



Variable Trust Control Setting for Autonomous Vehicle Highway Navigation and Improved User Experience

James E. Pickering¹  · Jisun Kim¹ · Joshua D'Souza¹ · Keith J. Burnham¹

Received: 10 October 2024 / Accepted: 14 January 2025
© The Author(s) 2025

Abstract

This paper addresses the development of a model-based design approach to enhance the acceptance of safe manoeuvrability of autonomous vehicles (AVs) on highways. A variable trust control setting (TCS) is introduced that empowers users to 'feel in control' of the AV, potentially increasing confidence in, and acceptance of, the technology. This setting is grounded in deontological ethics and utilises virtual boundaries (VBs) to guide driving decisions, i.e., the distance between two AVs interacting with one another. The approach is simulated using a dynamic bicycle model that represents each AV, controlled through an adaptive model-predictive control (MPC) algorithm. The paper outlines the MPC approach, the dynamic bicycle model, and the associated velocity control algorithm. Metrics are introduced to quantify safety of specific AV manoeuvres during interactions with other AVs, enabling the examination of various scenarios. A novel simulation package has been developed to investigate the impact of the proposed variable TCS, focusing on how VBs and steering limitations influence the safety and comfort of AVs during overtaking manoeuvres. The findings demonstrate the effectiveness of this approach, showing that it could potentially allow users to actively manage the safety and comfort aspects of AV operation.

Keywords Control engineering · Autonomous vehicles (AVs) · Model-predictive control (MPC) · Navigation algorithms · Variable trust control setting (TCS) · Vehicle safety

Introduction

Literature Review and Outline of Problem

The UK Government has developed a 25-year strategy, detailed in [1], with the aim to foster a greener and more sustainable nation. In 2022, the Government published a document focusing on responsible innovation in autonomous vehicles (AVs), emphasising ethics, safety, and transparency in the development of navigation algorithms, see [2]. Concurrently, the UK is endeavouring to become a leader

in the development and implementation of AVs [3]. Such developments have led to the increasing adoption of cyber-physical systems (CPSs) in both AVs and broader transportation system networks. CPSs are complex, multidimensional systems that integrate sensors and actuators through wireless communication networks and computing, thereby connecting the cybernetic and physical environments [4]. This approach shows great potential for improving the safety, environmental sustainability, and efficiency of transportation systems [5].

The above introduces the possibility for a fleet of AVs within which an individual AV can 'behave' differently based on a predefined user setting. If there is low trust in AV technology, then the AV does not enter the proximity of other AVs. It is conjectured that this may increase the adoption of the technology. Due to user defined variability, the on-board navigation algorithms will need to be programmed to consider user preferences and factor-in safety and ethical considerations. This will ensure that AV navigation algorithms are justiciable, reasonable and understandable by the users. Simulation models can be used to explore novel navigation algorithms as they are safer and less expensive,

✉ James E. Pickering
j.pickerig1@aston.ac.uk

Jisun Kim
j.kim23@aston.ac.uk

Joshua D'Souza
j.dsouza@aston.ac.uk

Keith J. Burnham
k.burnham@aston.ac.uk

¹ Aston University, Birmingham B4 7ET, UK

see [6] and [7]. Determining how an AV will perform in simulation is a crucial step as it enables different navigation algorithms to be explored and any potential defects to be highlighted and considered at the design stage.

Trust is an important element for users and occupants of AVs due to the likelihood of entering vulnerable situations whereby the users entrust faith in the system [8]. Trust can have an impact on the decision of users to embrace the deployment of such automation [9]. Trust is defined as the “attitude that an agent will help achieve an individual’s goals in a situation characterised by uncertainty and vulnerability” [10]. It is also described as the “willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trusting party, irrespective of the ability to monitor and control that other party” [11]. For there to be trust in the automation, the multifaceted construct that embraces performance, process and purpose must be established. Performance is related to consistency, stability, and desirability of automation. Process indicates the operators’ knowledge of the underlying algorithms that govern behaviour of the system. Purpose represents the producers’ intention in creating the system [12]. The introduction of AVs is expected to be accompanied by user uncertainty, resulting in varying levels of trust among users. This variability in trust will raise questions about how AVs behave in specific scenarios, such as, for example, executing a lane change to exit at a junction. Consequently, it is anticipated that during the initial phase of AV deployment, user trust levels could differ significantly. This implies the possibility of mismatch between the capability of AVs and users’ expectations that are shaped based on their previous driving experiences. The mismatch could lead to reduced task effectiveness and efficiency that can be detrimental in the driving context. This warrants the need to calibrate trust in AVs [13]. Hence, it is proposed that trust should be a variable within the design of AV navigation algorithms. Trust in AVs is a crucial aspect which may limit their adoption and integration into society. Trust has been addressed as a factor that needs calibration, see [14] and [15]. The methods utilised for calibrating trust have been used for designing human-machine-interfaces (HMIs). Also, this is to enhance system transparency and to provide information on system capabilities and limitations, see [14] and [15]. However, these approaches do not consider the behaviour of AVs and their related control. Recent literature has highlighted the multidimensional nature of trust, encompassing factors such as reliability, safety, transparency, and user experience. In [16], the authors emphasise the significance of reliability in shaping trust perceptions, as any errors in performance could lead to critical safety issues, thus negatively affecting users’ initial level of trust. However, reliability in this context is contingent upon robust technical infrastructure and effective maintenance

protocols. Safety is another fundamental determinant of trust in AVs. Research in [17] indicates that users prioritise safety features and accident prevention mechanisms when assessing their trustworthiness and also when evaluating the potential AV features that would be most attractive to the public. The perceived level of safety is influenced by the vehicle’s decision-making processes, collision avoidance strategies, and response to unforeseen circumstances. Transparency emerges as another key factor in fostering trust between users and AVs. In [18], the authors state that gaining proper trust would require transparent communication of AVs’ intentions, capabilities, and limitations. Providing users with insights into the vehicle’s algorithms, sensor data interpretation, and decision rationale could enhance their understanding and acceptance of autonomous technologies.

Scenario Considered in this Research

In this research, the scenario involves the use of a two-lane highway, with vehicles on the approach to a roundabout, see Fig. 1. The highway and roundabout considered in this research is representative of a typical configuration in the UK. The highway section model excludes details such as gradient, camber and wheel slip. The highway speed limits shown in Fig. 1 are, 31.29m/s (70mph) on the main stretch of the highway, 22.35m/s (50mph) as the roundabout is approached and 13.41m/s (30mph) for the first exit of the roundabout. In Fig. 1, the initial scenario consists of two AVs, denoted AV_a and AV_b , for Autonomous Vehicle a and Autonomous Vehicle b , respectively. In this scenario, AV_b

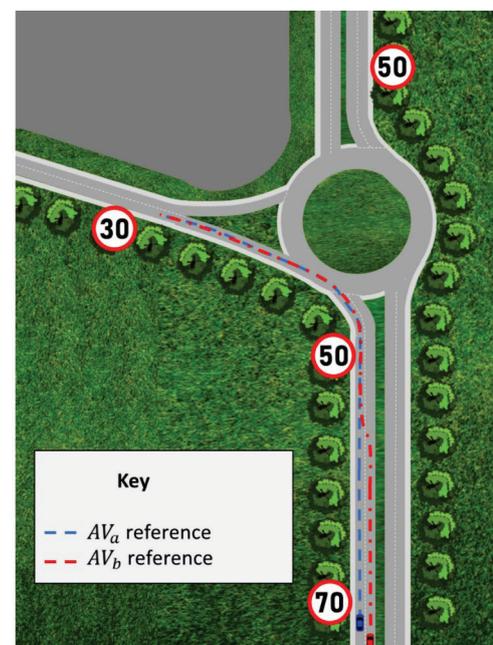


Fig. 1 Typical Highway and Roundabout Road Set-Up in the UK

is in the right-hand lane and is performing an overtaking manoeuvre on AV_a , before both AVs then plan to exit the highway via the first exit of the roundabout. The reference (waypoints) for each of the AVs are given, i.e., blue for AV_a and red for AV_b .

To provide a baseline for comparison, as detailed above, the initial problem considers the interaction of the two AVs when performing the overtaking manoeuvre on the approach to a roundabout. The aim is to develop navigation algorithms that consider deontological ethical principles, based on metrics involving safety. This is then used to develop a variable trust control setting (TCS) and to evaluate the effectiveness of the algorithm via a derived safety measure, or metric. Research presented in [19] has involved the development of a simulation package that estimates the collision probability and harm. In this research, a similar approach has been taken in the development of the proposed variable TCS. In [20], a coordinated control approach using model predictive control (MPC) for ethical manoeuvres of AVs is developed – a similar approach is adopted in this research. In addition to this, the effect of passenger comfort will also be investigated, i.e., assessing the effect that an increasing/decreasing passenger comfort has on the variable TCS in an increasing/decreasing manner.

Hypothesis, Contributions and Outline of Paper

Based on the premise alluded to in Section [Literature Review and Outline of Problem](#), the paper is motivated by the following hypothesis: Incorporation of user defined TCSs into the navigation algorithms of AVs can significantly enhance user trust and adoption rates of the technology, by dynamically adjusting the vehicle's operational behaviour according to individual comfort and safety preferences. These adjustments will lead to measurable improvements in user satisfaction, and overall trust in AV technology. This hypothesis is grounded in the idea that personalising AV behaviour based on user preferences and trust levels, as determined through user settings, will directly impact the perception and effectiveness of AVs in real-world scenarios. By aligning the operational characteristics of AVs with user expectations and ethical considerations, it is anticipated that users will be more inclined to trust and adopt this technology, thus enhancing the overall safety and efficacy of autonomous transportation systems.

The main contributions of this article are as follows:

- i. Introduction of a variable TCS that allows users to dynamically control the virtual boundaries (VBs) and maximum steering rate of AVs. This feature uniquely enables users to adjust how the AV operates based
- ii. Empirical investigation into how variations in the trust setting affect risk metrics such as the relative collision metric (RCM) and duration of risk imposed (DRI). This study is among the first to quantitatively link trust variation with measurable safety outcomes in AVs.

The article is organised as follows. The simulation model is formulated in Section [Simulation Model](#), with the simulation results presented and discussed in Section [Results and Discussion](#). Conclusions of the research as well as an outlook are presented in Section [Conclusions and Future Outlook](#)

Simulation Model

A simulation model considering AV_a and AV_b navigation is developed using a CPS approach, see Fig. 2a, with further details given in the following sub-sections.

The simulation model is based on the following initial assumptions:

- i. AVs can communicate (e.g., known position and velocity) using vehicle-to-vehicle (V2V) communication.
- ii. AVs are equipped with onboard sensors (e.g., cameras) to allow other AVs in each localised vicinity to be identified.
- iii. AVs adhere to maximum speed limits of $26.82m/s$ ($60mph$) in the left-hand lane and $31.29m/s$ ($70mph$) in the right-hand (overtaking) lane.

Rule-Based Trajectory Algorithms and Operation of Virtual Boundaries

This section presents the concept of VBs and outlines their application in the design of AV deontological navigation algorithms with a variable TCS, previously developed in [21]. These algorithms expand upon the framework previously established in [22].

Considering the scenario introduced in Section [Scenario Considered in this Research](#), the navigation algorithms are designed for the following situations:

1. Highway lane behaviour (longitudinal control)
2. Lane changing behaviour (lateral control)
3. Entering a decision point (roundabout junction)

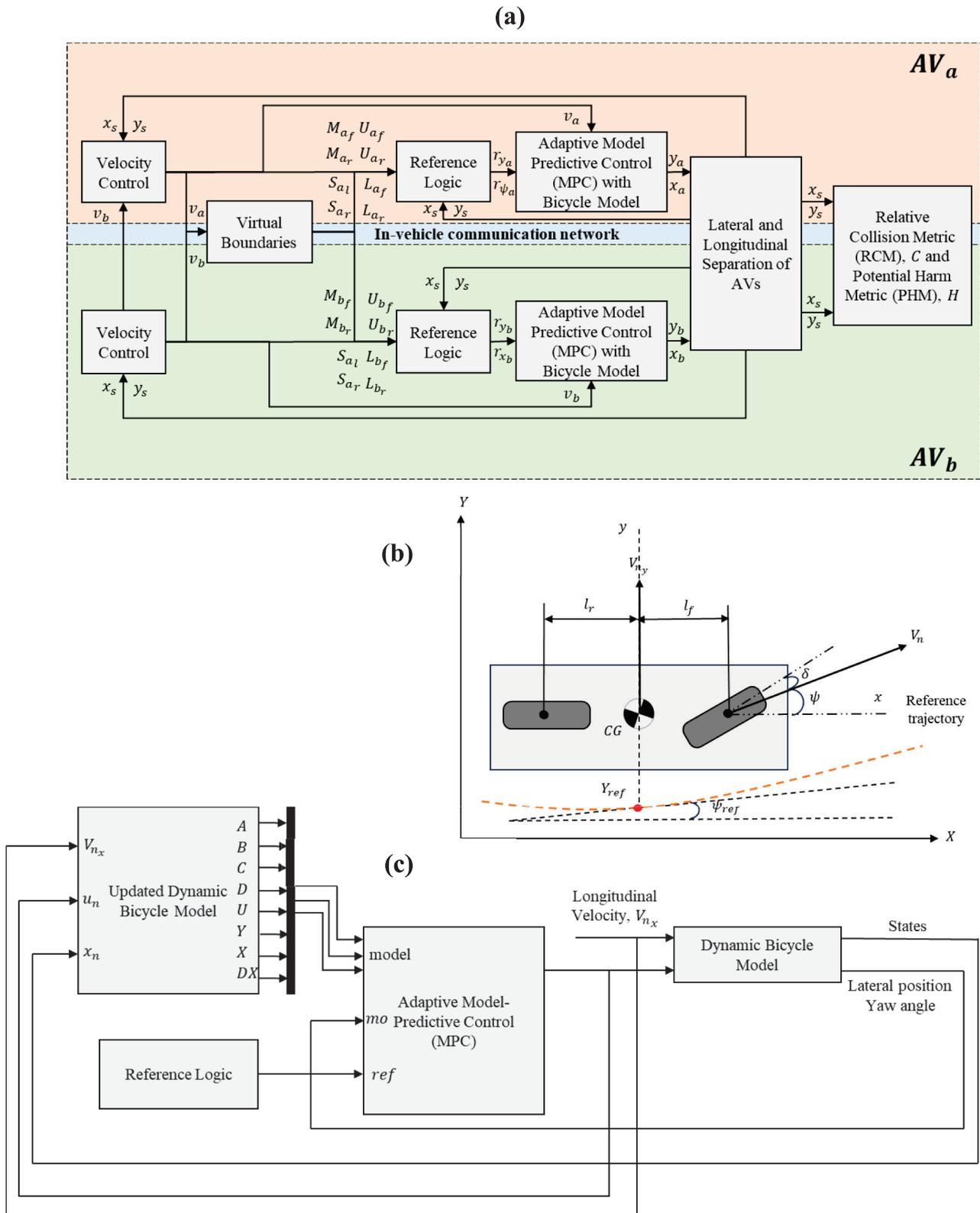


Fig. 2 Architecture Framework of the Simulation Model (a), Dynamic Bicycle Model (b) and Adaptive Model-Predictive Control (MPC) (c)

Virtual Boundaries

The VBs are used to provide a framework for the safe operation of AVs, whilst considering the navigation algorithm design. The VBs represent distances set-up from the centre of each AV, denoted C_n (where the subscript n denotes the Vehicle ID, i.e., Vehicle a or b), to quantify the potential risks, see Fig. 3a.

The VBs to the side of an AV are set-up to ensure that AVs travel in the centre of a given lane, see Fig. 3. The VBs to the side of a vehicle are given by:

- Side denoted S_{nl} and S_{nr} for the left and right (where n is the Vehicle ID) of a given vehicle, respectively. Vehicles should avoid entering each other’s ‘Side’ VB. In Fig. 3a,

the S_{ar} distance, denoted d , for AV_a is given, denoted $d_{S_{ar}}$. The value given for both S_{nl} and S_{nr} is 2 m, i.e., half the width of the lane.

The VBs to the front and rear of a vehicle, which relate to the speed dependent braking distances, are given by the three-tiered layers:

- Lower Limit (LL) denoted L_{nf} and L_{nr} for the front and rear of a given vehicle, respectively. The LL is a VB that must be avoided by any AV – entering this will possibly involve a collision. Whilst unavoidable collisions are not considered here, research in this area may be found in [23] and [24]. In Fig. 3, the L_{af} distance, denoted d , for AV_a is denoted $d_{L_{af}}$.
- Middle Limit (ML) denoted M_{nf} and M_{nr} for the front and rear of a given vehicle, respectively. The preferred operation of an AV is to avoid entering this VB. In Fig. 3, the M_{ar} distance, denoted d , for AV_a is denoted $d_{M_{ar}}$.
- Upper Limit (UL) denoted U_{nf} and U_{nr} for the front and rear of a given vehicle, respectively. The UL is a VB that other AVs can manoeuvre into safely. However, the AVs should aim to leave as soon as possible. The UL VB considers a safety margin, i.e., sufficient braking distance that will maintain satisfactory comfort levels to the occupants on-board a given AV, and other AVs involved. In Fig. 3, the U_{ar} distance, denoted d , for AV_a is denoted $d_{U_{ar}}$.

The initial VBs for this research are given in Table 1. The VBs are based on the stopping distances between two vehicles (these being functions of velocity), recommended by the Department for Transport and the Driver and Vehicle Standards Agency (DVSA) in the UK, see [25].

As the VB distances are set-up from the centre of each vehicle, the front and rear lengths of the vehicle, denoted D , are defined, see Table 1. In this work, D is 1.4 m, i.e., so that the overall length of the vehicle is 2.8 m. In the simulation model, the UL VB is programmed to emulate the stopping

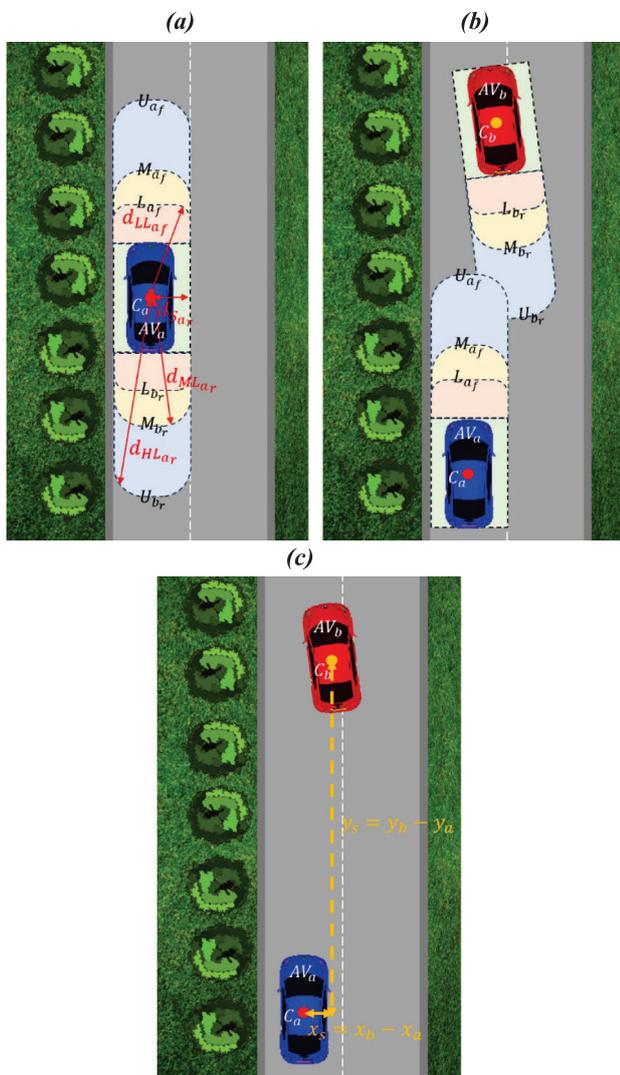


Fig. 3 Virtual Boundaries (VBs) for Highway Driving (a), VB Overlap for Deontological Algorithm Design when Overtaking (b) and Details of how the Separation Distances are Determined (c)

Table 1 Length of Virtual Boundaries (VBs) as Functions of the Autonomous Vehicle (AV) Velocity

Velocity [mph]	Virtual Boundaries		
	Lower Limit (LL), L_{nf} and L_{nr} [m]	Medium Limit (ML) M_{nf} and M_{nr} [m]	Upper Limit (UL) U_{nf} and U_{nr} [m]
0	$D + 0.00$	$D + 0.00$	$D + 0.00$
10	$D + 1.14$	$D + 2.29$	$D + 4.58$
30	$D + 5.72$	$D + 11.43$	$D + 22.86$
50	$D + 13.34$	$D + 26.67$	$D + 53.34$
70	$D + 24.00$	$D + 48.01$	$D + 96.01$

distances, with the ML and LL VBs initially being 50% and 25% of the stopping distances, respectively, see Table 1. These VB lengths are further investigated in Section [Results and Discussion](#) via an initial sensitivity analysis.

Deontological Ethics

In deontological ethics, an action is morally good if it follows a predetermined set of rules or moral values, see [26]. Considering Fig. 1, the AV being overtaken is subject to the actions and/or consequences of the AV performing the overtaking manoeuvre, with both AVs adopting the driving rules (DRs) as follows:

1. Highway lane behaviour: All AVs on the highway should maintain an acceptable distance to the AV ahead, behind and to the sides.
2. Lane changing: Lane changing of a given AV on a highway should be undertaken given there is acceptable space to minimise disruption to other AVs, and to avoid accidents.
3. Entering a decision point: When entering a decision point, an AV should only proceed forward if there are acceptable distances to other AVs ahead or to the sides.

In terms of the VBs, the ‘acceptable distance’ is based on the UL of the two AVs, see Fig. 3b. The DRs have the same maxim applied (a maxim is a rule that connects an action to the reasons for the actions), hence: act in such a way as to minimise impact to other road users (i.e., not effecting another AVs journey time, respecting human lives, and not causing damage to other AVs) and the environment.

Variable Trust Control (VTC) Setting

The proposed variable TCS, which is demonstrated in simulation in Section [Results and Discussion](#), is based on the VB lengths given in Table 1, and the maximum steering rate, see Section [Sensitivity Analysis involving Maximum Steering Rate](#).

It is proposed that an operator/user of the AV has an option to select from a range of TCSs. In this paper, five trust settings are considered. These are no trust, little trust, medium trust, medium-to-high-trust and complete trust, see Table 2. In this work it is assumed that the setting can be specified upon entering the AV or adjusted during usage, i.e., the setting can be adjusted if the users trust level of the AV technology changes. A setting of 0% implies no trust in the AV technology and 100% implies complete trust in the AV technology. In this work, the level of trust is associated with a perceived risk, as was considered in [14]. Essentially,

Table 2 Linguistic Terms and how these Correspond to the Variable Trust Control Settings (TCS), T

Trust [%]	Linguistic terms	Variable TCS, T [1.20–0.80]
0	No trust	1.20
25	Little trust	1.10
50	Medium trust	1.00
75	Medium to high trust	0.90
100	Complete trust	0.80

varying the level of trust in the AV will result in varying the distances of the VBs or the maximum steering rate.

The variable TCS values, denoted T , for the five settings are given in Table 2. These settings correspond to the user preferences (i.e., no trust, little trust, medium trust, medium-to-high-trust and complete trust). The associated values of variable TCSs, in the range 1.20 to 0.80 are used to adjust the VB lengths, with a medium trust setting of 1.00 corresponding to the nominal VB lengths given in Table 1 implying no adjustment. In the case of complete trust the VB lengths are reduced by a factor of 0.20 and in the case of no trust the VB lengths are increased by a factor of 0.20.

The above initial variable TCSs which affect VB lengths and steering rates are investigated further to assess vehicle comfort and safety via a sensitivity analysis in Section [Results and Discussion](#).

Control Strategy

This section introduces the dynamic bicycle model for representing AVs within a global coordinate system, focusing on controlling the steering angle to follow a reference trajectory. It also provides an outline of an adaptive model predictive control (MPC) algorithm that optimises real-time trajectories and a velocity control algorithm that regulates vehicle speeds based on interactions, ensuring safe navigation. In addition, the section introduces metrics for assessing collision risk and potential harm.

Dynamic Bicycle Model

A dynamic bicycle model is used to represent each of the AVs, within a global position given by X and Y , see Fig. 2b. The vectors from the bicycle model, denoted v_x and v_y , represent the longitudinal and lateral velocity, respectively. The vehicle’s longitudinal acceleration is assumed to be in a quasi-steady state, i.e., gradual longitudinal acceleration changes resulting in the approximation of steady state responses at each instantaneous state [27]. The aim of the dynamic bicycle model is to follow a reference trajectory denoted Y_{ref} , given by the red

circle on the orange line in Fig. 2b. The reference trajectory is obtained by controlling the steering angle, denoted δ . The yaw angle output, denoted ψ , and reference yaw angle, denoted ψ_{ref} , are determined with reference to the horizontal axis. On this basis, the dynamic model is initially adopted from [28] and is given by:

$$\frac{d}{dt} \begin{bmatrix} \dot{y} \\ \dot{\psi} \\ \dot{\psi} \\ y \end{bmatrix} = \begin{bmatrix} -\frac{2C_f+2C_r}{mV_x} & 0 & -V_x - \frac{2C_f l_f - 2C_r l_r}{mV_x} & 0 \\ 0 & 0 & 1 & 0 \\ \frac{2l_f C_f - 2l_r C_r}{I_z V_x} & 0 & -\frac{2l_f^2 C_f + 2l_r^2 C_r}{I_z V_x} & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y} \\ \dot{\psi} \\ \dot{\psi} \\ y \end{bmatrix} + \begin{bmatrix} \frac{2C_f}{m} \\ 0 \\ \frac{2l_f C_f}{I_z} \\ 0 \end{bmatrix} \delta \tag{1}$$

where V_x denotes the longitudinal velocity at the vehicle’s centre of gravity (CoG), m denotes the vehicle mass, I_z denotes the vehicle’s yaw moment of inertia, l_f denotes the longitudinal distance from the CoG to the front tyres, l_r denotes the longitudinal distance from the CoG to the rear tyres, C_f denotes the cornering stiffness of the front tyres, C_r denotes the cornering stiffness of the rear tyres, δ denotes the vehicle’s input front steering angle and y denotes the vehicle’s lateral position output. Simulation model parameters are given in Table 3, including maximum acceleration and deceleration values, denoted a_{max} and a_{min} , respectively.

Considering Eq. (1), the state space model does not account for the global lateral position as a state variable, denoted Y . Therefore, this is determined using the following [29]:

$$\dot{Y} = \dot{x} \sin \psi + \dot{y} \cos \psi \tag{2}$$

Using small angle approximations i.e., $\sin \psi \approx \psi$ and $\cos \psi \approx 1$, and considering the notation $V_x = \dot{x}$, Eq. (2) can be rewritten as follows:

$$\dot{Y} = V_x \psi + \dot{y} \tag{3}$$

The state space model for the 2 degrees of freedom is then given by:

$$\frac{d}{dt} \begin{bmatrix} \dot{y} \\ \dot{\psi} \\ \dot{\psi} \\ Y \end{bmatrix} = \begin{bmatrix} -\frac{2C_f+2C_r}{mV_x} & 0 & -V_x - \frac{2C_f l_f - 2C_r l_r}{mV_x} & 0 \\ 0 & 0 & 1 & 0 \\ \frac{2l_f C_f - 2l_r C_r}{I_z V_x} & 0 & -\frac{2l_f^2 C_f + 2l_r^2 C_r}{I_z V_x} & 0 \\ 1 & V_x & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y} \\ \dot{\psi} \\ \dot{\psi} \\ Y \end{bmatrix} + \begin{bmatrix} \frac{2C_f}{m} \\ 0 \\ \frac{2l_f C_f}{I_z} \\ 0 \end{bmatrix} \delta \tag{4}$$

Using the dynamic bicycle model, \dot{X} is derived from the following equation of motion, see [29]:

$$\dot{X} = \dot{x} \cos \psi - \dot{y} \sin \psi \tag{5}$$

Using small angle approximations, and considering the notation $V_x = \dot{x}$, Eq. (5) can be rewritten as follows:

$$\dot{X} = V_x - \dot{y} \psi \tag{6}$$

Model-Predictive Control

An adaptive model-predictive control (MPC) algorithm is developed for each AV, see Fig. 2c. MPC is a predictive control method that utilises a dynamic model of the system to optimise future trajectories, enabling the vehicle to make informed decisions in real-time. MPC is able to take into account constraints on control action and system dynamics, making it particularly well-suited for the inherently dynamic nature of AV navigation, see [30] and [31].

By predicting future system states using a mathematical model and optimising control inputs to minimise deviations from a given desired reference trajectory, while respecting constraints, MPC is able to steer the AV towards the desired trajectory. At each time-step, an optimisation problem is solved to determine the appropriate control action, with the process continuously repeating using updated state information. This continuous prediction, feedback and adaptation ensures that the AV effectively follows the desired path.

The MPC algorithm was tuned manually to produce an acceptable value for the integral of absolute error (IAE) by adjusting the control and prediction horizons to be 3 and 20, respectively. The bicycle model adopted here for each of the vehicles is constructed based on the references (waypoints) given in Fig. 1. The constraints are: maximum steering of 30 degrees and for driving comfort the rate of change of steering angle is limited to 15 degrees/second. Further details of the MPC algorithm used in this research can be found in [32].

Velocity Control

In a scenario as in Fig. 1, the velocities of AV_a and AV_b would be influenced by the presence of other vehicles, therefore velocity control is required. For AV_a , the aim of the deontological algorithm (see Section **Deontological Ethics**) is to maintain a velocity of 26.82 m/s (60 mph)

Table 3 Simulation Model Parameter Values

Model parameter	Value [units]
C_f	19,000 [N/rad]
C_r	22,000 [N/rad]
m_n	2400 [kg]
v_n	0 to 31.29 [m/s]
l_f	1.2 [m]
l_r	1.6 [m]
I_z	1575 [kgm ²]
r	140 [Ns/m]
a_{max}	10[m/s ²]
a_{min}	-6 [m/s ²]

and to decelerate to 13.41 m/s (30 mph) when 274 m away from the roundabout. If at any point in time, the VB UL is not respected (i.e., from another vehicle overtaking), AV_a is programmed to decelerate to a lower velocity of 24.59 m/s (55 mph), see Fig. 4a. This velocity is held until the VB UL is again respected by other vehicles. Regarding the velocity control of AV_b with the deontological algorithm, an initial velocity of 31.29 m/s(70 mph) is held while overtaking AV_a in Lane 1 with the VB UL being respected, see Fig. 4b. Once this condition is met, AV_b will then decelerate to the driving speed of 26.82 m/s (60 mph) in Lane 1. In a similar manner to AV_a , AV_b will decelerate to 13.41 m/s(30 mph) as the roundabout is approached whilst constantly respecting the VBs.

Reference Logic

From Fig. 1, AV_a must remain in Lane 1 for taking the 1st exit at the roundabout. Similarly, AV_b also aims to take the 1st exit. Hence, decision making algorithms are required to account for the lane changing manoeuvre of AV_b into Lane 1. Because AV_b is programmed to follow the DRs, it is permitted to change lane only if the ML VBs of the two AVs are respected, see Fig. 5.

Relative Collision and Potential Harm Metrics

The lateral and longitudinal separation distances between two vehicles, as illustrated in Fig. 3c, are given, respectively, by:

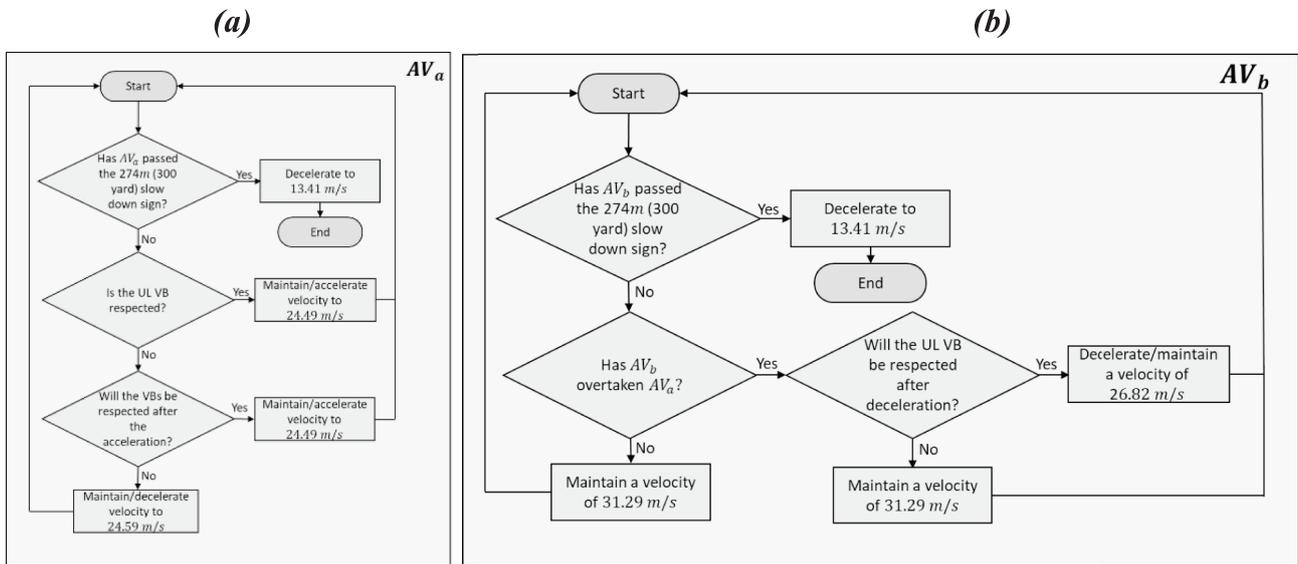


Fig. 4 Velocity Control Algorithm for AV_a with a Deontological Navigation Algorithm (a), AV_b with a Deontological Navigation Algorithm (b)

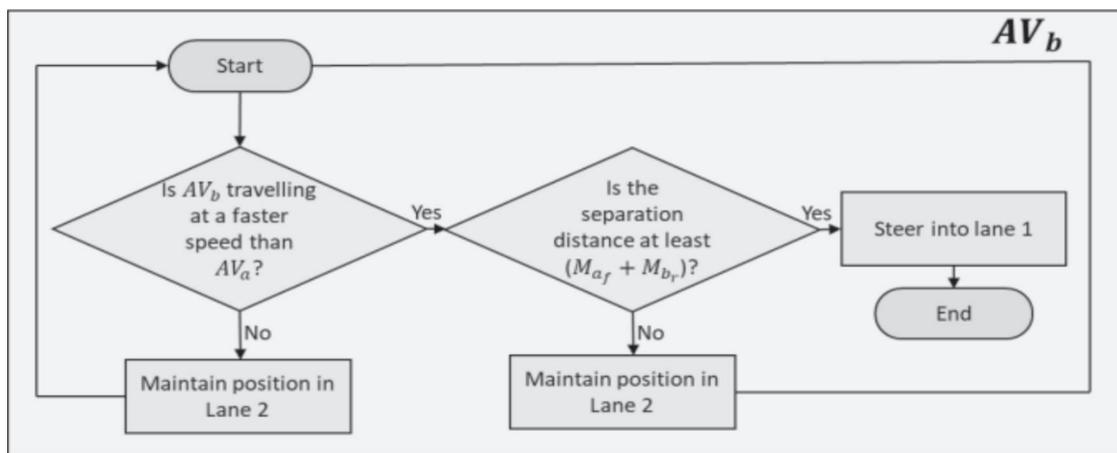


Fig. 5 Reference Logic for AV_b

$$y_s = y_b - y_a \tag{7}$$

$$x_s = x_b - x_a \tag{8}$$

where y_s and x_s denote the lateral and longitudinal separation distances between the vehicles, y_a and x_a denote the lateral and longitudinal position of AV_a at time t , and y_b and x_b denote the lateral and longitudinal positions of AV_b , respectively.

Equations (7) and (8) are used with the VBs to quantify the risk presented between the two vehicles. The RCM, denoted C , provides an approach to quantifying risk and is given by:

$$C = \left[1 - \frac{x_s}{U_{nf} + U_{nr}} \right] \left[1 - \frac{y_s}{S_{nl} + S_{nr}} \right] \tag{9}$$

Equation (9) outputs values between 0 and 1. A value 0 indicates zero risk, (no overlap of the VBs) hence, no risk of a collision. A value tending towards 1 indicates an increasing risk. If a value of 1 is recorded for the RCM, this is likely to result in a collision.

The potential harm metric (PHM), denoted H , provides an approach to quantifying risk and is given by:

$$H = \pm \left[\frac{\Delta v_n}{\Delta v_{max_n}} \right] \tag{10}$$

where Δv_n denotes the difference in the velocities of two vehicles, Δv_{max_n} denotes the maximum velocity difference between two vehicles and \pm is used to denote direction, i.e., positive for vehicles moving towards each other, negative for vehicles moving apart and a value of zero for vehicles travelling at the same velocities (i.e., identical magnitude and direction). The difference in vehicle velocities is commonly used for estimating harm, see [33] and [34]. Equation (10) provides a numerical value to determine risk, i.e., between 0 (no risk) and 1 (high risk).

Results and Discussion

The results are now detailed for the deontological algorithm in Section [Deontological Algorithm \(Nominal case with TCS, \)](#), as well as two initial sensitivity study analyses in Section [Sensitivity Analysis involving Variable Trust Control Setting \(TCS\)](#) for a variable TCS, and in Section [Sensitivity Analysis involving Maximum Steering Rate](#) for a maximum steering rate.

Deontological Algorithm (Nominal case with TCS, $T = 1.00$)

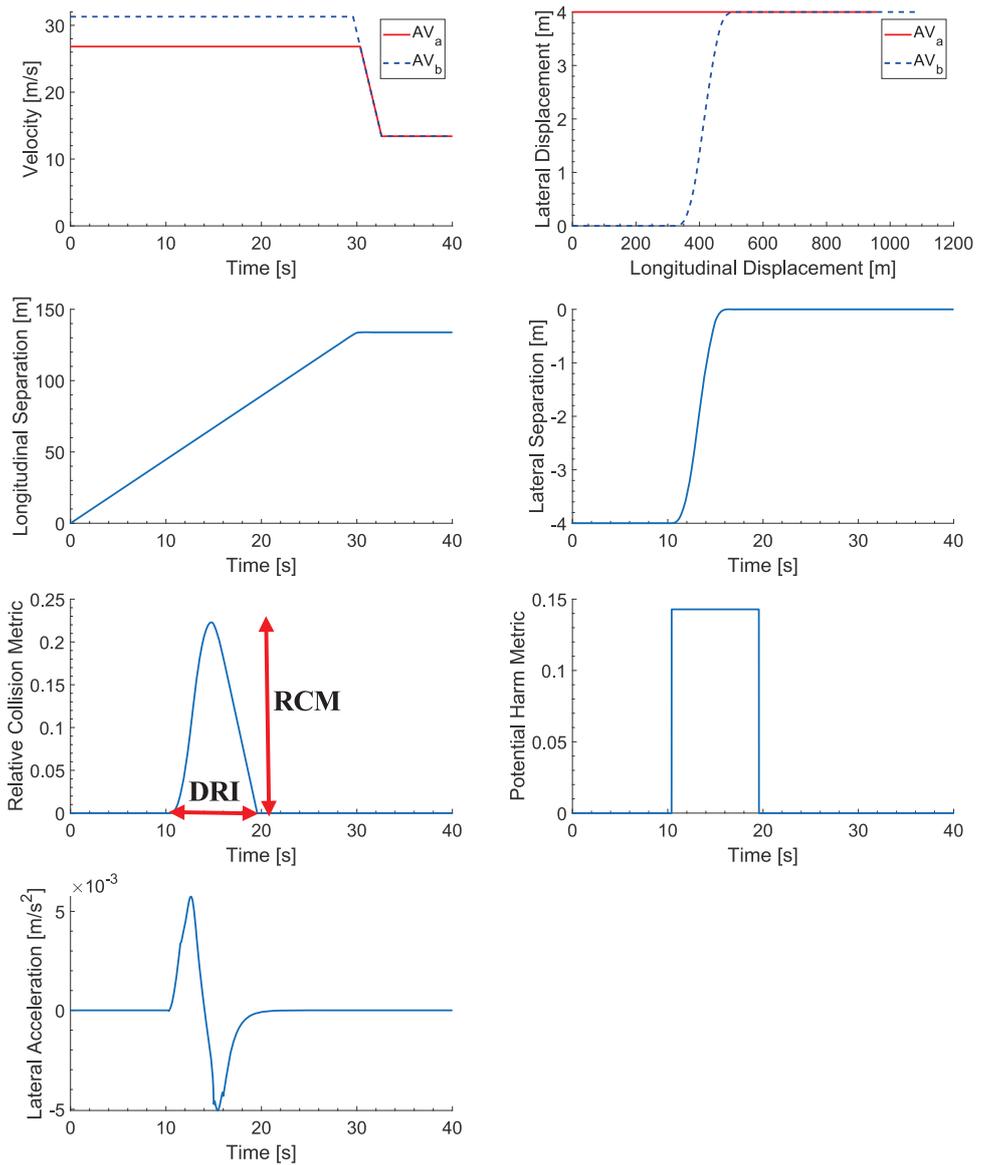
Simulation results for the deontological algorithm (as detailed in Section [Deontological Ethics](#)) corresponding to the nominal case of medium trust, with a TCS value of unity, are provided in Fig. 6. The graphical output of velocity versus time for the two AVs is given in Fig. 6 (Upper-Left). Recall from Fig. 1, AV_a is travelling in the left-hand-lane, initially at the maximum velocity (i.e., 60mph or 26.82m/s). AV_b is travelling at the maximum speed in the right-hand-lane, i.e., 31.29m/s (70mph). At approximately 30s, AV_b performs an overtaking manoeuvre on AV_a , with both AVs then travelling at their maximum speeds, i.e., in the left-hand-lane, 26.82 m/s and in the right-hand lane 31.29m/s. The speed of both AVs (with AV_b now being ahead of AV_a) then decreases as the speed limit approaching the roundabout sets-in, i.e., 13.41m/s (30mph). The lateral displacement versus longitudinal displacement graphical output for AV_a and AV_b is initially given, see Fig. 6 (Upper-Right). Note that the overtaking manoeuvre takes place at around 350m. The graphical outputs for the longitudinal and lateral separation of the two AVs are given in Fig. 6 (Upper-Middle-Left and Upper-Middle-Right, respectively). Note that the longitudinal separation distance remains constant once AV_b has completed the overtaking manoeuvre on AV_a . The graphical outputs for RCM and PHM for the two AVs are given in Fig. 6 (Lower-Middle-Left and Lower-Middle-Right, respectively). For the RCM, the peak RCM and DRI are labelled, with values of 0.22 and 9.15s being captured, respectively (see Table 4). Both values are relatively low as the rules within the deontological algorithm were respected. The RCM is high for a short duration (approximately 9.15s) due to the velocity control applied to AV_b , resulting in AV_b travelling at a velocity of 70mph when overtaking AV_a , and returning to the desired velocity of 60mph once AV_a has been overtaken. The PHM during the same time-period is also relatively low, with a positive value of 0.14, see Table 4. Finally, the lateral acceleration of the overtaking vehicle (i.e., AV_b) is given, see Fig. 6 (Lower-Left), with the values for the peak and positive mean acceleration recorded in Table 4. Note that a first order filter was applied to the acceleration versus time data with a filter transfer function of:

$$G(s) = \frac{1}{0.1s + 1} \tag{11}$$

Sensitivity Analysis involving Variable Trust Control Setting (TCS)

The results of the sensitivity analysis for the variable TCS, as detailed in Section [Variable Trust Control Setting](#), are

Fig. 6 Deontological Algorithm Design Results



presented in Fig. 7. The model has been used to examine the impact of variable TCSs, involving adjustments to the ML length of the deontological AV’s VBs, as shown in Table 2 of Section Variable Trust Control Setting. Figure 7

Table 4 Deontological Algorithm Design Results (Nominal Case, where $T = 1.00$)

	Maximum Relative Collision Metric (RCM), C	Maximum Potential Harm Metric (PHM), H	Duration of Risk Imposed (DRI) [Seconds]	Peak Acceleration a_{peak} [m/s ²]	Positive Mean Acceleration a_{mean_p} [m/s ²]
$T = 1.00$ (Medium Trust)	0.22	0.14	9.15	0.0056	0.0037

displays the RCM outcomes across various trust settings, with the PHM and DRI values being given in Table 5. As expected, if the level of trust increases, resulting in lower TCSs, hence smaller ML VBs, then the occupants’ risk increases. Conversely, lower values of trust, leading to higher TCSs, leading to a larger ML VBs, then the occupants risk decreases. Transitioning the trust setting from 'no trust' to 'complete trust' leads to a 19% increase in the RCM and a 22% increase in the DRI. Thus, the 'complete trust' setting exposes occupants to higher levels of risk over a longer duration. Hence, the more confident the user is with the technology, the more risks the user is prepared to take. This will ironically lead to a greater number of risk takers as the technology matures, with this leading to reduced safety. As anticipated, the peak PHM remains constant, due to the AVs' effective velocity control system.

Fig. 7 Sensitivity Analysis Results for the Variable Trust Control Setting (TCS)

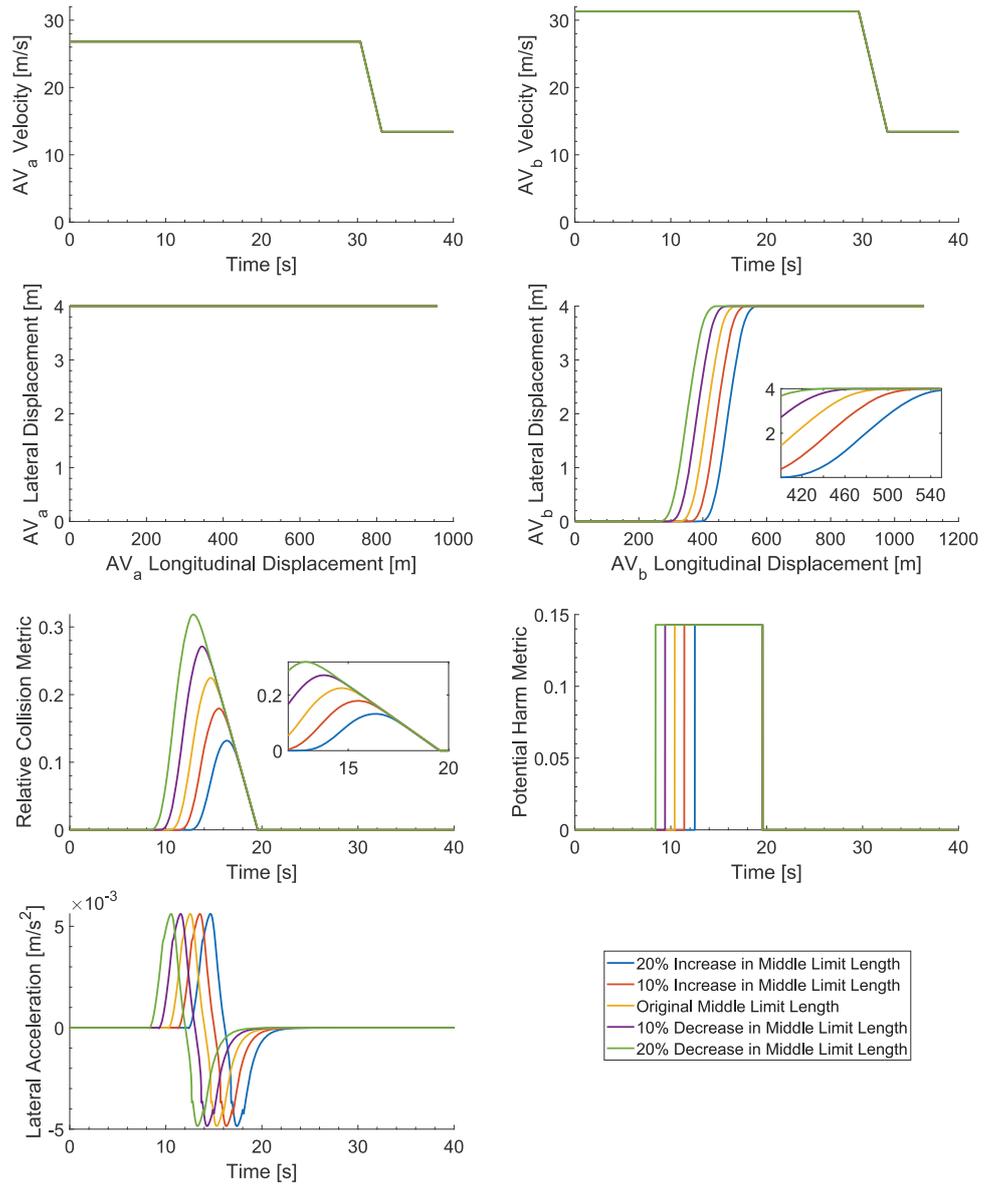


Table 5 Sensitivity Analysis Results for the Variable Trust Control Setting (TCS)

	Maximum Relative Collision Metric (RCM), C	Maximum Potential Harm Metric (PHM), H	Duration of Risk Imposed (DRI) [Seconds]	Peak Acceleration a_{peak} [m/s ²]	Positive Mean Acceleration a_{mean_p} [m/s ²]
1.20 (No Trust)	0.13	0.14	7.05	0.0056	0.0037
1.10 (Little Trust)	0.18	0.14	8.15	0.0056	0.0037
1.00 (Medium Trust)	0.22	0.14	9.15	0.0056	0.0037
0.90 (Medium-to-High-Trust)	0.27	0.14	10.15	0.0056	0.0037
0.80 (Complete Trust)	0.32	0.14	11.15	0.0056	0.0037

This ensures compliance with velocity limits during overtaking and maintaining lane discipline in both the left-hand and right-hand lanes. This result indicates that the peak acceleration values remain constant. Note that Row 3 in Table 5 corresponds to the results presented in Table 4.

Sensitivity Analysis Involving Maximum Steering Rate

The results of the sensitivity analysis for the maximum steering rate of change, set against a fixed ML VB (i.e.,

$T = 1.00$, medium trust), are presented in Fig. 8. The simulation model is now utilised to explore the user experience under AV operations that employ deontological algorithms with a variable TCS related to the maximum steering rate. Steering rates ranging from 0.10 rad/s and 0.30 rad/s are applied in 0.05 rad/s increments, representing levels from no trust to complete trust, respectively. Key findings relating to the RCM and DRI from Fig. 8 are presented in Table 6.

As the variable TCS decreases, indicating maximum trust and thus the highest steering rate, the risk imposed, as expected, indicates an increased risk to the occupants.

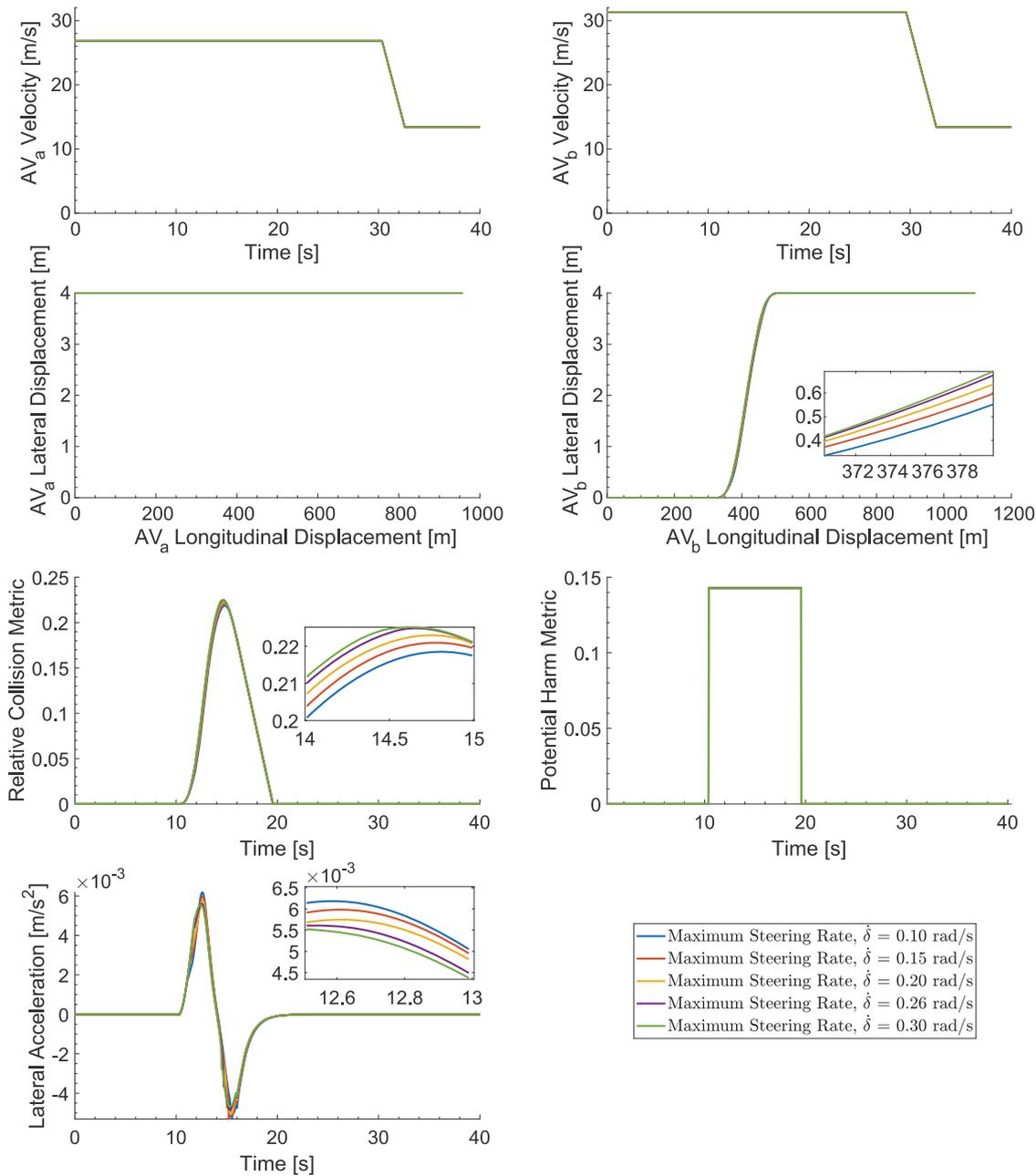


Fig. 8 Sensitivity Analysis Results for Maximum Steering Rate

Table 6 Sensitivity Analysis Results for Maximum Steering Rate

	Maximum Relative Collision Metric (RCM), C	Maximum Potential Harm Metric (PHM), H	Duration of Risk Imposed (DRI) [Seconds]	Peak Acceleration a_{peak} [m/s]	Positive Mean Acceleration a_{mean_p} [m/s]
± 0.10 rad/s (No Trust)	0.22	0.14	9.15	0.0062	0.0036
± 0.15 rad/s (Little Trust)	0.22	0.14	9.15	0.0060	0.0036
± 0.20 rad/s (Medium Trust)	0.22	0.14	9.15	0.0058	0.0037
± 0.26 rad/s (Medium-to-High-Trust)	0.22	0.14	9.15	0.0056	0.0037
± 0.30 rad/s (Complete Trust)	0.23	0.14	9.15	0.0055	0.0038

Conversely, as the TCS increases, indicating minimum trust with the lowest steering rate, the risk imposed decreases. Adjusting the trust setting from 'no trust' to 'complete trust' leads to a slight increase in the RCM, while the DRI remains unchanged. As expected, the peak PHM remains constant, owing to the effectiveness of the AV's velocity control system; both AVs adhere to speed restrictions during overtaking and maintain their position within their designated lanes. However, it should be noted that the peak acceleration, denoted as a_{peak} , diminishes with increasing trust. The positive mean acceleration, denoted a_{mean_p} , is also analysed, showing an increase as the trust level increases. This indicates that higher TCSs result in lower peak accelerations, but these accelerations prevail over a longer duration.

Conclusions and Future Outlook

This paper has presented a novel approach to enhancing the safe and ethical manoeuvrability of autonomous vehicles (AVs) on highways. It has described the integration of driving rules based on deontological ethical principles with the application of virtual boundaries (VBs) for AVs. An adaptive model-predictive control (MPC) algorithm, in conjunction with a dynamic bicycle model, has been utilised to model each AV and to facilitate the tracking of desired trajectories. The study has introduced a methodology for continuous risk evaluation, grounded on the interaction between two AVs. The findings demonstrate how the risk imposed can be effectively integrated into a novel user specified variable trust control setting (TCS) for use on AVs. It can be observed that increased user trust from 0 to 100%, thereby reducing the VB lengths of the AVs, results in an increased value of a relative collision metric (RCM) and an extended duration of risk imposed (DRI).

This observation suggests that the variable TCS could empower AV users to feel more in control, enabling them to engage more fully with the technology. This engagement is anticipated to foster greater confidence and better acceptance of the technology as it continues to evolve.

Adoption of a user-centric scheme for AVs, such as the variable TCS proposed here, is considered to represent a stepping-stone towards enhancing the overall comfort and perceived safety of AVs during their early release phase. However, as AV users become more confident with the technology, they are likely to become less risk averse and it is anticipated that the variable TCSs will allow closer running of AVs as the technology matures.

While the results are promising, they also indicate significant opportunities for further research. Future work could include extending the simulation environment to incorporate dynamically changing conditions, which would offer a more realistic approach to modelling. Additionally, employing high-fidelity proprietary tools, such as, e.g., CarMaker, could facilitate the implementation of developed algorithms in real-time scenarios. This would allow for more comprehensive testing and refinement of the control strategies under a wider variation of realistic driving conditions, thus further advancing the field of AV technology research.

Author Contributions Dr. James E. Pickering spearheaded the research project, focusing on its innovative aspects. Dr. Jisun Kim played a crucial role in exploring the trust elements within the study, bringing valuable insights to the forefront. Joshua D'Souza conducted the simulation work under the supervision of Prof Keith J. Burnham and Dr James E. Pickering. Prof Keith J. Burnham provided essential guidance and oversight, helping to define the scope of the research.

Funding This work was supported by the EPSRC Doctoral Training Program funding (Grant number is EP/W524566/1).

Data Availability Data can be provided upon request.

Declarations

Conflict of Interest There are no competing interests.

Research Involving Human and/or Animals Not applicable.

Informed Consent Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Environmental Improvement Plan 2023. Department for Environment, Food & Rural Affairs, UK. 2023.
- Responsible Innovation in Self-Driving Vehicles. Centre for Data Ethics and Innovation, UK. 2022.
- Dft. New Laws to Safely Roll out Self-Driving Vehicles Across British Roads, 2023. Department for Transport, The Rt Hon Jesse Norman MP, and The Rt Hon Mark Harper MP. 2023.
- Liu Y, Peng Y, Wang B, Yao S, Liu Z. Review on cyber-physical systems. *IEEE/CAA J Autom Sin*. 2017;4(1):27–40.
- Wang X, Zheng X, Chen W, Wang FY. Visual human–computer interactions for intelligent vehicles and intelligent transportation systems: the state of the art and future directions. *IEEE Trans Syst Man Cybern Syst*. 2020;51(1):253–65.
- Koopman P, Wagner M. Challenges in autonomous vehicle testing and validation. *SAE Int J Transport Saf*. 2016;4(1):15–24.
- Kalra N, Paddock SM. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transport Res Part A Policy Pract*. 2016;94:182–93.
- Körber M, Baseler E, Bengler K. Introduction matters: Manipulating trust in automation and reliance in automated driving. *Appl Ergon*. 2018;66:18–31.
- Lee JD, Moray N. Trust, self-confidence, and operators' adaptation to automation. *Int J Hum Comput Stud*. 1994;40(1):153–84.
- Lee JD, See KA. Trust in automation: Designing for appropriate reliance. *Hum Factors*. 2004;46(1):50–80.C.
- Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. *Acad Manag Rev*. 1995;20(3):709–34.
- Lee J, Moray N. Trust, control strategies and allocation of function in human- machine systems. *Ergonomics*. 1992;35(10):1243–70.
- Lee SC, Ji YG. Calibration of trust in autonomous vehicle. In: *Human-automation interaction: transportation*. Cham: Springer International Publishing; 2022. p. 267–80.
- Choi JK, Ji YG. Investigating the importance of trust on adopting an autonomous vehicle. *Int J Hum-Comput Interact*. 2015;31(10):692–702.
- Kraus J, Scholz D, Stiegemeier D, Baumann M. The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. *Hum Factors*. 2020;62(5):718–36.
- Zhang T, et al. The roles of initial trust and perceived risk in public's acceptance of Automated Vehicles. *Transport Res Part C Emerg Technol*. 2019;98:207–20. <https://doi.org/10.1016/j.trc.2018.11.018>.
- Howard D, Dai D. Public perceptions of self-driving cars: The case of Berkeley, California. In: *Transportation research board 93rd annual meeting* (Vol. 14, No. 4502, pp. 1–16). Washington, DC, USA: The National Academies of Sciences, Engineering, and Medicine. 2014.
- Xing Y, et al. Toward human-vehicle collaboration: Review and Perspectives on Human-centered collaborative automated driving. *Transport Res Part C Emerg Technol*. 2021;128: 103199. <https://doi.org/10.1016/j.trc.2021.103199>.
- Geisslinger M, Poszler F, Lienkamp M. An ethical trajectory planning algorithm for autonomous vehicles. *Nat Mach Intell*. 2023;5:1–8.
- Németh B. Coordinated control design for ethical maneuvering of autonomous vehicles. *Energies*. 2023;16(10):4254.
- D'Souza J, Kim J, Pickering J. Variable trust setting for safe and ethical algorithms for navigation of autonomous vehicles (C-NAV) on a Highway. In: *Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics-Volume 1: ICINCO* (pp. 88–96). SciTePress. 2023.
- Pickering JE, D'Souza J. Deontological Ethics for safe and ethical algorithms for navigation of autonomous vehicles (C-NAV) on a highway. In: *9th International Conference on Control, Decision and Information Technologies*, July 03 - 06 2023, Rome, Italy. 2023.
- Pickering JE. Systems modelling and ethical decision algorithms for autonomous vehicle collisions. PhD Thesis, University of Wolverhampton, UK. 2020.
- Gilbert A, Petrovic D, Pickering JE, Warwick K. Multi-attribute decision making on mitigating a collision of an autonomous vehicle on motorways. *Expert Syst Appl*. 2021;171: 114581.
- Driver & Vehicle Standards Agency and Great Britain: Department for Transport. *The Official Highway Code 2022*.
- Tännsjö T. *Understanding ethics*. Edinburgh University Press; 2013.
- Canudas-de-Wit C, Tsiotras P, Velenis E, Basset M, Gissinger G. Dynamic friction models for road/tire longitudinal interaction. *Veh Syst Dyn*. 2003;39(3):189–226. <https://doi.org/10.1076/vesd.39.3.189.14152>.
- Rajamani R. *Vehicle dynamics and control*. Springer Science & Business Media; 2011.
- Kong J, Pfeiffer M, Schildbach G, Borrelli F. Kinematic and dynamic vehicle models for autonomous driving control design. In: *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015; pp.1094–1099. <https://doi.org/10.1109/ivs.2015.7225830>.
- Wang H, Wang Q, Chen W, Zhao L, Tan D. Path tracking based on model predictive control with variable predictive horizon. *Trans Inst Meas Control*. 2021;43(12):2676–88.
- Xie Z, Huang X, Luo S, Zhang R, Ma F. A model predictive control trajectory tracking lateral controller for autonomous vehicles combined with deep deterministic policy gradient. *Trans Inst Meas Control*. 2023;46:01423312231197854.
- Ulusoy M. Designing an MPC controller with Simulink (<https://www.mathworks.com/matlabcentral/fileexchange/68992-designing-an-mpc-controller-with-simulink>), MATLAB Central File Exchange. 2023. Retrieved February 28, 2023.
- Jurewicz C, Sobhani A, Woolley J, et al. Exploration of vehicle impact speed-injury severity relationships for application in Safer Road Design. *Transport Res Proc*. 2016;14:4247–56.
- Evans K, de Moura N, Chauvier S, Chatila R, Dogan E. Ethical decision making in autonomous vehicles: the AV ethics project. *Sci Eng Ethics*. 2020;26(6):3285–312.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.