VENTURER Trial 2: Interactions Between Autonomous Vehicles and Other Vehicles on Links and at Junctions

Trial 2 Findings

November 2017

Delivered by:

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 48 pages including the covers.

Document history

Work Package: Seven			Document ref: WP7 D5			
Document Lead Organisation: Atkins			**Document Lead Name: Imogen Weight**			
Revision	**Purpose description**	**Originated**	**Checked**	**Reviewed**	**Authorised**	**Date**
Rev 1.0	Draft version for consortium review	IW	BS	RT/JB	CM	13/11/2017
Rev 1.1	Final version for release	IW	BS	RT	CM	24/11/2017

Acknowledgements

Partner	Authors and contributors
Atkins Ltd	Imogen Weight, Bella Slawin, Peter Blackley, Rebecca Tommey and Carolyn Mitchell
BAE Systems (Operations) Ltd	Mark Goodall and Gary Cross
Bristol Robotics Laboratory (BRL)	Tony Pipe and Martin Pearson
University of Bristol (UoB)	Arthur Richards and Tom Kent
University of the West of England (UWE) - Bristol	Jonathan Flower, Craig Williams, Chris Alford, Phil Morgan and John Parkin

Jonathan Flower and John Parkin from 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. In addition, trial design, analysis, commentary and discussion was provided by Craig Williams (Psychology - Centre for Health and Clinical Research, UWE), Chris Alford (Psychology – Psychological Sciences Research Group) and Phillip L. Morgan (School of Psychology, Cardiff University.)

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

ANOVA – ANalysis Of VAriance tests whether or not the means of several groups are equal.

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.

Bonferroni Post Hoc Analysis – a method to correct the problem of multiple comparisons with multiple null hypotheses.

Circuit – the complete set of four loops undertaken by the AV in the Trial 2 participant experiments.

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

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

Critical Gap – the time gap to an approaching vehicle at a junction that half of drivers would accept and half would reject.

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

Event – the scenario as experienced by the participant.

Experiment – the complete set of events, and the responses of participants to those events undertaken as part of Trial 2.

Headway – the time gap adopted by a vehicle following another vehicle.

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

Link – a road connecting two junctions.

Neutral Autonomous Mode (N) – the mode adopted by the Decision Making System which meant that the Autonomous Vehicle rejected the critical gap, that is to say, waited until the on-coming vehicle had passed before making a manoeuvre at a junction.

Neutral Prime Autonomous Mode (N') – the mode adopted by the Decision Making System which meant that the Autonomous Vehicle accepted the critical gap.

Pearson's Product Moment Coefficient – a measure of the strength of a linear association between two variables.

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. See also 'event': the word scenario is used in connection with the AV manoeuvre, and the word event is used in connection with the participants experience of the scenario.

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

Situation – a sub-set of the scenario and either being a scenario on a link or at a junction.

T-test – a statistical test used to determine whether two sets of data are significantly different from each other.

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

[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.

Contents

Headline Finding

The results outlined that, in order to foster public trust, it may be necessary for Autonomous Vehicles (AV) to operate more cautiously than the average human driver. This more cautious approach may have safety benefits in conditions with mixed autonomous vehicles and conventional vehicles. Under some traffic flow conditions, there may also be some net congestion reduction benefits.

Executive Summary

VENTURER Trial 2 explored interactions between an Autonomous Vehicle (AV) and other vehicles. The focus of Trial 2 was to review the vehicle, decision making systems, sensor performance and human reactions during two specific typical highway situations: navigating a T-junction; and navigating a parked vehicle on the carriageway.

These highways situations were selected by the VENTURER consortium at pre-trial meetings and were identified as being two of the most common situations encountered on a typical UK road. Demonstrations of safe interactions in these situations will be a crucial aspect in ensuring that AVs will be safe for deployment onto UK roads.

Within these two situations, there were ten different scenarios to ensure the experiments captured a wide range of potential real-world situations. The range of scenarios provides the VENTURER consortium with increased knowledge and data with regards to how an AV deals with these situations as well as providing insights into the users' experience, and the users trust in the AV.

Trial 2 used the BAE Systems Wildcat vehicle, and the VENTURER Simulator, both of which were controlled by a Decision Making System (DMS) developed as part of the VENTURER project. Two key decisions that need to be made are based on the time headway to a vehicle ahead and the time gap to an on-coming vehicle that is either accepted or rejected when making a manoeuvre at a junction. The critical gap is the time gap to an approaching vehicle at a junction that half of drivers would accept and half would reject. Trial 2 was undertaken in two phases:

- Phase 1 included participant experiments in both the Wildcat and the VENTURER Simulator when the DMS always rejected the critical gap (termed Neutral, N); and
- Phase 2 undertook further participant experiments only in the VENTURER Simulator to compare trust ratings in situations where the AV not only rejected the critical gap (N), but also accepted the critical gap (termed Neutral prime, N').

Alongside the trial experiments, ongoing technology development and validation occurred to prepare for future trials.

Phase 1

Trial 2, Phase 1 was successful with a total of 45 participants completing the Wildcat trial and 38 completing at least two laps in the VENTURER Simulator enabling reliable average data to be calculated. Overall, the assessment of trust was successful, with mean ratings in virtually all conditions exceeding seven out of ten, with ten representing the maximum possible trust rating. For both the Wildcat and VENTURER Simulator platforms, participants indicated that under the experimental conditions in which they were assessed, they trusted both platforms but overall, trust was higher in the VENTURER Simulator than the Wildcat.

The reliability and consistency of the trust measures in general were validated by the consistent pattern of significant associations, including some strong associations, between the participants' trust in automation and general technology. Phase 1 participant experiments found that the higher someone's general trust in automation and technology, the higher their trust ratings for the autonomous driving events experienced in both the Wildcat and VENTURER Simulator.

Overall, the associations were stronger for the Wildcat and this may reflect the perceptual evaluation of trust when being driven autonomously in a real vehicle when there is the potential for real detriment or harm if things went wrong. In the simulator, there is less potential for real detriment or harm and therefore less 'trust' required. Additionally, both platforms demonstrated a level of apparent 'jerkiness' and hesitancy which may have affected the trust ratings.

Phase 2

As with Phase 1, virtually all ratings of trust when the gap was rejected (N) were rated above a seven out of ten, further validating the reliability and consistency of the trust measures used in this study.

Trust ratings in Phase 2 experiments were consistently lower for the scenarios where the gap was accepted (N') by the autonomous mode as compared to being rejected (N). This suggests, for public acceptability reasons, that it may be necessary for AVs to operate more cautiously than the average driver. This more cautious approach would also likely have safety benefits in conditions with mixed AV and conventional vehicles. The result means that there might be congestion benefits resulting from the greater level of consistency in gap acceptance behaviour in some traffic conditions. These findings suggest a need for appropriately cautious decision making algorithms in order to foster public trust.

While the findings suggest that gap acceptance at the 50% threshold of acceptance by drivers reduces the trust the participants had in the VENTURER Simulator, it should be noted that some of the gap accepting manoeuvres in the simulator put the simulated AV in very close proximity to the oncoming vehicles. This was due to the AV driving quite slowly and pausing on turning. To confirm the findings, additional trials with variable gaps, and again with both gap acceptance and gap rejection, would be required.

Technology Development

The Wildcat operated successfully to schedule during participant experiments using a modified version of the DMS and sensor system. Different DMS behaviour was successfully deployed in Phase 2 when undertaking the N and N' scenarios. The trialling of the DMS during participant experiments was important both for Trial 2, and for preparation for future trials because it allowed for the testing of new algorithms on real data.

1. VENTURER Trials

1.1. Introduction

The VENTURER project is systematically assessing the responses of passengers and other road users, including pedestrians and cyclists, to Autonomous Vehicles (AVs), in a series of increasingly complex trials and demonstrations in urban settings. The trials and the data collected will provide a greater understanding of how AV technology performs, how people interact with AV technology, and will help inform the development of potential insurance models and the legal frameworks for AVs. Developing this understanding provides the first key step towards facilitating the deployment of AVs 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 AVs 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 (CAVs).
BAE Systems (Operations) Ltd	Providing the Wildcat vehicle and testing its autonomous ability.
Bristol Robotics Laboratory (BRL)	Developing and testing the 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	Providing access to public roads.
South Gloucestershire Council	Providing access to 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 AVs, analysing driver performance during the handover of an AV, and analysing interactions between AVs 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 insurance and legal aspects of autonomous technology. For each theme, there are specific objectives that have been agreed by the consortium for the trials:

- Systematically assess the responses of passengers, AV users, other road users and pedestrians to AVs through trials with the Wildcat and the VENTURER Simulator;
- 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; and
- Investigate the use of innovative sensor technology by collecting data on buses.

These objectives are being achieved by conducting three trials utilising the platforms available to VENTURER partners. This report summarises Trial 2.

1.3. Platforms

VENTURER is using three different platforms in the three trials, as follows:

- Road tests of autonomous technology using the Wildcat vehicle;
- Human factor experiments using the VENTURER Simulator; and
- The UWE STISIM Simulator.

Table 2 summarises 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, in Trial 3, the real-world performance of the situational awareness and decision making technology will be investigated using data collected from an operational bus. VENTURER also undertook a public acceptance demonstration using a pod between 03 and 06 August 2017.

1.4. Exploitation

VENTURER has made the West of England region 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. So far 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 technology and participant trials in simulation and on real roads;
- Develop the hardware and software for an AV simulator;
- Conduct research into public expectations, acceptance and response to AVs; and
- Analyse the legal and insurance implications of the deployment of AVs onto UK roads.

1.5. Purpose of this report

This report presents the method, findings and discussion of the Trial 2 experiments conducted using the Wildcat and the VENTURER Simulator. The findings and discussions focus specifically on the performance of the technology when interacting with other vehicles at junctions and on links and the reaction of a human participant to the decisions made by the vehicle during these manoeuvres.

1.6. Trial 2

The primary aim of Trial 2 was to demonstrate autonomous driving including manoeuvres along a link and at a junction and interactions with other vehicles. Trial 2 reviewed the vehicle, DMS, sensor performance and human reactions. The remainder of this document details the trial approach, discusses the results and summarises the findings. The structure of the report is outlined below:

- Section 2 – Trial 2 Overview
- Section 3 – Approach
- Section 4 – Trial 2 Technology
- Section 5 – Human Factors Experiments: Phase 1
- Section 6 – Human Factors Experiments: Phase 2
- Section 7 – Discussion

2. Trial 2 Overview

VENTURER Trial 2 – Autonomous driving including manoeuvres along links, at a junction and interactions with other vehicles.

Reviewing the performance of the technology and analysing the participant's response to a range of different scenarios and autonomous modes.

Trial 2 was undertaken using the **VENTURER Simulator** and the **Wildcat**.

2.1. Context

Trial 2 demonstrated potential interactions between a AV and other vehicles. The focus of Trial 2 was to review the vehicle, decision making systems, sensor performance and human reactions during the two specific typical highway situations, detailed as follows:

A. **Navigating a parked vehicle on the carriageway** – this is a typical issue in many UK cities that can result in significant tail backs and risky situations if inappropriate judgements are made. This is a difficult situation for an AV due to the requirement for the sensors to accurately detect the object, detect that it is stationary, determine if there are any oncoming vehicles, and then make the manoeuvre to overtake the stationary vehicle; and
B. **Navigating a T-junction** – this is a very common situation on UK roads. The AV must safely and efficiently handle the numerous scenarios that occur at T-Junctions. In particular, the vehicle's ability to make a judgement as to whether it is safe to make a turn or not.

These two situations were selected as it was agreed that overtaking a parked car on the carriageway is a typical circumstance on urban roads. An AV needs to be able to make appropriate decisions to ensure the passenger is safe. It is a complex situation for an autonomous vehicle because not only is the detection requirement quite complex, but also the decision is a complex one based on a determination of the on-coming vehicle's distance and speed. Manoeuvring at a T-junction is an extremely common urban road situation. The AV needs to safely and efficiently handle each one of the different possible movements, both with and without on-coming traffic. In particular, the vehicle's ability to decide whether it is safe to make a turn or not is critical.

2.2. Aim

Trial 2 aimed to validate decision-making strategies and sensor technology based on motion planning in dynamic multi-vehicle scenarios and to understand user responses to the emerging technology. To achieve this, Trial 2 included technology development and participant experiments to establish human acceptance in a range of AV scenarios.

2.3. Research Questions

The research questions have been split into Technology specific and Human Factors specific questions. These are summarised in Table 3 and are discussed in more detail in the relevant sections of the report.

Table 3: Overall Research Questions

Aspect	Component	Research Question
Technology	Decision Making System (DMS)	As the level of consideration is changed, did the vehicle behave as expected?
	Communications System, V2V (Vehicle to Vehicle)	Does the communications system accurately transmit data in a real-world CAV situation?
	Sensors	Was it possible to integrate sensors successfully within the system as a whole?
	Wildcat	Does the Wildcat perform to schedule, temporally and spatially?
Human Factors	Wildcat and VENTURER Simulator	Do respondents' trust scores vary dependent on the type of AV scenario they experience?
		How do these trust scores correlate with relevant validated psychometric test scores?
		How do respondents rate their general comfort, including in relation to nausea?
	Cross-platform comparison	Are trust scores significantly different depending on the platform?

It was originally expected that the Human Factors research questions would relate to the participants' feelings of physical and emotional comfort in the AV experiments. However, it was deemed unlikely that we would find notable variability in comfort between the situations being tested due to low speed movements in a closed environment of UWE campus roads. As such, the Human Factors research questions in Table 3 were developed to allow for the examination of the following:

- Different scenarios (link, junction turns and passing manoeuvres, with and without an oncoming vehicle);
- Different platforms (the Wildcat and the VENTURER Simulator); and
- Different approach of the AV DMS based on either accepting or rejecting the time gap to an oncoming vehicle, and for safety reasons was only undertaken in the VENTURER Simulator.

3. Approach

Trial 2 was conducted using the Wildcat and the VENTURER Simulator. Both platforms were used to provide a comprehensive and comparative assessment of the scenarios. Trial experiments were undertaken in two phases:

- Phase 1 included Human Factor experiments in the Wildcat and VENTURER Simulator; and
- Phase 2 undertook further participant experiments only in the VENTURER Simulator to test responses to acceptance as well as rejection of the critical gap.

Ongoing technology development and validation also occurred to explore the Trial 2 research questions and prepare for future trials.

3.1. Programme

Trial 2 was undertaken during June, July and August 2017, with testing and validation of the sensors, system and vehicle occurring prior to the main trials involving participants (Phase 1). Following the main trials, further technology testing occurred to implement the use of different autonomous modes (Phase 2).

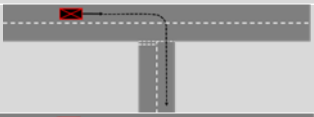
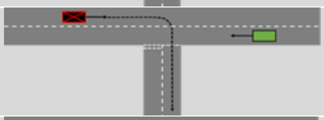
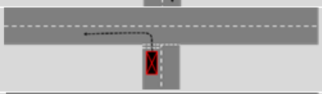
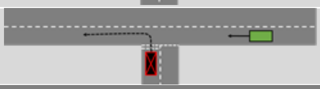
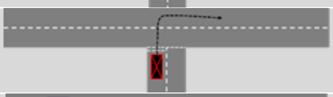
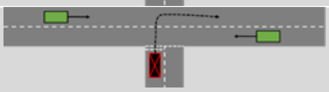
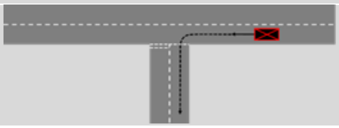
Table 4: Trial 2 Programme

Item	Dates	Wildcat	VENTURER Simulator
Pre-trial testing	02 June – 08 June 2017	✓	✓
Phase 1 Experiments: Participant Trials	09 June – 07 July 2017	✓	✓
Phase 2 Experiments: Participant Trials	07 July – 21 August 2017	-	✓
Ongoing Technology Development	02 June – 21 August 2017	✓	✓

3.2. Scenarios

Several scenarios on a link at a junction were developed as shown in Table 5.

Table 5: Scenario Summary

Scenario Type	Scenario Number	Scenario Description	Scenario Depiction
Link (A)	A1	Moving along an empty road at or below the speed limit.	
	A3	Overtaking a parked car while leaving a safe passing distance.	
	A4	Overtaking a parked car leaving a safe passing distance and waiting if necessary to leave a safe gap to an oncoming car.	
Junction (B)	B1	Turning right into the side road with no oncoming vehicle.	
	B2	Turning right into the side road with an oncoming vehicle.	
	B3	Turning left out of a side road with no oncoming vehicle.	
	B4	Turning left out of the side road with an oncoming vehicle from the right.	
	B5	Turning right out of the side road with no oncoming vehicle.	
	B6	Turning right out of the side road with an oncoming vehicle from the left and the right.	
	B7	Turning left into the side road with no oncoming vehicle.	

Note that there is no Scenario A2. This was to be a scenario involving a parked vehicle and communication between vehicles, but this was agreed to be deferred until Trial 3.

3.3. Location

The Wildcat and VENTURER Simulator are both housed in the Bristol Robotics Laboratory (BRL) Autonomous Driving Zone, T Block, University of the West of England (UWE), Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY.

The Wildcat scenarios were undertaken on UWE campus roads. A circuit comprises of four loops of travel, and included all scenarios being tested in the order shown in Table 6 and as depicted in Figure 1.

Figure 1: Trial Loops

Table 6: Scenario Ordering

Scenario Order	Scenario Number
1	A1
2	B3
3	B5
4	A1
5	A3
6	B4
7	B6
8	A4
9	B1
10	B7
11	B2

The trial loops were also replicated in the VENTURER Simulator for Phase 1 and a cut down version was replicated in the simulator for Phase 2 experiments, with the first loop (where there are no interactions with other cars) being omitted.

The features of this pattern of loops forming the overall circuit are as follows:

- It was possible to compare the results between:
  - Scenarios A1 versus A3 versus A4 on the same stretch of road;
  - Scenarios B3 versus B4 at the same left hand turn out;
  - Scenarios B5 to B6 at the same right hand turn out; and
  - Scenarios B1 versus B2 at the same right hand turn in.
- The complexity of the task for the AV increased over the duration of the whole circuit;
- There were variable lengths of time interval between events, which helped create the impression of the random arrival of events to participants;
- The start and finish were at the same point, so the Wildcat could be turned and the next circuit begun; and
- The stationary vehicles could be driven into the test area and parked at the appropriate point in the circuit.

3.4. Safety and Risk Management

The safety of participants, consortium members, other road users and the general public was the number one priority when conducting the trial.

The general safety principles were that:

- The public were kept away from the vehicle;
- There was only one person in the vehicle (excluding the trained driver) at a time;

- The Wildcat vehicle always had a trained safety driver;
- The vehicle was limited to a maximum speed of 20mph; and
- The trial adhered to best practice as set out in the DfT Code of Practice for Testing – the Pathway to Driverless Cars.[2]

A Safety Case was produced by BAE including associated risk assessments and documentation relating to Trial 2 (Wildcat Integration Management Plan and Hazard Analysis and Safety Case-1.0).

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

4. Trial 2 Technology

4.1. Overview

The technology specific research questions in Table 7 are discussed in the following sections.

Table 7: Technology Research Questions

Component	Research Question
Decision Marking System (DMS)	As the level of consideration is changed, did the vehicle behave as expected?
Communications System	Does the communications system accurately transmit data in a real-world CAV situation?
Sensors	Was it possible to integrate sensors successfully within the system as a whole?
Wildcat	Did the Wildcat perform to schedule, temporally and spatially?

4.2. Data Collection and Equipment

Trial 2 involved the use of two autonomous platforms:

- The Wildcat road vehicle; and
- The VENTURER Simulator.

Table 8 outlines which platform was used for each component of Trial 2.

Table 8: Platforms per Trial 2 Component

Trial 2 Component	Platform	
	Wildcat	VENTURER Simulator
Phase 1 Experiments	✓	✓
Phase 2 Experiments	-	✓
Ongoing Technology Development	✓	✓

4.2.1. Wildcat

The Wildcat was used during Phase 1 participant experiments and was also used for further technology validation. The Wildcat was not used during Phase 2.

Data from the Wildcat was collected and stored on one of the two on-board computer platforms and downloaded periodically and included information about acceleration, deceleration, steering angle, start-stop and detection of vehicles passing.

Figure 2 shows the bespoke Bowler Wildcat (registration D1 BAE).

Figure 2: Wildcat Autonomous Vehicle

4.2.1.1. Sensors

Sensors were fitted to the Wildcat during Trial 2. Some of these sensors were used in Trial 2 to detect the presence of external objects, but not to recognise a vehicle or its type:

- Front Radar – range up to 200m for vehicles;
- 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 – same specification as the front corner modules; and
- Two Blind Spot Detect modules – each with a single radar, range up to 100m for vehicles and a single camera.

Data was collected from the sensors on the Wildcat prior to and during Trial 2 participant experiments. The 4G data connection to the central controller was able to record from multiple sensors simultaneously for whatever period of time specified, meaning that the entire loops of the trial routes were able to be captured. However, during Trial 2, data collection was limited just to that necessary for system improvement.

4.2.1.2. Decision Making System (DMS)

The DMS was replicated for both the Wildcat and the VENTURER Simulator.

During the trial, all the data produced during the pre-trial testing and the participant trials was stored and logged. This allowed the playback of recorded data as well as the visualisation and analysis of anything that had happened during a run, enabling diagnosis and resolution of problems.

In Trial 2 pre-testing, it quickly became clear that a full capability DMS, complete with its requirement for comprehensive and fault-free identification of external objects would neither be possible nor necessary for conducting the planned Trial 2 participant experiments within the time constraints of the project plan.

As an alternative, the capability for the Wildcat to autonomously follow a GPS defined path was built. This included pre-ordained interactions with other vehicles and the capability to stop or revert to manual driving instantaneously at any time for safety purposes. For the participant experiments, the vehicle executed a taught and then fixed route, interacting appropriately with other vehicles at pre-planned intervals without the need to recognise which vehicles it was interacting with at any instant from its sensors alone. The vehicle was always in the direct or supervisory control of a trained driver from BAE Systems; the former for teaching a new route and the latter for safety motivated manual interventions.

This cut down version of the DMS used:

- The finite state machine – allowing the DMS to know what part of the trial it was on;
- Clear distance measurement – to stop behind the obstacle car for the avoidance manoeuvre;
- Car Crossing Detection – to give the clear-to-go signal at junctions; and
- On Coming Car Detection – to give the clear-to-go signal at overtake manoeuvres.

The DMS itself used the following planners:

- Centreline Planner – giving trajectories based on a pre-recorded GPS track, this was used for all parts of the trial except for the overtake manoeuvres; and
- Overtaking Manoeuvre Planner – planning the trajectory around a parked car.

This more scripted approach meant that the decisions made were all very constrained and led to a good deal of consistency between participants and multiple runs in differing conditions during the participant experiments.

4.2.2. VENTURER Simulator Platform

The VENTURER Simulator is shown in Figure 3.

Figure 3: VENTURER Simulator

The VENTURER Simulator software was set up to give 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 adjacent to the test circuit), but other parts such as foliage were rendered in a more generic manner.

The same cut down DMS software was used in the VENTURER Simulator as described for the Wildcat in Section 4.2.1.2.

Careful choreography of simulated traffic was implemented to avoid collisions. However, some parameters within the DMS profile were changed for the Phase 2 experiments to allow the car to act more assertively, namely the acceleration limits were increased slightly. Additionally, the constraints on lateral acceleration were relaxed slightly. This allowed for the car to pull out from the junctions faster and with less consideration for other vehicles.

4.3. Method and Discussion

Preparation for Trial 2 involved three phases:

1. Individual component testing, on and off vehicle;
2. Open-loop systems integration, with the vehicle being manually driven but monitoring what the various component outputs would be, and gradually coupling them to each other until a final command signal was being generated that could be compared with the driver's action; and
3. Closed-loop testing, to give autonomous operation.

During the progression from open to closed-loop testing, it became clear that there was a significant development period required to ensure that the overall system was error-free, since there were some system stability issues and there was also a certain amount of 'sensor noise' that slightly obscured the true detections.

Since timing and composition of the human participant experiments was already firmly in place, development of the full closed-loop system was halted and the simpler DMS and sensor sub-system described in Section 4.2.1.2 was developed. It was decided to do this as, while less capable than the Fusion Processing driven system, it was quicker to develop and fully able to fulfil the requirements of the participant experiments. The software/hardware configuration met the experimental requirements of Trial 2 and meant that further work could be carried out as part of other planned testing during the project, both before and during Trial 3.

4.3.1. Participant Experiments (Phase 1 and 2)

For the purposes of Phase 1 a 'neutral' level of AV assertiveness was replicated using the DMS settings. This meant that during the trial the AV would always give way to the auxiliary vehicles that it encountered during manoeuvres, or 'reject the gap'. Gap rejection was adopted in Phase 1 experiments as it was safer to have the Wildcat acting in this 'no go' mode, it assisted the choreography of the event to operate in this way as less precision was required by the auxiliary vehicle drivers and it allowed for more consistency in the trial which was critical when testing and comparing human experiences across multiple participants. This approach was termed 'N' for Neutral.

The inverse of the 'reject the gap' mode (N) is where the AV judges that it can make the move into the gap in front of an oncoming vehicle. The risk-free environment of the simulator allowed the introduction of this more 'assertive' AV mode, and was termed 'N-prime' (N'). During Phase 2 experiments, the simulator, under the direction of the same DMS, would pull out in front of other vehicles during the scenarios rather than giving way as was the case in N.

Table 9 outlines the headway and critical gap adopted for Trial 2.

Table 9: Headway and Critical Gap for Neutral AV mode

Description	Headway (car following)	Critical gap (gap acceptance at junctions)
Neutral	2.0 seconds	4.0 seconds

The headway is the time gap a driver leaves between them and the car they are following. Drivers may leave different headways depending on whether they are being passive, neutral or assertive[3]. The headway defined in Table 9 is a neutral headway, and hence that is why the term N has been used. The critical gap is the gap that 50% of drivers would accept[4].

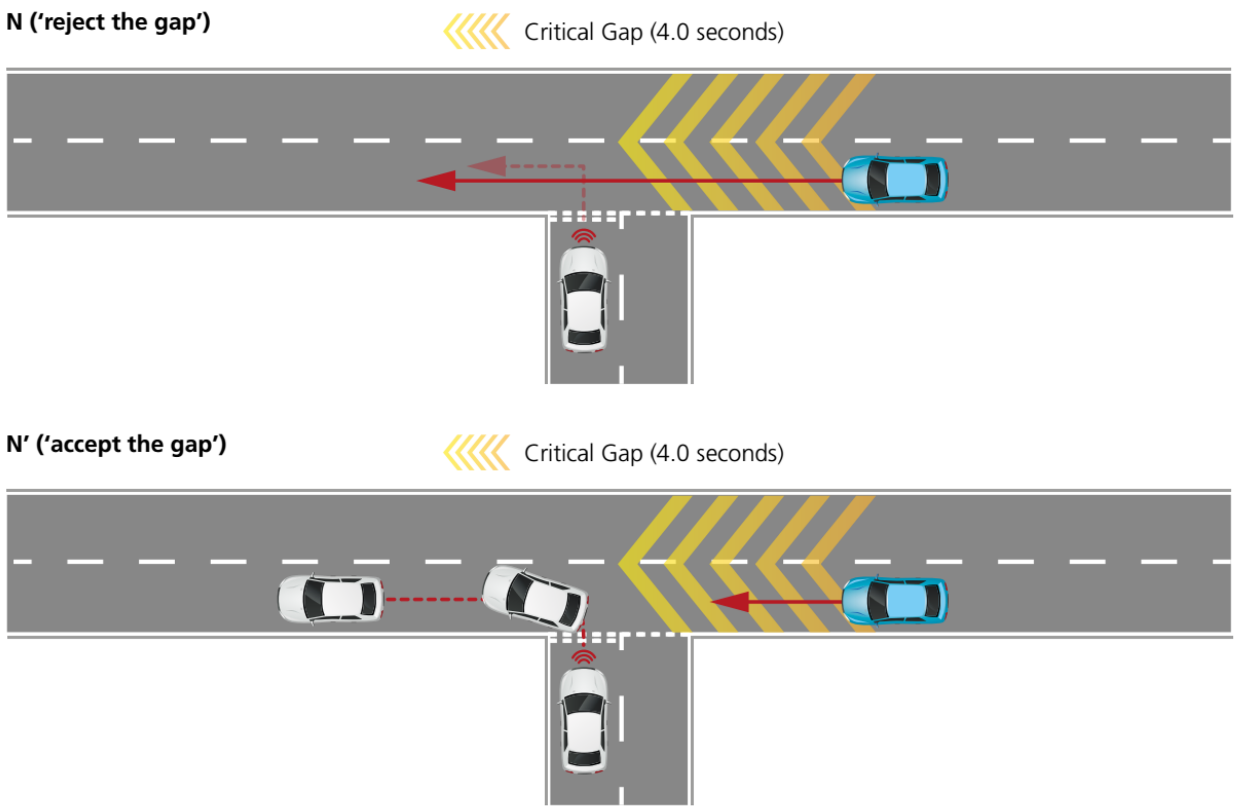
---

[3] Lewis-Evans, B., De Waard, D. and Brookhuis, K.A., 2010. That's close enough—A threshold effect of time headway on the experience of risk, task difficulty, effort, and comfort. Accident Analysis and Prevention [online]. 42 (6), pp.1926; 1926-1933; 1933.

[4] Ashalatha, R. and Chandra, S., 2011. Critical gap through clearing behavior of drivers at unsignalised intersections. KSCE Journal of Civil Engineering [online]. 15 (8), pp.1427-1434.

In all cases during Phase 1 and 2 experiments the gap offered to the AV was the critical gap for neutral driving.

Figure 4: N ('reject the gap') and N' ('accept the gap')



4.3.2. Ongoing Technology Validation

Aside from the participant experiments, further technology development was undertaken to validate the technology components and conclude the research questions outlined at the start of this document.

Continued open and closed-loop integration of the sensors and the full DMS enabled further testing and development of the more powerful perception capabilities of the Fusion Processing systems. During this testing, it was confirmed that Fusion Processing data was being sent through to the DMS. This allowed review and debugging of the stability and performance issues encountered prior to Phase 1 ready for incorporation of the full Fusion Processing sensor system for use in Trial 3.

Some measures that were taken to facilitate these advances include the implementation of a more flexible DMS framework, allowing expansion of the full-capability system. Fusion Processing also simplified the messaging structure from their units to support improved DMS triggering.

Further technical advances include the upgrading of the Wildcat's battery charging circuit that has reduced the battery changeover requirements. BAE also enhanced the Wildcat pilot to greatly reduce the 'jerky' steering observed by some participants during Phase 1 experiments.

4.3.3. Research Questions

Table 10 outlines the results of the technology-specific research questions during Trial 2.

Table 10: Technology Research Question Results

Component	Research Question	Result
Decision Making System (DMS)	As the level of consideration is changed, did the vehicle behave as expected?	Different levels of AV consideration were replicated successfully during N and N' experiments. This was only done in the VENTURER Simulator due to the potential safety implications of testing the more assertive N' mode in the Wildcat.
Communications System	Does the communications system accurately transmit data in a real-world CAV situation?	The communications system was not utilised during Trial 2.
Sensors	Was it possible to integrate sensors successfully within the system as a whole?	During ongoing technology development in Trial 2, it was confirmed that data was being sent from the sensors to the DMS however, further integration work will continue in order to enable sensor data to be used to inform decisions within Trial 3.
Wildcat	Does the Wildcat perform to schedule, temporally and spatially?	Although there were some issues of 'jerkiness' during the participant trials, the Wildcat was able to operate successfully to schedule during Phase 1 experiments using the cut-down version of the DMS and sensor system.

4.4. Limitations

Trial 2 had challenging time scales, with the trial experiments start date being delayed by approximately two weeks, due to system integration not being finished in time. There was also a one-week delay in finishing the trials.

The main difficulty was the fact that the participants had to be booked some time in advance of the trial taking place. Furthermore, the weather also made technology development more complicated: the hot weather affected sensing and power consumption, and heavy rain caused the integration activities.

4.5. Further Research

The trialling of the DMS during Phase 1 and 2 was important both for Trial 2 and for preparation for future trials as it allowed testing of new algorithms on real data without needing to be sat in the Wildcat.

While it was unfortunate that the full sensing suite was unable to be integrated for Trial 2, vehicle tracks during turning have been improved since the Trial 2 participant experiments. This in turn led to the development of a system that follows a line given by the planning sub-system more accurately, and at a higher speed.

There is a plan in place to increase the detection range of the autonomous subsystems as the plan for Trial 3 is to undertake testing at another location. Ongoing items of development that will be continued into Trial 3 include the analysis of the data gathered from Trial 2 as well as the development of Fusion Processing's lane detection sensor capability.

5. Human Factors Experiments: Phase 1

5.1. Overview

The focus of Trial 2 Phase 1 was undertaking experiments where participants were exposed to the scenarios described in Section 3, using the technology as described in Section 4.

All Phase 1 experiments replicated a 'neutral' level of AV assertiveness (Section 4.3.1). Both the Wildcat and the VENTURER Simulator were used in Phase 1 experiments. The findings from the Phase 1 experiments were compared with validated psychometric test scores, administered using Qualtrics software (an online battery of questionnaires), to determine possible associations between different individual characteristics (Section 5.6).

The overall Human Factors research questions are outlined in Table 11.

Table 11: Human Factors Research Questions

Component	Research Question
Wildcat and VENTURER Simulator	Do respondents' trust scores vary dependent on the type of AV scenario they experience?
	How do these trust scores correlate with relevant validated psychometric test scores?
	How do respondents rate their general comfort, including in relation to nausea?
Cross-platform comparison	Are trust scores significantly different depending on the platform?

5.2. Participants and Schedule

Table 12 summarises the number of participants that completed the Phase 1 experiments in the Wildcat and the VENTURER Simulator. There was one more participant in the VENTURER Simulator than the Wildcat as two participants were initially unable to undertake the Wildcat part of the experiment due to technical difficulties on the day. One participant was able to rebook and complete the Wildcat trial on another day, while the other could not.

Table 12: Phase 1 Experiments Participants

Platform	Total number of participants	Male	Female	Driving Experience > 40 years	Driving Experience 20-40 years	Driving Experience < 20 years
Wildcat	45	25	20	12	16	17
VENTURER Simulator	46	26	20	13	16	17

A quota sampling method was utilised to recruit 46 participants (Female = 20, Male = 26) from a recruitment pool of potential participants derived from a previous online survey concerning AVs, also run as part of the VENTURER project.

Participants' age ranged from 22–78 years and participant driving experience ranged from 2–60 years. Of this sample, eight participants fell under the category of 'older adults' with an age over 65 years. Four of the participants were categorised as novice drivers with driving experience of less than five years.

Participants were all required to complete a set of validated psychometric tests online using the Qualtrics software, which took 25 minutes. These were included to measure possible individual differences in driving experience, cognitive abilities linked with driving (e.g., perception and attention), personality, trust in technology and automation and factors such as risk taking, cognitive workload and impulsivity. The period of time driving was used as a proxy for driving experience. Participants were also asked if driving was a major part of their job. Full details of the Qualtrics questionnaire can be found in Appendix A.

Participants were also asked about their level of trust in how they think the two AV platforms would respond to events. These data show that participants range from AV sceptics to enthusiasts.

Both the Wildcat and the simulator trials took place during the same period, from 09 June to 07 July 2017, over a total of 14 working days. Participant experiments in the Wildcat and VENTURER Simulator occurred simultaneously to allow participants to experience scenarios multiple times across the two platforms.

The experiments were designed so that participants completed three circuits in the Wildcat and three in the VENTURER Simulator so that each scenario would be encountered three times per platform. This was required to achieve an average trust rating for each scenario. For data to be included in data analysis, participants had to complete a minimum of two circuits (of four loops each) to get an average trust rating. 37 participants completed at least two circuits in both the Wildcat and VENTURER Simulator. Table 13 provides details of participants who completed all three circuits, two circuits, and one circuit in both the Wildcat and VENTURER Simulator.

Table 13: Circuits completed by participants

Number of circuits	Wildcat: Number (and %) of participants who completed	VENTURER Simulator: Number (and %) of participants who completed
1	45 (97%)	41 (89%)
2	45 (97%)	38 (82%)
3	45 (97%)	35 (76%)*

It should be noted that the number of participants that completed the circuits declined in the simulator due to participants experiencing simulator sickness.

5.3. Experiment Design

A repeated measures design was used in which participants took part in all events across the two platforms. The study contained two independent variables (IVs). One IV was the autonomous platform with two levels: Wildcat and VENTURER Simulator. The other IV was the scenario the autonomous platform encounters, and responds to. There were a total of ten such scenarios (outlined in Table 5) encountered in the Wildcat and VENTURER Simulator. There was one numerical dependent variable (DV): trust ratings measured on a scale of 0 (completely do not trust) – 10 (completely trust).

A partial counterbalancing method was employed to control the order of the platforms to allow for comparisons of possible transfer or carry over effects between platforms e.g., participant one performed the Wildcat first then VENTURER Simulator second, participant two performed VENTURER Simulator first and then the Wildcat second.

5.4. Data Collection

Relative levels of the participants' experiences in both the Wildcat and simulator were recorded using an 11-point Likert scale (the scale (0-10) used by the participants when rating their responses) based on the following questions:

"You have just…

1. Driven on an empty road (A1i),
2. Turned out left with an empty road (B3),
3. Turned out right with an empty road (B5),
4. Driven on an empty road (A1ii),
5. Overtaken a parked car (A3),
6. Turned out left with oncoming traffic (B4),
7. Turned out right with crossing traffic (B6),
8. Overtaken a parked car with oncoming traffic (A4),

9. Turned right into an empty road (B1),
10. Turned left into an empty road (B7),
11. Turned right with oncoming traffic (B2),

…on a scale of 0-10, where 0 is 'no trust' and 10 is 'complete trust', rate how much you trusted the AV during the manoeuvre."

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

All human responses, for both the Wildcat and simulator experiments, were validated against self-reported scores for impulsivity, risk taking, self-control, personality and driving experience using Qualtrics software. This approach is similar to that undertaken in Trial 1[5].

5.4.1. Wildcat

In the Wildcat, questions were asked by a researcher in the Trial 2 base station (UWE's Conference Centre) over a radio connection and the researcher recorded the participants' responses using the 11-point Likert scale.

The participant sat in the right-hand seat of the Wildcat, which is left hand drive (i.e. the participant sat in the non-driving seat). The safety operator sat in the driving seat to the participant's left-hand side. The participant could ask for the experiment to stop at any time and the ambient temperature in the vehicle at the time of the test was recorded. Generally, the temperature was high, and this could explain some of the feelings of lack of comfort and nausea.

5.4.2. VENTURER Simulator

In the simulator experiments, participants' responses were collected by the researcher who asked the questions from the passenger seat in the simulator.

As for the Wildcat experiments, the participant also sat in the right-hand seat in the simulator but this time behind the steering wheel, thus retaining the same location within both test platforms and visual fields when viewing the environment. To create the sense of being a passenger in the simulator, and not having the sense of control of a driver, the participant was asked to keep their hands in their lap with feet firmly planted on the floor (so not as to interfere with the moving steering wheel or the pedals). The researcher started and stopped the vehicle from the passenger seat. The difference between the layout of the vehicles was not raised by any participant as being an issue that influenced their responses.

As in the Wildcat experiments, the participant could ask for the experiment to stop at any time.

5.5. Analysis and Results

This section describes the data and uses inferential statistics to test for differences between events and participants. T-test's, Analysis of Variance (ANOVA) and factorial repeated measures tests have been used to compare trust scores between scenarios. Full statistical analysis and results can be found in Appendix C.

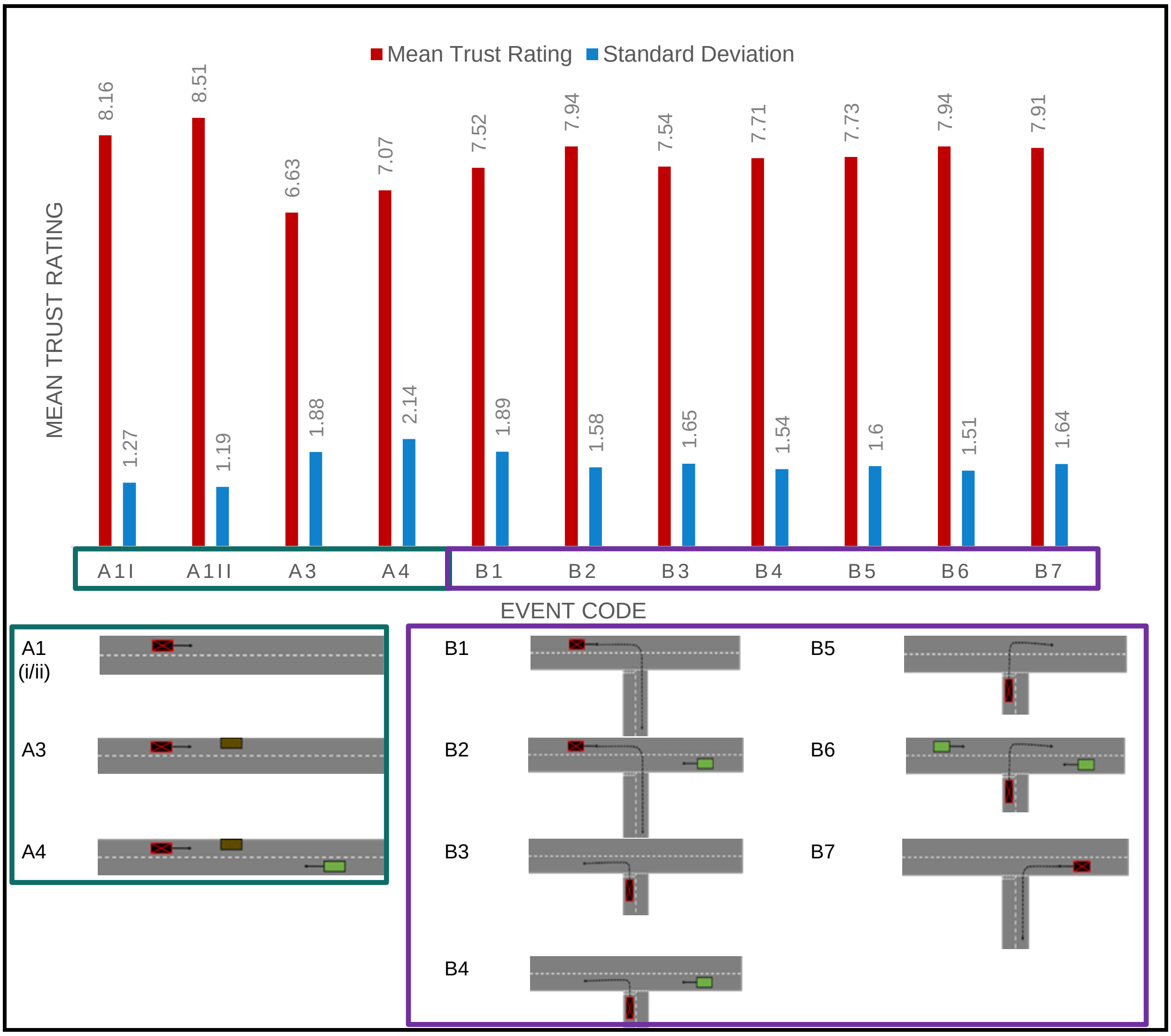
5.5.1. Wildcat Analysis and Results

Analysis of the trust scores gathered from the Wildcat experiments indicate differences between the events.

Figure 5 shows the Wildcat average trust ratings per scenario. The data within the figure is represented by the number of participants that completed the minimum of two circuits (N = 45).

[5] VENTURER Trial 1 Results: Planned Handover - http://www.venturer-cars.com/trial-1-results/

Figure 5: Wildcat Event Trust Ratings (Min = 0, Max = 10)



5.5.1.1. Link Events

Trust ratings on the link events in the Wildcat are higher for driving on an empty road compared to overtaking a parked car with and without an oncoming vehicle.

An unexpected finding is that participants trusted overtaking a parked car with an oncoming car more than overtaking a parked car with no oncoming car. This may be partially explained by technical issues experienced at this point on the circuit, where the vehicle waited before passing the stationary vehicle and sometimes would not progress without the safety driver taking control. These circumstances may have created a negative impact on participants, including trust ratings, which have been artificially reduced as a result. This issue will be addressed as part of the development towards Trial 3.

5.5.1.2. Turning Events

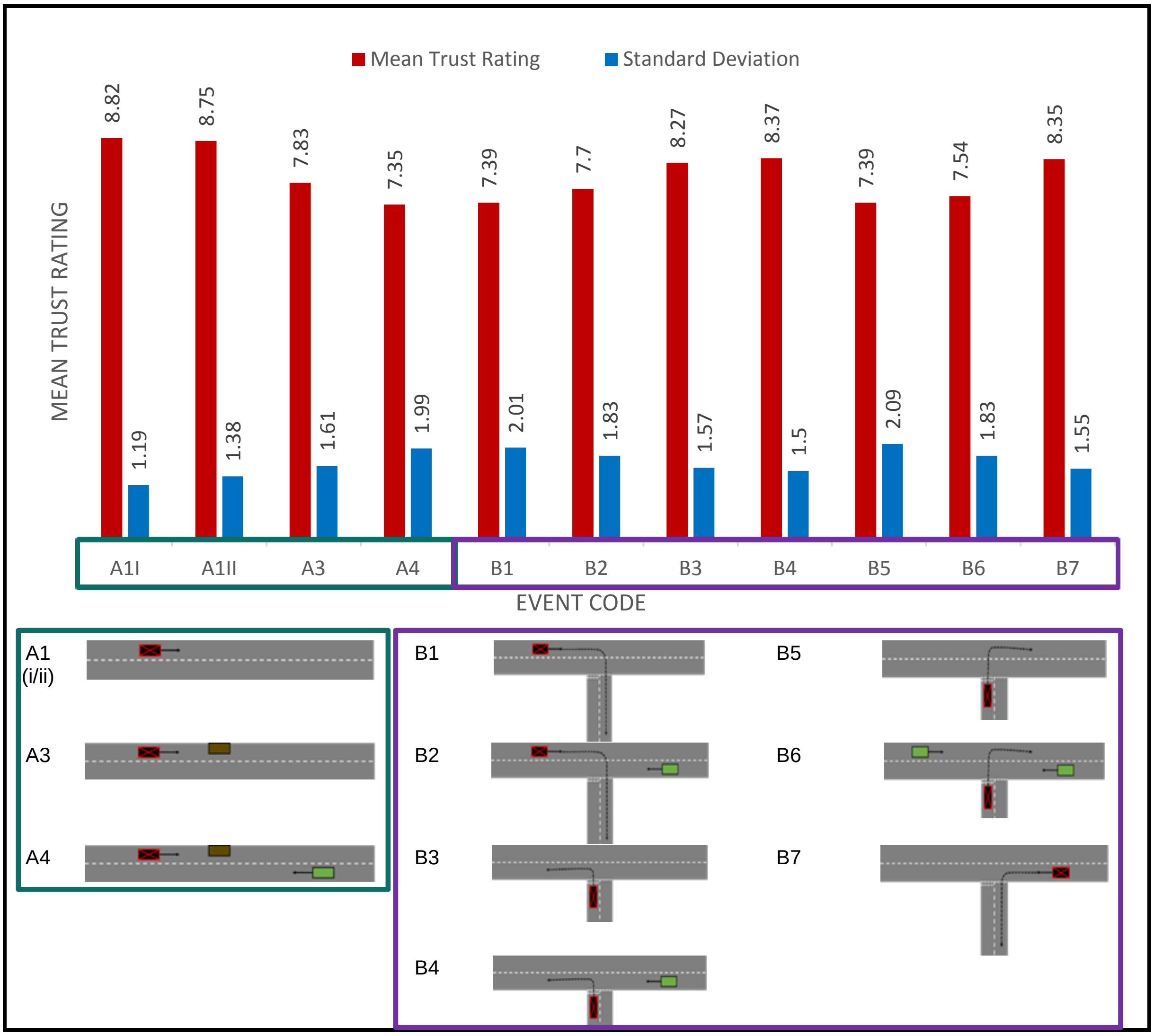
Trust ratings in the Wildcat were higher for right turning events with oncoming vehicles compared to the same events with no oncoming vehicles. This was not the case for left turns, where there was no significant difference observed between the turn with or without an oncoming vehicle.

5.5.2. VENTURER Simulator Analysis and Results

Analysis of the trust scores gathered from the VENTURER Simulator experiments indicate trends between the events.

Figure 6 shows the average trust ratings for the VENTURER Simulator. The data within the figure is represented by the number of participants that completed the minimum of 2 circuits on the autonomous platform (N = 38).

Figure 6: VENTURER Simulator Event Trust Ratings (Min = 0, Max = 10)



5.5.2.1. Link Events

For the VENTURER Simulator, trust ratings were higher for driving on an empty road compared to overtaking a parked car. Participants also trusted the VENTURER Simulator more when it was overtaking a parked car without an oncoming vehicle than with an oncoming vehicle.

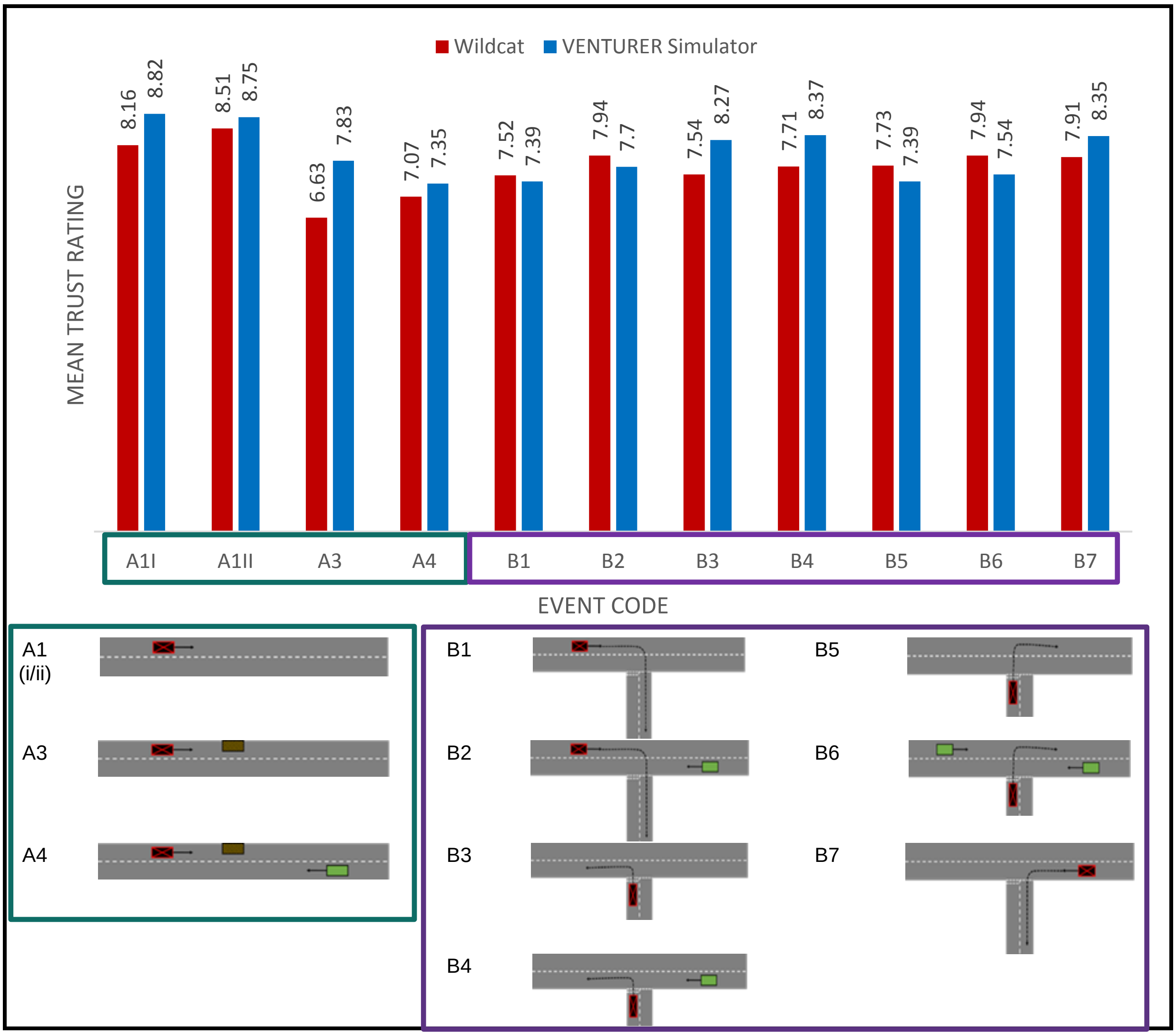
5.5.2.2. Turning Events

There was no significant difference in participants trust ratings for when the VENTURER Simulator was turning left with and without an oncoming vehicle. There was also no difference in trust ratings when the VENTURER Simulator was turning right out of the side road either with or without an oncoming vehicle. There were higher trust ratings when the VENTURER Simulator was turning right into the side road with an oncoming vehicle compared with no oncoming vehicle.

5.5.3. Cross Platform Comparison Analysis and Results

Figure 7 shows the mean trust scores for events in the Wildcat and the VENTURER Simulator.

Figure 7: Cross Platform Event Trust Ratings (Min = 0, Max = 10)



5.5.3.1. Link Events

When comparing the results observed across the autonomous platforms it appears that, regardless of the link event, trust ratings were higher in the VENTURER Simulator compared to the Wildcat. For both platforms the same ordering appears, with higher trust ratings for events with oncoming vehicles compared to the same events with no oncoming vehicles.

In the simulator and Wildcat there were higher trust ratings for driving on an empty road compared to overtaking a parked car, and overtaking a parked car with oncoming traffic. Appendix B outlines qualitative data linked to trust ratings and shows that the safety driver retaking control at any point was the comment raised by participants that had the biggest impact on trust ratings. Half of those reported were for the same event (A4, overtaking parked car with oncoming traffic). This could explain why events such as driving on an empty road consistently received higher trust ratings than scenario A4. It is important to note that due to the way that the trial was conducted, there was a lot of under reporting of qualitative data in the Wildcat compared with the simulator, this is outlined further in Section 5.7.

5.5.3.2. Turning Events

When comparing turning events across the autonomous platforms there are higher trust ratings for left turn events in the VENTURER Simulator compared to the Wildcat, but higher trust ratings for all other (right) turn events in the Wildcat compared to the VENTURER Simulator.

5.6. Psychometric Tests

Psychometric measures were used to examine possible associations between different individual characteristics (e.g., age, driving experience, trust in automation) and trust in the autonomous platform performing different events. The following sections outline the correlational analysis for age, driving experience (time driving licence has been held, and number of miles driven a year), trust in automation, and trust in general technology.

The relationship between each of the following aspects and trust ratings for each scenario occurring across both autonomous platforms was explored using a Pearson product-moment correlation coefficient. Full statistical analysis can be found in Appendix D.

5.6.1. Correlations Between Age and Trust Ratings

With the exception of event B3 (left turn out of side road with no oncoming vehicle), negative low strength associations were revealed between trust ratings for the events that occurred in the VENTURER Simulator and participant age, suggesting that as participant age increased, trust ratings decreased. Despite a relationship being present, it was not statistically significant.

Event B3 in the VENTURER Simulator revealed a significant medium negative correlation between participant age and trust ratings, in that as participant age increased trust ratings decreased. The coefficient of determination, which reveals how much variance is shared, indicates that participant age only explains 9.6% of the variation between scores for event B3 (turning left out of a side road with no oncoming vehicle), which highlights that 90.4% of variation may be explained by other factors. Such factors may be revealed by other individual factors, or the circumstantial data that was logged during the running of the trials and is presented in Appendix B.

For the Wildcat, there is an inverse correlation between age and trust for some events (e.g., A1i), and for the others when participant age went up so did the trust rating. However, none of these correlations were significant.

5.6.2. Correlations Between Driving Experience and Trust Ratings

Apart from event B3 (left turn out of side road with no oncoming vehicle), negative low strength associations were revealed between trust ratings for the events that occurred in the VENTURER Simulator and driving experience, suggesting that the longer a driving licence was held, the lower the trust ratings. Despite a relationship being present, they are not statistically significant.

Event B3 in the VENTURER Simulator revealed a significant medium negative correlation between driving experience and trust ratings: longer periods of driving licence holding are associated with lower trust ratings. The coefficient of determination, which reveals how much variance is shared, indicates that the time a driving licence was held only explains 11.6% of the variation between trust scores for event B3, which highlights that 88.4% of variations may be explained by other factors. Such factors may be revealed by other individual factors or the circumstantial data that was logged during the running of the trials, and presented in Appendix B.

There were also non-significant and low negative (e.g., A1i, A4, B7) and positive associations between driving experience and trust ratings for the events that occurred in the Wildcat.

5.6.3. Correlations Between Time Spent Driving per Year and Trust Ratings

Negative low strength associations were revealed between trust ratings for the events that occurred in the VENTURER Simulator and driving experience, with more miles being driven per year being associated with lower trust ratings. These were, however, not statistically significant.

Non-significant low negative and positive associations between driving experience and trust ratings were also identified for the events that occurred in the Wildcat. A higher number of miles driven per year is associated with lower trust ratings for some events (e.g., A1, A4, B7), but not others.

5.6.4. Correlations Between General Trust in Automation and Trust Ratings

With the exception of scenario A1 (driving along an empty road) significant positive low to medium strength associations were revealed between trust ratings for the events that occurred in the VENTURER Simulator and trust in automation scores. This suggests that, for higher scores on the trust in automation questionnaire, there were higher trust scores for the AV events. The coefficient of determination, which reveals how much variance is shared, indicates that between 8.2% - 17.7% of variation in event trust scores can be explained by scores in the trust in automation questionnaire. Between 82.3% - 91.8% may be explained by other factors. Such factors may be revealed by other individual factors or the circumstantial data that was logged during the running of the trials, detailed in Appendix B.

Significant positive medium strength associations were revealed between trust ratings for the events that occurred in the Wildcat and trust in automation scores, suggesting that for higher scores on the trust in automation questionnaire, there were higher scores of the trust in the event. The coefficient of determination, which reveals how much variance is shared, indicates that between 13.4% - 24.7% of variation in event trust scores can be explained by scores in the trust in automation questionnaire. Between 75.3% - 86.6% may be explained by other factors. Such factors may be revealed by other individual factors or the circumstantial data that was logged during the running of the trials, as detailed in Appendix B.

5.6.5. Correlations Between Trust in Technology and Trust Ratings

With the exception of event A1 (driving along an empty road), significant positive low to medium strength associations were revealed between trust ratings for the events that occurred in the VENTURER Simulator and trust in general technology scores, suggesting that for higher scores on the trust in general technology questionnaire there were also high scores for trust in the event. The coefficient of determination, which reveals how much variance is shared, indicates that between 7.8% - 29.3% of variation in event trust scores can be explained by scores in the trust in general technology questionnaire. Between 70.7% - 92.2% may be explained by other factors. Such factors may be revealed by other individual factors or the circumstantial data that was logged during the running of the trials, as detailed in Appendix B.

Significant positive low to medium strength associations were revealed between trust ratings for the events that occurred in the Wildcat and trust in automation scores, suggesting that for higher scores on the trust in general technology questionnaire, there were higher scores for the trust in the event. The coefficient of determination, which reveals how much variance is shared, indicates that between 8.4% - 24.5% of variation in event trust scores can be explained by scores in the trust in automation questionnaire. Between 91.7% - 75.5% may be explained by other factors. Such factors may be revealed by other individual factors or the circumstantial data that was logged during the running of the trials, as detailed in Appendix B.

5.7. Limitations

Whilst the overall reliability of both test platforms was good, a range of circumstantial occurrences were observed that could have had an impact on the results. For example, there were several technical failures in the Wildcat with the vehicle stopping before the overtaking manoeuvre with the parked car, with the 'safety driver' having to retake control.

Similarly, in the VENTURER Simulator, there was a tendency for the AV to stop on a particular section of the circuit, often near the point of the AV overtaking a parked car with oncoming traffic (A4). This meant a reboot of the simulator was required, and then a wait for the participant until they reached the same point on the circuit before the trial could continue. A further factor was the lack of smoothness, or 'jerkiness' at some junctions which may also have impacted results.

There was a lot of under-reporting of qualitative data in the Wildcat compared with the simulator. In the simulator, the data collector was physically in the car with the participant and could hear (and record) everything that they said, whereas in the Wildcat the only communication with the data collector was via the radio which meant that it was limited to the trust ratings. As such, the data in the tables in Appendix B comes from when a further member of staff was present with the Wildcat data collector and noticed that something happened and asked the safety driver for an explanation. Although safety drivers were asked to report any times when they re-took control or an incident happened, they did not do this, perhaps not understanding the importance of this information to the Human Factors study.

In relation to 'simulator sickness', of the 46 people who participated in Trial 2 (45 in the Wildcat), all participants completed 3 circuits in the Wildcat AV. However only 35 completed three circuits in the VENTURER Simulator, 38 completed two circuits and 41 completed only one circuit in the VENTURER Simulator (i.e. 5 failed to complete a circuit in the simulator). All of those that did not complete three circuits asked for the experiment to be ended prematurely due to nausea.

34 out of 46 people experienced an elevated feeling of nausea in the simulator after one familiarisation lap (compared with 6 out of 43 in the Wildcat). See the Table 14 below for details.

Table 14: Nausea Ratings (0 = not at all nauseous, 10 = completely nauseous)

Platform	Nausea Rating										
	0	1	2	3	4	5	6	7	8	9	10
Wildcat AV	37	2	0	1	1	0	1	0	1	0	0
VENTURER Simulator	12	5	5	12	2	4	3	2	1	0	0

Two people vomited before completing a circuit in the simulator. Levels of nausea generally increased after subsequent circuits (with people rating as high as 9), but this only started to be recorded half way through the trial (and only in the simulator).

Consideration is being given in the technological development leading to Trial 3 to deal with the simulator sickness issue, which is already a known issue in the field of inquiry[6].

[6] 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.

6. Human Factors Experiments: Phase 2

6.1. Overview

Phase 2 of the participant experiments were undertaken to introduce the N' AV mode. During Phase 2 experiments the simulator under the direction of the DMS, would pull out in front of other vehicles during the scenarios, accepting the critical gap, rather than giving way to oncoming vehicle, as was the case in N.

This experiment was included to explore how gap acceptance at the critical gap, rather than gap rejection, would affect participants' trust ratings.

6.2. Participants and Schedule

The participants involved in the VENTURER Simulator N' experiments (using simulated scenarios) are outlined in Table 15.

Table 15: VENTURER Simulator Trial 2 Phase 2 (N') Participants

Total number of participants	Male	Female	Did not specify gender	'Older Adults' (age ≥ 65 years)	Driving Experience < 5 years
41	25	14	2	12	0

As with Phase 1, participants were recruited initially from the VENTURER online survey respondents, but it was a condition of Phase 2 that participants had not previously participated in Phase 1. Participants were selected to aim at a 50/50 split in gender, and to generate a sample of participants with moderate driving experience. Participants' ages ranged from 23–81 years and participants' driving experience ranged from 6–63 years.

Table 16 provides details of participants who completed all circuits, two circuits, and one circuit in both modes, N and N'. Participants had to complete all circuits (to allow an average trust rating) for data to be included in the analysis.

Table 16: Circuits Completed by Participants in N'

Number of Circuits	Autonomous mode N: Number (and %) of participants who completed	Autonomous mode N': Number (and %) of participants who completed
1	40 (97%)	39 (95%)
2	37 (90%)	37 (90%)
34 participants (82.92%) completed all (two) circuits in both the autonomous modes.		

The trials lasted approximately 90 minutes for each participant. Half of the participants started with N' and the other half with N. The online Qualtrics questionnaire was self-administered using a laptop and consisted of a series of psychometric questions that took about 25 minutes to complete (as detailed in Appendix A). This questionnaire was completed by participants between circuits in N and circuits in N' in the simulator.

6.3. Experiment Design

A repeated measures design was utilised in which participants took part in all experimental conditions. The study contained two independent variables (IV). One IV was the autonomous mode with two levels: N and N', described in Section 4.3.1. The other IV was the event the autonomous platform encounters and responds to. There were a total of four events (described in Table 17) encountered in both N and N', and these included a range of passing and turning manoeuvres. During the N mode, all the turning manoeuvres were at a T-junction with the AV not accepting the critical gap and all turns where it had to give way (out of the side road and turning right into the side road). In the N' mode, the AV accepted the critical gap.

There was one numerical dependent variable (DV): trust ratings in the autonomous mode in which the event was experienced, measured on a scale of 0 (completely do not trust) – 10 (completely trust). A partial counterbalancing method was employed to control the order of the modes to allow for comparisons of possible transfer/carry over effects between modes (e.g. participant one performed the N mode first then N’ second, participant two performed N’ first and then N second). All trials were completed in the VENTURER Simulator.

Table 17: Phase 2 Event Descriptors

Scenario Type	Scenario Number	Scenario Description
Link (A)	A4	Overtaking a parked car leaving a safe passing distance and waiting if necessary to leave a safe gap to an oncoming car.
Junction (B)	B2	Turning right into the side road with an oncoming vehicle.
	B4	Turning left out of side road with an oncoming vehicle from the right.
	B6	Turning right out of the side road with an oncoming vehicle from the left and the right.

Table 5 has a full description of the events described in Table 17.

6.4. Data Collection

Participants completed three circuits in both autonomous modes so that each event was encountered three times for both of N and N’. This was required to achieve an average trust rating for each event.

The experimental design for Phase 2 was similar to the Phase 1 experiments (detailed in Section 4.3) in the VENTURER Simulator, except that participants sat in the passenger’s seat, with the experimenter sat in the driver’s seat. The wing mirrors and rear-view mirror were set up for the participant.

The familiarisation lap was shortened compared with Phase 1 to about one minute and involved overtaking a parked car. The initial question as to whether the participant felt nauseous was asked while the car was moving, prior to the first event.

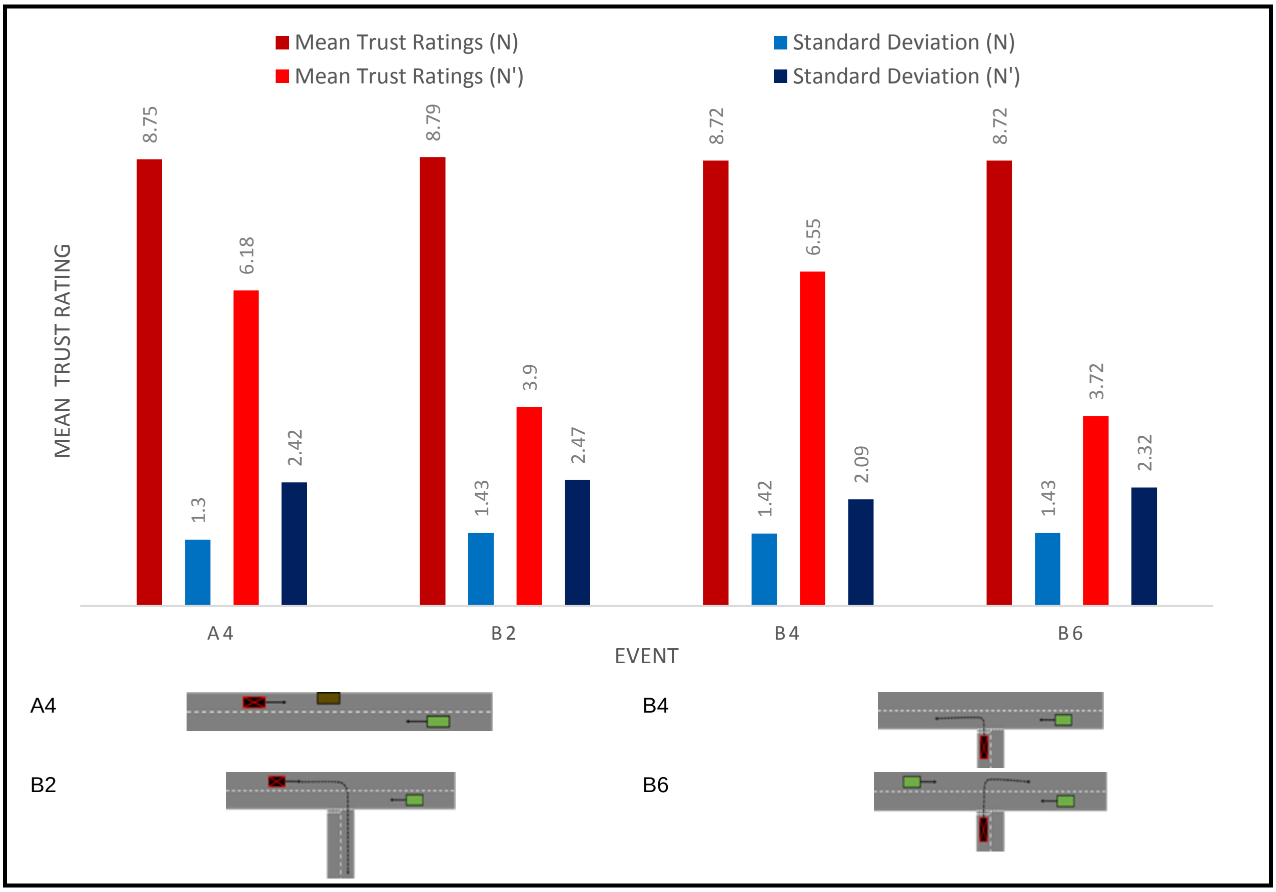
Again, all human responses were validated against self-reported scores for impulsivity, risk taking, self-control, personality and driving experience using Qualtrics software (as detailed in Appendix A).

6.5. Phase 2 Human Factors – Analysis and Results

6.5.1. Descriptive data

Figure 8 indicates the descriptive data for both the N and N’ mode. The descriptive data within the figure is represented by the number of participants that completed a minimum of 2 circuits on both autonomous modes (N = 34).

Figure 8: Descriptive Data for Autonomous Modes N and N' and Event Trust Ratings (Min = 0, Max = 10)



The first set of analysis involved comparing trust ratings in each autonomous mode for each event (e.g., B4 in N vs B4 in N'). All paired comparisons are reported as 2-tailed tests.

A paired samples t-test revealed a significant difference in trust ratings between the participants' results between N and N', with N' receiving the lower trust ratings on all events. This finding suggests that critical gap acceptance reduces the trust participants have in the VENTURER Simulator.

Following on from this, analyses were performed to compare trust ratings between events with each autonomous mode (e.g., B4 in N vs B6 in N).

A repeated measures ANOVA comparing trust ratings for events (B4, B6, A4, B2) in autonomous mode N and revealed no significant difference in trust ratings.

A repeated measures ANOVA comparing trust ratings for events (B4, B6, A4, B2) in autonomous mode N' and revealed a significant main effect of trust ratings on events that occurred in N'. Bonferroni post hoc analysis[7] revealed that certain scenarios (particularly right turn events) received lower trust ratings compared to other events.

6.6. Research Questions

Table 18 describes the conclusions drawn for each of the human factors research questions.

[7] Bonferroni Post Hoc Analysis – a method to correct the problem of multiple comparisons with multiple null hypotheses.

Table 18: Human Factors Research Question Results

Component	Research Question	Result
Wildcat	Do respondents' trust scores vary dependent on the type of AV scenario they experience?	Yes – Trust varied depending on the scenario.
	How do these trust scores correlate with relevant validated psychometric test scores?	The reliability and consistency of the trust measures in general has been validated by some strong associations between the participant's trust in automation and general technology. These associations were stronger for the Wildcat. There was no significant correlation in trust scores in events with either age or driving experience identified.
	How do respondents rate their general comfort, including in relation to nausea?	14% of participants experienced nausea after the familiarisation lap in the Wildcat.
VENTURER Simulator	Do respondents' trust scores vary depending on the type of AV scenario they experience?	Yes – Trust varied depending on the scenario.
	How do these trust scores correlate with relevant validated psychometric test scores?	The reliability and consistency of the trust measures in general has been validated by some strong associations between the participant's trust in automation and general technology however these associations were stronger for the Wildcat. There was no significant correlation in trust scores in events with either age or driving experience identified.
	How do respondents rate their general comfort, including in relation to nausea?	Three quarters of participants experienced elevated levels of nausea in the simulator, causing one quarter of participants to abort the trial before completion. Trust scores are yet to be analysed to see whether they were affected by nausea.
Cross-platform comparison	Are trust scores significantly different depending on the platform?	Overall, trust was higher in the VENTURER Simulator than the Wildcat.

7. Discussion

Many of the comparisons and findings of differences between scenarios, platforms and gap acceptance or rejection, may be related to the specific behaviour of the DMS in the scenarios tested. This trial demonstrates that the methodology can identify difference in trust by event.

7.1. General Validity of the Trial Results

Overall the assessment of trust has been successful. The experiments found that mean ratings under virtually all scenarios acting in the neutral (N) autonomous mode exceeded seven out of ten, with ten representing the maximum possible trust rating. Under all Phase 1 conditions, participants indicated that under the experimental conditions in which they were assessed, they trusted both the Wildcat and the VENTURER Simulator. The relative distribution of the data with standard deviations of two or less indicates that responses were overall consistent, whilst acknowledging individual variations in response style. The trust ratings were supported by ad hoc comments from participants, providing some cross-validation, including that they enjoyed the experience, with expressions of enthusiasm for future involvement in the project.

The reliability and consistency of the trust measures in general has been validated by the consistent pattern of significant associations, including some strong associations, between the participant’s trust in automation and general technology. The higher their general trust in automation and technology then the higher their trust ratings for the autonomous driving events experienced in both the Wildcat and VENTURER Simulator. These associations were stronger for the Wildcat which may reflect that the evaluation of trust when being driven autonomously in a real vehicle may cause detriment or harm if things went wrong, whereas in the simulator there is no potential for detriment and harm.

There was no significant correlation identified between trust scores and either age or driving experience. Perhaps contrary to expectations, this indicates that groups, for example older drivers, do not demonstrate a statistically lower or higher feeling of trust towards AV decisions.

7.2. Wildcat

Trust was higher on a road without other vehicles present compared to the event of overtaking a parked car, and this was whether or not there was an oncoming vehicle present at the overtaking manoeuvre decision point.

Trust was higher for the event of overtaking a parked car when there was an oncoming vehicle present as compared with the situation without an oncoming vehicle present. Similarly, trust was higher turning right into a side road and turning right out of a side when there was an oncoming vehicle present as compared with the situation without an oncoming vehicle present. This unusual finding may be explained by the occurrence of technical faults. However, for the Wildcat, the exact locations of technical faults were not consistently reported.

7.3. VENTURER Simulator

In contrast with the Wildcat, trust was higher for the event of overtaking a parked car without an oncoming vehicle present as compared with the situation with an oncoming vehicle being present. The mismatch between these findings may be the result of inconsistent technical faults between the two platforms.

In contrast with the Wildcat, trust was higher only for the condition of turning right into the side road with an oncoming vehicle present as compared with the situation without an oncoming vehicle.

7.4. Cross-platform Comparison

The pattern of responses in Phase 1 was generally consistent across platforms reflecting the general reliability and consistency of the results obtained. Where there were differences they could be partially explained by some of the issues raised by participants such as differences in visibility (due to foliage) between the real world and the simulated world. Additionally, slight inconsistencies in the real world of the arrival of auxiliary vehicles, compared to complete consistency in the simulator may have been a factor.

In keeping with general predictions, participants overall gave higher trust ratings when being driven in the VENTURER Simulator compared to the Wildcat on a real road. Similarly, participants' overall trust ratings were higher when being driven on an empty road when compared to passing a parked vehicle and when turning left or right.

There were some minor differences between platforms including a relatively lower increased rating in the VENTURER Simulator compared to the Wildcat for overtaking a parked car with oncoming vehicles; and overall lower ratings in the simulator for turning in right. A fuller evaluation of the observed differences in trust ratings will be possible when the more detailed qualitative data has been fully analysed.

7.5. Gap Rejection versus Gap Acceptance (N vs N')

Trust ratings in Phase 2 experiments were consistently lower for the N' mode in which the critical gap was accepted, as compared with the N mode, in which the critical gap was rejected.

For the N' autonomous mode, trust ratings were lower for right turns than compared with left turns, and overtaking a parked car with an oncoming vehicle.

While overall the findings suggest that critical gap acceptance (N') reduces the trust participants have in the VENTURER Simulator, it should be noted that the N' manoeuvres in the simulator, particularly the right turns, put the simulated AV in very close proximity to the oncoming vehicles. In some cases the AV then had to take avoiding action, such as stopping, due to driving slowly and pausing on turning. This will need to be further considered before further trials are conducted.

7.6. Application of Findings

The trust rating protocol used during the trials and validation against general trust measures has demonstrated the utility and suitability of these measures for further trials – including VENTURER Trial 3. Overall, these experiments have been a success.

Trust ratings were lower for all events when the AV accepted the gap that 50% of drivers would accept (N') as compared with when the AV rejected the gap that 50% of drivers would accept (N). This finding may indicate a need for AVs to operate more cautiously than the average human driver in order to foster public trust. The requirement for this more cautious approach could also likely have safety benefits in conditions with mixed AV and conventional vehicles. There might also be congestion benefits in some traffic flow conditions resulting from the greater level of consistency in gap acceptance behaviour.

7.7. Further Research

Triangulation of the preliminary results from the Wildcat road trials and VENTURER Simulator trials was important as this could give confidence in the findings with regards to potential generalisation to real world scenarios. The data collected provides a means of relating simulator scores to an equivalent real world event that has not been tested in the real world.

To fully understand the implications of the use of different AV modes on user acceptance, road safety, congestion and manufacturing standards, additional and more in depth exploration with variable headways and critical gaps should be included in future studies.

There were large number of participants affected by simulator sickness, and a proportion of those were not able to complete the trial. Technical developments are being put in place to assist in reducing the incidence of simulator sickness in Trial 3.

Appendix A

Listing of psychometric tests administered 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 year/month;

- 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 sleep disturbances, etc. (e.g. “During the past month, how many hours of actual sleep did you get at night?” “During the past month, how would you rate your overall sleep quality?”);

- Mood: Used VAS. 18 items (e.g. dreamy, attentive, bored, interested) (Bond & Lader 1974); and

- 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 measures were in addition to planned project activities and have not yet been coded or analysed. Such analysis would examine whether there are relationships between these factors and trust in the autonomous platform.

Appendix B

<u>Circumstantial Data Linked to Trust Ratings</u>

Table 19 and Table 20 outline the VENTURER simulator qualitative data made by participants or the researcher linked to the trust ratings given in the VENTURER Simulator.

Table 19: VENTURER Simulator Qualitative Data (linked to lower ratings)

Comments linked to lower ratings										
Event	**Wheels/ breaks squealed**	**Jittery/ heavy breaking/ jerks/ stops or turns too slowly**	**Additional movements not needed for manoeuvre**	**Steering wheel moves more than expected**	**No indicators**	**Sim crashed/ software glitch**	**Took too long**	**Did not slow to give way**	**Too close to other cars/ poor positioning**	**Total**
A1				4		1				**5**
A3		8	1		1	5	3			**18**
A4		11	7	1	1	1			2	**23**
B1	3	12		1			4			**20**
B2	5	8								**13**
B3		8		1	1	1		1		**12**
B4		3		1						**4**
B5	3	11		1				1		**16**
B6	1	6							1	**8**
B7					3					**3**
Total	**12**	**67**	**8**	**9**	**6**	**8**	**7**	**2**	**3**	

Table 20: VENTURER Simulator Qualitative Data (other comments)

Other comments not linked to lower ratings									
Event	**Screens and mirrors misaligned**	**Over cautious**	**Poor road positioning: didn't stop when it should; too close to centre line; took corner too wide**	**Sim crashed**	**Slow oncoming vehicles**	**More visibility in simulator (i.e. fewer hedges)**	**Too much unneeded brake or wheel screech**	**Blue line**	**Total**
Gen		1		2 (+8)		1	2 (+12)	1	**7**
A3		1							
A4	1								
B1		1							
B2		1	2						
B4		2	1		1				
B5		1							
B6		2			1				
Total	**1**	**9**	**3**	**2**	**2**	**1**	**2**	**1**	

Note: The manoeuvres that attracted the most participant comments that were linked to lower ratings were overtaking a parked car (with and without traffic, A3/A4) and turning right into and from an empty road. The simulated AV was considered to be too jerky in all of these manoeuvres. Additionally, overtaking a parked car with no oncoming traffic (A3) was the place where the simulator crashed the most. The jerky, stop-start ride was by far the most common comment that led to lower trust ratings from participants.

Some of the issues raised that had less bearing on the trust ratings could be used to improve the simulator experience going forward, including: the speed of auxiliary vehicles; reducing the squeal of the wheels and/or brakes; removing the blue line that followed the car; making the reduced visibility due to foliage match the real work. Additionally, some of the items that did impact on ratings could also be used to enhance the simulator including adding working indicators.

Table 21 and Table 22 outline the Wildcat AV qualitative data made by participants or the researcher linked to the trust ratings given in the Wildcat AV.

Table 21: Wildcat AV Qualitative Data (linked to lower ratings)

Comments linked to lower ratings					
Event	Safety driver took manual control	Vehicle stopped for manoeuvre & failed to continue in autonomous mode	Hit kerb	Too close to other cars/ poor positioning	Total
A3	1				**1**
A4	5	1	1	1	**8**
B1	1				**1**
B4	1				**1**
B7	2				**2**
Total	**10**	**1**	**1**	**1**	**13**

Table 22: Wildcat AV Qualitative Data (other comments)

Other Comments not linked to lower ratings							
Event	**Online tracking was lost and so data collector missed an event**	**Radio coms problem and so event was missed**	**Safety driver took manual control between or after events**	**Hit kerb between or after events**	**Wildcat had to restart circuit due to technical problems**	**Circuit had to be redone due to axillary vehicles missing their cues**	**Total**
Gen			3	1	2	1	**7**
A1	1						**1**
A4	1						**1**
B6		1					**1**

The safety driver retaking control at any point the biggest impact on trust ratings. Half of those reported were for the same event (A4, overtaking parked car with oncoming traffic). Due to the way that the trial was conducted there was a lot of under-reporting in the Wildcat compared with the simulator. In the simulator, the data collector was physically in the car with the participant and could hear (and record) everything that they said, whereas in the Wildcat the only communication with the data collector was via the radio which meant that it was limited to the trust ratings. As such, the data in Table 21 and Table 22 mainly comes from when a further member of staff was physically present with the Wildcat data collector and noticed that something happened and asked the safety driver for an explanation. Although safety drivers were asked to report any times when they re-took control or an incident happened, they did not do this, perhaps not understanding the importance of this information to the Human Factors study.

Appendix C

Wildcat Analysis and Results

Descriptive Data Analysis

Link events (A1 vs A3 vs A4)

Trust ratings on the straight road events in the Wildcat AV are higher for driving on an empty road (A1, mean = 8.16), compared to overtaking a parked car (A3, mean = 6.63), and overtaking a parked car with an oncoming car (A4, mean = 7.07). As indicated by these means, an interestingly unusual outcome is that participants trusted overtaking a parked car with an oncoming car (A4) more than overtaking a parked car with no oncoming car (A3). This unexpected result may be partially explained because there were some technical issues at this point on the circuit, where the vehicle waited before passing the stationary vehicle and sometimes would not progress without the safety driver taking control. Therefore, this will have created a negative impact on participants, including trust ratings which have therefore been artificially reduced due to a confounding effect.

Junction events (B3 vs B4, B5 vs B6, B1 vs B2)

Interestingly, trust ratings in the Wildcat AV were higher for turning events with oncoming vehicles (B4, mean = 7.71; B6, mean = 7.94; B2, mean = 7.94) compared to the same events with no oncoming vehicles (B3, mean = 7.54; B5, mean = 7.73; B1, mean = 7.52).

Inferential Data Analysis

Wildcat link events: A1 vs A3 vs A4 (repeated measures)

A repeated measure ANOVA revealed a significant main effect of the level of trust for straight road events, $F(1.56, 68.71) = 35.72$, $MSE = 1.22$, $p < 001$, $\eta p2 = .400$. Bonferroni post-hoc analysis revealed that higher trust ratings were given to being driven in the Wildcat on an empty road (A1i) compared to overtaking a parked car (A3) ($p < .001$), and when overtaking a parked car with oncoming traffic (A4) ($p < .001$). Furthermore, and quite surprisingly, results revealed higher trust ratings for overtaking a parked car with oncoming traffic (A4) compared to just overtaking a parked car (A3) ($p = .011$).

Wildcat junction events: B3 vs B4, B5 vs B6, B1 vs B2 (paired samples t-tests)

All paired comparisons are reported as 2-tailed. A paired samples t-test revealed no significant difference in trust ratings between turning out left with no oncoming vehicle (B3) and turning out left with oncoming traffic (B4), $t(44) = 1.40$, $p = .168$. Another paired-samples t-test revealed that average trust rating was significantly higher for turning right out of the side road with crossing traffic (B6) compared to turning out right into an empty road (B5), $t(44) = 2.68$, $p = .010$. Furthermore, average trust ratings were also significantly higher for turning right into the side road with oncoming traffic (B2) compared to turning in right with no oncoming traffic (B1), $t(44) = 3.39$, $p < .001$.

VENTURER Simulator Analysis and Results

Descriptive Data Analysis

Link events (A1 vs A3 vs A4)

For the VENTURER Simulator, trust ratings were higher for driving on an empty road (A1, mean = 8.82) compared to overtaking a parked car (A3, mean = 7.83), and overtaking a parked car with an oncoming car (A4, mean = 7.35). The order of trust scores for these three scenarios is different in the simulator from that in the Wildcat.

Comparing link events across autonomous platforms, it appears that, regardless of the event, trust ratings were higher in the VENTURER Simulator compared to the Wildcat.

Junction events (B3 vs B4, B5 vs B6, B1 vs B2):

The same ordering appears in the VENTURER Simulator as the Wildcat, with higher trust ratings for events with oncoming vehicles (B4, mean = 8.37; B6, mean = 7.54; B2, mean = 7.70) compared to the same events with no oncoming vehicles (B3, mean = 8.27; B5, mean = 7.39; B1, mean = 7.39).

Comparing turning events across autonomous platforms, there are higher trust ratings for events B3 and B4 (turn left out of side road with and without oncoming vehicle) in the VENTURER Simulator compared to the Wildcat, but higher trust ratings for events B1, B2, B5, and B6 (all other turn events) in the Wildcat compared to the VENTURER Simulator.

Inferential Data Analysis

VENTURER Simulator Link Events: A1 vs A3 vs A4 (repeated measures)

A repeated measures ANOVA revealed a significant main effect of the level of trust on road events that occurred within the Simulator, $F(1.37, 50.92) = 31.01$, $MSE = 1.17$, $p < .001$, $\eta p2 = .417$. Bonferroni post-hoc analysis revealed that higher trust ratings were given for being driven on an empty road (A1i) compared to overtaking a parked car (A3) ($p < .001$) and overtaking a parked car with oncoming traffic (A4) ($p < .001$). Furthermore, significantly higher trust ratings were also given when overtaking a parked car (A3) compared to overtaking a parked car with oncoming traffic (A4) ($p < .001$).

VENTURER Simulator junction events: B3 vs B4, B5 vs B6, B1 vs B2 (paired samples t-tests)

A paired samples t-test revealed no significant difference between turning out left with an empty road (B3) compared to turning out left with oncoming traffic (B4), $t(37) = 1.40$, $p = .168$. There was also no significant difference between turning out right with an empty road (B5) and turning out right with crossing traffic (B6), $t(37) = 1.57$, $p = .124$. However, there was a significantly higher rating of trust for turning right into the side road with oncoming traffic (B2) compared to turning right into the side road without oncoming vehicle (B1), $t(37) = 2.72$, $p = .010$.

Cross Platform Comparison Analysis and Results

Factorial Repeated Measures

Cross Platform A1 vs A3 vs A4

A factorial 2 (platform: Wildcat, VENTURER Simulator) x 3 (link events: A1i, A3, A4) repeated measures ANOVA revealed a significant main effect of platform, $F(1, 36) = 33.22$, $MSE = .889$, $p < 001$, $\eta p2 = .480$, a significant main effect of event, $F(1.25, 45.30) = 33.02$, $MSE = 2.01$, $p < .001$, $\eta p2 = .478$, and a significant interaction between platform and event, $F(2, 72) = 14.74$, $MSE = .331$, $p < .001$, $\eta p2 = .290$.

Bonferroni post hoc analysis revealed that trust ratings for the Simulator were significantly higher than trust ratings for the Wildcat ($p < .001$) for all link road events. For the combined platforms (Simulator and Wildcat) there were significantly higher trust ratings for driving on an empty road (A1i) compared to over taking a parked car (A3) ($p < .001$), and overtaking a parked car with oncoming traffic (A4) ($p < .001$). There was no significant difference between trust ratings for overtaking a parked car (A3) and overtaking a parked car with oncoming traffic (A4).

Cross Platform B3 vs B4

A factorial 2 (autonomous platform: Wildcat AV and VENTURER Simulator) x 2 (Event: B3 and B4) repeated measures ANOVA revealed a significant main effect of platform with the VENTURER Simulator rated higher, $F(1, 36) = 19.36$, $MSE = 1.14$, $p < .001$, $\eta p2 = .350$, but no statistically significant difference of event type on trust ratings, $F(1, 36) = 3.05$, $MSE = .214$, $p = .089$, $\eta p2 = .078$, and no significant interaction between platform and event, $F(1, 36) = .584$, MSE $.201$, $p = .450$, $\eta p2 = .016$.

Cross Platform B5 vs B6

A factorial (autonomous platform: Wildcat AV and VENTURER Simulator) x 2 (Event: B5 and B6) repeated measures ANOVA revealed no significant main effect of platform on trust ratings, $F(1, 36) = 2.06$, $MSE = .908$,

p = .159, η_p^2 = .054, a significant main effect of event on trust ratings, $F(1, 36) = 4.82$, MSE = .169, p = .035, η_p^2 = .118, and a non-significant interaction, $F(1, 36) = .042$, MSE = .161, p = .839, η_p^2 = .001. Bonferroni post hoc analysis revealed that trust ratings were higher for turning out right with crossing traffic (B6) compared to turning out right on an empty road (B5) regardless of the platform they were in (p = .035).

Cross Platform B1 vs B2

A factorial (autonomous platform: Wildcat AV and VENTURER Simulator) x 2 (Event: B1 and B2) repeated measures ANOVA revealed a non-significant main effect of platform on trust ratings, $F(1, 36) = .155$, MSE = .947, p = .696, $\eta p2$ = .004, a significant main effect of event on trust ratings, $F(1, 36) = 15.46$, MSE = .273, $p < .001$, $\eta p2$ = .300, and a non-significant interaction between platform and event, $F(1, 36) = .037$, MSE = .292, p = .725, $\eta p2$ = .003. Bonferroni post hoc analysis revealed that trust ratings were higher when turning in right with oncoming traffic (B2) compared to turning in right on an empty road (B1) ($p < .001$) regardless of the platform it is being performed on.

Appendix D

Table 23: Correlations between age and trust ratings across each autonomous platform and all events

Event (VENTURER Simulator)	A1i	B3	B5	A1ii	A3	B4	B6	A4	B1	B7	B2
Pearsons Correlation (r)	-.155	-.310	-.106	-.157	-.233	-.215	-.072	-.058	-.042	-.184	-.063
Significance p Value (1 tailed)	.177	.029	.263	.174	.080	.098	.334	.365	.401	.134	.353
Event (Wildcat)	**A1i**	**B3**	**B5**	**A1ii**	**A3**	**B4**	**B6**	**A4**	**B1**	**B7**	**B2**
Pearsons Correlation (r)	-.082	.161	.095	.063	.060	.118	.097	-.018	.045	-.051	.036
Significance *p* Value (1 tailed)	.295	.146	.268	.340	.349	.220	.262	.452	.385	.369	.406

Correlation strength (Cohen, 1988)		Significance p value	
Low (r = .10 - .29)		None significant (> .05)	
Medium (r = .30 - .59)		Significant (< .05)	
Strong (r = .50 – 1.0)		Significant (< .01)	

Table 24: Correlations between number of years driving and trust ratings across autonomous platforms and events

Event (VENTURER Simulator)	A1i	B3	B5	A1ii	A3	B4	B6	A4	B1	B7	B2
Pearsons Correlation (r)	-.159	-.340	-.143	-.164	-.247	-.234	-.120	-.088	-.081	-.200	-.103
Significance p Value (1 tailed)	.171	.018	.197	.163	.067	.078	.237	.301	.314	.115	.270
Event (Wildcat)	**A1i**	**B3**	**B5**	**A1ii**	**A3**	**B4**	**B6**	**A4**	**B1**	**B7**	**B2**
Pearsons Correlation (r)	-.096	.125	.071	.044	.014	.075	.063	-.050	.021	-.051	.011
Significance *p* Value (1 tailed)	.266	.207	.321	.386	.464	.312	.340	.372	.446	.370	.472

Correlation strength (Cohen, 1988)		Significance p value	
Low (r = .10 - .29)		None significant (> .05)	
Medium (r = .30 - .59)		Significant (< .05)	
Strong (r = .50 – 1.0)		Significant (< .01)	

Table 25: Correlations between time spent driving per year and trust ratings across autonomous platforms and events

Event (VENTURER Simulator)	A1i	B3	B5	A1ii	A3	B4	B6	A4	B1	B7	B2
Pearsons Correlation (r)	-0.097	-0.103	-0.074	-0.019	-0.235	-0.125	-0.084	-0.141	-0.079	-0.178	-0.098
Significance p value (1 tailed)	0.285	0.271	0.331	0.456	0.08	0.231	0.311	0.202	0.321	0.146	0.282
Event (Wildcat)	**A1i**	**B3**	**B5**	**A1ii**	**A3**	**B4**	**B6**	**A4**	**B1**	**B7**	**B2**
Pearsons Correlation (r)	-0.044	0.052	0.054	0.051	0.027	0.005	0.02	-0.098	0.09	-0.056	0.011
Significance p value (1 tailed)	0.388	0.369	0.365	0.37	0.431	0.488	0.449	0.263	0.281	0.359	0.472

Correlation strength (Cohen, 1988)		Significance p value	
Low (r = .10 - .29)		None significant (> .05)	
Medium (r = .30 - .59)		Significant (< .05)	
Strong (r = .50 – 1.0)		Significant (< .01)	

Table 26: Correlations between trust in automation and trust ratings across autonomous platforms and events

Event (VENTURER Simulator)	A1i	B3	B5	A1ii	A3	B4	B6	A4	B1	B7	B2
Pearsons Correlation (r)	.207	.287	.320	.412	.355	.276	.354	.392	.366	.309	.421
Significance p Value (1 tailed)	.106	.040	.025	.005	.014	.047	.015	.007	.012	.029	.004
Event (Wildcat)	**A1i**	**B3**	**B5**	**A1ii**	**A3**	**B4**	**B6**	**A4**	**B1**	**B7**	**B2**
Pearsons Correlation (r)	.366	.459	.491	.401	.395	.455	.390	.434	.497	.395	.424
Significance *p* Value (1 tailed)	.007	.001	.000	.003	.004	.001	.004	.001	.000	.004	.002

Correlation strength (Cohen, 1988)		Significance p value	
Low (r = .10 - .29)		None significant (> .05)	
Medium (r = .30 - .59)		Significant (< .05)	
Strong (r = .50 – 1.0)		Significant (< .01)	

Table 27: Correlations between trust in technology and trust ratings across autonomous platforms and events

Event (VENTURER Simulator)	A1i	B3	B5	A1ii	A3	B4	B6	A4	B1	B7	B2
Pearsons Correlation (r)	.257	.280	.477	.244	.454	.333	.299	.421	.541	.401	.493
Significance p Value (1 tailed)	.059	.044	.001	.070	.002	.020	.001	.004	.000	.006	.001
Event (Wildcat)	**A1i**	**B3**	**B5**	**A1ii**	**A3**	**B4**	**B6**	**A4**	**B1**	**B7**	**B2**
Pearsons Correlation (r)	.289	.374	.399	.350	.445	.423	.352	.509	.436	.495	.412
Significance *p* Value (1 tailed)	.027	.006	.003	.009	.001	.002	.009	.000	.001	.000	.002

Correlation strength (Cohen, 1988)		Significance p value	
Low (r = .10 - .29)		None significant (> .05)	
Medium (r = .30 - .59)		Significant (< .05)	
Strong (r = .50 – 1.0)		Significant (< .01)	

Carolyn Mitchell

Interim Practice Director

Carolyn.Mitchell@atkinsglobal.com