	24	Self-report data of the 'time sufficiency question'
Did not answer the two questions	1	24	Self-report data
Participant wore glasses, Software crashed, tracking/calibration problems	13	13*24	Eye-tracking data

Note. Reconstruction data encompasses positions and speeds of placed cars.

3.2. Tasks performance measures

Task performance in terms of the error in the number of placed cars (Fig. 6) and the total distance error (Fig. 7) shows a clear improvement with video length. In 20/21 of the 24 group A versus group B comparisons, the longer videos featured better task performance (Table 2). However, there appears to be a saturation effect of video length, whereby for the error in the number of placed cars there was no statistically significant difference anymore between 7 s and 20 s videos. Similarly, for the total distance error, there was no significant difference between 12 s and 20 s videos.

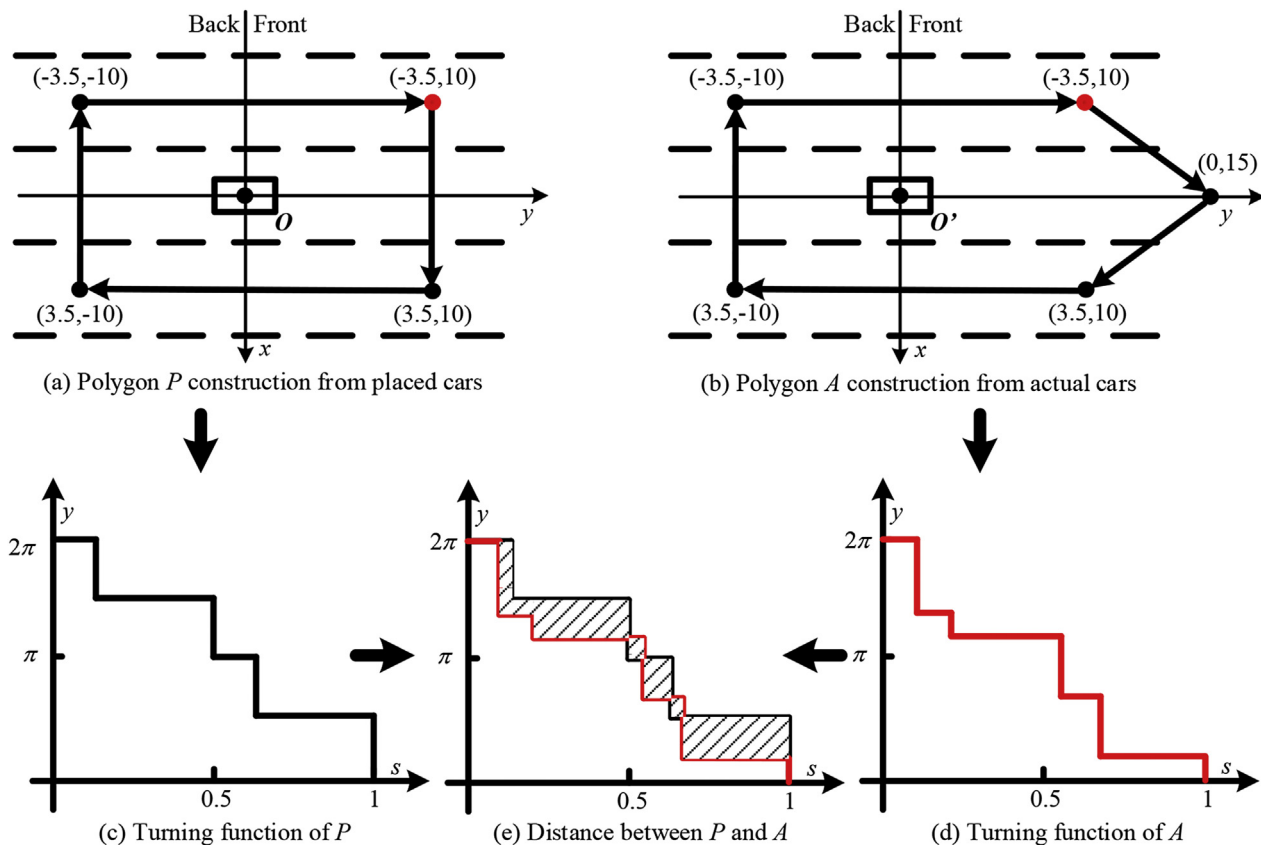
The total speed error shows a qualitatively different pattern than the total distance error, with no apparent saturation as a function of increasing video length (Fig. 8). A statistically significant improvement is observed even from 12 s videos to 20 s videos.

Moreover, the effects of video length are robust: in the group A versus group B comparison, 23 out of 24 videos with the same endpoint showed a better score for the longer video (Table 2).

The geometric difference score shows a weaker overall effect size (η^2) than the other measures, which may be because these scores exhibit strong situation-specific effects, with some situations yielding a notably better score than others (Fig. 9). Nonetheless, the A versus B comparison are consistent in the sense that longer-lasting videos yielded a lower difference score in 20 out of 24 situations. There was also no statistically significant difference anymore between 7 s and 20 s videos (Table 2).

3.3. Eye tracking

The distribution across the four areas of interest (i.e., a measure

**Fig. 3.** Procedure to determine the geometric difference between actual and placed cars. The lane width was 3.5 m.

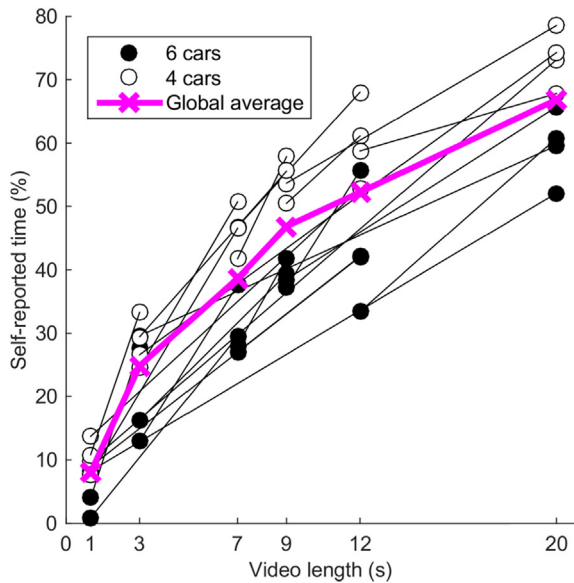


Fig. 4. Mean self-reported time sufficiency across the participants, for each of the 24 situations. The A and B groups are connected by a line.

of attention distribution) and the glance frequency to the mirrors (i.e., a measure of gaze activity) as a function of observation time are shown in Fig. 10. It can be seen that at the beginning (2–4 s), participants distributed their attention approximately equally between the front view and the mirrors. Near the end (7–20 s), however, participants were more likely to glance at the road than at the mirrors. Furthermore, in agreement with Over et al. (2007) and Unema et al. (2005), we found a decrease of glance frequency with increasing observation time.

3.4. Correlation between video length and dependent measures, partialling out objective scenario difficulty

For all dependent measures, differences occurred between

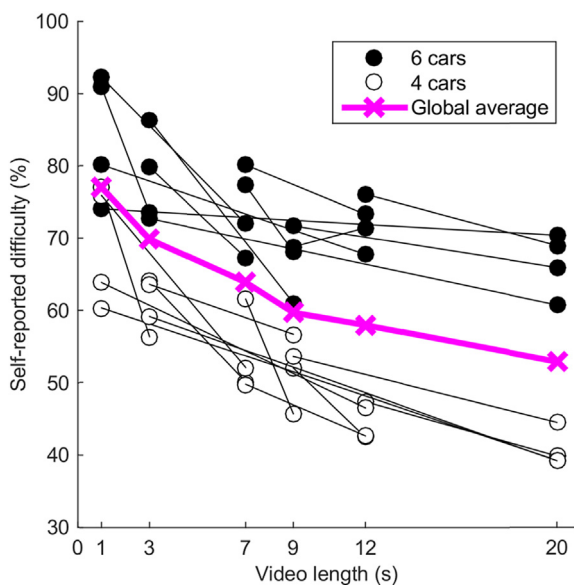


Fig. 5. Mean self-reported difficulty across the participants, for each of the 24 situations. The A and B groups are connected by a line.

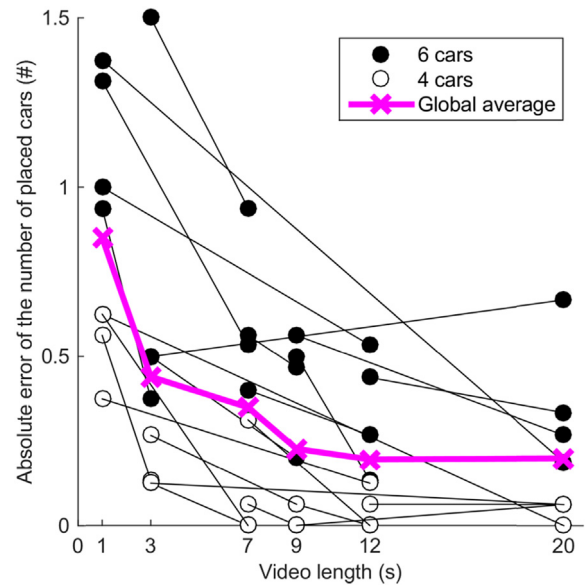


Fig. 6. Mean absolute error of the number of placed cars, for each of the 24 situations. The A and B groups are connected by a line.

situations, even when the video length and traffic density were the same (Figs. 4–9). Several characteristics of surrounding traffic might influence task performance. In particular, participants underestimated the distance to objects, and had more difficulty in estimating the position and speed of a car that is further away.

For this reason, we calculated an objective index of difficulty of the video, by summing the distances between the endpoints of 4/6 cars and the ego-vehicle. Table 3 shows the Spearman rank-order correlations between the mean score on the dependent measure per situation and the objective difficulty of the situation ($N = 24$). It can be seen that objective difficulty accounts for some of the situation-specific effects, in particular for the total distance error and total speed error. Furthermore, Table 3 shows the correlations between the dependent measures and video length. It can be seen

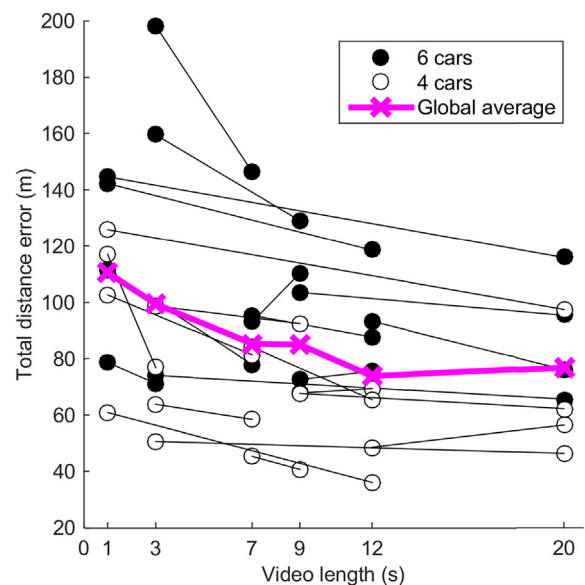


Fig. 7. Mean total distance error, for each of the 24 situations. The A and B groups are connected by a line.

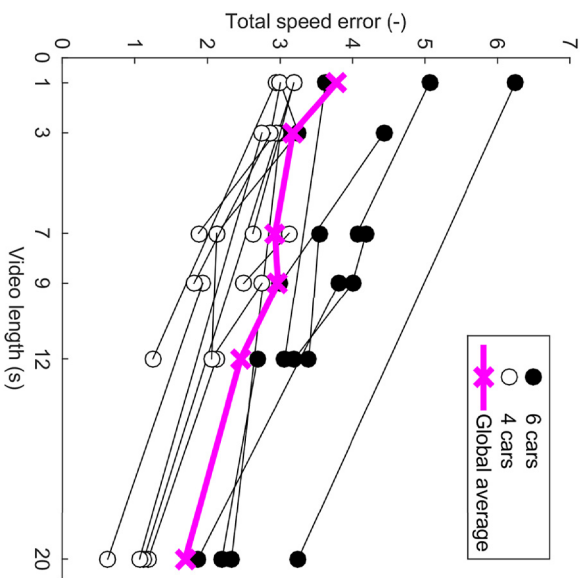


Fig. 8. Mean total speed error, for each of the 24 situations. The A and B groups are connected by a line.

that the strongest effect was observed for self-reported time sufficiency, whereas the weakest effect occurred for the geometric similarity, essentially mirroring the findings in Table 2. The third column of Table 3 shows again the correlations between the dependent measures and video length, but now after partialling out the objective difficulty. It can be seen that the partial correlations are slightly stronger than the zero-order correlations, in particular for the total distance error and total speed error. These findings indicate that when removing the effect of objective situation difficulty, the effects of video length becomes even stronger.

3.5. Effect of personal characteristics and driving experience

Previous research showed that in situations of transitions of

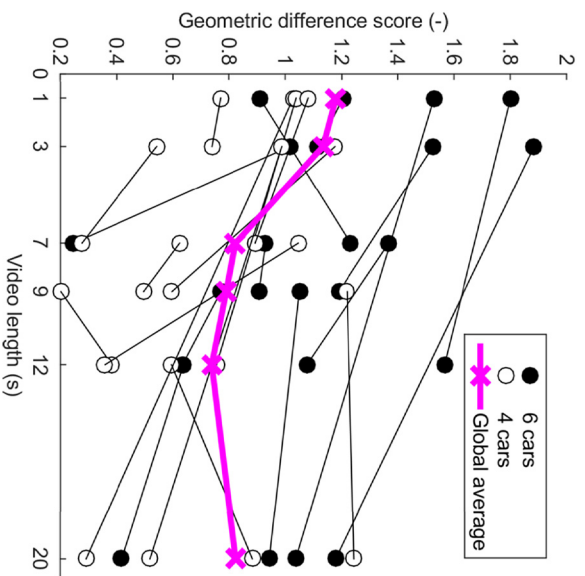


Fig. 9. Mean geometric difference score, for each of the 24 situations. The A and B groups are connected by a line.

Table 2

Result of the repeated measures ANOVA, and effect sizes (d_z) of paired comparisons between video lengths.

Measure	df	F^a	η^2	Paired comparisons between video lengths															
				1	1	1	1	1	3	3	3	3	7	7	7	9	9	12	A
				vs. 3	vs. 7	vs. 9	vs. 12	vs. 20	vs. 7	vs. 9	vs. 12	vs. 20	vs. 9	vs. 12	vs. 20	vs. 12	vs. 20	vs. 20	vs. B ^b
Time sufficiency	5,145	146.4	0.835	−1.69	−2.41	−2.35	−2.95	−4.93	−1.04	−1.44	−1.82	−3.01	−0.65	−1.27	−2.91	−0.48	−1.39	−1.07	24
Difficulty	5,150	24.9	0.454	0.56	0.95	1.12	1.32	1.33	0.51	0.87	0.90	1.03	0.34	0.65	0.79	0.26	0.45	0.33	1
Error in the number of placed cars	5,155	20.5	0.398	0.68	0.96	1.37	1.54	1.58	0.11	0.54	0.67	0.69	0.44	0.59	0.54	0.15	0.09	−0.07	3
Total distance error	5,155	11.4	0.269	0.39	0.88	0.95	1.46	1.19	0.16	0.18	0.58	0.52	0.05	0.58	0.36	0.60	0.32	−0.19	4
Total speed error	5,155	30.9	0.499	0.55	0.74	0.74	1.38	2.40	0.26	0.19	0.66	1.69	−0.03	0.56	1.47	0.49	1.43	0.78	1
Geometric difference	5,155	6.8	0.181	−0.01	0.45	0.74	0.61	0.51	0.56	0.71	0.76	0.66	0.12	0.20	0.04	0.10	−0.06	−0.15	4

Boldface = the effect between these two video lengths was statistically significant according to a paired-samples t -test with Bonferroni correction ($\alpha = 0.05/15$).

^a $p < .001$ for each of the six measures according to a repeated measures ANOVA.

^b This column shows for how many of the 24 situations the mean score in the longer video is greater than the mean score in the shorter video. This comparison was made for the participants of group A versus the participants of group B. The videos of groups A and B had identical endpoints of the cars.

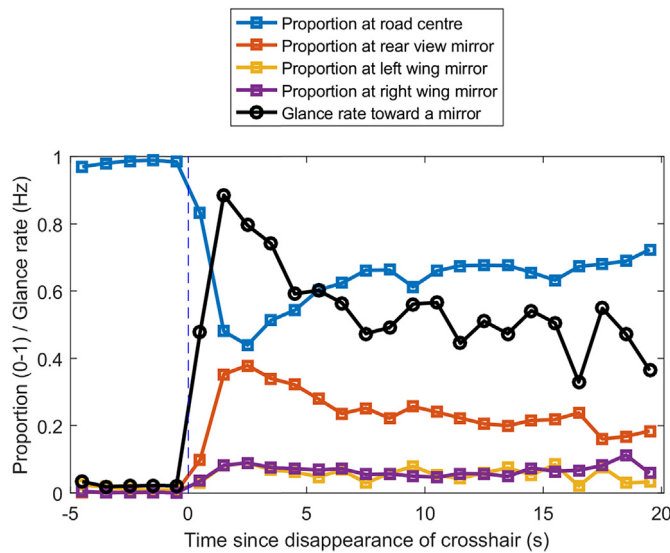


Fig. 10. Average net dwell time proportion across four areas of interest (road centre, rear view mirror, left wing mirror, right wing mirror) in 1 s increments. Also shown is the glance frequency towards one of the three mirrors. Note that these averages get noisier with increasing observation time because less and less data are available. That is, for the first second, data are available for 1, 3, 7, 9, 12, and 20 s videos. For the last 8 s, only data from 20 s videos are available.

control, experienced drivers are quicker to achieve SA than novices do (Wright et al., 2016). It has also been found that experienced drivers perform better than novices when asked to estimate the number of cars around them (Baumann et al., 2008).

We performed a correlational analysis to explore whether driving experience relates to performance at the SA task. Table 4

shows that participants who drive more frequently had a statistically significantly lower total distance and speed error. It is further interesting to observe that the subjective measures (time sufficiency and difficulty) exhibited a statistically significant correlation with each other, but did not correlate significantly with any of the objective measures. This indicates that participants who thought they did not have enough time or reported that the task was difficult did not necessarily perform poorly at the reproduction task. Finally, the eye-scanning measures were not significantly correlated with the performance measures.

4. Discussion and conclusion

The aim of this research was to assess the effect of viewing time on SA operationalized as reproduction performance of a traffic situation. We applied three complementary types of statistical analyses: (1) a main effect of video length, (2) a between-subjects comparison of participants; the videos for these two groups were identical except that one of the two videos started earlier than the other, and (3) a correlation between video length and task performance while removing the effect of objective situation difficulty. The first analysis is useful for gauging the overall effect of video length, whereas the latter two methods account for situation-specific variance that may exist and thus serve as a confirmation.

The self-report measures confirmed that participants found the task more demanding if less time was available. This manipulation check validates our experiment and shows that the amount of time was manipulated in a range for which participants experience strong and consistent effects in time and difficulty.

Concerning the accuracy of the number of cars and the positions of the placed cars, the largest effects of video length were obtained up to 7 s and 12 s, respectively. A 7 s threshold is in line with previous driving simulator research on take-over requests, in which

Table 3

Spearman rank-order correlations between dependent measures, and measures of objective situation difficulty and video length ($N = 24$ situations).

Measure	ρ with objective difficulty (group A/group B)	ρ with video length (group A/group B)	Partial ρ with video length (group A/group B)
Time sufficiency (%)	-0.10/-0.05	0.91/0.93	0.90/0.93
Difficulty (%)	0.42/0.43	-0.66/-0.40	-0.70/-0.49
Error in the number of placed cars (#)	0.20/0.52	-0.51/-0.52	-0.49/-0.62
Total distance error (m)	0.67/0.78	-0.35/-0.23	-0.46/-0.69
Total speed error (-)	0.52/0.31	-0.51/-0.57	-0.68/-0.73
Geometric difference (-)	0.35/0.28	-0.25/-0.39	-0.32/-0.45

Table 4

Spearman rank-order correlation matrix among personal characteristics and dependent measures.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age (years)													
2. Gender (0 = female, 1 = male)	-0.03												
3. License age (years)	0.26	0.11											
4. Licensure (years)	0.56	-0.15	-0.48										
5. Driving Frequency (0 = Every day, 5 = never)	0.26	-0.23	-0.12	0.38									
6. Yearly mileage (0 = 0 km, 10 = more than 100,000 km)	-0.06	0.09	0.28	-0.29	-0.57								
7. Time sufficiency (%)	0.08	0.14	0.30	-0.08	0.20	-0.06							
8. Difficulty (%)	-0.22	-0.32	-0.24	-0.14	0.09	0.00	-0.55						
9. Error in the number of placed cars (#)	0.29	-0.25	0.21	0.07	0.32	0.03	-0.07	0.13					
10. Total distance error (m)	0.08	-0.32	-0.13	0.11	0.41	-0.03	-0.14	0.21	0.57				
11. Total speed error (-)	0.30	-0.21	-0.20	0.33	0.35	-0.35	-0.21	0.10	0.25	0.32			
12. Geometric difference (-)	0.06	-0.16	-0.07	0.21	0.34	-0.09	0.19	-0.14	-0.01	0.35	0.37		
13. Mean mirror glance frequency (Hz)	-0.24	-0.40	-0.05	0.09	0.23	-0.14	-0.13	-0.05	0.13	0.25	0.10	0.17	
14. Road centre gaze proportion (0-1)	0.26	0.40	0.08	-0.03	-0.01	-0.18	0.23	-0.21	-0.16	-0.31	0.09	-0.22	-0.79

Note. Correlations that are significantly ($p < 0.05$) different from 0 are indicated in boldface. $N = 34$ for measures 1–6, $N = 30$ for measure 7, $N = 31$ for measure 8, $N = 32$ for measures 9–12, $N = 21$ for measures 13 and 14.

a lead time of 7 s was found to be sufficient for taking over control in a basic traffic scenario (e.g., Gold et al., 2013). Improvements in relative speed perception, on the other hand, were obtained up to 20 s. The apparent lack of saturation of the accuracy of the speed estimation can be explained by the fact that humans have to deduce speed from *changes* in a scene. In other words, people first need time to scan the environment to establish where the cars are, and only then can use their time for tracking these cars. This pattern was also reflected in the eye-tracking data, showing a decrease of glance frequency with viewing time. Future research may explore this aspect further; for example, in our study the video lengths were presented in randomized order; if participants know how much time they have on beforehand, they may exhibit a different viewing behaviour (Huebner and Gegenfurtner, 2010).

Additionally, we applied a measure of geometric difference between the placed and actual cars. This method establishes the accuracy of placed cars in relative terms and may be particularly suited for assessing whether participants have perceived the overall layout of a situation. One possible weakness of the geometric difference method as well as the total distance error may be that longitudinal distance dominates lateral distance, and so these methods are not particularly sensitive to mistakes whereby the participant placed a car in the wrong lane. Gugerty (1997) solved this issue by applying more weight to lateral errors than to longitudinal errors. The uniqueness of our approach as compared to Gugerty's is that it did not make use of any such weighting factors; therefore, we expect that our non-parametric method can be applied to other spatial memory or SA studies without adjustment. A follow-up analysis showed a statistically significant difference between 1 s and 3 s videos regarding the total lateral distance error, but no further improvements for 3 s videos and beyond. The lack of statistically significant effects beyond 3 s can be explained by the fact that once a participant has identified a car, he/she can easily remember whether this car was in the left, middle, or right lane, because these are only three discrete categories.

Endsley's model of SA includes (1) perception, (2) comprehension, and (3) projection as three ascending levels (Endsley, 2015). One may argue that our experiment focused predominantly on level 1 SA. Indeed, participants were asked to reproduce the situation without having to comprehend the relevance of the cars in the environment. In reality, certain cars may be more relevant than others when it comes to safety margins and controllability. For example, in real traffic, cars in the back may often be safely ignored, whereas cars in front could be on a collision course or bear direct consequence for future action. Moreover, participants in our experiment did not perform driving-related decisions or actions, which may normally be performed simultaneously with the assessment of the situation and therefore interact with the time required to obtain SA. For example, Gugerty (1997) found that if drivers were in control of the driving task (with keyboard arrows) they remembered hazardous car better as compared to when they were in passenger mode, whereas Mackenzie and Harris (2015) showed that participants were slower to detect hazards when driving themselves as opposed to passively viewing a video.

Even though our experiment emphasized level 1 SA, a unique aspect of our research is that it also measured an important facet of level 3 SA by means of the assessment of relative speed. In fact, by having knowledge of the distance and speeds of objects, it is possible to project how a situation will unfold. Recent research indicates that queries regarding level 3 SA awareness ('what happens next') may be particularly valid in the sense that they discriminate between inexperienced and experienced drivers (Jackson et al., 2009). These observations are in line with our results, which showed that participants who drove more frequently performed better at the distance and speed estimation tasks

(Table 4). However, our study was of a simple and highly controlled design in which cars did not change lanes or speed, and level 3 SA was probed indirectly; participants were not directly asked what would happen next. Follow-up research could assess such aspects of level 2 and 3 SA in greater depth. For example, it is possible to use eye tracking in scenarios involving transfer of control to measure the detection of latent hazards (Samuel et al., 2016; Wright et al., 2016) or hazard precursors, which are cues in the environment that place critical demands on the driver's understanding of an unfolding situation (Garay-Vega and Fisher, 2005; Underwood et al., 2011). Moreover, research could include lane changes, and vehicle acceleration and deceleration, such as cars who are braking for an emergency. It should be noted, however, that humans are unable to perceive acceleration directly but rather infer acceleration from changes in speed over time (Brouwer et al., 2002; Gottsdanker et al., 1961).

For all dependent measures, large differences were found between situations, even if these situations had the same video length (see Figs. 4–9). Several characteristics of the traffic might influence task performance. As illustrated by the results in Table 3, interpreting the behaviour of a car that is further away becomes increasingly harder as distance increases. Moreover, we showed that the higher the traffic density, the harder it is to reproduce the traffic situation. These results indicate that it is not possible to give a generic recommendation of the lead time that is required for issuing a take-over request; the required time strongly depends on the complexity of the traffic situation. This finding is in line with Gold et al. (2016) who showed that the higher the traffic density, the longer the take-over time, defined as the first measurable steering or braking response after receiving a take-over request. However, the difference between our study and most of the available empirical research on this topic is that we obtained systematic insight into the cognitive aspects of building up SA as a function of time. With a few notable exceptions (e.g., Samuel et al., 2016; Wright et al., 2016), most of the available research adopts a behaviourist approach by quantifying reaction times and steering/braking responses in take-over situations (e.g., Gold et al., 2013; see De Winter et al., 2014; for a review).

Several limitations have to be taken into account when interpreting the results. First, the traffic situations were relatively simple and did not involve features such as lane changes, curves, or decelerating vehicles. Second, the monitor provided a small field of view and low immersion. The experiment did not allow for head movement, and certain monocular and binocular depth cues were lacking in the computer animations. These hardware features may affect both task performance and eye activity. Third, this study was conducted with participants at an engineering university. Engineering students are not representative of the general population and are known to have above-average spatial skills (Wai et al., 2009). It has been previously shown that older drivers showed similar take-over times as younger drivers (Körber et al., 2016), despite the fact that biological age has a strong negative correlation with memory and spatial task performance (Salthouse, 2009). Possibly, having many years of driving experience may protect against age-related cognitive decline. The topic of individual differences in SA awareness is a promising topic of further research (Gugerty and Tirre, 2000). Fourth, in our experiment participants were not sitting in an actual automated car and knew they would be required to reproduce the traffic layout. Furthermore, no drowsiness or secondary tasks were induced, which are conditions that may have important effects on how quickly drivers gain SA, and how effectively they take-over control (Borowsky and Oron-Gilad, 2016; Feldhütter et al., 2016; Gibson et al., 2016; Neubauer et al., 2014; Schmidt et al., 2016; Schömig et al., 2015; Zeeb et al., 2016). Because of these factors, our results may underestimate the

time required for obtaining SA. However, despite the obvious differences between an automated driving task and the present experiment, it should be noted that the tasks are similar: In automated driving, the driver may also be prompted to monitor the traffic situation (without necessarily touching the steering wheel) to regain SA. Fifth, inherent to SAGAT-type methods, participants might have started to forget the driving situation when completing the reproduction task. For example, an analysis by Gugerty (1998) showed that participants tended to forget the location of cars during the time it took them to report the locations of the cars. One improvement would be to replace the slider bars with a less time-consuming interface where participants can drag the cars directly with the use of a mouse or by touch.

In conclusion, this research showed that participants need a few seconds in order to estimate the basic topology of a situation, but substantial improvements in speed estimation were still achieved between 12 and 20 s videos. These findings may have important consequences for the development of automated cars, in particular automated driving systems for which evidence starts to grow that humans are not well adapted to a task where they have to be able to regain control in a limited time frame (Casner et al., 2016; Norman, 2015; Poulin et al., 2015; U.S. Senate Committee on Commerce, Science and Transportation, 2016). Future research could use the method applied in this study in an interactive driving simulator or head-mounted display with integrated eye-tracker.

Acknowledgements

The research presented in this paper is supported by the Marie Curie Initial Training Network (ITN) HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817).

References

- Arkin, E.M., Chew, L.P., Huttenlocher, D.P., Kedem, K., Mitchell, J.S., 1991. An efficiently computable metric for comparing polygonal shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 13, 209–216.
- Baumann, M., Franke, T., Krems, J.F., 2008. The Effect of Experience, Relevance, and Interruption Duration on Drivers' Mental Representation of a Traffic Situation. *Human Factors for Assistance and Automation*. Shaker Publishing, Maastricht, The Netherlands, pp. 141–152.
- Borji, A., Itti, L., 2013. State-of-the-art in visual attention modelling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 185–207.
- Borowsky, A., Oron-Gilad, T., 2016. The effects of automation failure and secondary task on drivers' ability to mitigate hazards in highly or semi-automated vehicles. *Adv. Transp. Stud.* 1, 59–70.
- Brouwer, A.M., Brenner, E., Smeets, J.B., 2002. Perception of acceleration with short presentation times: can acceleration be used in interception? *Percept. Psychophys.* 64, 1160–1168.
- Casner, S.M., Hutchins, E.L., Norman, D., 2016. The challenges of partially automated driving. *Commun. ACM* 59, 70–77.
- Clark, H., Feng, J., 2015. Semi-autonomous vehicles examining driver performance during the take-over. *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 59, 781–785.
- Conover, W.J., Iman, R.L., 1981. Rank transformations as a bridge between parametric and nonparametric statistics. *Am. Statistician* 35, 124–129.
- Coster, X.A., 2015. A Method for Determining How Much Time People Need to Understand Traffic Situations (MSc thesis). Delft University of Technology, The Netherlands.
- Crundall, D., Chapman, P., Phelps, N., Underwood, G., 2003. Eye movements and hazard perception in police pursuit and emergency response driving. *J. Exp. Psychol. Appl.* 9, 163–174.
- Desmond, P., Hancock, P., Monette, J., 1998. Fatigue and automation-induced impairments in simulated driving performance. *Transp. Res. Rec. J. Transp. Res. Board* 1628, 8–14.
- 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. *Transp. Res. Part F Traffic Psychol. Behav.* 27, 196–217.
- Durso, F.T., Truitt, T.R., Hackworth, C.A., Crutchfield, J.M., Nikolic, D., Moertl, P.M., Manning, C.A., 1995. Expertise and chess: a pilot study comparing situation awareness methodologies. In: Garland, D., Endsley, M. (Eds.), *Proceedings of the International Conference on Experimental Analysis and Measurement of Situation Awareness*, pp. 295–303.
- Durso, F.T., Bleckley, M.K., Dattel, A.R., 2006. Does situation awareness add to the validity of cognitive tests? *Hum. Factors* 48, 721–733.
- Endsley, M.R., 2000. Direct measurement of situation awareness: validity and use of SAGAT. In: Endsley, M.R., Garland, D.J. (Eds.), *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 147–174.
- Endsley, M.R., 2015. Situation awareness: operationally necessary and scientifically grounded. *Cognition, Technol. Work* 17, 163–167.
- 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. *Behav. Res. Methods* 39, 175–191.
- Feldhütter, A., Gold, C., Schneider, S., Bengler, K., 2016. How the duration of automated driving influences take-over performance and gaze behavior. In: *Gesellschaft für Arbeitswissenschaft e. V. (Hrsg.), 62. Kongress der Gesellschaft für Arbeitswissenschaft*. GfA Press, Dortmund, Germany.
- Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., Menke, L., 2016. Which eye tracker is right for your research? Performance evaluation of several cost variant eye trackers. *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 60, 1240–1244.
- Garay-Vega, L., & Fisher, D. L. (2005). Can novice drivers recognize foreshadowing risks as easily as experienced drivers. *Proceedings of 3rd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Rockport, ME.
- Gardner, A.K., Kosemund, M., Martinez, J., 2015. Examining the feasibility and predictive validity of the SAGAT tool to assess situational awareness among surgical trainees. *J. Am. Coll. Surg.* 221, e9–e10.
- Garland, D.J., Endsley, M.R., 1995. *Experimental Analysis and Measurement of Situation Awareness*. EmbryRiddle: Aeronautical University Press, Daytona Beach, FL.
- Gasser, T., Westhoff, D., 2012. BASt-study: Definitions of Automation and Legal Issues in Germany. TRB Road Vehicle Automation Workshop, Irvine, CA, USA. Retrieved from: <http://onlinepubs.trb.org/onlinepubs/conferences/2012/Automation/presentations/Gasser.pdf>.
- Gibson, M., Lee, J., Venkatraman, V., Price, M., Lewis, J., Montgomery, O., Foley, J., 2016. Situation awareness, scenarios, and secondary tasks: measuring driver performance and safety margins in highly automated vehicles. *SAE Int. J. Passeng. Cars-Electronic Electr. Syst.* 9, 237–242.
- Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. "Take over!" How long does it take to get the driver back into the loop? *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 57, 1938–1942.
- Gold, C., Körber, M., Lechner, D., Bengler, K., 2016. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Hum. Factors J. Hum. Factors Ergonomics Soc.* 58, 642–652.
- Gottsdanker, R., Frick, J.W., Lockard, R.B., 1961. Identifying the acceleration of visual targets. *Br. J. Psychol.* 52, 31–42.
- Gugerty, L.J., 1997. Situation awareness during driving: explicit and implicit knowledge in dynamic spatial memory. *J. Exp. Psychol. Appl.* 3, 42–66.
- Gugerty, L., 1998. Evidence from a partial report task for forgetting in dynamic spatial memory. *Hum. Factors J. Hum. Factors Ergonomics Soc.* 40, 498–508.
- Gugerty, L.J., Tirre, W.C., 2000. Individual difference in situation awareness. In: Endsley, M.R., Garland, D.J. (Eds.), *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 249–276.
- Gugerty, L., 2011. Situation awareness in driving. In: Lee, J., Rizzo, M., Fischer, D., Caird, J. (Eds.), *Handbook for Driving Simulation in Engineering, Medicine and Psychology*. CRC Press, Boca Roca, FL (pp. 19–265–19–272).
- Gutzwiler, R.S., Clegg, B.A., 2013. The role of working memory in levels of situation awareness. *J. Cognitive Eng. Decis. Mak.* 7, 141–154.
- Henderson, J.M., 2003. Human gaze control during real-world scene perception. *Trends Cognitive Sci.* 7, 498–504.
- Horswill, M.S., McKenna, F.P., 2004. Drivers' hazard perception ability: situation awareness on the road. In: Banbury, S., Tremblay, S. (Eds.), *A Cognitive Approach to Situation Awareness*. Ashgate, Aldershot, United Kingdom, pp. 155–175.
- Hosking, S.G., Liu, C.C., Bayly, M., 2010. The visual search patterns and hazard responses of experienced and inexperienced motorcycle riders. *Accid. Anal. Prev.* 42, 196–202.
- Huebner, G.M., Gegenfurtner, K.R., 2010. Effects of viewing time, fixations, and viewing strategies on visual memory for briefly presented natural objects. *Q. J. Exp. Psychol.* 63, 1398–1413.
- Irwin, D.E., Zelinsky, G.J., 2002. Eye movements and scene perception: memory for things observed. *Percept. Psychophys.* 64, 882–895.
- Itti, L., Koch, C., 2001. Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203.
- Jackson, L., Chapman, P., Crundall, D., 2009. What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics* 52, 154–164.
- Kerschbaum, P., Lorenz, L., & Bengler, K. (2015). A transforming steering wheel for highly automated cars. *Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1287–1292). Seoul.
- Körber, M., Weißgerber, T., Kalb, L., Blaschke, C., Farid, M., 2015. Prediction of take-over time in highly automated driving by two psychometric tests. *Dyna* 82, 195–201.
- Körber, M., Gold, C., Lechner, D., Bengler, K., 2016. The influence of age on the take-over of vehicle control in highly automated driving. *Transp. Res. Part F Traffic Psychol. Behav.* 39, 19–32.
- Kyriakidis, M., Happee, R., De Winter, J.C.F., 2015. Public opinion on automated driving: results of an international questionnaire among 5000 respondents.

- Transp. Res. Part F Traffic Psychol. Behav. 32, 127–140.
- Loft, S., Morrell, D.B., Huf, S., 2013. Using the situation present assessment method to measure situation awareness in simulated submarine track management. *Int. J. Hum. Factors Ergonomics* 2, 33–48.
- Loft, S., Bowden, V., Braithwaite, J., Morrell, D.B., Huf, S., Durso, F.T., 2015. Situation awareness measures for simulated submarine track management. *Hum. Factors J. Hum. Factors Ergonomics Soc.* 57, 298–310.
- Lu, Z., Happee, R., Cabral, C.D.D., Kyriakidis, M., De Winter, J.C.F., 2016. Human factors of transitions in automated driving: A general framework and literature survey. *Transp. Res. Part F Traffic Psychol. Behav.* 43, 183–198.
- Mackenzie, A.K., Harris, J.M., 2015. Eye movements and hazard perception in active and passive driving. *Vis. Cogn.* 23, 736–757.
- Martelaro, N., Sirkin, D., & Ju, W. (2015). DAZE: a real-time situation awareness measurement tool for driving. *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 158–163.
- Melcher, V., Rauh, S., Diederichs, F., Bauer, W., 2015. Take-over requests for automated driving. *Procedia Manuf.* 3, 2867–2873.
- Merat, N., Jamson, A.H., Lai, F.C., Daly, M., Carsten, O.M., 2014. Transition to manual: driver behaviour when resuming control from a highly automated vehicle. *Transp. Res. Part F Traffic Psychol. Behav.* 27, 274–282.
- Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*.
- Moore, K., Gugerty, L., 2010. Development of a novel measure of situation awareness: the case for eye movement analysis. *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 54, 1650–1654.
- Neubauer, C., Matthews, G., Saxby, D., 2014. Fatigue in the automated vehicle do games and conversation distract or energize the driver? *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 58, 2053–2057.
- Norman, D.A., 2015. The human side of automation. In: *Road Vehicle Automation*, vol. 2. Springer International Publishing, pp. 73–79.
- Oestmann, J.W., Greene, R., Kushner, D.C., Bourgouin, P.M., Linetsky, L., Llewellyn, H.J., 1988. Lung lesions: correlation between viewing time and detection. *Radiology* 166, 451–453.
- Over, E.A.B., Hooge, I.T.C., Vlaskamp, B.N.S., Erkelens, C.J., 2007. Coarse-to-fine eye movement strategy in visual search. *Vis. Res.* 47, 2272–2280.
- Pierce, R.S., 2012. The effect of SPAM administration during a dynamic simulation. *Hum. Factors J. Hum. Factors Ergonomics Soc.* 54, 838–848.
- Poulin, C., Stanton, N.A., Cebon, D., Epple, W., 2015. Responses to: autonomous vehicles. *Ingenia* 62, 8–11.
- Pylyshyn, Z.W., Storm, R.W., 1988. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197.
- Rayner, K., 2009. Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506.
- SAE International, 2014. *Taxonomy and Definitions for Terms Related to On-road Motor Vehicle Automated Driving Systems (Standard No. J3016)*. SAE International. Retrieved from: http://standards.sae.org/j3016_201401/.
- Salmon, P., Stanton, N., Walker, G., Green, D., 2006. Situation awareness measurement: a review of applicability for C4i environments. *Appl. Ergon.* 37, 225–238.
- Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D., Ladva, D., Rafferty, L., Young, M., 2009. Measuring situation awareness in complex systems: comparison of measures study. *Int. J. Industrial Ergonomics* 39, 490–500.
- Salthouse, T.A., 2009. When does age-related cognitive decline begin? *Neurobiol. Aging* 30, 507–514.
- Samuel, S., Borowsky, A., Zilberstein, S., Fisher, D.L., 2016. Minimum Time to Situation Awareness in Scenarios Involving Transfer of Control from Automated Driving Suite. *Transportation Research Board 95th Annual Meeting*, 16–3788.
- Schmidt, J.S., Stolzmann, W., Karrer-Gauß, K., 2016. Experimental Evaluation of Different Request Intervals for a Driver Alertness Device for Conditionally Automated Driving with Induced Drowsiness. *Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel*.
- Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., Othersen, I., 2015. The interaction between highly automated driving and the development of drowsiness. *Procedia Manuf.* 3, 6652–6659.
- Stanton, N.A., Salmon, P.M., Walker, G.H., 2015. Let the reader decide: a paradigm shift for situation awareness in socio-technical systems. *J. Cognitive Eng. Decis. Mak.* 9, 44–50.
- Tass International, 2015. *PreScan: Simulation of ADAS & Active Safety*. Retrieved from: <https://www.tassinternational.com/prescan>.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381, 520–522.
- Underwood, G., Chapman, P., Bowden, K., Crundall, D., 2002. Visual search while driving: skill and awareness during inspection of the scene. *Transp. Res. Part F Traffic Psychol. Behav.* 5, 87–97.
- Underwood, G., Crundall, D., Chapman, P., 2011. Driving simulator validation with hazard perception. *Transp. Res. Part F Traffic Psychol. Behav.* 14, 435–446.
- Underwood, G., Ngai, A., Underwood, J., 2013. Driving experience and situation awareness in hazard detection. *Saf. Sci.* 56, 29–35.
- Underwood, S.E., 2014. *Automated Vehicles Forecast Vehicle Symposium Opinion Survey*. Presentation at the 2014 Automated Vehicles Symposium, San Francisco, CA.
- Unema, P.J., Pannasch, S., Joos, M., Velichkovsky, B.M., 2005. Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. *Vis. Cogn.* 12, 473–494.
- U.S. Senate Committee on Commerce, Science, & Transportation, 2016, March 15. *Hands off: the Future of Self-driving Cars (Testimony of M. L. Cummings)*.
- Van den Beukel, A.P., Van der Voort, M.C., 2013. The Influence of Time-criticality on Situation Awareness when Retrieving Human Control after Automated Driving. *16th International IEEE Annual Conference on Intelligent Transportation System, The Hague, the Netherlands*, pp. 2000–2005.
- Velichkovsky, B.M., Rothert, A., Kopf, M., Dornhöfer, S.M., Joos, M., 2002. Towards an express-diagnostics for level of processing and hazard perception. *Transp. Res. Part F Traffic Psychol. Behav.* 5, 145–156.
- Wai, J., Lubinski, D., Benbow, C.P., 2009. Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835.
- Wright, T., Samuel, S., Borowsky, A., Zilberstein, S., Fisher, D.L., 2016. Experienced drivers are quicker to achieve situation awareness than inexperienced drivers in situations of transfer of control within a Level 3 autonomous environment. *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 60, 270–273.
- 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. *Accid. Anal. Prev.* 92, 230–239.