The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance

de Winter, Joost; Stanton, NA; Price, J.S.; Mistry, H.

DOI
10.1504/IJVD.2016.076736

Publication date
2016

Document Version
Final published version

Published in
International Journal of Vehicle Design: the journal of vehicle engineering and components

The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance

J.C.F. de Winter*

Faculty of Mechanical, Maritime and Materials Engineering,
Department of BioMechanical Engineering,
Delft University of Technology,
Mekelweg 2, 2628 CD Delft, The Netherlands
Email: j.c.f.dewinter@tudelft.nl
*Corresponding author

N.A. Stanton, J.S. Price and H. Mistry

Transportation Research Group,
Faculty of Engineering and the Environment,
Boldrewood Innovation Campus,
University of Southampton,
Burgess Road, Southampton, SO16 7QF, UK
Email: n.stanton@soton.ac.uk
Email: harveymistry@gmail.com
Email: j.s.price@soton.ac.uk

Abstract: Until automated cars function perfectly, drivers will have to take over control when automation fails or reaches its functional limits. Two simulator experiments ($N = 24$ and 27) were conducted, each testing four automation levels ranging from manual control (MC) to highly automated driving. In both experiments, participants about once every 3 min experienced an event that required intervention. Participants performed a secondary divided attention task while driving. Automation generally resulted in improved secondary task performance and reduced self-reported physical demand and effort as compared to MC. However, automated speed control was experienced as more frustrating than MC. Participants responded quickly to the events when the stimulus was salient (i.e., stop sign, crossing pedestrian, and braking lead car), but often failed to react to an automation failure when their vehicle was driving slowly. In conclusion, driving with imperfect automation can be frustrating, even though mental and physical demands are reduced.

Keywords: workload; automated driving; secondary task; level of automation; critical events; ACC; adaptive cruise control; automated steering; human factors.

Reference to this paper should be made as follows: de Winter, J.C.F., Stanton, N.A., Price, J.S. and Mistry, H. (2016) 'The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance', *Int. J. Vehicle Design*, Vol. 70, No. 4, pp.297–324.

Biographical notes: Joost C.F. de Winter received his MSc in Aerospace Engineering and PhD (cum laude) from the Delft University of Technology, Delft, The Netherlands, in 2004 and 2009, respectively. He is currently an Assistant Professor at the Department of Mechanical Engineering, Delft University of Technology. His research interest is human factors, including the study of individual differences and driver behaviour modelling. He is also involved in the Marie Curie ITN projects HFauto: Human Factors of automated driving, and Motorist: MOTOrcycle Rider Integrated Safety.

Neville A. Stanton, PhD, DSc, is both a Chartered Psychologist and a Chartered Engineer and holds the Chair in Human Factors in the Faculty of Engineering and the Environment at the University of Southampton. He has Psychology, Applied Psychology and Human Factors Engineering. His research interests include modelling, predicting and analysing human performance in transport systems as well as designing interfaces between humans and machines. He has been working on cockpit design in automobiles and aircraft over the past 25 years, in a variety of automation projects. He has published over 30 books and 240 journal papers on ergonomics and human factors, and is currently an Editor of the peer-reviewed journal *Ergonomics*.

Joshua S. Price received his BE in Engineering Science from the School of Civil Engineering and the Environment, University of Southampton in 2011. In his BE dissertation, he investigated the impact of increasing levels of automation on urban driving performance using a driving simulator. He is now pursuing an Engineering Doctorate under the supervision of Prof. N.A. Stanton and Dr. B.J. Waterson (University of Southampton) and Mr. I. Snell (Siemens). In his research, he aims to investigate the benefits a human factors design may have when applied to the next generation of traffic management systems.

H. Mistry received his ME in Civil Engineering from the Faculty of Engineering and the Environment, University of Southampton in 2013. For his third-year dissertation, he explored “The impact of automation technology on driver performance and workload” whereas his fourth-year design project concerned the “Feasibility study of wave energy in Alderney”. Since graduating, he has been working as a Structural Engineer, first at Bombardier Transportation and currently at Buxton Associates. He is actively involved in the Institute of Structural Engineers, advancing the interests of young members and to promote structural engineering to the wider community in London and the surrounding counties.

1 Introduction

In the last decades, the driving task has become increasingly automated. Adaptive cruise control (ACC), a system that partly automates longitudinal vehicle motion, was first introduced in the 1990s. Our field is now transitioning towards highly automated driving (HAD), a concept in which both longitudinal and lateral control are taken over by the automation (Akamatsu et al., 2013; Gasser and Westhoff, 2012; Kircher et al., 2014; Merat and Lee, 2012). Considering that over 50% of road traffic accidents are attributable to human errors such as misjudgement, inattention and distraction (Klauer et al., 2006; Storie, 1977; Treat et al., 1979), HAD could provide enormous benefits to society.

Previous studies in driving simulators and real cars have shown that ACC contributes to a reduction of workload (measured with self-reports or secondary tasks) compared to manual control (MC) (Carsten et al., 2012; Nowakowski et al., 2010; Rudin-Brown and Parker, 2004; Young and Stanton, 2004; for a review see De Winter et al., 2014). HAD reduces self-reported workload to an even greater extent (Damböck et al., 2013; Young and Stanton, 2007a). For example, a driving simulator study by De Waard et al. (1999) found that the mean score on the Rating Scale Mental Effort after automated platooning with 1 s headway was 12 on a range from 0 to 150 (corresponding to 'almost no effort'), substantially lower than the score of 37 (corresponding to 'some effort') for MC. Similarly, two simulator studies conducted by Young and Stanton (2004; see also Young, 2000) measured workload using the NASA Task Load Index (TLX) and found that the overall workload for HAD was 12%, considerably lower than the scores for MC (57%), ACC (44%) and automated steering (AS; 41%) (0% = very low, 100% = very high).

In principle, driving in a highly automated vehicle is undemanding. Drivers can comfortably read a book, watch a movie or even fall asleep, and empirical research indicates that these non-driving activities indeed occur in HAD (Carsten et al., 2012; Llaneras et al., 2013; Omae et al., 2005). However, there is still an important role for the human in HAD, because there are situations where the driver will have to resume MC. Such situations may occur because of functional limits of the HAD system or because of sensor/computer failure. Ironically, HAD systems that intend to improve safety can actually be very dangerous, especially if a driver has to reclaim control within a short time span (Flemisch et al., 2008; Stanton et al., 1997). Hence, automation systems that reduce workload may actually evoke temporal bursts of high workload. The phrase "a high degree of automation tends to reduce supervisory work to 99% boredom and 1% sudden terror" (Bibby et al., 1975, p.4) captures the irony that automation makes the routine tasks easier and the non-routine tasks harder (see also Bainbridge, 1983; Hancock and Krueger, 2010; Hollnagel and Woods, 2005).

The NASA TLX, which is the most widely used workload questionnaire, measures workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort and frustration. These six items share positive correlations and thereby represent the overall workload construct (Hart and Staveland, 1988). However, each item also carries uniqueness, and can diagnose the operator's state in specific task environments. Many researchers who have studied the effect of automated driving system reported the overall workload score only, not the scores on the individual items (e.g., Saxby et al., 2013; Young and Stanton, 2007a). That is, much of the previous research does not recognise that workload is a multi-dimensional construct (and see Tsang and Velazquez, 1996; Yeh and Wickens, 1988).

In the present study, we investigated driver workload and response times to critical events in two similar but not identical driving simulator experiments. The main difference between the two experiments was that the first was conducted in a low fidelity simulator, whereas the second was conducted in a medium fidelity simulator. In each of the two experiments, participants drove four short sessions in an urban environment, each session with a different level of automation, ranging from MC to HAD. The reason to investigate multiple levels of automation was based on the earlier studies stating that automation should not be seen as a dichotomous 'all-or-none phenomenon' (Sheridan and Parasuraman, 2005). Studying multiple levels of automation provides a more comprehensive picture on human–automation interaction than comparing full automation with MC only (e.g., Endsley, 1999).

Our experiment differs from previous research on the same topic in the sense that the critical event rate was high: Human intervention was required about once every 3 min.

Previous research has mostly studied the drivers' response to rare (e.g., once a session) critical events. In such studies, drivers showed severely delayed response times compared to manual driving, especially when drivers were engaged in a secondary task that required them to take the eyes off the road (e.g., Bianchi Piccinini et al., 2015; De Waard et al., 1999; Flemisch et al., 2008; Hoedemaeker and Brookhuis, 1998; Körber et al., 2015; Louw et al., 2015; Rudin-Brown and Parker, 2004; Stanton et al., 1997, 2001; Stockert et al., 2015; Strand et al., 2014; see De Winter et al., 2014, for a review). A driving simulator study by Flemisch et al. (2008), for example, found that all participants veered off the road when the automation failed right before a curve while the participants were performing a task on an in-vehicle display. These previous studies are informative about what could happen in worst-case scenarios, and accordingly provide important knowledge for the design of takeover warning systems. However, in practice, as can be seen in traffic jam assistance systems that automate both longitudinal and lateral control, modern automated driving systems require almost constant eyes on the road (e.g., De Winter et al., 2014). Indeed, automated driving systems often reach their functional limits (e.g., when exceeding the speed envelope, when the camera system cannot track the lane, when there are sharp curves or when the sensor system does not detect objects on the road, such as crossing animals or pedestrians).

2 Methods

2.1 Experiment 1

2.1.1 Driving environment

The environment consisted of a 3.7 m wide single-lane two-way road with 12 traffic lights (four pedestrian crossings and eight junctions) and a mixture of mildly curved and straight sections. The traffic lights were spaced between 270 and 730 m apart. Three traffic lights were green; the remaining nine turned red when the driver approached. Traffic conditions were reasonably busy, with crossing cars at intersections and 20 oncoming cars. The mean distance between oncoming cars was 445 m ($SD = 280$ m). There were no vehicles in the driver's lane. Participants drove 7050 m (about 12 min of driving time) in each of the four sessions.

2.1.2 Automation conditions

The four automation conditions were:

- *Manual control (MC).* The driver was in control of steering (with the steering wheel) and speed (with the accelerator and brake pedal).
- *Partially automated steering (PAS).* The driver had full control of speed throughout the session. The automation controlled the steering when driving on junctions (including when approaching and leaving the junctions). Hence, the driver had full control of both speed and steering when travelling in between junctions. Accordingly, the PAS system resembled a type of automation that takes over steering control in demanding traffic scenarios (i.e., on/near junctions).

- *Partially automated driving (PAD).* The automation controlled both steering and speed when driving on junctions (including approaching and leaving the junction). The driver had full control of speed and steering when travelling in between junctions.
- *Highly automated driving (HAD).* The automation controlled both steering and speed throughout the session.

The automated speed (in PAD and HAD) maintained the vehicle at 48 km/h (30 mph) in between junctions, reduced speed on the approach to red light junctions with about 1.5 m/s^2 to 3 km/h (it did not stop completely), and increased speed with about 1.5 m/s^2 to 48 km/h when leaving these junctions. The automated steering (in PAS, PAD and HAD) kept the lateral position with respect to the lane centre at 0 m. The travelled distance under PAS/PAD control was 130 m per junction. All four sessions were driven using automatic transmission.

2.1.3 Critical events during driving

In the PAS, PAD and HAD conditions, a failure rate of the automation was introduced by giving control to the driver on approach of two out of 12 junctions. Specifically, for PAD and HAD, the vehicle automatically slowed down to 11 km/h, and the automation then turned off between 12 and 20 m before the junction so that unless the driver intervened, the vehicle would cross a red light. With PAS, the failure of the AS occurred at about 9 m in front of the junction. The automation failures did not have an immediate consequence on lateral performance, other than that the car could slowly drift away from the lane centre.

In addition, a critical event was included in the form of a pedestrian (in the MC, PAS and HAD conditions) or a Dalmatian dog (in the PAD condition) walking onto the road. This event was introduced to mimic a situation where the sensors did not detect a relatively small object. The pedestrian/dog was initially standing 8 m left of the centreline of the two-way road. Once the participant's vehicle was 7 s away, the pedestrian/dog began to walk towards the direction of the road at a speed that would cause a collision if the participant did not respond. The participant had to brake or swerve in order to avoid a crash. The HAD system was disabled for this event without a warning sign and was reactivated 100 m after deactivation. The locations of the three events (i.e., two automation failures and one pedestrian/dog) were different for each automation condition. The mean (*SD*) distance to the pedestrian/dog when the pedestrian/dog began to walk was 95 m (10), 94 m (9), 92 m (11) and 93 m (0) for the MC, PAS, PAD and HAD conditions, respectively.

Participants could not overrule the automation. Instead, the automation deactivated and reactivated at set points along the route. In other words, when a critical event had just occurred and manual intervention was necessary, drivers had full MC. Conversely, when no intervention was required, it was not possible to disable the automation.

2.1.4 Secondary divided attention task

A secondary divided attention task was used to assess workload. This task consisted of a diamond shown on both sides of the middle screen. At pre-determined travelled distances in the scenario (including points where the participant was approaching or leaving a junction), the left diamond changed into a left-pointing equilateral triangle or the right diamond changed into a right-pointing equilateral triangle. At these moments, the participants were required to press a button on the steering wheel as quickly as possible. There was one button for when the left diamond changed into a triangle and a second for when the right diamond changed into a triangle. The triangle changed back to a diamond when the participant pressed any of the two buttons, or when the maximum response time of 5 s had elapsed. The triangles were activated based on the travelled distance, according to a script that was different for each of the four conditions. The left- and right-pointing triangles were activated in no discernable order. There were 26 divided attention events per session, occurring on average every 242 m (min = 30 m, max = 710 m).

The reaction time was determined between the moment a diamond changed into a triangle and the participant pressed a button. If the participant pressed the wrong button, the event was labelled as an incorrect response. If the participant did not press a button within 5 s, the event was labelled as a miss. The software sometimes could not determine the response to the divided attention task, either because the divided attention task and a critical event occurred at overlapping periods or because a new divided attention event started before the participant had pressed a button (this could happen when the participant responded late and the distance between events was small, e.g., less than 67 m at an assumed constant speed of 48 km/h). Therefore, it was decided to delete events for which at least one participant had a missing value. Accordingly, the number of divided attention events included in the analysis were 23, 24, 25 and 25 for the MC, PAS, PAD and HAD conditions, respectively.

2.1.5 Simulator hardware and software

The simulator consisted of three linked monitors (Figure 1). A game controller steering wheel and pedals were used to control the virtual vehicle. Rear view and side mirrors where rendered on the screen, allowing the driver to see behind them. STISIM Drive M500W (build 2.10.07) software was used. Speakers positioned on the desk provided vehicle sound.

2.2 Experiment 2

2.2.1 Driving environment

The road environment was the same as in Experiment 1, except that lane width was 4.5 m instead of 3.7 m (there was no particular reason for selecting another road width; it was the idiosyncratic consequence of our scenario scripting efforts). Participants drove a shorter part of the route as compared to Experiment 1 (3760 m in each session; about 6 min of driving time). Accordingly, participants encountered six traffic lights (two pedestrian crossings and four junctions). Two of the traffic lights were green; the remaining four turned red. In each session, a vehicle appeared behind the participant and then overtook the participant early in the session, after which it continued driving with 50 km/h in front of the participant's vehicle. There were six oncoming cars, with a mean distance of 590 m ($SD = 432$ m) between these cars. There were no vehicles on the participant's lane, except the participant's own vehicle and the lead car. The lead car stopped at the red light junctions.

Figure 1 Driving simulator used in Experiment 1. The photo illustrates the hardware setup; the virtual environment in the actual experiment was not identical to the one shown in the figure (see online version for colours)

2.2.2 Automation conditions

- *MC*. Same as in Experiment 1.
- *ACC.* The automation controlled speed throughout the session while driving behind the lead car. The automation maintained headway between 15 m and 30 m. The participant was in control of the steering. Note that in reality, ACC systems control headway with respect to the car in front. In the present study, such ACC functionality was emulated by keeping the speed of the lead car and the participant's car at set values.
- *AS***.** The automation controlled steering throughout the session. The participant had control of speed.
- *HAD.* The automation controlled both steering and speed throughout the session. The automation maintained headway between 15 m and 30 m.

The MC condition used a manual transmission, whereas the three other conditions used automatic transmission. The reason for selecting a manual transmission in the MC condition (as opposed to the automatic transmission in the MC condition in Experiment 1) was that the hardware of the medium fidelity simulator in Experiment 2 facilitated the use of a realistic manual transmission. The automated speed (in ACC and HAD) maintained the vehicle at 50 km/h, reduced speed on the approach to red light junctions with about 1.5 m/s^2 to 3 km/h (it did not stop completely), and increased speed with about 3.5 m/s^2 to 50 km/h when leaving these junctions. The automated steering (in AS and HAD) kept the lateral position at 0 m.

2.2.3 Critical events during driving

Two critical events were scripted:

- a red 'stop' image displayed on the screen for 5 s; the participant was required to initiate an emergency stop and
- the lead car suddenly braked, and the participant was required to make an emergency stop in order to avoid crashing into the vehicle.

The lead car's deceleration rate was a constant 6.9 m/s^2. That is, it took 2 s to decelerate from 50 to 0 km/h. The vehicle pulled away once the participant had come to a stop. The automation (ACC, AS and HAD) was disabled for these two events and reactivated between 30 and 50 m afterwards.

2.2.4 Secondary divided attention task

The secondary task was the same as in Experiment 1 (a diamond was shown on both sides of the front projection), with the difference that participants were required to press buttons on the vehicle centre console instead of on the steering wheel. The divided attention events occurred on average every 187 m (min = 30 m, max = 600 m). After applying the same filtering criteria as used in Experiment 1, the number of divided attention events included in the analysis were 14, 14, 15 and 15 for the MC, ACC, AS and HAD conditions, respectively.

2.2.5 Simulator hardware and software

The simulator (Figure 2) consisted of a Jaguar XJ saloon car linked to a STISIM Drive M500W (build 2.10.09) with comprehensive vehicle dynamics model and active steering feedback. Three projectors provided a 135° field of view and a fourth projector provided the rear view. Two liquid crystal displays were used for the side mirrors. A surround system provided aural feedback.

2.3 Participants (Experiments 1 and 2)

The participants were mostly students recruited from the Faculty of Engineering and the Environment of the University of Southampton. They did not receive financial compensation for participating. Details about the participants are provided in Table 1. Experiments 1 and 2 were conducted with different participants. Overall, 61% of the participants were males. They all had their driving licence for at least two years, but had fairly low mileages (75% of participants reported a mileage of less than 10,000 miles per year). The participants in Experiment 2 were significantly younger than the participants in Experiment 1 ($p = 0.013$ according to an independent samples Student's t-test, $df = 49$).

Figure 2 Driving simulator used in Experiment 2. The photo illustrates the hardware setup; the virtual environment in the actual experiment was not identical to the one shown in the figure (see online version for colours)

Table 1 Participant details of Experiments 1 and 2

	Experiment 1	*Experiment 2*
Number of participants	24	27
Number of males	15	16
Mean age	27.42 (*SD* = 11.61, min = 19, max = 53)	21.52 (*SD* = 2.53, min = 18, max = 30)
Mean years holding a full driving licence	9.38 (*SD* = 11.10, min = 2, max = 35)	4.35 (*SD* = 2.39, min = 2, max = 12)
Number of participants per mileage category (miles)		
0–5000	13	7
5000–10,000	6	12
10,000–15,000	3	5
15,000–20,000	1	1
20,000+	1	2

2.4 Procedure (Experiments 1 and 2)

Participants were invited, one at a time, to the simulator laboratory. Participants were welcomed and read an information sheet that explained the background and procedures of the experiment. In Experiment 2, the information sheet had also been emailed to all participants on beforehand. All participants provided written informed consent.

The information sheet mentioned that participants would be driving four sessions, each with a different level of automation in any order. The information sheet of Experiment 1 stated that they would be driving with “Manual Control – Driver responsible for all tasks", "Assistive Technology – Control of steering shared between vehicle and driver", "Advanced Technology – Control of speed and steering shared between vehicle and driver" and "Full Automation – Vehicle controls all driving tasks" (corresponding to MC, PAS, PAD and HAD, respectively). The information sheet of Experiment 2 stated that they would be driving with "Manual Control – Driver has full control of the vehicle; responsible for all tasks", "Active Steering – Vehicle controls steering, driver controls all other tasks", "Adaptive Cruise Control – Vehicle controls speed, driver controls all other tasks" and "Full Automation – Vehicle controls all driving tasks" (corresponding to MC, AS, ACC and HAD, respectively). In other words, participants were provided with knowledge about the types of systems they would be driving with, but not with details about the functionality of these systems.

The information sheet mentioned that all conditions used automatic transmission (Experiment 1) or that all conditions except MC used automatic transmission (Experiment 2). It further informed participants that automation was imperfect as follows: "Please note that automation may fail at any time" (Experiment 1) or "There is the possibility of a critical event upon which the automation will deactivate and you will be required to take a suitable approach to the situation" (Experiment 2). We provided participants with this information, because, in reality, drivers of new vehicles will probably also have at least rudimentary knowledge about the fact that their automated driving system has functional limits.

Next, the participants signed the consent form and completed a short questionnaire to establish their driving experience. Afterwards the experimenter explained the controls of the simulator, and made clear that driving was the primary task. Participants were also told that they should only complete the secondary divided attention task if they felt that they could so without affecting their driving performance. Regarding Experiment 2, where a lead car was present, participants were told to follow this car at a comfortable distance.

The participants undertook a short (approximately 5 min) practice drive without automation, in order to get used to the driving controls and the divided attention task. Next, participants undertook the four sessions in counterbalanced order. Immediately after completing each session, participants completed a NASA-TLX on paper (Experiment 1) or on a computer (Experiment 2; Sharek, 2011), in which they had to rate mental demand, physical demand, temporal demand, effort and frustration from 'very low' to 'very high' and performance from 'perfect' to 'failure'.

Participants noted down any comments on the automation condition, after each session (Experiment 1), or at the end of all four sessions (Experiment 2). Specifically, in Experiment 1, the form read:

- "manual drive comments:"
- "assistive automation comments:"
- "advanced automation comments:"
- "full automation comments:".

Similarly, in Experiment 2, the form stated "Comments on each phase:" followed by

- "manual control:"
- "assisted steering:"

- “adaptive cruise control:”
- “full automation:”
- “any other comments:”.

On both forms, there were a few centimetres of space below each condition to write down any comments.

2.5 Dependent measures (Experiments 1 and 2)

The following measures of workload were determined per participant per session.

- Ratings on each of the six TLX items (%). The workload items were expressed on a scale from 0 to 100%, which is common practice (Hart and Staveland, 1988).
- Percentage of correct responses, percentage of incorrect responses and percentage of misses on the divided attention task (%). These three percentages add up to 100%.
- Mean reaction time on the divided attention task (s). Misses and incorrect responses were excluded. In addition, we calculated the ‘adjusted’ reaction time task by imputing the maximum possible reaction time of 5 s for missing or incorrect responses. The adjusted reaction time was used to combine the number of misses and reaction time into a single measure.

We calculated the following commonly used measures of driving performance, to describe speed and lane keeping performance. For these three measures, the first 60 s were excluded to remove lead-in effects.

- Mean speed (km/h).
- Maximum speed (km/h).
- Standard deviation of lateral position (SDLP; m). SDLP is a measure of swerving along the average lateral position in the session.

In addition, we calculated the brake reaction times, defined as the time difference between stimulus presentation and first brake pedal movement, to establish how quickly participants responded to the critical events. Note that brake reaction time is sometimes called ‘total braking time’ in the literature (Liebermann et al., 1995; Young and Stanton, 2007b). For the PAD and HAD automation failure before the red light in Experiment 1, we counted the number of red light violations.

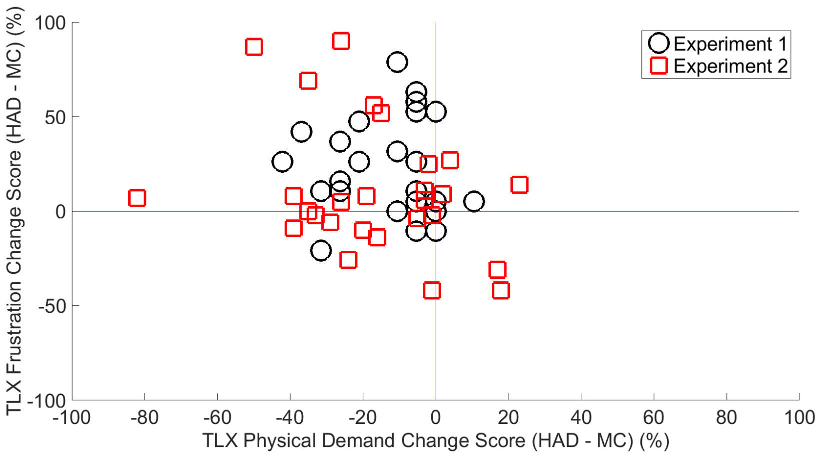
The results were statistically analysed by means of a repeated measures analysis of variance (ANOVA). An alpha value of 0.05 was used. Pairwise comparisons were conducted using a Tukey–Kramer test.

3 Results

3.1 Self-reported workload

Table 2 shows the means and standard deviations of the TLX items. A reduction of self-reported workload with increasing level of automation is observed for mental demand, physical demand, temporal demand and effort in Experiment 2 and to a lesser extent in Experiment 1. For Experiment 1, PAD and HAD were rated as significantly more frustrating than MC and PAS. For Experiment 2, HAD was rated as significantly more frustrating than AS. Figure 3 provides an illustration of individual differences of the TLX physical demand and frustration items. It can be seen that most of the participants reported decreased physical demand but increased frustration for HAD with respect to MC.

Figure 3 Frustration percentage in highly automated driving (HAD) minus frustration percentage in manual control (MC) vs. physical demand percentage in the HAD condition minus physical demand percentage in the MC condition (*N* = 24 for Experiment 1; *N* = 27 for Experiment 2). It can be seen that most participants are located in the second quadrant. In other words, HAD resulted in higher frustration but lower physical demand than MC. TLX = NASA task load index. The mental demand and frustration items were measured on a scale from 'very low' (0%) to 'very high' (100%) (see online version for colours)



3.2 *Performance on the divided attention task*

One participant in the MC condition of Experiment 2 did not react to any of the stimuli (for unknown reasons). Hence, a reaction time could not be established, and the mean reaction time and mean number of correct/incorrect/missed responses was based on *N* = 26 instead of *N* = 27.

Table 2 shows that for Experiment 2, a higher level of automation was associated with faster reaction times on the divided attention task (all combinations of adjusted reaction times were statistically significant, except ACC vs. AS). For Experiment 1, HAD yielded significantly faster adjusted reaction times than MC and PAS.

Table 3 provides Pearson correlation coefficients between the mean scores on the TLX items on the one hand, and the mean raw reaction times, mean adjusted reaction times, and the percentage of correct response on the divided attention task on the other. The correlation coefficients are reported for the eight experimental conditions of Experiments 1 and 2 combined, and for Experiments 1 and 2 separately. The correlation of 0.93 between TLX mental demand and the raw reaction time is illustrated in Figure 4.

Table 2 Means (standard deviations in parentheses) of dependent measures for Experiments 1 and 2

	Experiment 1 (N = 24)						*Experiment 2 (N = 27)*					
	1. MC	*2. PAS*	*3. PAD*	*4. HAD*	*p value*	*sig. pairs*	*1. MC*	*2. ACC*	*3. AS*	*4. HAD*	*p value*	*sig. pairs*
TLX Mental demand (%)	48 (20)	46 (19)	47 (19)	40 (28)	0.087		55 (20)	35 (17)	39 (20)	25 (20)	<0.001	12, 13, 14, 34
TLX Physical demand (%)	30 (18)	29 (19)	25 (19)	17 (16)	<0.001	14,24,34	35 (20)	29 (17)	24 (16)	18 (16)	<0.001	13, 14, 24
TLX Temporal demand (%)	28 (15)	28 (15)	29 (14)	22 (17)	0.107		42 (17)	33 (19)	31 (16)	24 (17)	<0.001	12, 13, 14
TLX Performance (%)	26 (14)	30 (19)	34 (23)	35 (20)	0.221		42 (19)	45 (16)	35 (16)	43 (28)	0.093	
TLX Effort (%)	39 (21)	37 (20)	40 (24)	27 (24)	0.007	14,34	49 (20)	37 (17)	34 (21)	27 (18)	<0.001	12, 13, 14, 24
TLX Frustration (%)	22 (18)	29 (21)	42 (25)	45 (27)	<0.001	13,14,23,24	35 (27)	44 (25)	31 (20)	46 (29)	0.009	34
DA RT raw (s)	1.95 (0.48)	1.94 (0.39)	1.89 (0.44)	1.81 (0.49)	0.351		1.99 (0.30)	1.67 (0.47)	1.57 (0.36)	1.39 (0.31)	<0.001	12, 13, 14, 24

Table 2 Means (standard deviations in parentheses) of dependent measures for Experiments 1 and 2 (continued)

	Experiment 1 (N = 24)						Experiment 2 (N = 27)					
	1. MC	*2. PAS*	*3. PAD*	*4. HAD*	*p value*	*sig. pairs*	*1. MC*	*2. ACC*	*3. AS*	*4. HAD*	*p value*	*sig. pairs*
DA RT adjusted (s)	2.37 (0.72)	2.43 (0.84)	2.33 (0.62)	1.98 (0.60)	0.006	14, 24	2.29 (0.60)	1.80 (0.61)	1.84 (0.61)	1.49 (0.39)	<0.001	12, 13, 14, 24, 34
DA correct (%)	85 (14)	82 (21)	84 (12)	93 (8)	0.015	24	85 (16)	94 (9)	90 (13)	96 (8)	<0.001	12, 14
DA incorrect (%)	0 (1)	1 (2)	1 (2)	1 (2)	0.435		3 (5)	1 (4)	1 (3)	1 (2)	0.034	14
DA miss (%)	15 (13)	17 (22)	15 (12)	6 (8)	0.012	24	12 (15)	5 (8)	9 (13)	3 (7)	0.003	12, 14
Mean speed (km/h)	34.8 (1.6)	34.2 (1.7)	33.8 (1.6)	34.0 (0.4)	0.009	13	33.7 (1.5)	34.6 (0.7)	32.9 (0.8)	34.2 (0.6)	<0.001	12, 13, 23, 34
Max speed (km/h)	58.9 (8.0)	61.3 (9.9)	59.9 (9.5)	48.9 (1.8)	<0.001	14, 24, 34	62.8 (12.2)	50.3 (1.3)	57.6 (7.7)	50.2 (1.1)	<0.001	12, 13, 14, 23, 34
SDLP (m)	0.23 (0.05)	0.20 (0.04)	0.20 (0.04)	0.16 (0.10)	<0.001	14, 24, 34	0.27 (0.07)	0.31 (0.12)	0.02 (0.02)	0.10 (0.08)	<0.001	13, 14, 23, 24, 34

TLX = NASA Task Load Index, DA = Divided attention task, RT = reaction time, SDLP = standard deviation of lateral position, MC = manual control, PAS = partially automated steering, PAD = partially automated driving, HAD = highly automated driving, ACC = adaptive cruise control, AS = automated steering. N = 26 for DA RT raw, DA RT adjusted, and DA misses/incorrect in the MC condition in Experiment 2. "Sig. pairs" refers to the experimental conditions that are significantly different from each other (e.g., 14 means that Condition 1 and Condition 4 were significantly different).

Figure 4 Mean raw reaction time on the divided attention task vs. mean mental demand on the NASA task load index (TLX). Circles = Experiment 1 (average across 24 participants); Squares = Experiment 2 (average across 27 participants, but $N = 26$ for MC). The mental demand item was measured on a scale from 'very low' (0%) to 'very high' (100%). TLX = NASA task load index, MC = manual control, PAS = partially automated steering, PAD = partially automated driving, HAD = highly automated driving, ACC = adaptive cruise control, AS = automated steering (see online version for colours)

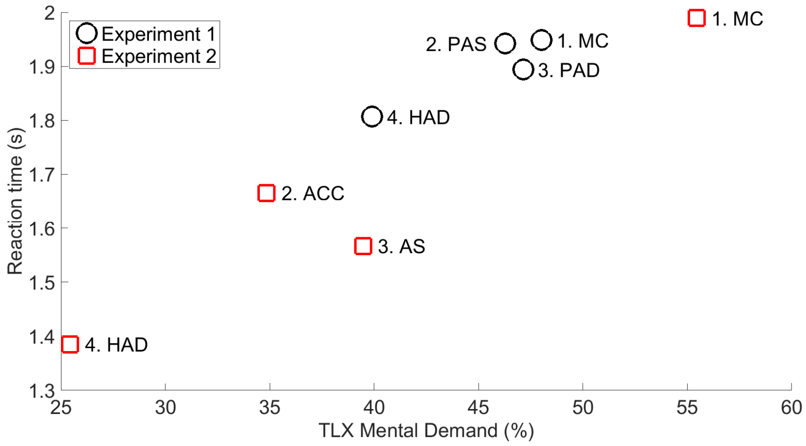


Table 3 Pearson correlation coefficients (*r*) between mean TLX item scores and mean reaction time on the divided attention task

	Raw reaction time			
	Experiments 1 and 2 combined (N = 8 sessions)		*Experiment 1 (N = 4 sessions)*	*Experiment 2 (N = 4 sessions)*
	r	*p value*	*r*	*r*
TLX Mental demand (%)	0.93	0.001	0.92	0.95
TLX Physical demand (%)	0.64	0.087	1.00	0.97
TLX Temporal demand (%)	0.36	0.374	0.85	1.00
TLX Performance (%)	–0.52	0.184	–0.83	–0.08
TLX Effort (%)	0.68	0.061	0.88	1.00
TLX Frustration (%)	–0.48	0.229	–0.88	–0.45
	Adjusted reaction time			
	r	*p value*	*r*	*r*
TLX Mental demand (%)	0.91	0.002	0.94	0.99
TLX Physical demand (%)	0.63	0.093	0.95	0.94
TLX Temporal demand (%)	0.30	0.469	0.93	0.99
TLX Performance (%)	–0.67	0.069	–0.66	–0.13
TLX Effort (%)	0.67	0.070	0.92	0.99
TLX Frustration (%)	–0.61	0.109	–0.72	–0.62

Table 3 Pearson correlation coefficients (*r*) between mean TLX item scores and mean reaction time on the divided attention task (continued)

	Divided attention task, percentage correct responses			
	r	*p value*	*r*	*r*
TLX Mental demand (%)	–0.87	0.005	–0.91	–0.99
TLX Physical demand (%)	–0.66	0.075	–0.91	–0.81
TLX Temporal demand (%)	–0.36	0.377	–0.95	–0.91
TLX Performance (%)	0.66	0.074	0.55	0.36
TLX Effort (%)	–0.71	0.050	–0.92	–0.91
TLX Frustration (%)	0.68	0.062	0.62	0.78

TLX = NASA task load index, DA = secondary divided attention task. The correlation coefficients were calculated among the means of the experimental conditions (see Figure 4, for an illustration).

3.3 Vehicle centred measures

The mean speed was similar for both experiments and for all conditions, averaging at values between 32.9 (AS in Experiment 2) and 34.8 km/h (MC in Experiment 1, see Table 2). Standard deviations of mean speed and maximum speed were small (<1 km/h) for ACC and HAD, because speed was under automatic control for most of the driving time. Furthermore, a lead car was present in Experiment 2, setting an upper limit to the participants' driving speed. SDLP values were the lowest for the AS and HAD conditions, because the automation held the lateral position of the car at the lane centre for the entire driving time except for the occasions per session during which the participant drove manually in response to the critical event.

3.4 Brake and steering reaction to the critical events

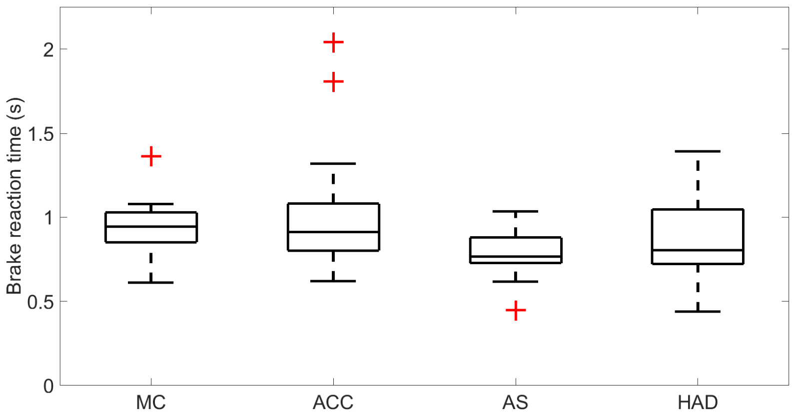
In Experiment 1, eight participants driving in the PAD condition made a red light violation at one of the two junctions where automation failed, and a further two participants made a red light violation at both junctions. For HAD, eight participants made a red light violation. In comparison, for the MC and PAS sessions (during which speed was controlled by the driver and which therefore had 12 opportunities for crossing a red light), there were four and six participants, respectively, who made at least one red light violation.

All participants, except one participant driving with PAS, avoided collision with the pedestrian/dog. The participant who crashed with the pedestrian/dog did not press the brakes, but applied an evasive steering correction. An evasive steering correction was defined here as a lateral motion of more than 2 m from the lane centre. Almost all participants (21, 22, 23 and 19 participants in the MC, PAS, PAD and HAD conditions) braked in response to the crossing pedestrian/dog and did not make an evasive steering action. A small number of participants made an evasive manoeuvre but did not press the brakes (2, 1, 1 and 1 participants in the MC, PAS, PAD and HAD conditions). There was one participant in the MC condition and four participants in the HAD condition who made both an evasive manoeuvre and braked. Furthermore, there was one participant in the PAS condition who did not apply the brakes and did not make an evasive manoeuvre.

The mean brake reaction times since the moment the pedestrian/dog started to walk for the MC, PAS, PAD and HAD conditions were, respectively, 3.74 s ($SD = 0.85$ s, $N = 22$), 3.69 s ($SD = 0.90$ s, $N = 22$), 4.48 s ($SD = 0.67$ s, $N = 23$) and 3.83 s ($SD = 0.70$, $N = 23$). According to a repeated measures ANOVA, these reaction times were significantly different from each other ($p = 0.001$, after excluding three participants who did not brake in one or more conditions). A *post hoc* analysis indicated significantly longer brake reaction times for PAD compared to MC, PAS and HAD.

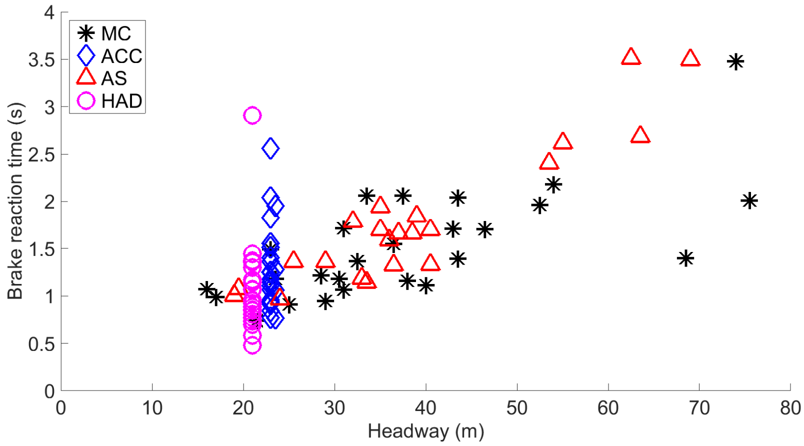
In Experiment 2, the mean (*SD*) brake reaction times to the stop sign were similar for all automation conditions: 0.93 s (0.15), 1.00 s (0.32), 0.79 s (0.13) and 0.87 s (0.24), for MC, ACC, AS and HAD, respectively ($N = 27$; $p = 0.002$ according to the ANOVA; ACC significantly longer than AS), see Figure 5 for a boxplot.

Figure 5 Boxplot of brake reaction times to the stop sign (Experiment 2). MC = manual control, ACC = adaptive cruise control, AS = automated steering, HAD = highly automated driving. $N = 27$ for all conditions (see online version for colours)



There were five collisions with the decelerating lead vehicle in Experiment 2, all occurring in the ACC condition. The mean (*SD*) brake reaction times were 1.53 s (0.58), 1.29 s (0.42), 1.79 s (0.73) and 1.01 s (0.45), for MC, ACC, AS and HAD, respectively. Headway plays an important moderating role in explaining these reaction times: One participant in the MC condition and five participants in the AS condition did not press the brakes at all; they avoided collision because their headway was large (>80 m) when the lead car started to decelerate. The Pearson correlation coefficients between headway and brake reaction time were 0.71 for MC ($N = 26$, $p < 0.001$) and 0.93 for AS ($N = 22$, $p < 0.001$), indicating that participants with larger headways had longer reaction times (Figure 6). The mean (*SD*) headway when the lead car started to decelerate was 40 m (18), 23 m (0), 41 m (17) and 21 m (0) for the MC, ACC, AS and HAD conditions, respectively (Figure 6).

Figure 6 Brake reaction time vs. headway at the moment the lead car started to decelerate (Experiment 2). MC = manual control (*N* = 26), ACC = adaptive cruise control (*N* = 26), AS = automated steering (*N* = 22), HAD = highly automated driving (*N* = 27) (see online version for colours)



3.5 Participants' comments on the automation conditions

Regarding Experiment 1, 10 participants remarked that they did not notice any activity of the PAS system. Eighteen participants made a negative remark about PAD: they mentioned lack of trust, pointed out the automation failure or stated they were annoyed, uncertain or irritated. Only one participant stated that the PAD system was helpful. Fourteen participants made a similar negative remark about HAD, and another five participants mentioned that HAD made them distracted, bored or lose concentration. Some participants (two for PAS, one for PAD and four for HAD) noted down that they disliked the lack of feedback about the automation status. Two participants who drove with PAD and two who drove with HAD stated that they shadowed the controls regardless of automation status. This could also be seen in the recorded data, as some participants pressed the brakes and/or turned the steering wheel even when automated speed was active.

Regarding Experiment 2, 11 participants stated that AS was the easiest or preferred condition. Twelve participants who drove with ACC stated that the system made them feel stressed or uneasy, and six participants mentioned lack of trust regarding whether the vehicle would brake automatically or not. Seventeen participants stated that the HAD system made them complacent or distracted, and four participants mentioned lack of trust during braking. Participants commented that the traffic lights in particular made them feel uncomfortable due to an uncertainty of whether they would stop in time.

4 Discussion

4.1 Effects of automation on driver performance and workload

We performed two experiments, each testing four levels of automation in short driving sessions (about 12 and 6 min per session, in Experiments 1 and 2, respectively). When the automation was active (i.e., the vast majority of the time for ACC, AS and HAD, and only on junctions for PAS and PAD), the technology ensured a constant speed in between junctions, safe negotiation of junctions and pedestrian crossings, and/or zero lane centre error. It is therefore not surprising that the automation contributed to statistically significantly lower self-reported mental demand (Experiment 2), physical demand (Experiments 1 and 2), temporal demand (Experiment 2) and effort (Experiments 1 and 2), as well as faster raw (Experiment 2) and adjusted (Experiments 1 and 2) reaction times to the divided attention task compared to MC. The mean reaction times on the secondary task and four items of the TLX (mental demand, physical demand, temporal demand and effort) were positively correlated (Table 3; Figure 4), suggesting that both the reaction times and these self-report workload items tap the same underlying construct.

Participants reported a relatively high frustration level for PAD (Experiment 1), ACC (Experiment 2) and HAD (Experiments 1 and 2), and expressed dislike of these conditions. Our results correspond to Omae et al. (2005), who asked participants whether they wanted to use automatically driven vehicles if they were required to supervise the system; 23 of 30 drivers answered 'no' to this question. Our results also mirror what some car manufacturers and human factors researchers have mentioned all along, namely that pleasure is an important component of the driving experience and that humans should not be forced into a monotonous supervisory role (e.g., Hancock, 2015; Neubauer et al., 2011; Walker et al., 2001). It was commented by 40% of participants in Experiment 2 that the AS condition was the easiest and/or preferred condition, possibly because the critical events did not require an evasive steering action. Human factors experiments in the medical and nuclear power plant domains have also shown that poorly designed automation can yield elevated levels of frustration (Lin et al., 2010; Luz et al., 2010).

Our results clearly differ from previous research into ACC and HAD. Specifically, previous research indicates that the TLX frustration item behaves in line with the other five TLX items, with all the TLX items showing either significantly lower (Flemisch et al., 2008; Ma, 2006) or about the same (Bjørkli et al., 2003; McDowell et al., 2008; Nilsson, 1995; Nilsson and Nåbo, 1996; Peters, 2001; Törnros et al., 2002) ratings between manual and automated driving. In these previous studies, the automation required either no intervention or included an occasional 'surprise' event that required a human response, but drivers were not put into the role of supervisor of automation.

Our study is unique in the sense that the automation required manual intervention about once every 3 min. We believe that an automated driving system that requires frequent human intervention is realistic from a technological point of view, considering that sensor systems cannot detect all possible hazards. Formal definitions of levels of driving automation, as proposed by BASt, SAE and NHTSA, make explicitly clear that at intermediate levels of automation, constant human alertness to the roadway is required (Gasser and Westhoff, 2012; Smith, 2013).

In retrospect, it is not surprising that automation in some cases can contribute to frustration. More generally, it is known that vigilance tasks in which humans are required to detect infrequent events are typically experienced as stressful or frustrating (Szalma et al., 2004; Warm et al., 2008). Driving in a highly automated car that requires frequent intervention may be similarly frustrating, as it requires sustained attention and puts participants 'on-guard' in order to detect when to resume MC. Furthermore, it may be frustrating for people to not be able to take over control. For example, Comte (2000) found that intelligent speed adaptation (ISA) yielded a score of 60% on the TLX frustration item compared to only 15% for manual driving, presumably because the ISA prevented the participants from overtaking other vehicles.

4.2 Response to critical events

In both experiments, there were critical events that required intervention with the brake pedal and/or steering wheel. Participants in Experiment 1 successfully avoided crashing into the crossing pedestrian (in MC, PAS and HAD) or the crossing dog (in PAD). The one collision that did occur was in the PAS condition that was never operative in between junctions, so this collision was probably a conventional driver error related to perception or action execution (for a review of human errors in driving, see Stanton and Salmon, 2009). The mean brake reaction times to the pedestrian/dog were about 4 s since their initial movement, for all four conditions. These seemingly long reaction times can be explained by the fact that the walking pedestrian/dog was not a threat until it actually entered the road, and participants likely delayed their response until braking was required. The low standard deviations of the brake reaction times (about 0.8 s for all four conditions) are consistent with the idea that participants reacted rapidly as soon as the situation demanded an intervention. The reaction times in the PAD condition were significantly longer than in the other three conditions, which may be caused by the fact that the PAD condition featured a dog rather than a pedestrian. It is possible that participants were less willing to perform an emergency brake for an animal than for a human. The fact that the crossing pedestrian/dog event occurred at different points along the route for each of the four conditions may have also had an influence on the brake reaction times.

Many participants in Experiment 1 made a red light violation after the PAD/HAD failed. The automation provided no warning feedback upon failure. However, in principle, the drivers could have been primed through sound cues: Once the automation switched off, the acceleration ceased, and so there should have been a noticeable change in vehicle sound. The minimum feedback that could have been provided was an indication of the status of the automation or an alert in the event of automation failure. We doubt, however, whether feedback *per se* could alleviate much of the frustration, considering that warning/feedback systems themselves are known to be a source of mental demand and frustration, potentially leading to automation disuse (Parasuraman and Riley, 1997; Stanton et al., 2011; Tango et al., 2011). In our experiment, most participants did not notice the PAS and kept turning the steering wheel as they did during manual driving, a phenomenon possibly also caused by lack of automation-status feedback. Although a warning signal may have been useful in case the automation fails or malfunctions, a warning *cannot* be provided in other types of critical events (e.g., the pedestrian/dog scenario in Experiment 1), because the automation cannot have knowledge of events that are beyond its programming logic or detection capabilities. In other words, in a real car, no warning can be provided to situations that are not detected by the sensor system.

Participants in Experiment 2 reacted rapidly to the stop sign, with mean brake reaction times of about 1.0 s for all four conditions. This is considerably faster than the 3.5 s reaction time to an unexpected automation failure reported by Young and Stanton (2007b). Green (2000) argued that the level of expectation is a primary factor influencing brake reaction times, with an average brake reaction time of 0.7 s for situations that are entirely expected up to 1.5 s for scenarios in which an object suddenly enters to road (for a review, see Summala, 2000). Thus, brake reaction times in our experiment represent those in normal driving, and suggest that participants in our study were vigilant and prepared for intervention. The brake reaction times in the AS condition were significantly faster than the ACC condition. A possible explanation is that participants in the AS condition were engaged in the longitudinal control loop (i.e., they had their feet close to the pedals and paid attention to the road, as in the MC condition), yet freed from the physical and cognitive demands of steering (i.e., they did not have to turn the wheel or pay attention to lateral disturbances, which is similar to the HAD condition). This could mean that participants in the AS condition were optimally prepared to react to events that require the longitudinal response of braking.

Participants in Experiment 2 reacted rapidly to the braking lead car, with reaction times of 1.01 for HAD and 1.29 s for ACC. The MC and AS conditions yielded longer mean brake reaction times (1.53 and 1.79 s), likely because participants in these two conditions drove with a longer mean headway than participants driving with the ACC and HAD systems, and so were able to delay their braking response (cf. Figure 6). These results may have implications for the validity of brake reaction times in general. Drivers clearly do not brake immediately, but decide to wait until the situation becomes urgent, which means that longer brake reaction times do not imply worse driving *per se*.

In summary, the responses to the critical events suggest that participants were alert throughout the experiment and prepared to intervene whenever necessary. These findings differ considerably from those of previous research, in which drivers of automated cars were exposed to emergency situations that were a surprise to them (e.g., Flemisch et al., 2008). We argue that our study represents a realistic case of how future automation may be used in city environments.

4.3 Limitations of this research

There has been considerable debate about whether driving simulators can provide valid outcomes regarding research into human factors of automated driving (e.g., Farber, 1999; Neale and Dingus, 1998; Neubauer et al., 2010), and the present study sheds further light on this topic. Our results showed that the experimental effects were *relatively* similar between the two experiments. For example, both the low and medium fidelity driving simulator studies showed that automation improves reaction times to the secondary task. This suggests that even low fidelity simulators are useful for this type of experimental research. However, we suspect that in *absolute* terms, the results cannot be immediately translated to real world driving. For example, the SDLP in the manual condition was 0.23 m in Experiment 1 and 0.27 m in Experiment 2. These numbers are considerably greater than the SDLP values of 0.15–0.20 m typically observed in on-road experiments (Godthelp et al., 1984; Veldstra et al., 2015; Verster and Roth, 2011). Such discrepancy in SDLP between simulation and reality has also been observed in previous research (Helland et al., 2013) and may be owing to the imperfect visual, tactile and vestibular cues offered by driving simulators (Kemeny and Panerai, 2003). The results in Table 2 showed that effects were more often statistically significant in the medium fidelity setup than in the low fidelity setup, which tentatively suggests that the medium fidelity setup is more sensitive to experimental manipulations than the low fidelity configuration (for similar results, see Santos et al., 2005). Note, however, that Experiments 1 and 2 were different in various ways, including the type of transmission, length of the sessions, lane width, types of critical events, presence of a lead car, types of automated driving systems, methods of collecting the TLX ratings and the age of participants). It should also be remembered that any simulator, be it a low fidelity (Experiment 1) or a medium fidelity one (Experiment 2), provides only an abstraction of real driving.

A second limitation is that the participants in our study were not provided with any training in the use of the automation systems. Participants had to learn how to respond to automation failures during the experiment itself. Perhaps, drivers will find PAD and HAD less frustrating if they have the opportunity beforehand to develop an appropriate mental model about automation failure modes. Providing participants with such training could have facilitated appropriate reliance (see also Bahner et al., 2008; Lee and See, 2004; Parasuraman and Riley, 1997). However, prior vigilance research indicates that ratings of frustration rise linearly with time, in conditions where one has to monitor a display for three quarters of an hour (Dember et al., 1993; Szalma et al., 2004). Accordingly, participants may consider long driving sessions to be more frustrating than short sessions.

Another limitation of the present research is that the tested systems may be seen as ‘unrealistic’. The automated driving systems in our experiment were tested in a low speed city environment featuring traffic lights and vulnerable road users. It may be the case that automation technology will be first deployed in simpler environments, such as highways. However, we argue that automation systems might also be used in complex environments, exactly because drivers need help in such situations. On highways, automation may be more of a comfort feature rather than fulfilling an actual safety need.

The inability to override the automated driving systems may also be regarded as unrealistic. However, we argue that this feature is actually realistic, because future human–machine interfaces may be designed in such a way as to “prevent unwarranted human driving” (Tsao et al., 1997, p.11). Furthermore, our approach of pre-programmed automation is useful with regard to experimental control. Previous driving simulator research shows that when automation can be enabled and disabled at will, a large share of drivers choose to disable the automation and drive manually instead (Neubauer et al., 2012). Letting drivers enable or disable automation at their own discretion compromises the experimental results if the goal of the study is to compare automated driving with manual driving.

A fifth limitation is that participants in this research were recruited from the university community with relatively little driving experience. University students are not representative of the general population (Henrich et al., 2010) and are known to have relatively strong cognitive and spatial abilities (Wai et al., 2009). Furthermore, young people in general have higher tendencies for sensation seeking and committing traffic violations than older persons (Zuckerman et al., 1978). Thus, it remains to be investigated whether these results are generalisable to the overall driving population.

Sixth, our experiment had involved only 25–50 min of driving per person, equivalent to most previous driving simulator research on automated driving (see De Winter et al., 2014, for an overview of experiment durations). Hence, our research design does not allow us to draw conclusions about the long-term effects of automated driving on workload, brake reaction times and drivers’ opinions.

A seventh and final limitation is that drivers in our experiments had to perform an artificial secondary task while driving. In real car driving, drivers also perform secondary tasks, such as using a mobile phone, route navigation device or car radio. However, our secondary task stimulated the participants to keep focused on the visual projection, and allowed us to obtain a quantitative indication of drivers’ spare attentional capacity. It is possible that the secondary and primary task interfered with each other. Indeed, previous research showed that executing a spatial secondary task while driving yielded increased NASA TLX scores, and somewhat deteriorated drivers' lane keeping performance compared to driving without the secondary tasks (Young and Stanton, 2007c).

4.4 Implications of this research

What implications do the present findings have for automation design? Our research suggests that having no automation is less frustrating than automation that requires frequent human intervention. This is a reason for concern, because ratings of frustration are probably closely associated with willingness to buy. One recommendation would be to only bring to the market automation that has extremely high reliability such that the driver is relieved from the supervisory task. The same approach is now followed by Google, who recently removed the steering wheel and pedals from their prototypes of their driverless cars because human test drivers were observed to be no reliable backup (Teller, 2015). However, the Google car is still a prototype, and a fully automated car without steering wheel may remain technologically unfeasible for commercial applications (Shladover, 2015). A parallel may be drawn with aviation here, where pilots are still needed to fly the aircraft despite high levels of automation. Having a possibility for voluntary (de-)activation of automation may also help to reduce frustration (cf. Comte, 2000, 2001). Another option is to provide salient feedback about automation status, or to develop so-called 'human centred automation', which involves and informs the human driver, a message expressed by others (Billings, 1991; Young et al., 2007) in an aviation context. If the human driver is continuously involved and aware of the automation's capacities and limitations, frustration ratings may reduce, while the rapidity and quality of manual takeovers may improve (see Louw et al., 2015 for a discussion).

References

Akamatsu, M., Green, P. and Bengler, K. (2013) 'Automotive technology and human factors research: past, present, and future', *International Journal of Vehicular Technology*, Vol. 2013, pp.1–27.

Bahner, J.E., Hüper, A.D. and Manzey, D. (2008) 'Misuse of automated decision aids: Complacency, automation bias and the impact of training experience', *International Journal of Human-Computer Studies*, Vol. 66, No. 9, 688–699.

Bainbridge, L. (1983) 'Ironies of automation', *Automatica*, Vol. 19, No. 6, pp.775–779.

Bianchi Piccinini, G.F., Rodrigues, C.M., Leitão, M. and Simões, A. (2015) 'Reaction to a critical situation during driving with adaptive cruise control for users and non-users of the system', *Safety Science*, Vol. 72, pp.116–126.

Bibby, K.S., Margulies, F., Rijnsdorp, J.E. and Withers, R.M.J. (1975) 'Man's role in control systems', *Proceedings of the 6th IFAC Congress*, 24–30 August, Boston, MA.

Billings, C.E. (1991) 'Toward a human-centered aircraft automation philosophy', *The International Journal of Aviation Psychology*, Vol. 1, No. 4, pp.261–270.

Bjørkli, C.A., Jenssen, G.D., Moen, T. and Vaa, T. (2003) 'Adaptive Cruise Control (ACC) and driver performance: effects on objective and subjective measures', *Proceedings of the 10th World Congress on Intelligent Transportation Systems*, 16–20 November, Madrid, Spain, pp.16–20.

Carsten, O., Lai, F.C., Barnard, Y., Jamson, A.H. and Merat, N. (2012) 'Control task substitution in semiautomated driving. Does it matter what aspects are automated?', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 54, No. 5, pp.747–761.

Comte, S.L. (2000) 'New systems: new behaviour?', *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 3, No. 2, pp.95–111.

Comte, S.L. (2001) *Intelligent Speed Adaptation: Evaluating the Possible Effects of an Innovative Speed Management System on Driver Behaviour and Road Safety*, PhD thesis, The University of Leeds, Leeds, UK.

Damböck, D., Weißgerber, T., Kienle, M. and Bengler, K. (2013) 'Requirements for cooperative vehicle guidance', *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013)*, 6–9 October, The Hague, the Netherlands, pp.1656–1661.

De Waard, D., Van der Hulst, M., Hoedemaeker, M. and Brookhuis, K.A. (1999) 'Driver behavior in an emergency situation in the Automated Highway System', *Transportation Human Factors*, Vol. 1, No. 1, pp.67–82.

De Winter, J.C.F., Happee, R., Martens, M. and 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*, Vol. 27, Part B, pp.196–217.

Dember, W.N., Warm, J.S., Nelson, W.T., Simons, K.G., Hancock, P.A. and Gluckman, J.P. (1993) 'The rate of gain of perceived workload in sustained attention', *Proceedings of the Human Factors Society 37th Annual Meeting*, Vol. 37, No. 19, pp.1388–1392.

Endsley, M.R. (1999) 'Level of automation effects on performance, situation awareness and workload in a dynamic control task', *Ergonomics*, Vol. 42, No. 3, pp.462–492.

Farber, E.I. (1999) 'Comments on "Driver behavior in an emergency situation in the Automated Highway System"', *Transportation Human Factors*, Vol. 1, No. 1, pp.83–85.

Flemisch, F.O., Kelsch, J., Löper, C., Schieben, A., Schindler, J. and Heesen, M. (2008) 'Cooperative control and active interfaces for vehicle assistance and automation', *FISITA World Automotive Congress*, 14–19 September, Munich, Germany.

Gasser, T.M. and Westhoff, D. (2012) 'BASt-study: Definitions of automation and legal issues in Germany', *Presentation at the Workshop on the Future of Road Vehicle Automation*, 24–26 July, Irvine, CA, http://onlinepubs.trb.org/onlinepubs/conferences/2012/Automation/presentations/Gasser.pdf

Godthelp, H., Milgram, P. and Blaauw, G.J. (1984) 'The development of a time-related measure to describe driving strategy', *Human Factors*, Vol. 26, No. 3, pp.257–268.

Green, M. (2000) '"How long does it take to stop?" Methodological analysis of driver perception-brake times', *Transportation Human Factors*, Vol. 2, No. 3, pp.195–216.

Hancock, P.A. (2015) 'Automobility: the coming use of fully-automated on-road vehicles', *IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 9–12 March, Orlando, FL, pp.137–139.

Hancock, P.A. and Krueger, G.P. (2010) *Hours of Boredom, Moments of Terror: Temporal Desynchrony in military and Security Force Operations*, Report, Center for Technology and National Security Policy, National Defense University, Washington DC.

Hart, S.G. and Staveland, L.E. (1988) 'Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research', in Hancock, P.A. and Meshkati, N. (Eds.): *Human Mental Workload*, North Holland Press, Amsterdam, The Netherlands, pp.139–183.

Helland, A., Jenssen, G.D., Lervåg, L.E., Westin, A.A., Moen, T., Sakshaug, K., Lydersen, S., Mørland, J. and Slørdal, L. (2013) 'Comparison of driving simulator performance with real driving after alcohol intake: a randomised, single blind, placebo-controlled, cross-over trial', *Accident Analysis and Prevention*, Vol. 53, pp.9–16.

Henrich, J., Heine, S.J. and Norenzayan, A. (2010) 'The weirdest people in the world?', *Behavioral and Brain Sciences*, Vol. 33, Nos. 2–3, pp.61–83.

Hoedemaeker, M. and Brookhuis, K.A. (1998) 'Behavioural adaptation to driving with an adaptive cruise control (ACC)', *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 1, No. 2, pp.95–106.

Hollnagel, E. and Woods, D.D. (2005) *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*, CRC Press, Taylor & Francis Group, Boca Raton, FL.

Kemeny, A. and Panerai, F. (2003) 'Evaluating perception in driving simulation experiments', *Trends in Cognitive Sciences*, Vol. 7, No. 1, pp.31–37.

Kircher, K., Larsson, A. and Hultgren, J.A. (2014) 'Tactical driving behavior with different levels of automation', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 1, pp.158–167.

Klauer, S.G., Dingus, D.R., Neale, T.A., Sudweeks, J. and Ramsey, D.J. (2006) *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis using the 100-car Naturalistic Study Data*, Report No. DOT HS 810 594, National Highway Traffic Safety Administration, Washington DC.

Körber, M., Schneider, W. and Zimmermann, M. (2015) 'Vigilance, boredom proneness and detection time of a malfunction in partially automated driving', *International Conference on Collaboration Technologies and Systems (CTS)*, 1–5 June, Atlanta, GA, pp.70–76.

Lee, J.D. and See, K.A. (2004) 'Trust in automation: Designing for appropriate reliance', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 46, No. 1, pp.50–80.

Liebermann, D.G., Ben-David, G., Schweitzer, N., Apter, Y. and Parush, A. (1995) 'A field study on braking responses during driving. I. Triggering and modulation', *Ergonomics*, Vol. 38, No. 9, pp.1894–1902.

Lin, C.J., Yenn, T-C. and Yang, C-W. (2010) 'Evaluation of operators' performance for automation design in the fully digital control room of nuclear power plants', *Human Factors and Ergonomics in Manufacturing and Service Industries*, Vol. 20, No. 1, pp.10–23.

Llaneras, R.E., Salinger, J. and Green, C.A. (2013) 'Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway', *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 17–20 June, Bolton Landing, NY, pp.92–98.

Louw, T., Kountouriotis, G., Carsten, O. and Merat, N. (2015) 'Driver distraction during vehicle automation: How does driver engagement affect resumption of control?', *4th International Driver Distraction and Inattention Conference*, 9–11 November, Sydney, New South Wales, https://www.researchgate.net/publication/280309727_Driver_Distraction_During_Vehicle_Automation_How_Does_Driver_Engagement_Affect_Resumption_Of_Control?ev=prf_pub

Luz, M., Mueller, S., Strauss, G., Dietz, A., Meixensberger, J. and Manzey, D. (2010) 'Automation in surgery: The impact of navigated-control assistance on the performance, workload and situation awareness of surgeons', *Proceedings of the Human Factors and Ergonomics Society*, Vol. 54, No. 12, pp.889–893.

Ma, R. (2006) *The Effects of In-Vehicle Automation and Reliability on Driver Situation Awareness and Trust*, PhD thesis, North Carolina State University, Raleigh, NC.

McDowell, K., Nunez, P., Hutchins, S. and Metcalfe, J.S. (2008) 'Secure mobility and the autonomous driver', *IEEE Transactions on Robotics*, Vol. 24, No. 3, pp.688–697.

Merat, N. and Lee, J.D. (2012) 'Preface to the special section on human factors and automation in vehicles designing highly automated vehicles with the driver in mind', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 54, No. 5, pp.681–686.

Neale, V.L. and Dingus, T.A. (1998) 'Commentaries in: human factors issues for automated highway systems (AHS)', *Intelligent Transportation Systems Journal: Technology, Planning, and Operations*, Vol. 4, pp.111–119.

Nilsson, L. (1995) 'Safety effects of adaptive cruise controls in critical traffic situations', *Proceedings of Steps Forward. Intelligent Transport Systems World Congress,* 9–11 November, Yokohama, Japan, pp.1254–1259.

Nilsson, L. and Nåbo, A. (1996) *Evaluation of Application 3: Intelligent Cruise Control Simulator Experiment*, Report No. 266, VTI Särtryck, Sweden.

Neubauer, C., Matthews, G., Langheim, L. and Saxby, D. (2012) 'Fatigue and voluntary utilization of automation in simulated driving', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 54, No. 5, pp.734–746.

Neubauer, C., Matthews, G., Saxby, D. and Langheim, L. (2010) 'Simulator methodologies for investigating fatigue and stress in the automated vehicle', *Advances in Transportation Studies*, pp.7–18.

Neubauer, C., Matthews, G., Saxby, D. and Langheim, L. (2011) 'Individual differences and automation choice in simulated driving', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55, No. 1, pp.1563–1567.

Nowakowski, C., O'Connell, J., Shladover, S.E. and Cody, D. (2010) 'Cooperative adaptive cruise control: Driver acceptance of following gap settings less than one second', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 54, No. 24, pp.2033–2037.

Omae, M., Hashimoto, N., Sugamoto, T. and Shimizu, H. (2005) 'Measurement of driver's reaction time to failure of steering controller during automatic driving', *Review of Automotive Engineering*, Vol. 26, No. 2, pp.213–215.

Parasuraman, R. and Riley, V. (1997) 'Humans and automation: Use, misuse, disuse, abuse', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 39, No. 2, pp.230–253.

Peters, B. (2001) 'Adaptation evaluation: an adaptive cruise control system used by drivers with lower limb disabilities', *IATSS Research*, Vol. 25, pp.51–60.

Rudin-Brown, C.M. and Parker, H.A. (2004) 'Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies', *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 7, No. 2, pp.59–76.

Santos, J., Merat, N., Mouta, S., Brookhuis, K. and De Waard, D. (2005) 'The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies', *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 8, No. 2, pp.135–146.

Saxby, D.J., Matthews, G., Warm, J.S., Hitchcock, E.M. and Neubauer, C. (2013) 'Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts', *Journal of Experimental Psychology: Applied*, Vol. 19, No. 4, pp.287–300.

Sharek, D. (2011) 'A useable, online NASA-TLX tool', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55, No. 1, pp.1375–1379.

Sheridan, T.B. and Parasuraman, R. (2005) 'Human-automation interaction', *Reviews of Human Factors and Ergonomics*, Vol. 1, No. 1, pp.89–129.

Shladover, S.E. (2015) 'Road vehicle automation history, opportunities and challenges', *Mini-Seminar 'Developments Selfdriving Vehicles in USA'*, 9 November, Delft, the Netherlands.

Smith, B.W. (2013) *SAE Levels of Driving Automation*, http://cyberlaw.stanford.edu/loda (Accessed 19 November, 2015).

Stanton, N.A., Dunoyer, A. and Leatherland, A. (2011) 'Detection of new in-path targets by drivers using Stop & Go adaptive cruise control', *Applied Ergonomics*, Vol. 42, No. 4, pp.592–601.

Stanton, N.A. and Salmon, P.M. (2009) 'Human error taxonomies applied to driving: a generic driver error taxonomy and its implications for intelligent transport systems', *Safety Science*, Vol. 47, No. 2, pp.227–237.

Stanton, N.A., Young, M. and McCaulder, B. (1997) 'Drive-by-wire: the case of driver workload and reclaiming control with adaptive cruise control', *Safety Science*, Vol. 27, Nos. 2–3, pp.149–159.

Stanton, N.A., Young, M.S., Walker, G.H., Turner, H. and Randle, S. (2001) 'Automating the driver's control tasks', *International Journal of Cognitive Ergonomics*, Vol. 5, No. 3, pp.221–236.

Stockert, S., Richardson, N.T. and Lienkamp, M. (2015) 'Driving in an increasingly automated world – approaches to improve the driver-automation interaction', *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015),* 26–30 July, Las Vegas, NV, pp.4226–4233.

Storie, V.J. (1977) *Male and Female Car Drivers: Differences Observed in Accidents*, TRRL Research Report No. 761, Transport Research Laboratory, Crowthome, Berkshire.

Strand, N., Nilsson, J., Karlsson, I.C. and Nilsson, L. (2014) 'Semi-automated versus highly automated driving in critical situations caused by automation failures', *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 27, Part B, pp.218–228.

Summala, H. (2000) 'Brake reaction times and driver behavior analysis', *Transportation Human Factors*, Vol. 2, No. 3, pp.217–226.

Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., Meier, A. and Eggemeier, F.T. (2004) 'Effects of sensory modality and task duration on performance, workload, and stress in sustained attention', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 46, No. 2, pp.219–233.

Tango, F., Minin, L., Aras, R. and Pietquin, O. (2011) 'Automation effects on driver's behaviour when integrating a PADAS and a distraction classifier', in *Digital Human Modeling*, Springer Berlin Heidelberg, pp.503–512.

Teller, A. (2015) 'Moonshots and reality', *Keynote presentation given at South by Southwest (SXSW)*, 17 March, Austin TX.

Törnros, J., Nilsson, L., Ostlund, J. and Kircher, A. (2002) 'Effects of ACC on driver behaviour, workload and acceptance in relation to minimum time headway', *9th World Congress on Intelligent Transport Systems*, 14–17 October, Chicago, IL.

Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stansifer, R.L. and Castellan, N.J. (1979) *Tri-Level Study of the Causes of Traffic Accidents: Executive Summary*, Report No. DOTHS034353579TAC(5), National Highway Traffic Safety Administration, Washington DC.

Tsang, P.S. and Velazquez, V.L. (1996) 'Diagnosticity and multidimensional subjective workload ratings', *Ergonomics*, Vol. 39, No. 3, pp.358–381.

Tsao, H.S., Hall, R.W. and Shladover, S.E. (1997) *Automated Highway Systems Operating Strategies and Events: A Driver's Perspective*, California PATH Working Paper No. UCB-ITS-PWP-97-3, University of California, Berkeley, Richmond, CA.

Veldstra, J.L., Bosker, W.M., De Waard, D., Ramaekers, J.G. and Brookhuis, K.A. (2015) 'Comparing treatment effects of oral THC on simulated and on-the-road driving performance: testing the validity of driving simulator drug research', *Psychopharmacology*, Vol. 232, No. 16, pp.2911–2911.

Verster, J.C. and Roth, T. (2011) 'Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP)', *International Journal of General Medicine*, Vol. 4, pp.359–371.

Wai, J., Lubinski, D. and Benbow, C.P. (2009) 'Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance', *Journal of Educational Psychology*, Vol. 101, No. 4, pp.817–835.

Walker, G.H., Stanton, N.A. and Young, M.S. (2001) 'Where is computing driving cars?', *International Journal of Human-Computer Interaction*, Vol. 13, No. 2, pp.203–229.

Warm, J.S., Parasuraman, R. and Matthews, G. (2008) 'Vigilance requires hard mental work and is stressful', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 50, No. 3, pp.433–441.

Yeh, Y.Y. and Wickens, C.D. (1988) 'Dissociation of performance and subjective measures of workload', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 30, No. 1, pp.111–120.

Young, M.S. (2000) *Attention, Automaticity, and Automation: New Perspectives on Mental Underload and Performance*, PhD thesis, University of Southampton, Southampton, UK.

Young, M.S. and Stanton, N.A. (2004) 'Taking the load off: investigations of how adaptive cruise control affects mental workload', *Ergonomics*, Vol. 47, No. 9, pp.1014–1035.

Young, M.S., Stanton, N.A. and Harris, D. (2007) 'Driving automation: learning from aviation about design philosophies', *International Journal of Vehicle Design*, Vol. 45, No. 3, pp.323–338.

Young, M.S. and Stanton, N.A. (2007a) 'What's skill got to do with it? Vehicle automation and driver mental workload', *Ergonomics*, Vol. 50, No. 8, pp.1324–1339.

Young, M.S. and Stanton, N.A. (2007b) 'Back to the future: Brake reaction times for manual and automated vehicles', *Ergonomics*, Vol. 50, No. 1, pp.46–58.

Young, M.S. and Stanton, N.A. (2007c) 'Miles away: determining the extent of secondary task interference on simulated driving', *Theoretical Issues in Ergonomics Science*, Vol. 8, No. 3, pp.233–253.

Zuckerman, M., Eysenck, S.B.G. and Eysenck, H.J. (1978) 'Sensation seeking in England and America: Cross-cultural, age, and sex comparisons', *Journal of Consulting and Clinical Psychology*, Vol. 46, No. 1, pp.139–149.