VENTURER Trial 3: Interactions between Autonomous Vehicles and Pedestrians and Cyclists

Trial 3 Findings

Notice

This document and its contents are intended solely for information and use in relation to VENTURER.

VENTURER assumes no responsibility to any other party in respect of or arising out of or in connection with this document and/or its contents.

This document has 42 pages including the covers.

Document history

Work Package: Seven			Document ref: WP7 D12(a/b)			
Document Lead Organisation: Atkins			**Document Lead Name: Imogen Weight**			
Revision	**Purpose description**	**Originated**	**Checked**	**Reviewed**	**Authorised**	**Date**
Rev 1.0	Draft Report	IW	BS	PB/RH	CM	01/06/18
Rev 2.0	Final Report	IW	JM	PB/RH	CM	11/06/18

Acknowledgements

Partner	Authors and contributors
Atkins Ltd	Imogen Weight, Peter Blackley, Rebecca Tommey and Carolyn Mitchell
Bristol Robotics Laboratory (BRL)	Tony Pipe
University of the West of England (UWE) - Bristol	Fiona Crawford, Jonathan Flower, John Parkin and Graham Parkhurst

The University of the West of England - Bristol (UWE) Centre for Transport and Society were the primary collators of the human factors elements of this report.

Discussion on the technology was co-ordinated by BRL. BAE Systems provided the Wildcat vehicle and implemented the Wildcat experiments in partnership with UWE.

The VENTURER Simulator was developed by Williams Advanced Engineering.

Atkins co-ordinated the experiments and the production of this report

Glossary

Actors – scenarios involving the AV interacting with a pedestrian or cyclist were undertaken using trained members of the trial team following a script. These were termed 'actors'.

Autonomous Vehicle (AV) – a vehicle which uses a range of advanced vehicle systems, enabling it to operate with no driver intervention. Currently, most autonomous vehicles require some form of driver control. The Wildcat used in VENTURER's third trial is classed as an AV.

Connected Vehicle (CV) – connectivity allows vehicles to communicate with other vehicles and infrastructure, and hence providing information for the driver for example on road, traffic, and weather conditions.

Connected and Autonomous Vehicles (CAV) – autonomous vehicles that communicate with each other and the wider world.

Decision-Making System (DMS) – the software that manages the movements of an autonomous vehicle.

Experiment – the complete set of scenarios, and the responses of participants to those events undertaken as part of Trial 3.

Likert Scale – the 11-point scale (0-10) used by the participants when rating their responses.

Scenario – the interaction or manoeuvre undertaken by the AV at a specified point on the circuit. The scenarios comprise of two types of situation: A, manoeuvres undertaken on links, and B, manoeuvres undertaken at a junction.

Simulator Sickness[1] – a syndrome, similar to motion sickness, which can cause symptoms such as nausea and dizziness in a simulated environment.

Trial – refers to the overall stage of the VENTURER project and includes technology and human factors elements.

VENTURER Simulator – the immersive CAV simulator was provided by Williams Advanced Engineering and was housed at and run by the Bristol Robotics Laboratory throughout the trial.

Wildcat – an adapted Bowler off-road vehicle, heavily modified from a Land Rover Defender, and further customized by BAE Systems to include on-board computing, navigation, and automated driving capability.

---

[1] Johnell, O., Brooks, R. R., Goodenough, M. C., Crisler, N. D., Klein, R. L., et al. (2010). Simulator sickness during driving simulation studies. Accident Analysis & Prevention, 42(3), 788-796.

Table of contents

Table of Tables

Table of Figures

Headline Findings

The findings in Trial 3 indicate that exposure to AVs in safe environments yields generally high trust ratings. A possible implication of the high levels of trust given to AVs operating in test environments is that members of the public might be too trusting of CAVs in their early stages of development. This becomes particularly problematic when considered against the backdrop of on-going debates around liability at different SAE levels of autonomy. The distinctions between the SAE levels can be small and the potential for confusion in terms of a vehicle's capabilities will be more acute during the early stages of CAV development and deployment.

AXA and Burges Salmon's third VENTURER report, acknowledges that more work needs to be done to explore ways in which the risk of consumer confusion could be mitigated or prevented by design and to ensure that drivers are educated as to the capabilities and terms of use of the specific CAVs they may drive. Beyond this, as indicated by Trial 3 findings, CAVs may also need to be clearly identifiable to other road users, such as pedestrians and cyclists, particularly in their early stages of deployment.

Executive Summary

The ability for CAVs to interact safely with a range of road users, particularly in urban settings where interactions are most common, is a crucial step towards enabling the deployment of CAVs onto the UK's public roads. Trial 3 focused on measuring participants' trust in AV behaviour during scenarios where the vehicle had to consider other moving and static road users on straight roads, roads with bends, roads with occluded forward visibility; and at T-junctions. The objective of these experiments was to test responses of observing cyclists, pedestrians and drivers by asking the participants to rate their trust in an AV as they observed it interacting with pedestrian and cyclist actors in several typical highway scenarios.

Participants were asked, thinking of themselves in the role of drivers, pedestrians and cyclists whether they trusted the behaviour of the AV in a range of seven typical highway scenarios, some of which involved the AV interacting with cyclist and pedestrian actors (trained members of the trial team). The participants experienced the same scenarios undertaken by the real world Wildcat and through a simulated environment in the VENTURER Simulator enabling a comparison between the two platforms. In the simulator only, there was also opportunity for a comparison between trust ratings given to the vehicle acting in autonomous mode compared to when the vehicle was driven manually by the driver-participant.

The findings in Trial 3 were consistent with those of Trial 2, which gives further confidence in the experimental design used in these VENTURER trials. The reliability and consistency of the trust measures obtained in Trial 3 are also supported by the correlation between participants' trust in automation and their trust ratings. With people who reported high levels of trust in general automation also providing high trust scores for the AV.

The experiments found that mean ratings for all participants combined, under all scenarios in autonomous mode, were at least 7.3 out of ten, with ten representing the maximum possible trust rating. This is consistent with VENTURER's Trial 2 experiments in which virtually all scenarios received trust ratings of at least seven out of ten. Additionally, in the simulator, trust scores given to a human driver were not significantly higher than those given to the AV for all scenarios, and where a statistically significant difference was noticed, the difference was only small. This shows that exposure to AVs in safe environments yields generally high trust ratings.

Although overall trust was higher for the Wildcat than for the VENTURER Simulator, only two of the scenarios exhibited a statistically significant difference between the two platforms. Where statistical differences were identified they are likely to be explained by specific aspects of the execution of the scenarios or experimental design. These Trial 3 results therefore support the view that much useful research and development can be undertaken in simulation in addition to conducting more resource intensive real world experimentation.

There were no statistically significant differences in trust ratings between cyclist, driver or pedestrian participants reported for any of the scenarios in either the Wildcat or the simulator (a significance level of 0.05). This could indicate that there is no specific need to differentiate messaging around CAVs for different audiences.

1. VENTURER Trials

1.1. Introduction

The VENTURER project is assessing the responses of passengers and other road users to Connected and Autonomous Vehicles (CAVs) in a series of increasingly complex trials and demonstrations in urban settings. The trials and the data collected provide an understanding of how CAV technology performs, how people interact with CAV technology, and they will help inform the development of potential insurance models and the legal frameworks for CAVs. Developing this understanding provides a first step towards addressing some of the barriers and issues in the deployment of CAVs on UK roads.

VENTURER is a broad and experienced partnership of public and private sector organisations that are utilising their expertise to help investigate the issues around the deployment of CAVs on the UK road network. The consortium is made up of the organisations shown in Table 1.

Table 1: VENTURER Consortium

Partner	Role in delivering the trials
Atkins Ltd	Co-ordinating the trials, ensuring the scenarios fulfil the requirements of all partners and programme management.
AXA UK Ltd	Providing technical support and data analysis. Gathering data to develop new insurance models for connected and autonomous vehicles.
BAE Systems (Operations) Ltd	Providing the Wildcat vehicle and testing its autonomous ability.
Bristol Robotics Laboratory (BRL)	Developing and testing the Decision-Making System (DMS) and facilitating technology integration activities.
Fusion Processing Ltd	Developing and testing the sensors.
First Bus Ltd	Providing a bus as a means of collecting data.
Bristol City Council	Supporting trials on public roads.
South Gloucestershire Council	Supporting trials on public roads.
University of Bristol (UoB)	Developing and testing the data algorithms, especially in relation to communications.
University of the West of England (UWE) – Bristol	Analysing perceptions of CAVs, analysing driver performance during the handover of a CAV, and analysing interactions between CAVs and other road users.
Williams Grand Prix Engineering Ltd (WGPE)	Developing the simulator and providing technical support during the trials.
Burges Salmon LLP	Supporting partner, providing legal expertise.

1.2. Objectives

The VENTURER trials and demonstrations focus on three themes: technology, human factors and the insurance and legal aspects of autonomous technology. The following specific objectives relating to these themes were agreed by the consortium at the outset of the project:

- Systematically assess the responses of passengers, AV users, other road users and pedestrians to AVs through trials with the Wildcat and in simulation;
- Establish a realistic simulation environment of roads around Bristol. This can be used in the trials as a test bed for our own and other AV technologies, and for public acceptance studies;
- Develop an understanding of insurance and legal implications of increased vehicle autonomy;
- Investigate the use of world-leading sensors on the Wildcat to detect, track and predict road user and pedestrian behaviour;
- Understand how decision-making algorithms can best use this information for safety and comfort;
- Undertake pod demonstrations to review public acceptance of AVs;
- Investigate the use of innovative sensor technology by collecting data on buses; and
- Support the UK's wider CAV sector messages on working to build a commercially sustainable CAV capability.

These objectives are being achieved by conducting three trials utilising the platforms available to VENTURER partners.

1.3. Platforms

VENTURER has used three different platforms throughout the three trials:

- Road tests of autonomous technology using the Wildcat autonomous vehicle (AV);
- Human factor experiments using the VENTURER Simulator; and
- Handover trials in the UWE STISIM Simulator.

Table 2 outlines the use of the three platforms in the three trials.

Table 2: VENTURER trials and platforms

VENTURER Trials	Platform		
	UWE STISIM Simulator	Wildcat	VENTURER Simulator
Trial 1 Summer 2016	✓	✓	-
Trial 2 Spring/Summer 2017	-	✓	✓
Trial 3 Winter 2017/18	-	✓	✓

In addition, during winter 2018 the performance of the situational awareness and decision-making technology will be investigated in real-time using data passed via a communications link from a bus. VENTURER also undertook a public acceptance demonstration using an autonomous pod between 03 and 06 August 2017 at Millennium Square, Bristol.

1.4. Exploitation

VENTURER has contributed to the West of England becoming a recognised centre for its range of CAV related capabilities that are either not available elsewhere in the UK, or are more advanced than in other regions.

The VENTURER consortium will continue to exploit the capabilities of its partners with continued collaboration throughout and after the three-year project. During the project, the VENTURER partners have demonstrated the ability to:

- Successfully develop an AV Decision-Making System (DMS) integrated with sensor technology;
- Successfully manage and deliver safe technology and participant trials in simulation and on local roads;
- Develop the hardware and software for an AV simulator;
- Conduct research into public expectations, acceptance and response to CAVs; and
- Analyse the legal and insurance implications of the deployment of CAVs onto UK roads.

1.5. Purpose and structure of this report

This report presents the method, findings and discussion of VENTURER's third and final trial. The aim of Trial 3 was to enable and examine interactions between Autonomous Vehicles (AVs) and pedestrians and cyclists. The findings and discussion in this report focus specifically on participant reactions to these interactions.

The remainder of this report is structured as follows:

- Section 2 – Trial 3;
- Section 3 – Approach;
- Section 4 – Technology;
- Section 5 – Human Factors; and
- Section 6 – Discussion.

2. Trial 3

Trial 3, completed in April 2018, focused on investigating road user behaviour and trust when AVs form part of the traffic in urban conditions. The ability of AVs to interact safely with a range of road users, particularly in urban settings where interactions are most common, is a crucial step towards enabling the deployment of CAVs onto the UK's public roads.

An Autonomous Vehicle (AV) is a vehicle which uses a range of advanced vehicle systems, enabling it to operate with no driver intervention in some or all driving situations.

Connected and Autonomous Vehicles (CAV) are autonomous vehicles that communicate with each other and the wider world including infrastructure.

The Wildcat used in VENTURER's third trial is classed as an AV and is referred to accordingly throughout this report.

2.1. Aims

The overall aim of Trial 3 was to assess driver, cyclist and pedestrian responses to an AV's manoeuvres on the UWE campus.

2.2. Objectives

Trial 3 focused on measuring participants' trust in AV behaviour during scenarios where the vehicle had to consider other moving and static road users on straight roads, roads with bends, roads with occluded forward visibility; and at T-junctions. The objective of these experiments was to test responses of observing cyclists, pedestrians and drivers by asking the participants to rate their trust in an AV as they observed it interacting with pedestrian and cyclist actors in a number of typical highway scenarios.

The objectives for each theme in Trial 3 are outlined in sections 2.2.1 and 2.2.2.

2.2.1. Technology

The objectives for the technology were to:

- Facilitate the experiments by providing the technology;
- Assess the performance of the technology components in the Wildcat and VENTURER Simulator in a range of scenarios.

The outputs of Trial 3 demonstrate the progression of VENTURER's technology throughout the project.

2.2.2. Human Factors

The objectives for Trial 3 in terms of human factors were to:

- Collect and analyse data on participant trust in AV behaviour;
- Compare participant trust scores across AV platforms; and
- Compare participant trust scores between participant roles (cyclist, driver and pedestrian).

The human factors outputs of Trial 3 will inform testing and manufacturing standards for AV development in the UK.

2.3. Research Questions

This report presents the findings from Trial 3 participant experiments with the aim of addressing the research questions identified from the literature review provided in Table 3.

Table 3: Human factors research questions

Human Factors Research Questions
How does the trust in the AV vary depending on the manoeuvre the AV is undertaking?
How does trust in the AV vary depending on whether the observer is a pedestrian, cyclist or driver-passenger in the AV?
How does trust in the AV vary depending on whether the AV is the real world Wildcat or the VENTURER Simulator?
In the simulator environment, how do the observing pedestrian and cyclist (as passengers in the AV at that time) rate the events when the vehicle is being driven by a human as compared with when the vehicle is in AV mode?
Does trust in the AV correlate with age, driving experience or trust in automation?

Participants were asked, thinking of themselves in the role of drivers, pedestrians and cyclists whether they trusted the behaviour of the AV. Age and role (cyclist, driver, and pedestrian) have also been included in the analysis.

3. Approach

Trial 3 was designed to derive trust scores in both the real world and in a simulated environment, for three categories of participant: drivers; cyclists and pedestrians.

3.1. Programme

Trial 3 was undertaken during early 2018. Due to a range of technical and logistical challenges the participant experiments were undertaken in two phases with a six week break to allow for resolution of issues, discussed further in section 4. The Trial 3 delivery programme is shown in Table 4.

Table 4: Trial 3 delivery programme

Item	Dates
Recruit participants	October – December 2017
Pre-trial testing	08 – 24 January 2018
Participant experiments	24 January – 09 February 2018
Participant experiments	26 March – 13 April 2018

3.2. Participants

Three categories of participant were involved in the Trial 3 participant experiments:

- Cyclists – who observed the Wildcat from the footway of the UWE carpark with their bicycle, and sat in the front passenger seat in the VENTURER Simulator with a bicycle mounted on the back of the vehicle;
- Drivers – who sat inside the Wildcat and in the driving seat of the VENTURER Simulator; and
- Pedestrians – who also observed the Wildcat from the footway of the UWE carpark and sat in a rear passenger seat in the VENTURER Simulator.

These were roles assigned to participants for their involvement in the experiments and were based on their self-reported primary mode of travel for all journeys.

Participants were recruited through the Centre for Transport and Society's (UWE) contact list as well as through local cycling and pedestrian groups such as Sustrans and Living Streets. UWE were responsible for participant recruitment with support from Atkins.

Participants' age ranged from 18 to 79 years and of those who had passed their driving test, driving experience ranged from 10 months to 59 years. In total, 24 (18%) participants fall under the category of 'older adults' being 65 years or older. In both the Wildcat and the simulator, participants were asked a pre-question about their level of trust in how they think the AVs will respond to events. The responses to this show that participants ranged from AV sceptics to enthusiasts.

For data to be included in the analysis, participants had to complete a minimum of two full circuits. In total, 76 participants (57% of the total participants) completed at least two circuits in both the Wildcat AV and VENTURER simulator. Table 5 shows the breakdown of participants who completed at least two circuits in each part of the trial according to participant role.

Table 5: Number of participants

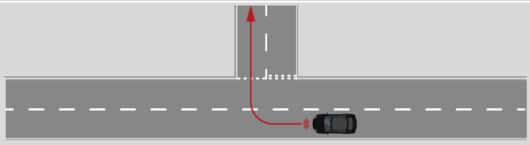
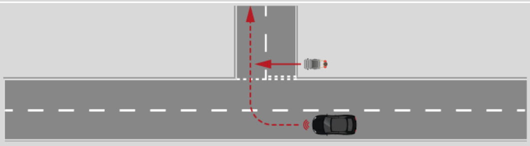
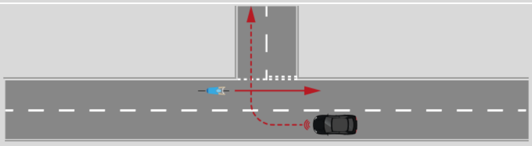
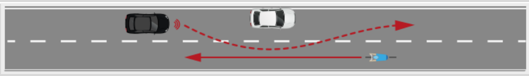
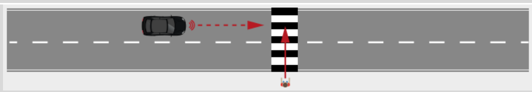
Platform	Total	Cyclists	Drivers	Pedestrians
Wildcat	95	33	35	27
VENTURER Simulator	110	37	41	32
Both platforms	76	26	29	21

In the Wildcat, not all participants were able to complete three circuits due to technical issues which either cut the session short or resulted in a late start, meaning there was insufficient time to complete the circuits. In the simulator there was only one occasion in which an error resulted in participants being unable to complete three circuits, the majority of people who did not complete three circuits dropped out early due to nausea.

3.3. Scenarios

Seven scenarios typical of those encountered on urban roads were tested during the participant experiments, these are described in Table 6. Scenarios either occurred on a road link (A scenarios) or at a T-junction (B scenarios).

Table 6: Trial 3 participant experiment scenarios

Scenario	Name	Diagram
A1	Proceeding along a road with a zebra crossing, but no-one crossing.	
B1	Turning right into an empty road.	
B2	Turning right with pedestrian crossing side road.	
B3	Turning right into side road with on-coming cyclist.	
A2	Overtaking a parked car with an on-coming cyclist.	
A3	Overtaking a parked car with no on-coming traffic.	
A4	Proceeding along a road with pedestrian crossing zebra crossing.	

The AV used in Trial 3 was the Wildcat and all scenarios were also replicated in the VENTURER Simulator. All scenarios including the AV interacting with a pedestrian or cyclist were undertaken using members of the VENTURER team following a script, termed ‘actors’.

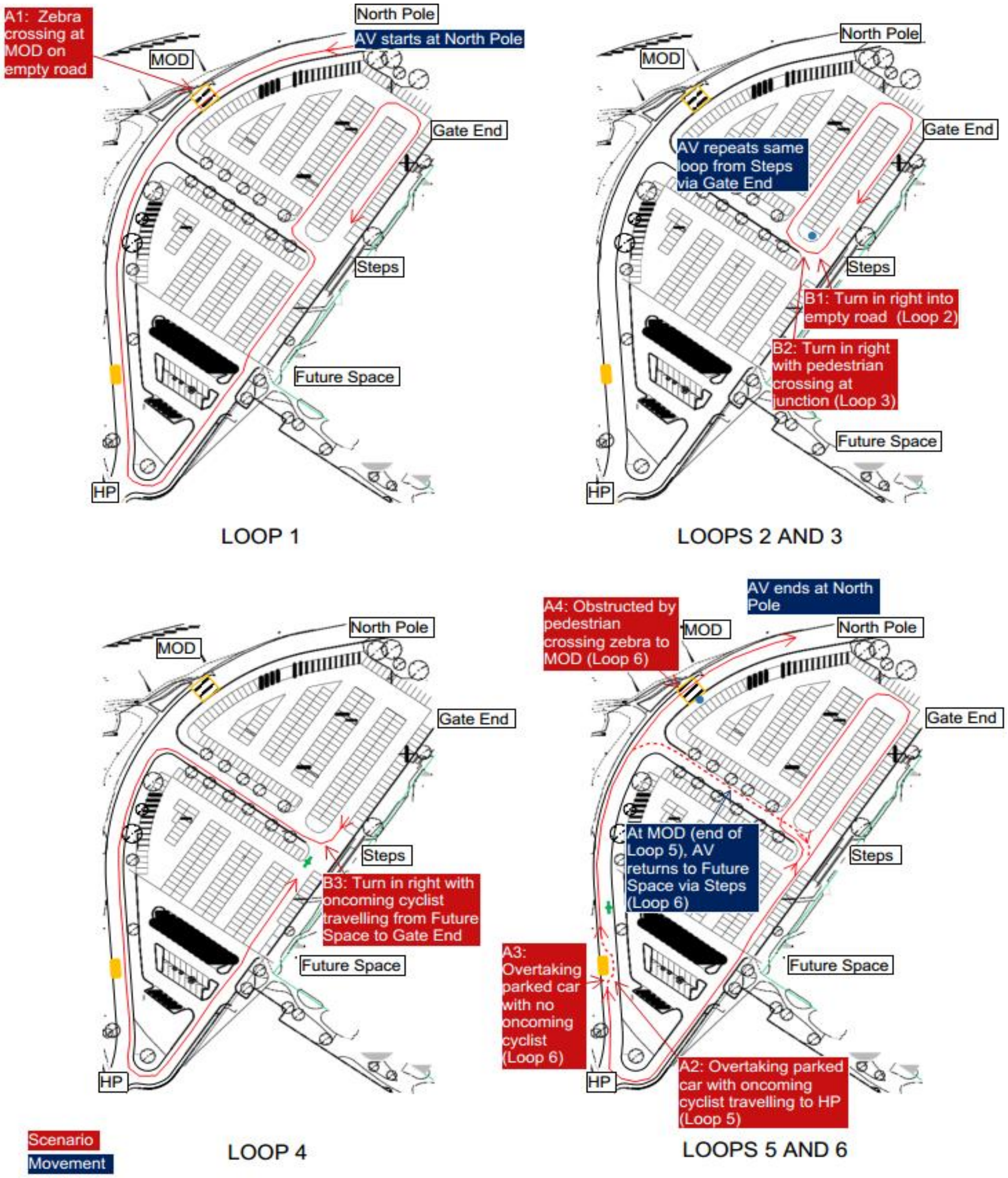
3.4. Location

The Wildcat and VENTURER Simulator are both housed at UWE's Bristol Robotics Laboratory (BRL) Autonomous Driving Zone.

The Wildcat scenarios were undertaken on UWE campus roads and were designed to be undertaken at a maximum speed of 20mph which is the speed limit of many roads in Bristol city centre and on UWE campus.

A complete circuit, including all seven scenarios can be completed in the six loops shown in Figure 1. Variable lengths of time between scenarios helped to create the impression of the random arrival of events to participants.

Figure 1: Experiment circuit



3.5. Experimental Design

The Dependent Variable (DV) recorded during the trial was each participants' trust rating in the vehicle in terms of how it responded to each scenario recorded using an 11- point Likert scale (the scale 0-10).

Participants were asked to respond to the relevant experience question on a clipboard immediately following each scenario encountered in both the Wildcat and the simulator.

A partial counterbalancing method was employed when deciding the order of the platforms, i.e. half of the participants experienced the trial in the Wildcat first, and the other half undertook the simulator trial first. This allowed cross-platform comparisons by eliminating the issue of carry over effects between platforms.

In the simulator, the participants undertook two circuits in autonomous mode followed by a third circuit which was manually driven by the driver participant. The driver participant therefore did not provide trust ratings for that third circuit.

All participant responses were validated against self-reported scores for impulsivity, risk taking, self-control, personality and driving experience using Qualtrics software[5]. Qualtrics is a software for collecting and analysing data.

3.6. Safety

As in all VENTURER trials, the safety of consortium members, participants and the general public was the number one priority when conducting the experiments. A detailed risk assessment and safety case was produced in accordance with best practice as set out in the DfT Code of Practice for Testing – the Pathway to Driverless Cars[2].

The general safety principles were that:

- The participants could ask for the experiments to stop at any time;
- The public were kept away from the vehicle by marshals managing traffic flow;
- The Wildcat vehicle was always under the supervisory control of a trained safety driver;
- There was only one person in the vehicle (excluding the trained safety driver) at a time;
- The vehicle was limited to a maximum speed of 20mph; and
- The trial adhered to the DfT's Code of Practice guidance.

Additional safety features of the trial included the use of a pram being pushed by the pedestrian, and a bicycle trailer being hauled by the cyclist. These were used by the actors and created a larger object for the Wildcat to detect hence reducing the possibility of the Wildcat colliding with an actor.

[2] Pathway to Driverless Cars, DfT https://www.gov.uk/government/publications/driverless-cars-in-the-uk-a-regulatory-review

4. Technology

The experimental design was based on using two platforms, the Wildcat and the VENTURER Simulator. Following on from Trial 2, the Trial 3 human factors experiments used a similar technology specification consisting of the aspects outlined in the following sections.

The main objective for the technology was to facilitate the participant trials and hence, enable the human factors research. The technology components used in Trial 3 are described below.

4.1. Wildcat

The Wildcat was used during all participant experiments. Data from the Wildcat was collected and stored on one of the two on-board computer platforms and downloaded periodically. This data included information about acceleration, deceleration, steering angle, start-stop and, most importantly in the context of Trial 3, detection of the approach and passing of other road users including pedestrians and cyclists.

Figure 2 shows the Wildcat (registration D1 BAE), an adapted Bowler off-road vehicle, heavily modified from a Land Rover Defender, and further customized by BAE Systems to include on-board computing, navigation, and automated driving capability.

Figure 2: Wildcat

The Wildcat was in the direct or supervisory control of a trained safety driver from BAE Systems at all times; the former for teaching a new route and the latter for safety motivated manual interventions. As in all other trials, there was a capability to stop or revert to manual driving instantaneously at any time for safety purposes.

4.2. Sensors

The sensors detailed in In the context of the carefully-designed experiments in Trial 3, recognition was not strictly necessary as the object behaviour was partially predictable. The interacting cyclist and pedestrian (actors) followed defined, scripted movements, but with potential variations in timing. Therefore, the combination of both sensor suites was used to ensure the highest degree of reliability needed for repeatable participant experiences. In practice, each triggered scenario in the experiment sequence used an individually-tuned combination of all the sensors determined by testing. Naturally, a production system would need to employ redundant sensors and cross-validate their outputs to achieve integrity and fault tolerance. However, for the trials, safety was assured by the presence of the trained driver able to revert to manual control, so the sensors were tuned to maximize repeatability instead.

Table 7 were fitted to the Wildcat during Trial 3. They consisted of primary sensing and perception units from Fusion Processing and the BAE Systems backups. The backups were supplied in the trial because sensing and perception have proven to be one of the most challenging technology sub-components of CAV research in general, and this project has been no exception in that respect.

The primary sensor suite, furnished by Fusion Processing, used a range of multi-modal sensors to both sense and recognise objects. The backup system, employing the same BAE Systems sensors as deployed in Trial 2, could only sense the presence of objects and could not recognise them.

In the context of the carefully-designed experiments in Trial 3, recognition was not strictly necessary as the object behaviour was partially predictable. The interacting cyclist and pedestrian (actors) followed defined, scripted movements, but with potential variations in timing. Therefore, the combination of both sensor suites was used to ensure the highest degree of reliability needed for repeatable participant experiences. In practice, each triggered scenario in the experiment sequence used an individually-tuned combination of all the sensors determined by testing. Naturally, a production system would need to employ redundant sensors and cross-validate their outputs to achieve integrity and fault tolerance. However, for the trials, safety was assured by the presence of the trained driver able to revert to manual control, so the sensors were tuned to maximize repeatability instead.

Table 7: Sensors

Sensor type	Description	Owner	Primary/Backup
Front radar	Range up to 200m for vehicles.	Fusion Processing	Primary
Front camera	Forward facing video camera, dual lens.		
Two front corner modules	Each with dual radar, dual camera sensor utilising radars and two cameras, 100m range for vehicles.		
Two side modules			
Two blind-spot detect modules	Each with a single radar, range up to 100m for vehicles and a single camera.		
Velodyne HDL-64 3D LiDAR	Roof mounted with ~50m range or higher.	BAE Systems	Backup
SICK 2D LiDAR	Front mounted with ~20m range.		

4.3. Decision-Making System (DMS)

To achieve safe automated driving in the presence of pedestrians and cyclists, the trial was designed to deploy predictable behaviours, with the Wildcat actions being determined by triggers received from the sensors based on the situation in the wider environment. For example, given prior knowledge that *something* was due to cross in front of the Wildcat at a certain point in the trial, the DMS would await sensor indications that the pedestrian had passed in front of the vehicle and moved safely to the far side of the road. This allowed the trial to accommodate inevitable variations in the timing of the movements, while maintaining the effect that the Wildcat waits for the pedestrian.

This approach represents a pragmatic balance between genuine automated driving responses and predictable behaviour which is essential for the trials setting, where the safety driver is relied upon to prevent collisions and must therefore be able to identify unexpected behaviour rapidly.

The role of the DMS was to manage the schedule of the trial, executing a portion of the predefined route and then awaiting the appropriate trigger before proceeding. Timeouts, with appropriate warnings to the safety driver, were deployed to achieve smooth running.

As a step up from Trial 2, the DMS was upgraded to employ a behaviour tree as its core decision-making engine. Behaviour trees are a way to structure the switching between different tasks in an autonomous agent and are useful for scenarios in which sequences or priorities of actions need to be altered during the development process. They are popular in computer games for enabling a level of artificial intelligence in virtual characters, and are now being proposed to enable autonomy in real world robots[3].

This DMS was replicated for both the Wildcat and the VENTURER Simulator, with both employing the standard Robot Operating Systems (ROS) interface[4]. ROS provides an open-source standard interface for sharing information between components in a robotic system. For the simulator, the trigger inputs were replicated, derived from the world state information available from its software.

4.4. VENTURER Simulator

The VENTURER Simulator is shown in Figure 3. The simulator was provided by Williams Advanced Engineering and was housed at and run by the Bristol Robotics Laboratory throughout the trial.

[3] Behaviour Trees http://www.csc.kth.se/~petter/Publications/ogren2012bt.pdf

4 Robot Operating System (ROS) http://www.ros.org/

Figure 3: VENTURER Simulator

The software was set up to provide the same environment as that experienced in the real world in the Wildcat in terms of layout of the road and certain distinctive landmarks (e.g. the Exhibition Centre building), but other parts such as foliage were rendered in a more generic manner.

4.5. Outcomes

The objectives for the technology were to:

- Facilitate the experiments by providing the technology;
- Assess the performance of the Wildcat and VENTURER Simulator in a range of scenarios.

Table 8 summarises the successes of Trial 3 in relation to the Wildcat and the Simulator.

Table 8: Technology component performance

Component	Success criteria	Result
Wildcat	The Wildcat performed to schedule, temporally and spatially.	Most issues of 'jerkiness' observed during Trial 2 were removed for Trial 3. Despite the GPS unit failure, the Wildcat was able to operate successfully to schedule during both parts of the main experimental programme of Trial 3.
Simulator	The simulator performed adequately to facilitate the participant trials.	Useful progress in reducing participant nausea was achieved. The Trial 3 experiments were conducted without significant incident enabling results comparable with the Wildcat.

4.6. Limitations

4.6.1. GPS unit failure

In February 2018 the Wildcat's GPS unit, which had been highly reliable up to that point, malfunctioned in such a way that the vehicle would lose a sense of position at certain parts of the route. Despite best efforts, it was not possible to repair this unit in situ. Although this was an uncontrollable error, it was one of the most influential factors resulting in the six-week delay to the Trial 3 programme.

The delay was approved by the VENTURER consortium, and BAE Systems staff were given time to resolve the problem, with alternative plans formed that ranged from off-site repair of the unit to the purchase of a new unit. An intermediate solution was found which involved making use of another unit already in BAE System's possession.

Consequently, in March 2018, the trial recommenced after thorough testing of the replacement unit. This was achieved in a creditably short timescale by BAE staff, and experiments were resumed with a delay that could be incorporated into the overall programme. The replacement GPS unit worked accurately and consistently throughout the remainder of the experiments, although at a lower performance specification than the original.

4.6.2. Simulator sickness

Despite improvements to the simulator screens, shortening of the amount of time spent by participants in the simulator and trial staff being briefed on how to reduce the impact of nausea, the effect of simulator sickness still impacted Trial 3 significantly.

Participants were asked to rate their feelings of nausea on a scale from 0-10 before commencing the experiments and throughout their participation. The nausea ratings for the driver participants in the Wildcat were found to be relatively low, both during and at the end of the trial. However, in the simulator almost a third of participants (32%) rated their nausea at 5 or higher after the first circuit, and this increased to 41% after the third circuit.

Although some people reported the nausea subsiding as the trial continued, for others it got worse. The data after the third circuit is also somewhat biased as people who dropped out of the simulator part of the trial early due to nausea did not answer this question.

While a worthwhile decrease in the occurrence and severity of participant's nausea, further improvements will be the focus of further post-project work given the project partners' increased understanding of the issue.

4.7. Further research

There are two main areas of further research that have been identified during the technology-related parts of VENTURER that pertain directly to Trial 3, as well as having wider implications for the whole project, and CAV research in general. These are summarised below:

1. Enhancing the VENTURER Simulator to minimise the impact of simulator sickness using learnings from Trials 2 and 3; and
2. Continued development of AV sensing and perception to improve the speed and accuracy of this.

5. Human Factors

The human factors experimentation has not been linked so much with the performance of humans, but is rather a study in risk perception. The objectives for Trial 3 in terms of human risk perception were to:

- Collect and analyse data on participant trust in AV behaviour;
- Compare participant trust scores across AV platforms; and
- Compare participant trust scores between participant types (cyclist, driver and pedestrian).

The findings of the Trial 3 participant experiments are introduced and discussed in the following sections. Results from the statistical analyses of the Wildcat, simulator and cross-comparison experiments can be found at Appendix B, Appendix C and Appendix D respectively.

5.1. Wildcat experiments: Analysis and Results

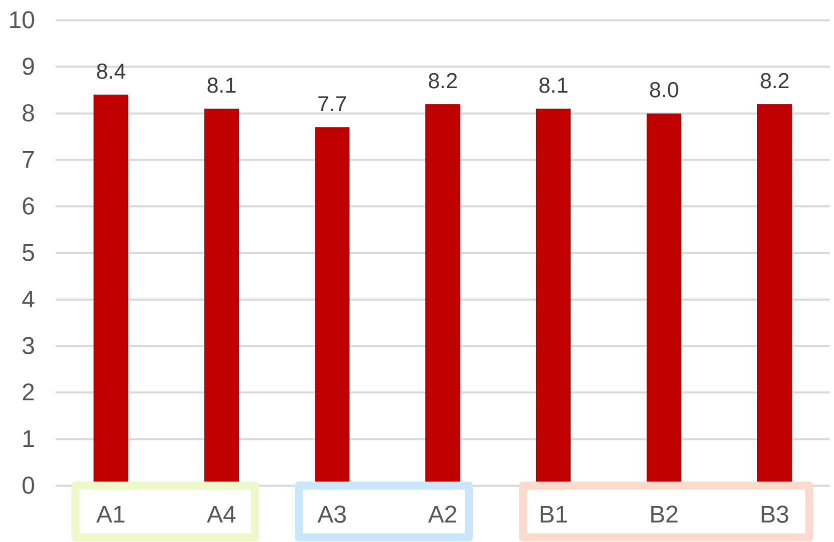
Participants were asked to rate their trust in the Wildcat in terms of how it responded to each of the seven scenarios on a scale from 0 (no trust) to 10 (complete trust). Overall, the trust scores relating to the Wildcat were relatively high with the most commonly given trust scores being nine out of ten.

5.1.1. Comparison by scenario

Figure 4 contains the average trust ratings for each event observed for the Wildcat. These averages have been calculated using data from participants who completed at least two circuits (n = 95) and only data from complete circuits have been included.

For clarity the scenarios have been ordered in the table to enable direct comparison between scenario types, i.e. scenarios at a zebra crossing, overtaking a parked car or at a T-junction.

Figure 4: Average trust ratings for Wildcat – by scenario



Scenario	Manoeuvre	Interaction
A1	Approaching a zebra crossing	None
A4		Pedestrian
A3	Overtaking a parked car	None
A2		Cyclist
B1	Right turn	None
B2		Pedestrian
B3		Cyclist

Figure 4 shows that average trust ratings for the manoeuvres without interactions varied, with lower ratings given to overtaking a parked car (A3), slightly higher ratings for the right turn (B1) and the highest ratings were for crossing an empty zebra crossing (A1).

The two manoeuvres involving an interaction with a pedestrian (A4 and B2) scored similarly in terms of trust in the AV (8.1 and 8.0 respectively). The two manoeuvres involving an interaction with a cyclist (A2 and B3) also scored similarly in terms of trust in the AV (both 8.2).

Overall, interactions involving pedestrians scored slightly lower than those involving no interaction, whereas interactions involving cyclists scored slightly higher.

For the scenarios at a zebra crossing (A1 and A4) there was a statistically significant difference between trust scores depending on whether the scenario included an interaction with a pedestrian. For A1 versus A4, the presence of a pedestrian decreased the trust score at the pedestrian crossing. The presence of a pedestrian at the zebra crossing is likely to act as a reminder of the risk involved in the scenarios presented which could result in a decrease in trust.

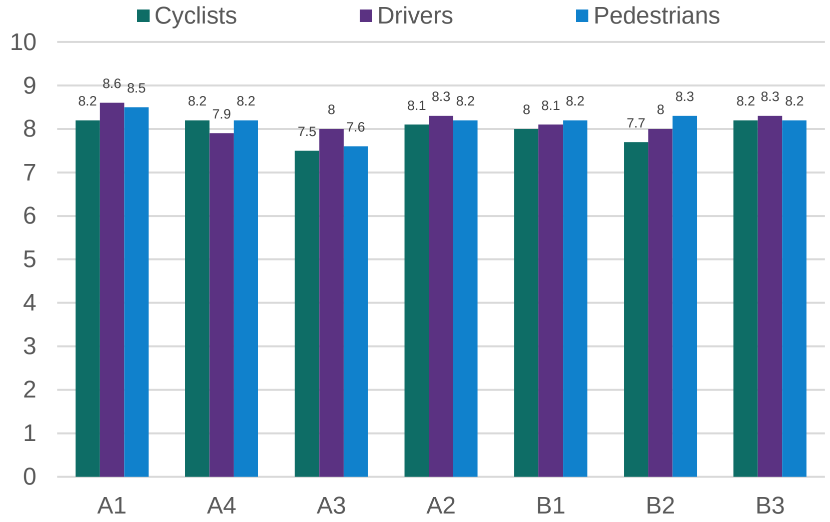
There was also a statistically significant difference found between scenarios A3 and A2, overtaking a parked car with or without an on-coming cyclist. Here, the presence of an on-coming cyclist increased the trust score when overtaking a parked car. This is similar to the result from Trial 2 relating to an on-coming car. In both cases, the increased trust could be due to the vehicle undertaking a smoother movement as the manoeuvre occurs at lower speeds than when there is no on-coming traffic. Another possible explanation is that a successful interaction with another road user raises trust in the AV, and when there is no interaction the observer is left wondering how the AV would have responded had another road user been present.

There is no statistically significant difference in trust ratings for a right-hand turn into a side road depending upon whether there was a pedestrian crossing the junction, there was an on-coming cyclist along the main road, or the manoeuvre did not involve an interaction.

5.1.2. Comparison by participant category

The average trust ratings according to the role of the participant are shown in Figure 5. For the Wildcat experiments there was an additional differentiating factor between participants with different roles as the driver participants were inside the vehicle, whereas the cyclist and pedestrian participants were observing the events from outside of the vehicle. Trust ratings for at least two complete circuits were collected for 35 drivers, 33 cyclists and 27 pedestrians.

Figure 5: Average trust ratings for Wildcat – by participant category



Scenario	Manoeuvre	Interaction
A1	Approaching a zebra crossing	None
A4		Pedestrian
A3	Overtaking a parked car	None
A2		Cyclist
B1	Right turn	None
B2		Pedestrian
B3		Cyclist

Figure 5 shows that overall, the cyclist participants were slightly less trusting of the Wildcat than the pedestrian and driver participants, with an average trust rating across all events of 8.0 compared to 8.2 for both driver and pedestrian participants. The exception to this was scenario A4, where several driver participants reported jerky braking, although very few of the cyclist or pedestrian participants observed this from outside of the vehicle. This may have contributed to the lower trust ratings for this event from driver participants.

Interestingly, for the two scenarios involving an interaction with a pedestrian actor (A4 and B2), the driver participants reported lower average trust in the manoeuvre than the pedestrian participants. It is possible that this is because the driver participants consider the pedestrian actor to be a more vulnerable or unpredictable road user than the pedestrian participants do, or that the difference in the location of the point of observation meant that different attributes of the interactions were being noticed.

In contrast, for scenarios involving an interaction with the cycling actor (A2 and B3), driver participants rated the events higher than the cyclist participants. This could reflect the cyclist participants' understanding of the necessary safety margins for cyclists being different and larger than those assumed by drivers. It should be noted that some cyclist participants reported having been involved in collisions with vehicles and hence have experience of safety margins having been pushed beyond the limit.

Analysis of standard deviations suggested that the cyclist participants were slightly more consistent in terms of their trust in the Wildcat compared to the other participant categories. For pedestrian participants, more variability in trust scores was observed for the scenario crossing an empty zebra crossing (A1). Comments from participants suggest that this could be due to uncertainty about what the Wildcat would have done, had a pedestrian been present.

Despite this, none of the differences in trust ratings between participant categories was statistically significant. None of the interaction terms are statistically significant either. Therefore, the participant role does not have a significant effect on the relative trust ratings between similar events with and without a pedestrian or cyclist interaction. This suggests that neither the usual transport mode of the observer, nor their viewpoint (either within the vehicle or on footways) had an impact on their trust in the Wildcat.

5.2. VENTURER Simulator experiments: Analysis and Results

In total, there were 32 more participants in the VENTURER Simulator than the Wildcat due to technical difficulties meaning that some participants only undertook the VENTURER Simulator part of the trial.

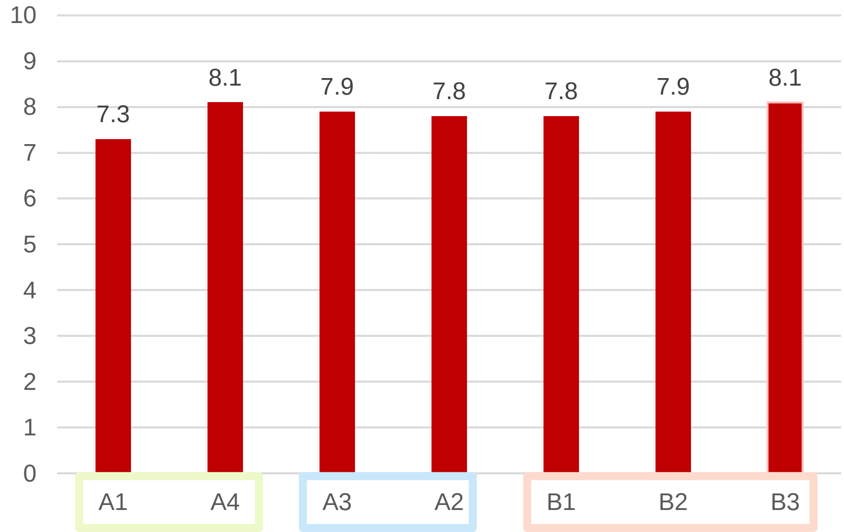
The researcher running the VENTURER Simulator kept a log detailing any unusual events during the trial. Using this log, four occasions during which the behaviour of the simulated pedestrian behaved erratically were identified. Each time, this severely impacted upon trust scores for this and subsequent events. Seeing as this was a result of a glitch in the set-up of the simulated pedestrians, data effected by the error was removed and is not included in the analysis below.

Participants were asked to rate their trust in the VENTURER Simulator in terms of how it responded to each event on a scale from 0 (no trust) to 10 (complete trust). Overall, the trust scores relating to the simulator driving in autonomous mode were relatively high, with the most common trust scores being eight out of ten.

5.2.1. Comparison by scenario

Figure 6 shows the average trust ratings for each scenario observed in the simulator. These averages have been calculated using data from participants who completed two circuits in AV mode (n = 110). Only data from complete circuits have been included.

Figure 6: Average trust ratings for VENTURER Simulator – by scenario



Scenario	Manoeuvre	Interaction
A1	Approaching a zebra crossing	None
A4		Pedestrian
A3	Overtaking a parked car	None
A2		Cyclist
B1	Right turn	None
B2		Pedestrian
B3		Cyclist

Figure 6 shows that the scenario in which the AV crossed an empty zebra crossing (A1) received an average trust rating of 7.3. This was the lowest average trust rating of all scenarios in either the simulator or the Wildcat. It is expected that this is the result of the scenario ordering, as participants always encountered this scenario first and it was shortly after the start of the simulation. Observations from participants in the simulator suggested that trust ratings may have been impacted by the rapid acceleration and the associated noise of the simulated vehicle setting off in autonomous mode.

The joint highest scoring scenarios in the simulator (A4 and B3) both involve an interaction with a moving road user.

Statistical analysis revealed that there is a statistically significant difference between trust ratings in the simulator when approaching a zebra crossing where a pedestrian is present (A4) compared to when no pedestrian is present (A1). This is likely to be due to the trial start effect described above.

There is no statistically significant difference between trust ratings when overtaking a parked car with or without an oncoming cyclist (A2 and A3). This could be because improvements made in the simulator since Trial 2 meant that the overtake was a much smoother manoeuvre, particularly when there was no on-coming traffic, where the AV is moving away slightly faster.

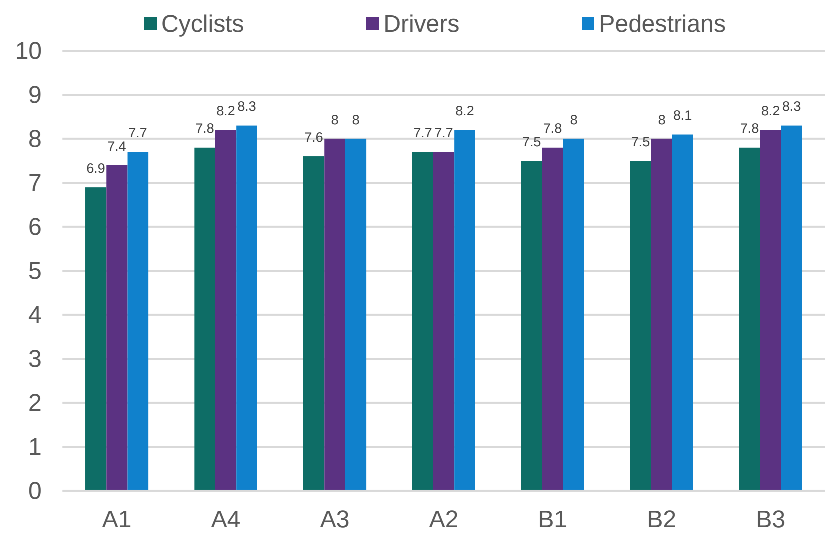
Statistical analyses revealed that trust ratings in the simulator for the right turn scenarios included significant differences. The right turn into an empty side road and into a side road with a pedestrian crossing the junction were not significantly different (B1 and B2). However, the trust ratings for these two scenarios were both significantly lower than for a right turn with an on-coming cyclist (B3). This could be because the presence of other traffic gives confidence that the AV is behaving properly.

5.2.2. Comparison by participant category

The trust ratings were also examined by comparing across the three participant roles, namely cyclist, driver and pedestrian. As the simulation only works from inside the vehicle, in the VENTURER Simulator all participants observed the scenarios from inside the vehicle and therefore the only experiment-related difference between participant types is their seat in the vehicle: the driver participant sat behind the steering wheel, the cyclist participant in the front passenger seat and the pedestrian in the back seat on the left-hand side.

As in any other car, the driver participant had the additional visual benefit of the wing and rear-view mirrors. The cyclist and pedestrian participants had a broadly similar experience as the positioning of the screens meant that the simulation could be observed from both the front and rear passenger seats.

Figure 7: Average trust ratings for VENTURER Simulator – by participant category



Scenario	Manoeuvre	Interaction
A1	Approaching a zebra crossing	None
A4		Pedestrian
A3	Overtaking a parked car	None
A2		Cyclist
B1	Right turn	None
B2		Pedestrian
B3		Cyclist

Although there were no statistically significant differences in trust scores given by the three participant groups for any scenario (at significance level of 0.05), there were some small differences in the average trust ratings for scenarios by participant role.

For every scenario, the average trust rating for cyclist participants was the lowest (or joint lowest) amongst the three participant types. It is possible that this is because cyclists are more aware of their vulnerability as a road user than the other participant types, particularly as some mentioned previous crashes they had been involved in.

Despite the differences in viewpoint between the driver and pedestrian participants, their trust ratings were relatively similar for five out of the seven scenarios. The two scenarios with more substantial differences in trust ratings between drivers and pedestrians involved the AV crossing an empty zebra crossing (A1) and overtaking a parked car with an on-coming cyclist (A2).

The lowest average trust ratings for all three participant categories were given to the scenario in which the AV crosses the empty zebra crossing (A1). However, the trust scores of the pedestrian participants were the highest on average for this event. This could be due to the difference in viewpoint, as the more limited view of the simulator screens from the rear seat could have resulted in pedestrian participants being less aware of jerky movements.

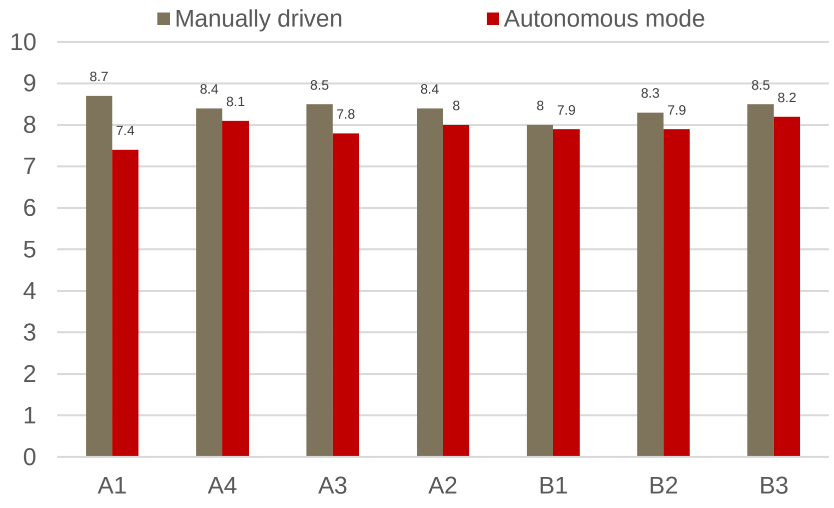
Overtaking a parked car with an on-coming cyclist (A2) is probably the scenario which pedestrians will be able to relate to the least as a road user, and therefore perhaps their trust scores were not influenced by anxiety related to previous experiences in this situation.

5.2.3. Comparison by autonomous or manual mode driving

As noted in section 0, when three simulator runs were completed, the last was manually driven by the driver participant so that human and automated driving could be compared for trust scores. As the driver participant was driving the simulator for the manually driven circuit, comparisons can only be made for pedestrian and cyclist participants.

Since only one manual circuit was undertaken, the requirement for at least two trust ratings for each person for each scenario does not apply in this subsection. In total, 66 people completed at least one circuit in autonomous mode and a complete manually-driven circuit in the simulator. It should also be noted that each set of participants were rating their trust in different drivers, and in most cases the driver was unknown to them prior to the trial.

Figure 8: Average trust ratings for VENTURER Simulator – by driving mode



Scenario	Manoeuvre	Interaction
A1	Approaching a zebra crossing	None
A4		Pedestrian
A3	Overtaking a parked car	None
A2		Cyclist
B1	Right turn	None
B2		Pedestrian
B3		Cyclist

Figure 8 shows that overall, trust was higher when the simulator was being manually driven and that the trust ratings in a human driver are higher for all seven scenarios, however not all of these differences are statistically significant.

The largest difference in trust scores between manual and autonomous mode relates to event A1 (crossing an empty zebra crossing), which could be due to the impact of the abrupt setting off process in autonomous mode, as discussed in section 5.2.1.

The trust ratings for both overtaking manoeuvres (A3 and A2) were also statistically significantly higher when the simulator was manually driven. This suggests that although manoeuvres were undertaken more smoothly in the VENTURER Simulator in Trial 3 compared to Trial 2, there may still be progress needed to ensure the movements of the simulated AV reflect the behaviours of human drivers.

Both scenarios involving an on-coming cyclist had statistically significant differences between manual and autonomous mode (A2 and B3), whereas neither of the events involving pedestrian interactions did (A4 and B2). This could reflect the fact that participants reported that the AV was overly cautious when interacting with the cyclist, but that this was less of an issue when interacting with a pedestrian or because pedestrians are considered to be less predictable and less likely to be wearing protective clothing, for example a helmet than cyclists.

It should be noted that seeing as the manually driven circuit always occurred after the two autonomous circuits, there may have been a presentation-order bias effect by which trust ratings improved with experience. There was a statistically significant increase in trust ratings in the second autonomous circuit in the simulator compared to the first autonomous circuit which could be due to trust growing in the AV as more manoeuvres are observed but could also be due to the increased familiarity with the testing environment and the events observed. If the latter is true, then we might expect the third circuit to have the highest trust ratings irrespective of the driver type (AV or human).

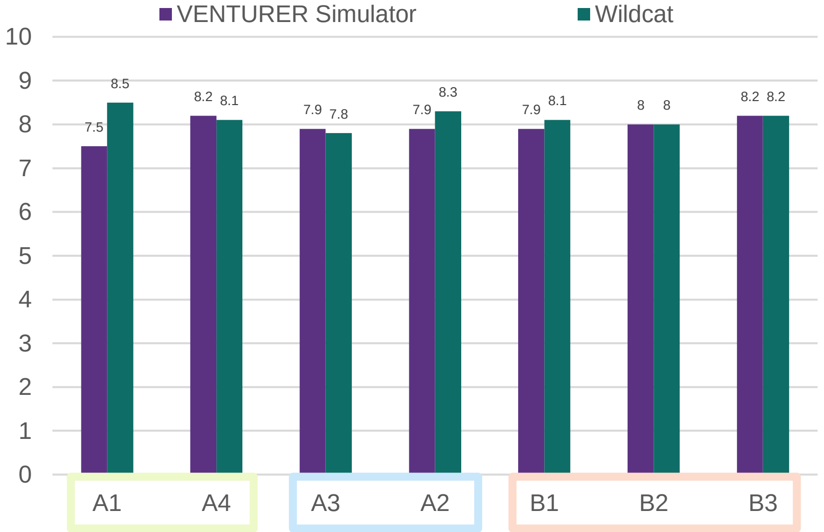
5.3. Cross platform: Analysis and Results

A total of 76 participants completed two full circuits in AV mode in the VENTURER Simulator and at least two full circuits in the Wildcat, enabling a robust comparison of trust ratings between platforms.

5.3.1. Comparison by scenario

The mean trust scores for each scenario and each platform are shown in Figure 9 for the subset of participants who completed two AV circuits in both parts of the trial (n = 76).

Figure 9: Cross platform comparison of average trust ratings – by scenario



Scenario	Manoeuvre	Interaction
A1	Approaching a zebra crossing	None
A4		Pedestrian
A3	Overtaking a parked car	None
A2		Cyclist
B1	Right turn	None
B2		Pedestrian
B3		Cyclist

Overall, the trust ratings were slightly higher for the Wildcat than the VENTURER Simulator, however there is not a consistent pattern of scores across scenarios. Excluding scenario A1, in which trust ratings were negatively impacted by the behaviour of the VENTURER Simulator, for the other six scenarios: two are rated higher for the Wildcat, two are rated higher for the simulator and two received the same trust ratings.

Statistical analyses showed that only two of the scenarios exhibited a statistically significant difference between the two platforms. These were crossing an empty zebra crossing (A1), which has been explained above, and overtaking a parked vehicle with an on-coming cyclist (A2).

Analysis of scenarios including crossing a zebra crossing revealed a statistically significant difference in the relationship between the scenarios with and without a pedestrian present for the different platforms (A4 vs A1). This is consistent with the relatively substantial decrease in trust scores for scenario A1 in the simulator due to the way the circuit begins in the simulator. This effect may be due to the fact it is the first feature encountered after the vehicle sets off somewhat aggressively.

For the two scenarios involving overtaking a parked car, there was a statistically significant difference in the relationship between the scenarios with and without an on-coming cyclist (A2 vs A3). These findings are likely to be because of the presence of the on-coming cyclist slowed the Wildcat sufficiently to reduce the jerkiness in the overtaking manoeuvre, thus increasing trust. As the movement was smoother in the VENTURER Simulator overall, there was no difference between trust scores with and without an on-coming cyclist.

Statistical analyses of events involving a right turn into a side road (B1, B2 and B3), identified a statistically significant difference by scenario, but not by platform. This demonstrates that a consistent relationship between the scenarios involving a right turn into a side road are observed in the Wildcat and the VENTURER Simulator.

5.3.2. Comparison by participant category

Statistical analyses to reveal the impact the different platforms had on trust ratings in the different participant categories found that none of the interactions were statistically significant. The participant role therefore, did not influence the impact the different platforms would have on their trust ratings in any of the scenarios.

5.4. Psychometric tests

All participants that took part in Trial 3 were required to complete several psychometric tests. These were administered using Qualtrics[5] as the survey platform. A total of 131 participants completed the Qualtrics questionnaire. The analysis of this data is outlined below.

5.4.1. Analysis and results

The psychometric measures were used to examine possible associations between different individual characteristics (e.g., age, driving experience and trust in automation) and trust in the autonomous platform performing different events.

When considering the relationship between participant age and trust ratings, statistical analyses indicated low strength associations, meaning that participant age is not a good predictor of their trust score. Correlation analyses for driving experience variables (number of years since passing a driving test and the average miles driven per year), showed the same pattern.

In contrast, there is a statistically significant correlation between self-reported trust in automation and the trust scores for all scenarios in the simulator and six of the seven scenarios in the Wildcat. The associations are all positive, indicating that higher levels of trust in automation generally are correlated with higher levels of trust during the trials.

The only scenario where the correlation coefficient was not statistically significant was the overtaking manoeuvre with an on-coming cyclist in the Wildcat (A2). It is possible that the trust scores for this scenario in the Wildcat were also driven by other factors such as individual gap acceptance preferences as some participants commented that the vehicle was giving too much space and time to the cyclist.

[5] www.qualtrics.com

6. Discussion

The trialling of CAV technology in the UK is being undertaken following a structured approach under the DfT's Code of Practice[2]. The safe and successful completion of the complex VENTURER trials is testament to the CAV research and development regime adopted in the UK and will ultimately contribute to the UK government's vision of bringing 'market-ready' CAVs to public roads in a safe manner that realises all possible safety benefits.

The findings in Trial 3 are consistent with those of Trial 2, which gives further confidence in the experimental design used in these VENTURER trials. The reliability and consistency of the trust measures obtained in Trial 3 are also supported by the correlation between participants' trust in automation and their trust ratings, with people who reported high levels of trust in general automation providing high trust scores.

Table 9 summarises the Trial 3 findings in relation to the human factors research questions outlined at the beginning of this report. The remainder of this section summarises further findings and considers their implications.

Table 9: Human factors research questions

Human Factors Research Questions	Narrative
How does the trust in the AV vary depending on the manoeuvre the AV is undertaking?	Trust ratings varied between participants and between manoeuvres. However, the differences were small and patterns were not consistent across the Wildcat and simulator experiments.
How does trust in the AV vary depending on whether the observer is a pedestrian, cyclist or driver-passenger in the AV?	There was no statistically significant difference in the trust ratings given by cyclist, driver or pedestrian participants. This was the case for all scenarios in both the Wildcat and the simulator.
How does trust in the AV vary depending on whether the AV is the real world Wildcat or the VENTURER Simulator?	For all scenarios combined, trust ratings in the Wildcat were slightly higher than the trust ratings in the VENTURER Simulator. However, the difference in trust scores between the platforms was only statistically significant for two of the seven scenarios.
In the simulator environment, how do the observing pedestrian and cyclist rate the scenarios when the vehicle is being driven by a human as compared with when the vehicle is in autonomous mode?	Trust ratings were slightly higher when the simulator was being manually driven as opposed to driving in autonomous mode. The differences in trust ratings were statistically significant for four out of the seven scenarios. It should be noted, however, that the trust scores are likely to have been influenced by the fact that the manually driven circuit was always performed last, when participants were most comfortable with the testing environment.
Does trust in the AV correlate with age, driving experience or trust in automation?	There were no statistically significant correlations between trust ratings in any of the scenarios and either age or driving experience indicating that groups such as older drivers, do not demonstrate a statistically lower or higher feeling of trust towards AV decisions. However, there were statistically significant correlations between the results from the 'trust in automation' survey and the trust scores for all scenarios in the simulator and six of the seven events in the Wildcat.

6.1. Rating of trust

The experiments found that mean ratings for all participants combined, under all scenarios in autonomous mode, were at least 7.3 out of ten, with ten representing the maximum possible trust rating. This is consistent with VENTURER's Trial 2 experiments in which virtually all scenarios received trust ratings of at least seven out of ten. Additionally, in the simulator, trust scores given to a human driver were not significantly higher than those given to the AV for all scenarios, and where a statistically significant difference was noticed, the difference was only small. This shows that exposure to AVs in safe environments yields generally high trust ratings.

There was also some evidence that the presence of another road user, in this case a cyclist, resulted in higher trust ratings. This suggests that not knowing how an AV would have behaved in a circumstance where there is a high degree of openness about what is about to happen, can be a barrier to achieving the highest trust ratings.

During Trial 3, participants' trust scores appeared to be more sensitive to the movement and noise of the AV rather than the perceived risk of the scenario. Therefore, during the early stages of CAV development, it is important to understand that aspects such as the smoothness of manoeuvres will have an impact on the public's perception of safety and trust. Undertaking the experiments with further refinements to automated control of the AV might result in even higher ratings.

6.2. Platform comparability

Although overall trust was higher for the Wildcat than for the VENTURER Simulator, only two of the scenarios exhibited a statistically significant difference between the two platforms. Where statistical differences were identified they are likely to be explained by specific aspects of the execution of the scenarios or experimental design.

These Trial 3 results therefore support the view that much useful research and development can be undertaken in simulation in addition to conducting more resource intensive real world experimentation.

6.3. Variability between participant category

There were no statistically significant differences in trust ratings between cyclist, driver or pedestrian participants reported for any of the scenarios in either the Wildcat or the simulator (a significance level of 0.05). This could indicate that there is no particular need to differentiate messaging around CAVs for different audiences.

According to these results, no matter if members of the public can drive cars, or are users of a range of other transport modes in their everyday life, cars-in-motion are highly familiar objects in the environment, and provided they operate in ways similar to human-driven vehicles, they are accepted as 'normal' cars.

6.4. Application of findings

A possible implication of the high levels of trust given to AVs operating in test environments is that members of the public might be too trusting of CAVs in their early stages of development. This becomes particularly problematic when considered against the backdrop of on-going debates around liability at different SAE levels of autonomy. The distinctions between the SAE levels can be small and the potential for confusion in terms of a vehicle's capabilities will be more acute during the early stages of CAV development and deployment.

AXA and Burges Salmon's third VENTURER report, acknowledges that more work needs to be done to explore ways in which the risk of consumer confusion could be mitigated or prevented by design and to ensure that drivers are educated as to the capabilities and terms of use of the specific CAVs they may drive. Beyond this, as indicated by Trial 3 findings, CAVs may also need to be clearly identifiable to other road users, such as pedestrians and cyclists, particularly in their early stages of deployment.

6.5. Further research

Ahead of the deployment of fully automated CAVs onto UK roads, it is crucial that they continue to be thoroughly tested in safe and controlled environments. Future research should consider how the use of incrementally complex trials akin to real world situations can be used to increase understanding of road users' trust in CAVs. More complex trials should involve combining scenarios, for example introducing multiple pedestrians, cyclists and motorised vehicles, and allow for scenarios involving closer proximity between CAVs and actors. It is also crucial that trust ratings generated when participants are interacting with the CAVs themselves are gathered and analysed, including in scenarios where participants are blind to whether a vehicle they are being asked to rate is an AV or not.

Furthermore, it is possible that the appearance of a very 'safe' environment is having an impact on trust scores and therefore future research should examine how trials can appear to be less closely-managed, whilst retaining the high levels of safety. This could involve marshals wearing more discrete clothing or being more physically removed towards the boundary of the testing area.

6.6. VENTURER Alliance

The full learnings and recommendations from the VENTURER project, including Trial 3, will be summarised in the project's Final Report, which will be released later in 2018. The VENTURER partners anticipate continuing to work with government and industry to support the future deployment of CAVs onto UK roads through the mechanism of the VENTURER Alliance.

Launched in July 2018, the VENTURER Alliance will leverage the expertise and capabilities acquired during the project's investigations to offer customers a diverse portfolio of services relating to the development and testing of driverless cars and their potential impact on UK society and infrastructure.

Appendix A. Qualtrics

Psychometric tests issued to participants using Qualtrics:

- A driving experience questionnaire: e.g., time since holding a full driving licence, miles driven annually, miles driven monthly, and driving frequency per week.
- Faith and Trust Stance in General Technology: 7 items (e.g., "I believe that most technologies are effective at what they are designed to do") measuring individuals' trust in technology.
- Trust in Automation: Trust in Automation Checklist *(Jian, Bisantz, & Drury, 2000).* 12 item scale (e.g., the system is reliable) measuring trust in the autonomous platform they have just experienced.
- Impulsivity: Barratt Impulsiveness Scale *(Patton, Stanford, & Barratt, 1995).* 30 items (e.g., "I do things without thinking", "I am more interested in the present than the future") measuring attention, motor, self-control, cognitive complexity, perseverance, and cognitive instability impulsiveness as well as attentional, motor, and non-planning impulsiveness;
- Self-control: Brief Self-Control Scale *(Tangney, Baumeister, & Boone, 2004).* Ten items (e.g., "I get distracted easily", "…I can't stop myself from doing something, even if…it is wrong");
- Risk taking: RT18 *(de Haan et al. 2011)* – a subjective risk taking scale including both risk taking assessment and behaviour. Eighteen items (e.g. "I often try new things just for fun or thrills, even if most people think it is a waste of time", "Would you enjoy parachute jumping?");
- Distractibility: Cognitive Failures Questionnaire *(Broadbent et al. 1982).* 25-items (e.g., "do you fail to notice signposts on the road?");
- Personality: Big-Five Personality Questionnaire *(Costa & McCrae 1992)*: 60 items to measure personality dimensions of extraversion, neuroticism, agreeableness, conscientiousness and openness to experience (e.g., "I have a clear set of goals and work toward them in an orderly fashion", "I would rather go my own way than be a leader of others");
- Sleep: Pittsburgh Sleep Quality Index *(Buysse et al. 1989)*: 19 items measuring e.g. subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency and sleep disturbances, etc. (e.g. "During the past month, how many hours of actual sleep did you get at night?");
- Mood: Used VAS. 18 items (e.g. dreamy, attentive, bored, interested) *(Bond & Lader 1974)*;
- Cognitive workload: NASA Task Load Index *(Hart & Staveland 1988).* Assesses workload on 7-point scales (e.g., "How mentally demanding was the task?" "How hurried or rushed was the pace of the task?").

The above scales are all valid instruments that are widely used in psychological and related human factors research. The references included indicate the source and authenticity of the scales.

Participants

Table 10: Psychometric test participants

		Gender		Time since passing driving test			
Participant role	**Number of participants**	**Female (%)**	**Male (%)**	**Not passed test**	**< 20 years**	**20-40 years**	**> 40 years**
Cyclist	45	33%	67%	1	12	17	15
Driver	49	31%	69%	0	13	20	16
Pedestrian	37	43%	57%	2	12	9	14
Total	**131**	**35%**	**65%**	**3**	**37**	**46**	**45**

Analysis

Table 11: Correlation of participant characteristics and VENTURER Simulator trust ratings

		A1	A2	A3	A4	B1	B2	B3
Age	**Pearson correlation coefficient**	-0.07	0.01	-0.05	-0.09	-0.02	-0.03	-0.03

	p-value	0.490	0.888	0.636	0.376	0.829	0.737	0.720
Years since driving test	Pearson correlation coefficient	-0.09	0.00	-0.06	-0.11	-0.06	-0.04	-0.07
	p-value	0.379	0.993	0.524	0.255	0.507	0.676	0.475
Average miles driven per year	Pearson correlation coefficient	-0.01	-0.02	-0.03	0.00	-0.01	0.02	0.00
	p-value	0.933	0.840	0.789	0.974	0.932	0.865	0.994
Trust in automation	Pearson correlation coefficient	0.28	0.41	0.36	0.36	0.44	0.41	0.34
	p-value	0.009**	<0.001**	0.001**	0.001**	<0.001**	<0.001**	0.002**

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

Table 12: Correlation analysis of participant characteristics and Wildcat trust ratings

		A1	A2	A3	A4	B1	B2	B3
Age	Pearson correlation coefficient	-0.03	-0.12	0.09	-0.03	0.02	-0.02	0.00
	p-value	0.751	0.268	0.397	0.755	0.831	0.861	0.969
Years since driving test	Pearson correlation coefficient	-0.07	-0.14	0.05	-0.06	-0.04	0.00	-0.04
	p-value	0.518	0.184	0.654	0.542	0.695	0.987	0.724
Average miles driven per year	Pearson correlation coefficient	-0.04	-0.07	-0.02	0.01	0.00	0.04	0.07
	p-value	0.739	0.491	0.844	0.911	0.971	0.715	0.492
Trust in automation	Pearson correlation coefficient	0.30	0.16	0.29	0.23	0.28	0.27	0.26
	p-value	0.005**	0.135	0.008**	0.033*	0.010*	0.012*	0.017*

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

Appendix B. Wildcat experiments

Participants

The participants involved in the Wildcat experiments are outlined in Table 13.

Table 13: Wildcat participants

Participant role	Total number of participants	Participants with two or more complete circuits	Gender		Average age
			Female	Male	
Cyclist	35	33	10	23	53
Driver	37	35	10	25	51
Pedestrian	29	27	12	15	48
Total	**101**	**95**	**32**	**63**	**51**

Analysis and results

Table 14: Average trust ratings in the Wildcat by scenario for all participants combined (n=95)

Manoeuvre	Interaction	Scenario	Mean score	Standard deviation
Approaching a zebra crossing	None	A1	8.4	1.35
	Pedestrian	A4	8.1	1.52
Overtaking a parked car	None	A3	7.7	1.52
	Cyclist	A2	8.2	1.40
Right turn	None	B1	8.1	1.31
	Pedestrian	B2	8.0	1.42
	Cyclist	B3	8.2	1.32

Table 15: Results of paired samples t-tests on Wildcat trust scores for all participants combined (n=95)

Manoeuvre	Scenario without interaction	Scenario with interaction	Interaction type	Sample mean of differences	t-value	p-value
Approaching a zebra crossing	A1	A4	Pedestrian	-0.32	-2.51	0.014*
Overtaking a parked car	A3	A2	Cyclist	0.46	4.52	< 0.001**

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

Table 16: Results of repeated measures ANOVA on Wildcat trust scores for all participants combined (n=95)

Manoeuvre	Scenarios	F-statistic	Degrees of freedom (conditions)	Degrees of freedom (error)	p-value
Turning right into a side road	B1 (empty side road); B2 (pedestrian crossing the junction); B3 (on-coming cyclist)	2.80	1.8	171.5	0.069

*: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)

Table 17: Average and standard deviation of trust ratings in the Wildcat by participant role (n_{cyc}=33, n_{driv}=35, n_{ped}=27)

			Mean trust rating			Standard deviation of trust ratings		
Manoeuvre	**Interaction**	**Scenario**	**Cyclists**	**Drivers**	**Peds**	**Cyclists**	**Drivers**	**Peds**
Approaching a zebra crossing	None	A1	8.2	8.6	8.5	1.12	1.33	1.62
	Pedestrian	A4	8.2	7.9	8.2	1.35	1.76	1.40
Overtaking a parked car	None	A3	7.5	8.0	7.6	1.52	1.61	1.38
	Cyclist	A2	8.1	8.3	8.2	1.47	1.42	1.34
Right turn	None	B1	8.0	8.1	8.2	1.18	1.38	1.41
	Pedestrian	B2	7.7	8.0	8.3	1.42	1.40	1.44
	Cyclist	B3	8.2	8.3	8.2	1.17	1.39	1.42

Table 18: Results of ANOVA on Wildcat trust scores by participant role and scenario type (n=95)

Manoeuvre	Interaction	Scenario	F-statistic	Degrees of freedom (conditions)	Degrees of freedom (error)	p-value
Crossing a pedestrian crossing	None	A1	0.56	2	92	0.57
	Pedestrian	A4	0.31	2	92	0.74
Overtaking a parked car	None	A3	1.15	2	92	0.32
	Cyclist	A2	0.23	2	92	0.80
Right turn	None	B1	0.13	2	92	0.88
	Pedestrian	B2	1.14	2	92	0.32
	Cyclist	B3	0.09	2	92	0.92

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

Table 19: Results of two factor mixed-design ANOVA by participant role and scenario in the Wildcat (n=93)

Scenario	Effect	F-statistic	df1	df2	p-value	Partial Eta Squared (η_p^2)
Approaching a zebra crossing A1: without a pedestrian present A4: with a pedestrian present	Event (A1 or A4)	5.89	1	92	0.017*	0.060
	Interaction term	1.76	2		0.177	0.037
Overtaking a parked car A3: with no on-coming traffic A2: with an on-coming cyclist	Event (A2 or A3)	21.27	1	92	<0.001**	0.188
	Interaction term	1.094	2		0.339	0.023
Turning right into a side road B1: no interaction, B2: with a pedestrian crossing the junction, B3: with an on-coming cyclist	Event[6] (B1, B2 or B3)	2.31	1.8	166.5	0.108	0.024
	Interaction term	1.71	3.6		0.157	0.036

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

[6] Mauchly test for sphericity indicated that corrections for departure from sphericity were required.

Appendix C. Simulator experiments

Participants

Table 20: VENTURER Simulator participants

Participant role	Total number of participants	Participants with two complete AV circuits	Gender		Average age
			Female	Male	
Cyclist	45	37	10	27	52
Driver	49	41	14	27	49
Pedestrian	39	32	15	17	48
Total	**133**	**110**	**39**	**71**	**50**

Analysis and results

Table 21: Average trust ratings in the VENTURER Simulator by scenario for all participants combined (n=110)

Manoeuvre	Interaction	Scenario	Mean score	Standard deviation
Crossing a zebra crossing	None	A1	7.3	1.84
	Pedestrian	A4	8.1	1.43
Overtaking a parked car	None	A3	7.9	1.51
	Cyclist	A2	7.8	1.54
Right turn	None	B1	7.8	1.64
	Pedestrian	B2	7.9	1.47
	Cyclist	B3	8.1	1.38

Table 22: Results of paired t-tests on VENTURER Simulator trust scores for all participants combined (n=110)

Manoeuvre	Scenario without interaction	Scenario with interaction	Interaction type	Sample mean of differences	t-value	p-value
Crossing a pedestrian crossing	A1	A4	Pedestrian	0.77	6.04	< 0.001**
Overtaking a parked car	A3	A2	Cyclist	-0.03	-0.29	0.773

*: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)

Table 23: Results of repeated ANOVA on VENTURER Simulator trust scores for all participants combined (n=110)

Manoeuvre	Scenarios	F-statistic	Degrees of freedom (conditions)	Degrees of freedom (error)	p-value
Turning right into a side road[7]	B1 (empty side road); B2 (pedestrian crossing the junction); B3 (on-coming cyclist)	10.41	2	218	< 0.001**

*: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)

Table 24: Average and standard deviation of trust ratings in the VENTURER Simulator by participant role (n_{cyc}=37, n_{driv}=41, n_{ped}=32)

			Mean trust rating			Standard deviation of trust ratings		
Manoeuvre	Interaction	Scenario	Cyclists	Drivers	Peds	Cyclists	Drivers	Peds
Crossing a zebra crossing	None	A1	6.9	7.4	7.7	2.06	1.66	1.75
	Pedestrian	A4	7.8	8.2	8.3	1.47	1.52	1.27
Overtaking a parked car	None	A3	7.6	8.0	8.0	1.35	1.75	1.34
	Cyclist	A2	7.7	7.7	8.2	1.36	1.82	1.34
Right turn	None	B1	7.5	7.8	8.0	1.47	1.81	1.60
	Pedestrian	B2	7.5	8.0	8.1	1.52	1.48	1.37
	Cyclist	B3	7.8	8.2	8.3	1.34	1.50	1.22

Table 25: Results of ANOVA on participant role for VENTURER Simulator

Manoeuvre	Interaction	Scenario	F-statistic	Degrees of freedom (conditions)	Degrees of freedom (error)	p-value
Approaching a zebra crossing	None	A1	1.79	2	107	0.17
	Pedestrian	A4	0.85	2	107	0.43
Overtaking a parked car	None	A3	0.89	2	107	0.41
	Cyclist	A2	1.10	2	107	0.34
Right turn	None	B1	0.95	2	107	0.39
	Pedestrian	B2	1.81	2	107	0.17
	Cyclist	B3	1.36	2	107	0.26

*: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)

[7] A Mauchly test for sphericity indicated that no corrections for departure from sphericity were required

Table 26: Results of two factor mixed-design ANOVA by participant role and scenario in AV mode in the VENTURER Simulator (n=108)

Scenario	Effect	F-statistic	df1	df2	p-value	Partial Eta Squared (η_p^2)
Approaching a zebra crossing A1: without a pedestrian present A4: with a pedestrian present	Event (A1 or A4)	34.73	1	107	< 0.001**	0.245
	Interaction term	0.878	2		0.419	0.016
Overtaking a parked car A3: with no on-coming traffic A2: with an on-coming cyclist	Event (A2 or A3)	0.016	1	107	0.901	0.000
	Interaction term	1.982	2		0.143	0.036
Turning right into a side road B1: no interaction, B2: with a pedestrian crossing the junction, B3: with an on-coming cyclist	Event (B1, B2 or B3)	9.798	2	214	< 0.001**	0.084
	Interaction term	0.423	4		0.792	0.008

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

Table 27: Average and standard deviation of trust ratings in the simulator by scenario in manual and autonomous mode (n=66)

			Manually driven		AV mode	
Manoeuvre	Interaction	Event	Mean score	Standard deviation	Mean score	Standard deviation
Approaching a zebra crossing	None	A1	8.7	1.26	7.4	1.83
	Pedestrian	A4	8.4	1.70	8.1	1.27
Overtaking a parked car	None	A3	8.5	1.45	7.8	1.30
	Cyclist	A2	8.4	1.51	8.0	1.30
Right turn	None	B1	8.0	1.48	7.9	1.38
	Pedestrian	B2	8.3	1.48	7.9	1.40
	Cyclist	B3	8.5	1.21	8.2	1.24

Table 28: Results of t-tests comparing trust in manual versus autonomous driving

Manoeuvre	Interaction	Event	Sample mean of differences	t statistic	p-value
Approaching a zebra crossing	None	A1	1.2	5.53	< 0.001**
	Pedestrian	A4	0.3	1.13	0.262
Overtaking a parked car	None	A3	0.6	2.83	0.006**
	Cyclist	A2	0.4	2.16	0.034*
Right turn	None	B1	0.1	0.33	0.741
	Pedestrian	B2	0.4	1.98	0.052
	Cyclist	B3	0.3	2.16	0.035*

*: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)

Appendix D. Cross platform comparison

Participants

Table 29: Wildcat and VENTURER Simulator participants

Participant role	Number of participants	Female	Male
Cyclist	26	6	20
Driver	29	10	19
Pedestrian	21	11	10
Total	76	**27**	**49**

Analysis and results

Table 30: Average trust ratings for the Wildcat and VENTURER Simulator for participants undertaking both sets of experiments (n=76)

			VENTURER Simulator		Wildcat AV	
Manoeuvre	**Interaction**	**Scenario**	**Mean score**	**Standard deviation**	**Mean score**	**Standard deviation**
Approaching a zebra crossing	None	A1	7.5	1.60	8.5	1.29
	Pedestrian	A4	8.2	1.36	8.1	1.52
Overtaking a parked car	None	A3	7.9	1.41	7.8	1.51
	Cyclist	A2	7.9	1.36	8.3	1.35
Right turn	None	B1	7.9	1.49	8.1	1.27
	Pedestrian	B2	8.0	1.33	8.0	1.42
	Cyclist	B3	8.2	1.32	8.2	1.33

Table 31: Matched pairs analysis of trust scores in the Wildcat and VENTURER Simulator (n=76)

Manoeuvre	Interaction	Scenario	Sample mean of differences	t statistic	p-value
Approaching a zebra crossing	None	A1	-0.9	-4.68	< 0.001**
	Pedestrian	A4	0.0	0.22	0.828
Overtaking a parked car	None	A3	0.2	0.92	0.359
	Cyclist	A2	-0.4	-2.17	0.033*
Right turn	None	B1	-0.2	-1.26	0.212
	Pedestrian	B2	0.0	-0.02	0.983
	Cyclist	B3	-0.1	-0.36	0.723

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

Table 32: Results of two-way repeated measures ANOVA by platform and scenario (n=76)

Event	Effect	F-statistic	df1	df2	p-value	Partial Eta Squared (η_p^2)
Approaching a zebra crossing A1: without a pedestrian present A4: with a pedestrian present	Platform (Wildcat or simulator)	7.29	1	75	0.009**	0.089
	Event (A1 or A4)	1.47	1	75	0.230	0.019
	Interaction term	33.87	1	75	< 0.001**	0.311
Overtaking a parked car A3: with no on-coming traffic A2: with an on-coming cyclist	Platform (Wildcat or simulator)	0.46	1	75	0.502	0.006
	Event (A2 or A3)	7.24	1	75	0.009**	0.088
	Interaction term	15.55	1	75	< 0.001**	0.172
Turning right into a side road B1: no interaction, B2: with a pedestrian crossing the junction, B3: with an on-coming cyclist	Platform (Wildcat or simulator)	0.36	1	75	0.550	0.005
	Event[8] (B1, B2 or B3)	5.47	1.75	130.95	0.007**	0.068
	Interaction term	1.56	2	150	0.213	0.020

**: significant using 0.05 significance level (2-tailed), **: significant using 0.01 significance level (2-tailed)*

Table 33: Results of two factor mixed-design ANOVA by platform and participant role for each scenario (n=76)

			Platform			Interaction term		
Manoeuvre	Interaction	Event	F-statistic	p-value	Partial Eta Squared (η_p^2)	F-statistic	p-value	Partial Eta Squared (η_p^2)
Approaching a zebra crossing	None	A1	20.54	< 0.001**	0.220	0.64	0.529	0.017
	Pedestrian	A4	0.01	0.925	0.000	0.78	0.461	0.021
Overtaking a parked car	None	A3	0.92	0.340	0.012	0.17	0.840	0.005
	Cyclist	A2	3.97	0.050	0.052	0.60	0.551	0.016
Right turn	None	B1	1.49	0.226	0.020	0.78	0.454	0.021
	Pedestrian	B2	0.01	0.905	0.000	0.29	0.750	0.008
	Cyclist	B3	0.13	0.722	0.002	0.52	0.595	0.014

[8] Mauchly test for sphericity indicated that corrections for departure from sphericity were required.

Carolyn Mitchell

VENTURER Project Manager

Carolyn.Mitchell@atkinsglobal.com