Moral disengagement mechanisms in interactions of human drivers with autonomous vehicles: Validation of a new scale and relevance with personality, driving style and attitudes.

Abstract

The introduction of autonomous vehicles (AVs) in the road transportation systems raises questions with respect to their interactions with human drivers', especially during the early stages. Issues such as unfamiliarity or false assumptions regarding the timid and safe behaviour of AVs could potentially result in undesirable human driver behaviours, for instance "testing" AVs or being aggressive towards them. Among other factors, morality has been determined as a source of aggressive driving behaviour. Following previous approaches on moral disengagement, the current paper argues that moral standards during interactions of human drivers with AVs could potentially blur, leading to the disengagement of self-regulation mechanisms of moral behaviour. The study investigates the impact of moral disengagement on the intention of human drivers to be aggressive towards AVs. To that end, an online survey was conducted including a newly developed survey of moral disengagement, adapted to the context of AVs. Moreover, measures of personality, driving style, attitudes towards sharing the road with AVs and perceived threats were collected. A confirmatory factor analysis provided support for the concept of moral disengagement in the context of AVs. Moreover, relationships between personality, driving style and attitudes towards sharing the road with AVs were found, via a structural equation modelling approach (SEM). The results could have implications in the future driver training and education programmes, as it might be necessary to not only focus on driving skills but also on the development of procedural skills that will improve the understanding of AVs' capabilities and ensure safer interactions. Efforts on improving attitudes towards AVs may also be necessary for improving human driver behaviour.

Keywords: Moral disengagement; Aggressive driving; Autonomous vehicles; Personality; Attitudes

1 Introduction

In the prospect of the deployment of autonomous vehicles (AVs), significant changes are expected in the future transportation systems. AVs have the potential to considerably improve road safety by minimising human error, reducing congestion and the need for parking spaces, allowing for the development of new mobility services, while they could also provide more environmentally friendly solutions (Fagnant & Kockelman, 2015). However, transportation systems that heavily or exclusively rely on fully autonomous vehicles are still a distant scenario. The introduction of AVs on the roads is anticipated to be a gradual process. During the early stages of fully autonomous vehicles, it is likely that they will be interacting with other vehicles of lower automation levels or even fully human driven vehicles.

The interaction of human drivers with fully autonomous vehicles is of particular interest as there is some level of uncertainty in predicting how the former will behave when encountering the latter, especially given that vehicles of Level 4 or 5 could potentially operate on the roads without any passengers on board. Also, fully autonomous vehicles are expected to have timid behaviour and comply with traffic rules and even have additional safety features to minimise the potential occurrence of dangerous situations or crashes. Manufacturers in the automotive industry are already aware or expect that human drivers could potentially try to take advantage of or "bully" fully automated vehicles. Volvo (Connor, 2016) announced that they are planning to leave their initial fleet of AVs unmarked to avoid competitive of "combative" behaviour of human drivers. In the same direction, the CEO of Mercedes-Benz USA has reported that unless AVs are programmed to be more aggressive, human drivers are going to bully them (Mitchell, 2016). Timid behaviour of fully autonomous vehicles is also an issue that has raised concerns in terms of slowing down traffic (Brooks, 2017). Moreover, AVs may not be following the "unwritten" traffic rules and may be subject to bullying behaviour from both pedestrians and human drivers. A series of experts has already raised the issue of increased jaywalking tendency when pedestrians are encountering AVs, since the former will always expect these vehicles to stop for them (Tabone et al., 2021), while similar concerns have been also reported by Ackermann et al. (2019).

Since automation is still at its infancy, the issue of aggressive behaviour against AVs still lacks validation, given that extensive empirical evidence is not yet available. Parkin et al. (2018) reported a series of worth noting interactions of AVs with other human road users including drivers, pedestrians, and cyclists. The authors suggested that the AVs reaction in handling antagonistic or aggressive behaviour of human drivers, are crucial issues to be investigated in the future research. However, an important question is whether this type of human behaviour consists a major problem. In a video analysis investigating interactions between AVs and other road users, Madigan et al. (2019) found only a few incidents of road users "testing" the behaviour of AVs. Nevertheless, the rate of occurrence would result in one incident every approximately 5 hours, which is still a concerning value especially considering that during the early stages, the novelty of these vehicles might highly trigger curiosity. Moore et al. (2020) conducted a Wizard-of-Oz trial and reported a small number of vehicle "testing" cases by pedestrians, verbal abuse and one case of "testing" by a driver. Within an effort to rationalise human behaviour, the authors concluded that curiosity could be a significant driving factor. Moreover, human road users could be assertive in an effort to show dominance, as the presence of AVs may raise uncertainty regarding the status of the former while sharing the road. Mirnig et al. (2020) tested a sample of six drivers interacting with a shuttle on a test-track but they did not report any events of risky human driving behaviour. However, the results of that study may have been affected by the low sample size, the lower speeds and controlled conditions due to

safety issues. In a simulator study conducted by TRL (2017), participants accepted smaller gaps in a crossing scenario with a higher proportion of AVs, while when AVs were more recognisable more false starts were also observed. Further qualitative analysis indicated that some of the participants felt greater confidence when interacting with an AV, because they assumed it would react quicker, and it was programmed to avoid crashes, hence they could accept shorter gaps. Even though this type of behaviour may not be the norm, it is still likely to occur in the future. Liu et al. (2020) conducted a questionnaire survey investigating intention to "bully" autonomous and conventional vehicles. Results indicated a higher tendency (despite the low scores overall) to be aggressive towards AVs while differences across countries were also found. In the same direction, Lee et al. (2021) reported that interactions with AVs can elicit higher levels of aggression in human drivers, compared to the same events taking place with other conventional vehicles. These findings were supported in a study conducted by Trende et al. (2019) who concluded that in motorway merging situations, time pressure increased gap-acceptance in interactions with AVs.

As there is still uncertainty around the behaviour of human drivers in their encounters with AVs of higher levels of automation, it is necessary to further investigate these interactions. With reference to previous research that has suggested a higher tendency of drivers to be more aggressive towards AVs, compared to other human drivers (Liu et al., 2020), it might be reasonable to assume that characteristics related to current driving behaviour patterns can also persist in behaviour towards AVs. Aggressive and risky driving behaviour have been linked to several individual characteristics such as personality, socio-demographics, attitudes towards driving, contextual, and environmental factors. In addition to the aforementioned variables, the concepts of morality and moral disengagement have been also incorporated in the research of aggressive driving behaviour (Cleary et al., 2016; Swann et al., 2017) and traffic safety (Otto et al., 2021). In particular, among other factors, moral disengagement has been found to have the highest impact on driving aggression, in the context of conventional driving (Swann et al., 2017). Moral disengagement refers at situations that an individual is disengaged from the moral self-regulation process and behaves inconsistently with the internal moral standards that obstruct commitment of wrong behavioural actions (Bandura, 2002). The concept of moral disengagement has been extensively used in research related to bullying behaviour (Thornberg & Jungert, 2013, 2014). AVs are also expected to bring changes in the road networks. Human drivers may be uncertain regarding the performance of the technology, in terms of driving behaviour. These issues may cause confusion in negotiations with AVs. Moreover, AVs still need to "prove" their efficiency, reliability and safety. Uncertainty regarding these aspects can raise scepticism and mistrust which may also extend to the behaviour of human drivers towards these vehicles via moral disengagement.

The potential of aggressive behaviour towards AVs could have significant negative safety implications hence, it is important to acquire a greater level of understanding about its determinants. Given that moral disengagement has been identified as a significant factor of aggressive driving behaviour, its relevance to AVs also needs to be investigated, as it could also persist during these interactions. Moreover, factors related to aggressive driving behaviour in the context of conventional vehicles such as personality (Akbari et al., 2019) or driving style (Taubman-Ben-Ari & Skvirsky, 2016) could still be important. Finally, uncertainty about the presence of AVs on the road could also affect human drivers' behaviour towards the latter and might need to be taken into account. The latter may be also related to the perceived risks of AVs. Considering the relevance of moral disengagement with bully or aggressive behaviour, the aim of the current paper is twofold, namely, (a) Develop and validate a moral disengagement scale, particularly in the context of human drivers and AVs interactions and (b),

investigate the relevance of the new moral disengagement scale with drivers' individual traits related to aggressive driving behaviour and attitudes towards AVs.

The remainder of the paper is organised as follows. Section 2 provides the theoretical background of the study with a focus on morality, moral disengagement, and factors related to aggressive driving behaviour. Section 3 presents the questionnaire scales used and an overview of the analyses. This is followed by Section 4, which presents the results. This section is further divided in the descriptive statistics, bivariate analysis, validation of the moral disengagement scale and its relationship with individuals' attributes using a structural equation model (SEM) approach. The paper concludes, with a discussion section of the main findings, implications, limitations of the study and future research directions.

2 Theoretical background

2.1 Moral behaviour and relevance to the driving behaviour context

Moral or ethical behaviour is predominantly regarded as behaviour consistent with generally accepted moral norms of behaviour (Reynolds & Ceranic, 2007). Moral norms are related to aspects of abuse, rights and justice (Sachdeva et al., 2011) and represent the perceived moral rules that would justify the performance (or not) of a specific action (Ajzen, 1991). In the context of driving, moral norms have been primarily considered in traffic safety research, that has applied the Theory of Planned Behaviour (TPB; Ajzen, 1991) to investigate among others, aggressive manoeuvres (Parker et al., 1995), speeding behaviour (Chorlton et al., 2012; Conner et al., 2007; Elliott & Thomson, 2010), phone use while driving (Benson et al., 2015; Gauld et al., 2017; Kim, 2018) or drink and drive behaviour (Moan & Rise, 2011). The role of moral norms is important as they guide moral reasoning and moral judgement behind ethical behaviour (Campbell & Kumar, 2012). The process of moral judgement is activated when an individual faces an issue of ethical nature (Rest, 1986). One of the most widely adopted approaches to investigate moral judgement has been Kohlberg's cognitive moral development theory (Kohlberg, 1969). Kohlberg analysed the reasoning of males, from middle school to young adulthood, during interviews about choices in hypothetical moral dilemmas, and concluded that moral reasoning was being developed over time.

Page et al. (2013) suggested that aspects from the research on moral judgement and reasoning could potentially provide insights about individual differences in driving behaviour and road safety research. Concepts revolving around morality, although under investigated, can be still found in traffic-related literature. Bianchi and Summala (2002) further built on Kohlberg's cognitive moral development theory to investigate the impact of moral judgement on aggressive driving behaviour but no significant relationships were found. The authors concluded that their outcomes were probably affected by the small variance of moral judgement in the sample. In a more recent study, Bailey et al. (2016) used the concepts of moral values and moral reasoning to investigate their impact on emotions and reported differences between moral judgement and emotions of anger provoked by the behaviour of others. Also, groups related to lower moral values were more likely to support aggressive driving behaviours. Comparable findings were also reported by Du et al. (2018) where a similar approach was used to investigate the impact of ethical position on perceived responsibility about the behaviour of other drivers, anger provoked, and ultimately aggressive driving. Veldscholten (2015) concluded that norm-complying driving behaviour is more related to the safety of the whole traffic system rather than egocentric motives of punishment avoidance. In the same study,

drivers' who supported arguments that justified antisocial behaviour based on concepts cognitive distortion only represented a small proportion of the sample. Finally, lower levels of moral reasoning were related to a higher number of accidents, higher driving speed, and higher extent of space-taking behaviour. More recently, van den Berg et al. (2020) used concepts from the Moral Foundation Theory (MFT) (Graham et al., 2013), however they authors reported that moral values may not be the main cause of aggressive driving behaviour.

2.2 Moral disengagement and driving behaviour

The concept of moral disengagement (MD) was introduced as a type of moral reasoning within Bandura's social cognitive theory of moral agency (Bandura, 1986; Bandura, 1990). Bandura (1986, 1990) focused on the issue of moral reasoning to unfold how overall well-behaved people may conduct bad actions. In the course of their life, people develop standards of right or wrong that direct their moral conduct. This self-regulatory process enables people to act in ways compliant with their moral standards. Bandura (2002) argued that self-regulation of behaviour is not influenced solely by moral reasoning, but the latter is also linked to moral behaviour via a series of self-regulation mechanisms. The author also debated that self-regulation does not develop a constant control system, but there are many psychological and social mechanisms via which self-sanctions can be disengaged. To that end, among individuals sharing the same moral standards, different activation or disengagement of these mechanisms can result in different types of negative actions.

The notion of moral disengagement is linked to aggressive behaviour in various areas and has been also incorporated in studies related to bullying. Acts of harassment or bullying are related to the dissuasive exercise of moral agency (Thornberg & Jungert, 2014) which is defined as an individual's ability to make moral judgements based on some notion of right and wrong and to be held accountable for these actions (Taylor, 2009). However, as reported previously, people may selectively disengage from moral agency and humane acts, and instead execute harmful acts towards others; a process else known as moral disengagement. Examples that have investigated the effect of moral disengagement on bullying behaviour extend from school bullying (Georgiou et al., 2020; Gini et al., 2014; Thornberg & Jungert, 2014) and cyberbullying (Cuadrado-Gordillo & Fernandez-Antelo, 2019; Hwang et al., 2020; Luo & Bussey, 2019) to bullying in prisons (South & Wood, 2006) and unethical behaviour in interactions at the workplace (Newman et al., 2019; Ogunfowora et al., 2021).

Moral disengagement is composed by a set of psychological mechanisms via which the moral standards of a person are biased. Bandura (2014) suggested that the mechanisms of moral disengagement are part of four behavioural loci and each locus follows the next one in a sequential process. Most studies, however, consider moral disengagement as a single construct (Lee et al., 2014). The structure of moral disengagement mechanisms along with their definitions are presented in Table 1.

More recently, the concept of moral disengagement was incorporated in the research of aggressive driving behaviour (Cleary et al., 2016; Swann et al., 2017) and traffic safety (Otto et al., 2021). Swann et al. (2017) adopted this approach in the context of aggressive driving behaviour by developing the Driving Moral Disengagement Scale. The authors concluded that moral disengagement could be a better predictor of aggressive driving behaviour, compared to driving anger. Sutton (2010) presented analyses focusing on the impact of video games on aggressive driving behaviour, also accounting for the impact of moral disengagement. The latter had a positive and significant correlation with self-reported driving errors and violations.

Table 1 Structure of moral disengagement mechanisms

Psychological mechanism - Locus	MD Mechanism	Description
Cognitive reconstructing of harmful behaviour (CR) – Behavioural locus	Moral justification	When a person is reframing an act as it is to be for the greater good
	Euphemistic labelling	When a person is labelling an act in a way to be presented as less harmful
	Advantageous comparison	When a person is comparing an act with more harmful acts to justify its appropriateness
Obscuring or minimising one's role in causing harm (OM) – Agency locus	Displacement of responsibility	When a person is mentally shifting responsibility of a harmful act to someone else
	Diffusion of responsibility	When a person is allocating the responsibility of an act across a group
Disregarding or distorting the impact of harmful behaviour (DC) – Outcome locus	Distortion of consequences	When a person is justifying an act in a sense that its consequences are not harmful
Blaming and dehumanising the victim (BD) – Victim locus	Attribution of blame	When a person is alleging that it is someone's own fault for experiencing harmful actions by others
	Dehumanisation	When someone is removing human qualities from the victim and instead treating the latter as animal or object

Adapted from (Lee et al., 2014; Newman et al., 2019)

Among the concepts presented in Table 1, there is a number of references in traffic behaviour literature related to the concept of dehumanisation. Turner et al. (1975) controlled for the dehumanisation condition in a naturalistic study by using a curtain to hide the driver of a vehicle that was used to obstruct participants and intended to induce aggressive responses. Dehumanisation was linked to higher aggression and lack of understanding of the other driver's behaviour. Delbosc et al. (2019) reported that drivers tend to be more aggressive towards cyclists, when perceived dehumanisation about the latter increases. Referring to the context of aggressive driving, Dula et al. (2011) reported that prejudice expressed in the form of negative attitudes towards the members of a group could lead to dehumanisation, while anonymity could also be another reason for easier dehumanising while driving (Denny, 2000). Michael (2020) suggested that dehumanisation of others, and consequently road rage, could be a repercussion of perceived power one gets when using a car. Lennon and Watson (2011) conducted a series of interviews and concluded that aggressive driving behaviour could be a form of expressing disapproval to specific behaviours of other drivers and an attempt to correct them, else referred by the authors as *"teaching them a lesson"*. This behavioural pattern shares similarities to the concept of moral justification where harmful acts are perceived to be *"for the greater good"*. Similar to the concept of dehumanisation, the notions of perceived anthropomorphism and human-likeness have been used as indicators of trust towards the performance of AVs (Waytz et al., 2014; Young & Monroe, 2019).

2.3 The role of personality and driving style

2.3.1 The impact of personality on aggressive driving

The impact of personality (and personality traits) on aggressive driving behaviour has been extensively examined by several research studies. The importance of personality has been already acknowledged in the past, as efforts to investigate the correlation between sensation-seeking (else excitement-seeking) and aggressive driving can be found since the 70s (Jonah, 1997). The Big Five Factor Model (FFM) of personality traits (Goldberg, 1990, 1993) or facets (more specific aspects of broader personality traits) of it, are commonly used concepts. The FFM consists of the extraversion, agreeableness, openness, conscientiousness, and neuroticism factors. Several researchers have used these factors to investigate their connection to aggressive driving behaviour. Jovanović et al. (2011) found that neuroticism and conscientiousness had a positive impact on driving-related anger while agreeableness had a negative impact on both driving anger and aggressive driving. Dahlen et al. (2012) found that agreeableness and driving anger - derived from the Driving Anger Scale (DAS) developed by Deffenbacher et al. (1994) - were linked to increased aggressive driving. The latter was further related to higher crash involvement and more tickets for traffic violations. Also, Riendeau et al. (2018) reported that the extraversion and neuroticism traits were positively related to unsafe overall performance in a driving simulator setting while conscientiousness was linked to lower risky behaviour only among middle-aged drivers. Moreover, sensation-seeking [using the scale by Zuckerman (1994)] was positively related to increased risky behaviour among younger drivers. Chraif et al. (2016) mentioned that emotional stability[1], agreeableness, and conscientiousness had a negative association with aggressive driving. Taubman-Ben-Ari and Yehiel (2012) related different personality traits with different driving styles. Some significant correlations regarded the negative association of conscientiousness with the reckless, anxious and angry driving styles while a positive correlation was found with the careful driving style. Similar correlation patterns were also observed regarding agreeableness and the aforementioned driving styles (except for the anxious style, which yielded an insignificant result). Finally neuroticism was positively correlated to the anxious driving style and openness was positively related to the careful style.

Except for the main five factors of the FFM, researchers have also considered the impact of their facets on driving behaviour. Oltedal and Rundmo (2006) found that aggression and excitement-seeking were positively related to risky driving and accident involvement while anxiety had a negative association to risky driving. The authors also included in their study the traits of normlessness (a measure of an individual's lack of respect for and obedience of norms) and irritability – derived from the DAS – reporting that both were positively related to risky driving and accident involvement as well. Ge et al. (2014) also used FFM facets and concluded in similar results with respect to the positive impact of anger and excitement-seeking on aggressive driving and drunk driving, while a negative effect of altruism on these was found. Machin and Sankey (2008) reported that excitement-seeking was positively related to accident involvement and speeding but negatively related to self-reported aversion to risk taking. The opposite results were found with respect to altruism. Shen et al. (2018) reported that in their models, altruism was positively related to prosocial driving behaviour while on the other hand, sensation-seeking was negatively related to prosocial behaviour. Their results were reversed when aggressive driving behaviour was investigated as the dependent variable. In a similar direction, Yang et al. (2013) developed regression models concluding that anger, sensation-seeking and normlessness positively contributed to violations, while altruism had an opposite effect. The authors also reported similar outcomes with respect to accident involvement, except for sensation-seeking which did not have a significant impact. Finally, focusing specifically on young drivers, Ulleberg and Rundmo (2003) found relationships between personality, risky

[1] Emotional stability is considered as the opposite of neuroticism.

driving and attitudes towards driving risk and safety. In particular, significant negative correlation occurred between altruism and anxiety with risk taking behaviour while on the other hand significant positive correlations were observed between normlessness, sensation-seeking and aggressiveness with risk taking behaviour. It is worth mentioning that similar types of correlation were also observed between the examined personality facets and attitudes towards traffic safety. The interested reader is referred to the meta-analysis of Akbari et al. (2019) for a more comprehensive review about the relation of personality to aggressive driving behaviour.

The role of personality has been also investigated in the context of AVs. Kraus et al. (2020) found a negative relationship between neuroticism and affinity to technology while extraversion, agreeableness and self-esteem were positively related to interpersonal trust, which in turn had a positive correlation with trust in AVs. Similarly, Charness et al. (2018) reported a positive correlation between conscientiousness and concerns for AVs. Additionally, they reported a negative relation of conscientiousness and a positive relation of emotional stability and openness with the eagerness to adopt. These results indicate that, apart from conventional driving behaviour, personality traits can be a predictor of attitudes towards AVs.

2.3.2 The impact of personality on moral disengagement

In relevance to the context of the present study, the effect of personality traits has been also considered on moral disengagement. Kuilman et al. (2019) developed a latent variable framework to investigate the effect of moral disengagement on moral reasoning of nurse practitioners and physician assistants. The authors considered two higher-order meta traits of personality based on the FFM. Both traits α (agreeableness, conscientiousness, emotional stability) and β (extraversion and openness) were negatively related to moral disengagement. Zhou et al. (2018) concluded in some similar results regarding cyberbullying. In particular, the authors reported negative correlations between extraversion, agreeableness and conscientiousness with mechanisms of moral disengagement, while a positive correlation occurred with respect to neuroticism. No significant correlations were found between openness and mechanisms of moral disengagement. Saidon et al. (2010) mentioned a negative association of conscientiousness and extraversion with moral disengagement while Rengifo and Laham (2022) reported a negative impact of openness, agreeableness and honesty. Wang et al. (2016) found that moral disengagement is negatively correlated with moral reasoning and positively related with Machiavellianism, which is a component of the Dark Triad of personality factors (together with narcissism and psychopathy). Jones et al. (2017) reported that narcissism is positively correlated with moral disengagement and both significantly affect antisocial behaviour in sports. Finally, Kapoor et al. (2021) reported a mediating effect of moral disengagement on the narcissism and psychopathy traits, regarding the intention to exaggerate on online reviews.

2.3.3 Hypotheses development regarding the effect of personality

With reference to existing studies investigating aggressive driving behaviour, the FFM facets of anxiety, excitement-seeking and altruism were included in the present study. Moreover, the trait of normlessness was also considered. Although neuroticism (and hence anxiety as one of its facets) has been linked to increase in moral disengagement, anxiety per se is related to less risky or aggressive driving. Given that interactions with AVs might introduce novel aspects in traffic negotiations it might be expected that people with increased anxiety might be less prone to exploiting the behaviour of AVs due to moral disengagement. Excitement-seeking (part of extraversion) has been linked to riskier driving behaviour. However, extraversion has been also

linked to higher trust towards AV technologies and reduced moral disengagement. Hence, it could be expected to observe a negative relation of excitement-seeking with moral disengagement towards AVs. Finally, based on literature related to risky and aggressive driving, a negative relationship is expected between altruism and moral disengagement while the opposite is anticipated regarding normlessness. Following the aforementioned statements, the following hypotheses were developed:

H1a: Anxiety is expected to have a negative association with moral disengagement
H1b: Excitement-seeking is expected to have a negative association with moral disengagement
H1c: Altruism is expected to have a negative association with moral disengagement
H1d: Normlessness is expected to have a positive association with moral disengagement

2.3.4 Driving style and hypotheses development

Except for the role of personality on driving related outcomes (such as violations, risky behaviour and others), self-reported scales have been also used as indicators or predictors of aggressive driving behaviour and driving styles. The Multi-dimensional Driving Style Inventory or MDSI (Taubman-Ben-Ari et al., 2004) has been widely used to identify how driving style is related to behaviour. van Huysduynen et al. (2018) concluded that responses in the MDSI were consistent with observed behaviour in a driving simulator experiment. Long and Ruosong (2019) positively related traffic violations and crashes with risky and angry driving styles and negatively with the careful driving style. A similar conclusion was reported by Taubman-Ben-Ari and Skvirsky (2016) as they found that involvement in severe crashes was positively correlated with the angry driving style and negatively related to patient and careful driving styles, while some similar findings were reported by Holman and Havârneanu (2015). Padilla et al. (2018) used the MDSI and found significant relationships regarding traffic offences. The relevance of specific driving styles with risky behaviour raises the question on how these would apply in interactions of human drivers with AVs. The occurrence of moral disengagement could lead to aggressive human behaviour or involvement in risky situations. Hence, on top of personality, driving style could also be investigated with respect to its impact. As reported later in Section 3.2.3, only some of the driving styles were included in the present analysis. Based on the selected driving styles, the following research hypotheses were developed:

H2a: Angry driving style is expected to have a positive association with moral disengagement
H2b: Anxious driving style is expected to have a negative association with moral disengagement
H2c: Careful driving style is expected to have a negative association with moral disengagement
H2d: High-velocity driving style is expected to have a positive association with moral disengagement
H2e: Patient driving style is expected to have a negative association with moral disengagement
H2f: Risky driving style is expected to have a positive association with moral disengagement

2.4 Attitudes towards AVs, perceived risks and threats

2.4.1 Attitudes towards sharing the road

Research investigating the general acceptance and intention to use AVs has highlighted the complexity of the issue. Except for factors such as travel cost or time, the technology introduces a series of new elements to be considered, such as attitudes towards the technology, social

norms, trust or perceived risks (Jing et al., 2020). Although the effect of the aforementioned aspects has been primarily investigated with respect to general acceptance, it has been suggested that issues such as attitude towards AVs could also form the behaviour of human drivers while interacting with these vehicles in the future (Strömberg et al., 2021). Attitude mainly refers to the propensity of an individual to favour or not a particular entity and have been highlighted as one of the most important determinants of AVs acceptance (Jing et al., 2020). Attitudes towards AVs are usually investigated from a generic point of view; researchers are interested in understanding the general opinion of the public. However, other studies have focused on regulatory preferences when sharing the road with AVs. Nair and Bhat (2021) developed models to examine preferences regarding separate infrastructure for AVs, restricted presence in certain locations or mandatory presence of a driver inside the vehicle at all times. Similarly, Rahman et al. (2021) conducted a qualitative study to investigate preferences of vulnerable road users when sharing the road with AVs. Preferences towards regulation may reflect attitudes of the general public about sharing the road with AVs. Given that factors related to general acceptance could also influence behaviour towards AVs, attitudes (towards sharing the road) could have a significant role.

2.4.2 Perceived risks and threat

Perceived risks in the context of AVs are in close relation to both attitudes and acceptance. Perceived risks can take several forms however, the most commonly considered are related to system malfunctions, data privacy, cybersecurity and legal liability (Golbabaei et al., 2020). Additional types of risks regarding AVs may refer to societal implications such as job losses (Pettigrew et al., 2018) while others have highlighted detrimental effects on public health due to higher use on AVs (Alonso Raposo et al., 2018; Hoadley, 2018). Wang et al. (2020) approached acceptance of AVs from a more ontological perspective and argued that on top of the typically examined risks, the use of technology could also be hindered by the perceived threat on the human drivers' role; adoption of the AV technology would not only replace the human driver but also minimise the psychological attachment to the driving experience. This type of threat, that is challenging one's role, is part of the theory of identity threat.

The concept of threat is very common in bullying behaviour literature. Relevant studies have highlighted that perceived threat of an out-group entity towards the group identity can lead to aggressive behaviour (Gini, 2007). Examples of this behaviour have been reported in the context of school bullying (Gini, 2006; Kuldas et al., 2021; Ojala & Nesdale, 2004) but also in human-robot interactions (Keijsers & Bartneck, 2018; Keijsers et al., 2021). Perceived threat has been also linked to the notion of dehumanisation, particularly in the context of immigrants (Louis et al., 2013; Pavetich & Stathi, 2021; Viki et al., 2013). Perceived threat is taking two main forms in the framework of bullying, aggressive behaviour and dehumanisation. Realistic threat is related to the physical and economic well-being of the group. Identity threats refers to perceived threat to the group's cultural values. The role of perceived threat has been also considered in technology-related research, mainly in the area of robot acceptance. For instance, Złotowski et al. (2017) reported that participants exposed to the idea of autonomous robots perceived higher threat (both realistic and identity) while they expressed more negative attitudes, compared to those exposed to the idea of non-autonomous robots. In the same direction, Yogeeswaran et al. (2016) examined the influence of background information about robots on perceived realistic and identity threat, concluding that the former can affect human attitude. Huang et al. (2021) investigated the intention to use hotel service robots and found that realistic and identity threat increased the negative attitude towards the robots, while the latter negatively influenced intention to use. In the context of AVs, realistic threat (as reported

in human-robot interactions literature) is comparable to perceived risks (i.e. safety, security and societal) that most studies have examined, whereas identity threat is yet under investigated.

In relation to the issue of perceived threat, there is some evidence in the AV literature, which highlights that perceived risk stemming from AVs could affect attitudes towards them. Jing et al. (2021) reported a negative impact of perceived risk on attitudes, channelled via perceived usefulness. Solbraa Bay (2016) also presented a variation of the Technology Acceptance Model and found a significant impact of perceived risk on attitudes about AVs. Moreover, Yuen et al. (2021) considered an indirect effect of perceived threat on attitudes towards AVs via trust. Nair and Bhat (2021) reported that decrease in perceived safety of drivers when sharing the roads with AVs, had a positive impact on supporting regulations that minimised interactions with them. Considering the impact of perceived threat from the domain of bullying, and the relevance of the latter to moral disengagement and aggressive behaviour, it could be expected that negative attitudes towards AVs could considerably affect moral disengagement in the context of AVs particularly.

2.4.3 Hypotheses development regarding attitudes and perceived threat

Following the idea that higher levels of AV acceptance might lead to more favourable opinion of human drivers to share the road and less aggressive behaviour during interactions, it is expected that positive attitudes could be negatively related to moral disengagement. Moreover, considering previous literature on the impact of perceived risk and threat on attitudes (in AV and robot-related literature), it is anticipated that increased perceived threat will be positively associated to attitudes against sharing the road. Therefore, the following research hypotheses are formed:

H3a: Perceived threat is expected to have a positive relationship with attitudes against sharing the road with AVs.
H3b: Attitudes against sharing the road with AVs are expected to have a positive relation to moral disengagement.

3 Method

3.1 Participants

The data were collected as part of a broader online survey that focused on the interactions of human drivers with AVs. The survey was developed using the Online surveys (Jisc) platform and took approximately 30 mins to complete. Respondents were required to be above 18 years old and hold a driving licence to participate. Respondents provided their consent prior participation while they were also informed that participation was voluntary and data were anonymous. Two participants were excluded from the analysis as they failed in at least two out of three attention check questions. Moreover, another participant was excluded due to invalid age response (below 18). In total, data from 424 participants were considered as valid for the analysis, as participants who failed to pass specific attention check questions were removed. The sample had a good balance in terms of gender; female participants were 215, male 201, while the remaining did not wish to state their gender. The average age of the sample was approximately 26.8 years old with a standard deviation of 8.3 years. Respondents were also asked to rate on a 1-10 scale their level of experience (1:Very inexperienced, 10: Very experienced) with advanced driver-assistance systems (ADAS) and their overall opinion about

ADAS (1:Very negative, 10: Very positive). In the latter, a "No opinion" option was also included. The average value of experience with ADAS was 4.28 (approximately 26% replied "Very inexperienced"). Overall, for those who expressed an opinion about ADAS, the average value was 4.72, while approximately 42% selected the "No opinion" option. It should be mentioned that questions related to sociodemographic characteristics were not mandatory to be replied. Some aggregated values regarding participants demographic characteristics are presented in Table 2.

Table 2 Sociodemographic characteristics of the sample

Variable	Category	Frequency	Percentage
Gender	Female	215	52%
	Male	201	39%
Age	<25	211	51%
	25–34	152	37%
	35–44	33	8%
	45–54	11	3%
	>54	7	2%
Driving frequency	Everyday	136	32%
	4-6 days a week	110	26%
	2-3 days a week	86	20%
	About once a week	41	10%
	Less often	51	12%

3.2 Measures

3.2.1 The AV moral disengagement scale

Moral disengagement was measured using the AV moral disengagement scale (AVMDS), a scale developed by the authors, following the example of the Driving Moral Disengagement Scale (DMDS) presented by Swann et al. (2017) to adapt the concept of moral disengagement in the driving behaviour context. Each item was developed to fall within one of the eight mechanisms of moral disengagement, as they have been defined in Table 1. In particular, the scale items presented situations relevant to aggressive driving including aggressive, risky and competitive behaviour or traffic violations during interactions with AVs. The general framework of the aforementioned behaviours mainly revolved around taking advantage of the safe behaviour of unoccupied (without any passengers on board) AVs. Each item included some type of justification, which was related to one of the eight mechanisms of moral disengagement. As an example, the mechanism of moral justification refers to reframing an action, as it is to be for the greater good. In the context of school bullying, Thornberg and Jungert (2014) used the item "*It's okay to hurt a person a couple of times a week if you do that in order to help your friends*". Swann et al. (2017) adapted moral justification in the context of driving behaviour as "*It's alright to deliberately hold someone up by going slow if it's for their own good*". In the present study, an example item to represent moral justification was "*It is fine to behave aggressively towards AVs to highlight potential problems in their behaviour*". In a different example, advantageous comparison refers at comparing an action with more harmful actions to justify it. Following the same studies, Thornberg and Jungert (2014) used the item "*Teasing a person now and then every week is not so bad if you compare it to hitting and kicking a person every day*", while Swann et al. (2017) used "*Speeding a little over the limit is not too serious compared to those that speed a lot over the limit*". In the present study, an

example item for advantageous comparison was "*It is not a big problem to behave aggressively towards AVs, thinking that other people behave aggressively towards human drivers*". The various components of moral disengagement were implemented (including example items) as follows:

- Moral justification: Reframing an act as it is to be for the greater good – understanding/highlighting the deficiencies in the behaviour of AVs so they can be improved (example item: *"It is fine to behave aggressively towards AVs to highlight potential problems in their behaviour")*
- Euphemistic labelling: Reasoning an act as part of the process or curiosity to understand the behaviour of AVs (example item: *"It would be fine to behave aggressively towards AVs out of curiosity to observe their behaviour"*)
- Advantageous comparison: Comparing an act with more harmful acts to justify its appropriateness (example item: *"It is not a big problem to behave aggressively towards AVs, thinking that other people behave aggressively towards human drivers"*)
- Displacement of responsibility: Mentally shifting responsibility of a harmful act to someone/something else (example item: *"Drivers cannot be blamed for driving competitively against AVs, if their friend pressured them into it"*)
- Diffusion of responsibility: Allocating the responsibility of an act across a group – what other drivers generally do (example item: *"It is ok to behave aggressively towards AVs, if the other drivers are doing the same as well"*)
- Distortion of consequences: Justifying an act in a sense that its consequences are not harmful – interaction with the AV is not dangerous (example item: *"It is fine to take risks when encountering an AV, as it is a machine supposed to always behave safely"*)
- Attribution of blame: Alleging that it is someone's own fault for experiencing harmful actions by others – It is own fault of the AV (example item: *"If human drivers take advantage of the overly cautions behaviour of AVs, it is probably because there is something wrong with it"*)
- Dehumanisation: Divesting human qualities from the AV and instead treating the latter as object – An AV is not a real human driver (example item: *"There is no problem if human drivers behave aggressively towards an AV, as it is a machine that cannot react"*)

The scale included 24 items in total. All items were applied using a 6-point Likert scale (Strongly disagree – Strongly agree). The full item list is appended in Table A.1.

3.2.2 Attitudes towards AVs, perceived risks and threats

Attitudes towards the presence of AVs (or sharing the road with them) were also included in the survey, to investigate the potential impact of opinions regarding this issue on moral disengagement. Items of that scale were developed based on reported regulations and preferences regarding the common presence of AVs with human drivers (Nair & Bhat, 2021) or vulnerable road users (Rahman et al., 2021). Examples of such preferences refer to separate infrastructure for AVs or restricting their presence in specific areas. The statements were developed so that respondents would express their opposition towards AVs on the roads in specific circumstances (hence attitudes against sharing the road). The full list of items is presented in Table A.2 of the Appendix. All items were applied using a 6-point Likert scale (Strongly disagree – Strongly agree).

Section 2.4.2 examined the relevance of perceived risk on attitudes and acceptance of AVs. Literature on aggressive (or bullying) behaviour, including human-robot interactions, has adopted the term of threat and investigated its impact on attitudes or acceptance. Perceived threat is commonly decomposed into realistic and identity threat. In the framework of human-robot interactions, realistic threat has been approximated via direct risks or robots as their impact on job security or human well-being. These aspects have been also considered as potential societal risks in the context of AVs, together with other safety and security risks. Identity threat has been captured via statements related to threats deriving from robots regarding the uniqueness and essence of humanity (Huang et al., 2021; Złotowski et al., 2017). The latter type of threat has received little consideration in the context of AVs (Wang et al., 2020).

The term of threat was adopted in the current study, to account for the broader impact of this issue on attitudes towards sharing the roads with AVs. Survey items related to realistic threat were included by using typical items from previous studies [as in Liu and Xu (2020)] and referred to accident occurrence, hacking and data privacy. Also, following studies in human-robot interactions (and given the relevance of these issues to the AV technology), two more items were included with respect to the impact of AVs on job losses and public health. Finally, four items related to identity threat were included. These items revolved around the issues of personal/human driving style, ban of human driving and communications between drivers. A 6-point (Strongly Disagree – Strongly Agree) Likert scale was used for both surveys. The full list of items is presented in Table A.3.

3.2.3 Personality and driving style

In order to capture respondents' personality, facets of the Big Five Factor Model were included in the survey. Following past studies related to aggressive driving behaviour, the facets of anxiety, anger, excitement-seeking and altruism were considered (Table A.4) while a 4-item scale of normlessness was also included (Table A.5). A short version (Konstabel et al., 2012) of the personality items was used in order to reduce the total survey length. Moreover, the Normlessness scale was used, as presented in Ulleberg and Rundmo (2003).

Driving style was captured via the use of the MDSI (Taubman-Ben-Ari et al., 2004). In order to reduce the total length of the survey, some of the driving styles (dissociative and distress reduction styles) were dropped from the survey. Moreover, from the remaining driving styles, only items that resulted in higher factor loadings in the original paper were considered. The items and driving styles included are presented in Table A.6.

3.3 Analyses overview

The analyses conducted in the current study can be segregated in three main components. In the initial stage, the questionnaire items were grouped into factors, and the values of mean and standard deviation were calculated to obtain a better picture of the data. The internal consistency of factors was evaluated via Chronbach's α statistic and a cut-off 0.6 value was considered as adequate. This was followed by bivariate (Pearson) correlation analysis to further examine the associations between the factors considered in the analysis. The third and final component of the analysis is divided in two parts. First, a confirmatory factor analysis (CFA) was applied to investigate the construct validity of the author-developed moral disengagement scale, which was adapted for the framework of human drivers and AVs interactions. Then, a structural equation model (SEM) was estimated to further investigate the relationship of

personality, driving style and attitudes variables with moral disengagement. Maximum likelihood estimation was used for both the CFA and SEM analyses with significance level criteria for $p<0.05$, $p<0.01$, and $p<0.001$. Further details regarding model specification and model fit can be found in Sections 4.3 (Table 10) and 4.4.1. The descriptive and bivariate analysis presented in Sections 4.1 and 4.2 were performed using ©IBM SPSS (Statistical Package for Social Sciences; IBM Corp, 2020), v27.0 while the CFA and SEM models were estimated using the lavaan package (Rosseel, 2012) in R software (Team, 2013).

4 Results

4.1 Descriptive analysis

4.1.1 Moral disengagement in interactions with AVs

The AVMDS was investigated both in terms of the MD factors and in terms of the behavioural loci (Table 3). The lower values reported in the moral disengagement items are consistent with similar studies on moral reasoning and disengagement, in the driving context (Swann et al., 2017; Veldscholten, 2015) and stated intention to be aggressive towards an AV (Liu et al., 2020). The results suggested that with respect to the behavioural locus, moral justification and euphemistic labelling had higher average scores, compared to advantageous comparison. On the other hand, the mechanisms of the victim locus had the highest average values, following these of moral justification and euphemistic labelling.

Table 3 Descriptive statistics and internal consistency of the MD factors

Psychological mechanism	**Mechanism**	Mechanism			Psychological mechanism		
		M	SD	Chronbach's α	M	SD	Chronbach's α
Cognitive reconstruing of harmful behaviour (CR) – Behavioural locus	Moral justification	2.30	1.57	0.561	2.08	1.44	0.827
	Euphemistic labelling	2.25	1.44	0.685			
	Advantageous comparison	1.70	1.19	0.635			
Obscuring or minimising one's role in causing harm (OM) – Agency locus	Displacement of responsibility	1.82	1.26	0.656	1.71	1.18	0.813
	Diffusion of responsibility	1.53	1.01	0.762			
Disregarding or distorting the impact of harmful behaviour (DC) – Outcome locus	Distortion of consequences	1.90	1.33	0.603	1.90	1.34	0.603
Blaming and dehumanising the victim (BD) – Victim locus	Attribution of blame	2.10	1.33	0.545	2.11	1.38	0.755
	Dehumanisation	2.12	1.41	0.657			

Given the lower internal consistency scores (Chronbach's α) of the eight MD mechanisms, the four broader psychological mechanisms reported by Bandura (2014) were instead considered in the next steps, regarding the construct validation of the scale (Fernandez-Antelo & Cuadrado-Gordillo, 2019; Lee et al., 2014).

4.1.2 Personality and driving style

The descriptive statistics of the driving style and personality results are presented in Table 4. It should be mentioned that correlation analysis showed that the item NL4 from the normlessness scale (Table A.5) was not correlated with the remaining items of the scale and was removed from the analysis. This specific item was also insignificant in the results of the model presented later in Section 4.4.2 and was removed, as it was not considered a significant indicator of normlessness. The results in Table 4 suggest that the sample consisted of drivers who stated higher levels of positive driving styles such as patient and careful, compared to negative driving styles, for instance risky or high-velocity. In terms of personality, lower levels of normlessness were observed, compared to altruism. The average values of excitement-seeking and neuroticism were in between the other two personality factors.

Table 4 Descriptive statistics of driving style and personality

	M	SD
MDSI		
Angry	2.11	1.36
Anxious	3.15	1.65
Careful	4.82	1.13
High-velocity	2.66	1.49
Patient	5.00	1.19
Risky	2.08	1.29
Personality		
Neuroticism	3.61	1.91
Excitement-seeking	4.42	1.88
Altruism	5.34	1.57
Normlessness	2.48	1.31

4.1.3 Attitudes and perceived risk

The results of attitudes towards sharing the road with AVs and perceived threat are presented in Table 5. Since several of the threat-related items were developed by adapting items from the human-robots interactions literature, their parsimony was initially investigated via the use of exploratory factor analysis (EFA). Although, the EFA results suggested a two-factor solution, items T4 and T5 (threats on job loss and active mobility/ public health) of Table A.3 were in the same factor with items T6-T9 which aimed at referring at identity threat (factor loadings presented in Table A.7). Hence, in the context of the current study, realistic threat represents safety and security risks related to the use of AVs, while identity threat includes aspects of perceived societal impacts (replacement of a human's role by a robot) and changes in the driving culture and the role of human as a driver. The same analysis was applied regarding attitudes however, only a single factor was identified. The internal consistency of all factors (Chronbach's α) was above 0.6 therefore, all items were retained.

Table 5 Descriptive statistics of attitude on sharing and perceived threat

Measure	M	SD	Chronbach's α
Attitudes on sharing	3.44	1.72	0.776
Realistic threat (T1-T3)	4.05	1.61	0.730
Identity threat (T4-T9)	3.52	1.68	0.805

4.2 Bivariate analysis

The research hypotheses developed in Sections 2.3 and 2.4 were initially examined via bivariate correlation analysis. Before conducting the correlation analysis, the values of survey items related to each factor presented in Tables 6-8 were averaged per individual.

Table 6 presents the correlations between moral disengagement and personality factors. The hypotheses developed regarding the aforementioned relationship were only partially confirmed. In particular, anxiety and excitement-seeking did not result in any significant correlations with any of the moral disengagement psychological mechanisms. On the other hand, altruism was negatively correlated to all moral disengagement mechanisms while normlessness was positively correlated with the latter. Hence, only hypotheses H1c and H1d were validated.

Table 6 Correlations between moral disengagement and personality factors

	Anxiety	Excitement-seeking	Altruism	Normlessness
CR	0.037	0.003	-.185**	.281**
OM	0.092	-0.023	-.214**	.219**
DC	0.041	0.006	-.194**	.210**
BD	0.071	-0.038	-.263**	.250**

***. Correlation is significant at the 0.01 level (2-tailed).*

Table 7 presents the correlations between moral disengagement and driving style factors. All driving styles led to significant correlations except for the anxious driving style. Moreover, the direction of correlation was consistent with the hypotheses developed in Section 2.3.4, excluding hypothesis H2b.

Table 7 Correlations between moral disengagement and driving style factors

	Angry	Anxious	Careful	High-velocity	Patient	Risky
CR	.239**	0.048	-.157**	.340**	-.156**	.270**
OM	.298**	0.036	-.262**	.375**	-.290**	.319**
DC	.295**	0.044	-.166**	.315**	-.196**	.277**
BD	.266**	0.030	-.188**	.331**	-.241**	.348**

***. Correlation is significant at the 0.01 level (2-tailed).*

Table 8 presents the correlations between moral disengagement with attitudes and perceived threats. Attitudes against sharing the road with AVs had positive correlations with all factors of moral disengagement. The same result also occurred with respect to identity threat. On the other hand, realistic threat had positive correlations only with the CR and BD factors. Based on the bivariate correlation analysis, hypothesis H3b was confirmed.

Table 8 Correlations between moral disengagement with attitudes and perceived threats

	Attitudes	Realistic threat	Identity threat
CR	.189**	.146**	.220**
OM	.169**	0.078	.146**
DC	.128**	0.041	.123*
BD	.233**	.165**	.214**

***. Correlation is significant at the 0.01 level (2-tailed).*

**. Correlation is significant at the 0.05 level (2-tailed).*

4.3 Moral disengagement measurement model

The construct validity of the scale was assessed via confirmatory factor analysis (CFA). The initial factors composing the mechanisms of moral disengagement indicated that the internal consistency for some of them was below the minimum acceptable 0.6 cut-off value (Gallais et al., 2017; van Griethuijsen et al., 2014) of the Cronbach's α criterion. As in previous studies, the initial eight factors were reconsidered to follow the four psychological mechanisms of moral disengagement. The proposed final model is illustrated in Figure 1. Out of the total 24 items of the initial scale, three were removed (details in Table A.1) either due to lower factor loadings or correlation with the other items of each factor. This process was followed to ensure adequate fit of the final model. The Goodness-of-fit indices are presented in detail in Table 10. The cut-off criteria of the indices were based on existing literature (Abduh & Abdul Razak, 2012; West et al., 2012; Yang et al., 2019). The power analysis of the model was conducted using the "semPower" R package (Jobst et al., 2021; Moshagen & Erdfelder, 2016). Based on the sample size (N=424) and the degrees-of-freedom (dof=183), the model resulted in achieved power > .99 for α=.05 and power =.80, using a critical RMSEA=.05 as the effect measure.

All factor loadings were significant and had positive signs (Table 9). All items retained in the final specification had values above 0.3. The latter suggests their suitability to represent the intended factors (Della Vedova et al., 2022). Moreover, moral disengagement is usually considered as a single factor in existing studies (Bandura et al., 1996). The correlations across the latent variables were all significant and (apart from one) above 0.8, supporting the hypothesis that the factors can be used to represent a single construct also in the current MD scale. Lower correlation values, such as between CR and DC, may imply that although both factors represent moral disengagement, unique differences in individual attributes are captured by each latent variable (Lee et al., 2014). The composite reliability (CR) and average variance explained (AVE) were also calculated as measures of convergent validity. The results suggest that all CR values were above the 0.6 cut-off threshold (Fornell & Larcker, 1981). On the other hand, only the OM factor passed the recommended 0.5 AVE value. Given that AVE is a more conservative measure of reliability and the validity of the measurement model may still be considered acceptable based on the CR criterion (Fornell & Larcker, 1981; Lam, 2012).

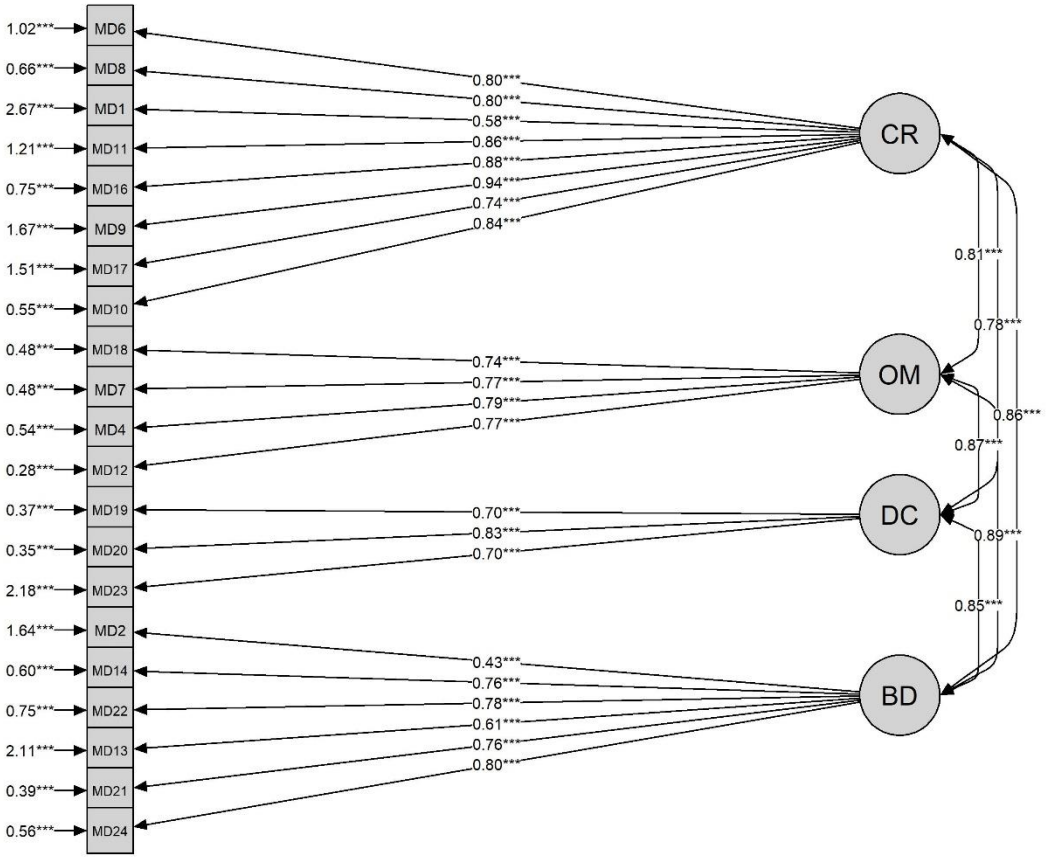

Figure 1 Measurement model of moral disengagement

Table 9 Parameter estimates of the moral disengagement measurement model

	Estimate	Standardised estimate	p-value
CR			
MD6	0.80	0.62	0.000
MD8	0.80	0.70	0.000
MD1	0.58	0.34	0.000
MD11	0.86	0.62	0.000
MD16	0.88	0.72	0.000
MD9	0.94	0.59	0.000
MD17	0.75	0.52	0.000
MD10	0.84	0.75	0.000
CR=.81; AVE=.34			
OM			
MD18	0.75	0.73	0.000
MD7	0.77	0.74	0.000
MD4	0.79	0.73	0.000
MD12	0.77	0.82	0.000
CR=.84; AVE=.57			
DC			
MD19	0.70	0.75	0.000
MD20	0.83	0.81	0.000
MD23	0.70	0.43	0.000

CR=.63; AVE=.36			
BD			
MD2	0.43	0.32	0.000
MD14	0.76	0.70	0.000
MD22	0.78	0.67	0.000
MD13	0.61	0.39	0.000
MD21	0.76	0.77	0.000
MD24	0.80	0.73	0.000
CR=.74; AVE=.33			

Table 10 Goodness-of-fit indices of the measurement model

Measure	Value	Criteria	
		Acceptable	Good
χ2-test			
χ2/df	3.04	<5.0	< 3.0
Absolute fit			
RMSEA	0.069	<0.1	< 0.08
SRMR	0.06	<0.1	< 0.05
GFI	0.884	>0.7	>0.9
Incremental fit			
AGFI	0.854	>0.7	> 0.85
CFI	0.903	>0.7	> 0.9
TLI	0.889	>0.8	>0.9

4.4 Moral disengagement and individual characteristics –Structural Equation Model approach

4.4.1 Model specification

Structural equation modelling uses several types of models to illustrate relationships among observed variables. The main objective of this method is to provide quantitative support for a theoretical model assumed by the researcher (Schumacker & Lomax, 2004). Structural equation models are composed of two main parts. The latent variable model captures the relationship between endogenous (dependent) and exogenous (independent) latent variables. The measurement model expresses the relationship between latent and observed variables. The main formula (Equation 1) of a latent variable model (Schumacker & Lomax, 2004) is

$$\eta = B\eta + \Gamma\xi + \zeta \quad (1)$$

where η is an (m × 1) vector of the endogenous variables, ξ is an (n × 1) vector of the exogenous latent variables, and ζ is an (m × 1) vector of random disturbance. The m and n indicators denote the number of the endogenous and exogenous latent variables respectively. The elements of the B and Γ matrices are the coefficients of the model. In particular, the B matrix is an (m × m) coefficient matrix of the latent endogenous variables and the Γ matrix is an (m × n) coefficient matrix for the latent exogenous variables. The main formulae of the measurement model are

$$x = \Lambda_x\xi + \delta \quad (2)$$

for the exogenous variables (Equation 2) and

$$y = \Lambda_y \eta + \varepsilon \quad (3)$$

for the endogenous variables (Equation 3), where the observed variables are indicated by the vectors y (p × 1) and x (q × 1). The p and q indicators denote the number of the endogenous and exogenous indicator (observed) variables respectively. The matrix Λ_y (p × m) denotes the coefficients of the y elements while the matrix Λ_x (q x n) indicates the coefficients of the x elements. The measurement errors for y are denoted by the (p × 1) vector ε and for the x by the (q × 1) vector δ.

The present approach aimed in confirming the presence of relationships between individual traits from the existing literature related to aggressive driving, with moral disengagement in the context of AVs. Moreover, based on the literature review, it was expected that perceived threat and attitudes towards AVs could also have an impact on moral disengagement.

4.4.2 Parameter estimates

The significant results of the structural model component are presented in Table 11. The detailed results of the measurement models are presented in Table A.8 while the model is illustrated in Figure 2. The item names follow the codes presented in Tables A.1 to A.6 in the Appendix. The parameter estimates were all significant at the 0.05 level (or higher) and had expected signs. The personality and reported driving styles were anticipated to be related thus, correlations among these variables were considered in the final model. Moreover, due to high correlation, two items of the normlessness scale were allowed to correlate. The covariance results are presented in Table A.9. The Goodness-of-fit indices presented in Table 12, suggest an acceptable model fit. Based on the sample size (N=424) and the degrees-of-freedom (dof=1160), the model resulted in achieved power > .99 for α=.05 and power =.80, using a critical RMSEA=.05 as the effect measure.

Table 11 Structural model parameter estimates

Paths			Estimate	p-value
Attitudes	←	Threat	1.195	0.000
Moral disengagement (MD)	←	Altruism (ALTR)	-0.216	0.006
Moral disengagement (MD)	←	Normlessness (NORML)	0.353	0.000
Moral disengagement (MD)	←	Attitudes (ATT)	0.117	0.003
Moral disengagement (MD)	←	High-velocity (HIGH_VEL)	0.310	0.000
Moral disengagement (MD)	←	Risky (RISKY)	0.229	0.002

The perceived threat latent variable had a significant positive association with attitudes against sharing the road with AVs. Also, moral disengagement was found to be positively related to the normlessness trait and negatively related to altruism. Moreover, the high-velocity and risky driving styles had positive relationships with moral disengagement. Hence, it may be concluded that lower levels of occurrence of these driving behaviours could be linked to lower likelihood of moral disengagement when interacting with AVs. Implications of these findings are discussed in Section 5.1.

Table 12 Goodness-of-fit indices

Measure	Value
χ2-test	
χ2/df	1.95
Absolute fit	
RMSEA	.047
SRMR	.077
GFI	.818
Incremental fit	
AGFI	.800
CFI	.858
TLI	.850

5 Discussion

Aggressive driving and "road rage" have been acknowledged as areas that require psychological interventions (Galovski & Blanchard, 2004). Research in the field of traffic psychology has demonstrated significant contributions in investigating and determining the elements of aggressive, risky and aberrant driving behaviour. In the light of technological advances and the gradual introduction of automated systems on the roads, terms as "road rage" and "aggressive driving" may need to be revised, in order to account the new types of interactions that are expected to take place. In addition, potential gaps in safety need to be identified and addressed in a proactive rather than a reactive manner.

It has been argued that risk-averse and timid behaviour of AVs might spark the competitive behaviour of other human agents, motorised or not. These road users could potentially engage in risky behaviour either as a result of the safe behaviour of AVs or in order to "stand their ground" against machines taking over the roads. The issue of road rage against AVs is not hypothetical, as similar cases have been already reported (Wong, 2018), while naturalistic studies have suggested the occurrence of incidents that involved "testing" of the AV behaviour (Madigan et al., 2019; Moore et al., 2020). These early results indicate that these incidents are not frequent and only represent a vast minority. However, they could still pose a threat towards the safety of road users and require further investigation to be addressed appropriately. Despite its low frequency, the propensity to be aggressive towards an AV is higher, compared to another human driver (Liu et al., 2020). The current paper investigated the potential manifestation of moral disengagement, in interactions of human drivers with AVs, and its relevance to personality, driving style and attitudes towards sharing the road. To that end, a survey scale was developed that aimed in capturing the basic mechanisms of moral disengagement, adapted in the context of AVs.

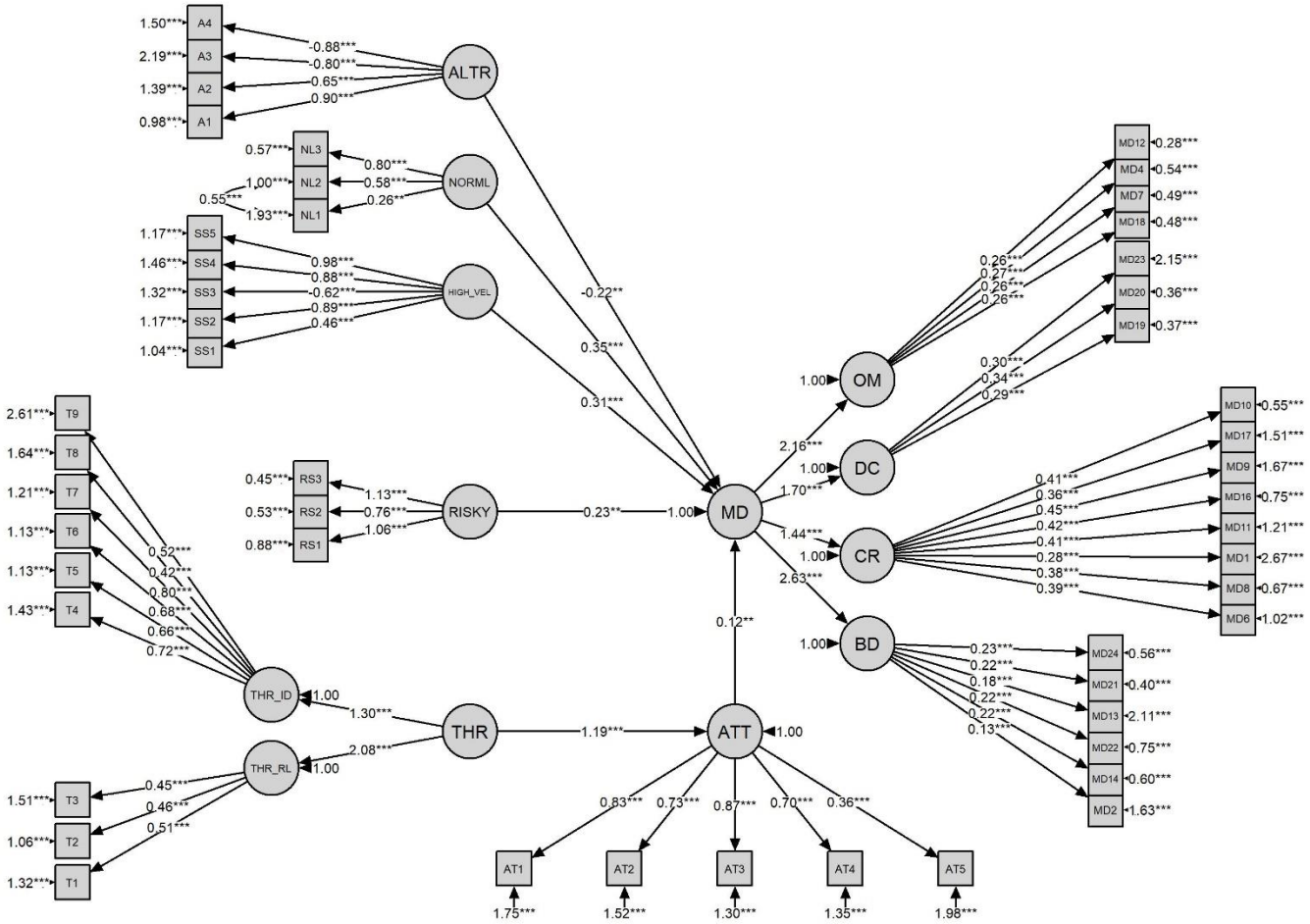

Figure 2 The moral disengagement SEM model

5.1 Summary of findings

5.1.1 Descriptive statistics and bivariate analysis

The average value of the responses related to the moral disengagement mechanisms was investigated as a first step of the descriptive statistics analysis. The results suggested that the behavioural loci of moral justification and euphemistic labelling had higher average scores, compared to advantageous comparison. This could be an indication that human drivers are more likely to disengage their self-regulations mechanisms when this is in the interest of "correcting" or better understanding the behaviour of AVs, whereas the same is not as likely when aggressive behaviour is compared to more harmful actions. The same finding might also apply with respect to the agency locus, in particular for the diffusion of responsibility mechanism. On the other hand, the outcome locus had somewhat higher average and standard deviation values, implying that it could be a more likely factor of moral disengagement. Finally, the mechanisms of the victim locus had the highest average values, following these of moral justification and euphemistic labelling. This finding could suggest that human drivers may perceive AVs as less of human and blame the latter (and their behaviour), if human drivers are aggressive to them. Finally it is worth mentioning that regarding perceived threat, items related to realistic threat had the highest average value, denoting that aspects related to safety and security are relatively more important compared to social or identity implications.

The bivariate correlation analysis provided further insights regarding the hypotheses formed in Sections 2.3 and 2.4. In particular, only the correlations of moral disengagement with altruism (H1c) and normlessness (H1d) were significant. Under the assumption that moral disengagement towards AVs could lead to actual aggressive behaviour against these vehicles, this finding suggests that some personality traits remain relevant as potential determinants of aggressive driving behaviour. The insignificant correlation regarding anxiety might imply that higher levels of this personality trait do not affect moral mechanisms of drivers. The same finding may also hold regarding excitement-seeking. Kraus et al. (2020) reported a positive relation between extraversion (of which excitement-seeking is a facet) and trust towards AVs. It is hence likely that excitement seekers may be less prone to being aggressive towards AVs therefore, no significant relation to moral disengagement was found. With respect to driving styles, all correlations with moral disengagement were significant except for the anxious driving style (H2b). The results suggested that driving style may be used as a possible determinant of moral disengagement. In particular, the angry, high-velocity and risky styles [else mentioned as negative driving styles by Herrero-Fernandez (2021)] were also positively associated to moral disengagement towards AVs. On the other hand, the opposite type of correlation occurred regarding the patient and risky driving styles (positive driving styles). Hence, potential aggressive behaviour during interactions with AVs could be identified via styles related to conventional driving. Finally, attitudes against sharing the road with AVs were positively related to the psychological factors of moral disengagement, supporting hypothesis (H3b).

5.1.2 Moral disengagement structural equation model

The CFA suggested that the construct validity of the AVMDS could be supported for the four broader psychological mechanisms of moral disengagement. This specification was preferred in the present study as the internal consistency for two out of the eight moral disengagement mechanisms resulted in Chronbach α values below the 0.6 cut-off value (but above 0.5). This four-factor model was further examined via SEM analysis to investigate the relationship of

moral disengagement with individual characteristics related to driving behaviour and attitudes towards AVs. The parameter estimates of the SEM model indicated that the perceived threat latent variable had a significant and positive association with attitudes against sharing the road with AVs. The latter was found to have a significantly positive relation with moral disengagement. These findings suggest that perceived threat due to technological deficiencies of AVs and potential broader societal impacts might be related to decreased willingness of human drivers to drive among these vehicles and interact with them. Hence, negative opinions about the presence of AVs could be linked to the self-regulation mechanisms of human drivers and encourage aggressive behaviour. Moral disengagement was also found to have a positive association with the normlessness trait and a negative association with altruism. As mentioned in Section 2.3, the effect of personality traits has been already investigated in the context of driving and aggressive behaviour. The present findings suggest that these relationships may also extend to the context of AVs. Considering indications that human drivers could potentially be more aggressive towards AVs compared to human drivers (Liu et al., 2020), personality as an individual characteristic may also be related to human driving behaviour towards AVs. Charness et al. (2018) have already reported some relevance between personality and attitudes towards AVs. The results with respect to moral disengagement suggest that these effects could also apply in human driver behaviour while interacting with these vehicles. Moreover, the high-velocity and risky driving styles also had a positive relationship with moral disengagement. Hence, it could be concluded that lower levels of occurrence of these driving behaviours could result in lower likelihood of moral disengagement when interacting with AVs.

The results of the SEM model further supported the research hypotheses related to personality traits from the bivariate analysis. However, only the negative driving styles resulted in significant parameter estimates, regarding their relation to moral disengagement. A potential interpretation of this finding could be related to the self-report nature of the survey items. The descriptive statistics of the MDSI factors showed higher average values with smaller variance in the responses related to the careful and patient driving styles. This tendency in the (stated) driving styles might not allow for significant results regarding these driving styles to be observed. Another potential source for this outcome could be the selection of the MDSI items, as not all of the questions from the original scales were retained, to reduce the total length of the survey. The significant parameter estimate that suggested a significant positive relation between perceived threats and attitudes against sharing the road with AVs, supported hypothesis (H3a). The latter had a significant parameter estimate with respect to moral disengagement, further supporting (H3b).

5.2 Implications of the study

The findings of the present study suggested a relevance between factors related to aggressive driving behaviour and the potential distortion in moral norms during interactions with autonomous vehicles. The concept of morality has been already investigated in the area of driving behaviour in interactions with conventional vehicles and studies have confirmed, to some extent, the relationship between these aspects. The results of the current analysis indicated that the concept of moral disengagement potentially holds also in interactions of human drivers with AVs. Moreover, moral disengagement in that context was related to factors known to be influencing conventional driving behaviour, such as personality, hence, they remain relevant. Although current findings require further investigation and validation, it might be the case that human drivers in the future might attempt to be competitive or aggressive against AVs, on the pretext of improving their behaviour or simply because of an erroneous belief about their role on the road. The results of the SEM analysis indicated that the AVMDS

might be adopted in the future as a tool for further investigating how morality, moral norms and disengagement relate to interactions and negotiations of human drivers with AVs.

Except for theoretical implications in the research of morality, human driving behaviour and AVs, the findings of the current study may have implications on future approaches regarding driver training. Redshaw (2001) approximated driver training from a social perspective. The author suggested that driver training practices which emphasise on the practical skills neglect the considerable significance of motivation and attitudinal factors (Saffron, 1982). With reference to Redshaw (2001), "*... dealing with behaviour, beliefs and attitudes in a group context is important. If individuals can see that others just like them are also undergoing the same process they will have more courage and conviction to follow it through themselves*". On the same matter, Redshaw (2006) mentioned that assessment of information received while driving is significantly affected by social beliefs and values as well as assumptions, attitudes and expectations. The author argued that these factors should receive attention during the training, so novice drivers learn on how to reflect on these aspects which otherwise become systematic with the more technical part of the driving task. Veldscholten (2015) followed the idea of self-reflection and assessment via the concept of moral reasoning and concluded that these aspects can influence driving style. This particular "social approach" in driver training is in line with some of the findings from the moral disengagement scale, for instance blaming the AVs and their behaviour, if they are victims of aggression from human drivers. Also, if the role and motivation of AVs is clear, their surrounding human traffic could potentially decrease aggressive behaviour related to the reconstruction of harmful behaviour, which was also higher, compared to the other behavioural loci in the present study.

In a literature review study, Merriman et al. (2021) identified areas where potential interventions may be needed for automated vehicle driver training. Perhaps, interventions in training programs should not only regard how to use AVs but also how human drivers should interact safely and efficiently with them. The development of mental models should extend from creating accurate expectations about the behaviour of AVs as a user, to a broader perspective; human drivers need to be aware of the range in the behaviour of AVs (and the rationale behind) as well as the role of each individual agent during their interactions. An approach like this could tackle phenomena of "testing" the behaviour of AVs, attempting to "correct" their behaviour or engaging in risky manoeuvres from the training stage, rather than allowing human drivers to undertake a decision-making process, which may not be appropriately calibrated for these specific interactions. Existing research has highlighted that commitment in moral mechanisms and information about the negative consequences of deficient behaviour could decrease aggressive behaviour (Du et al., 2018; Swann et al., 2017). With reference to Merriman et al. (2021), attitudes towards driving and personality can affect procedural skills while driving. The results of the current paper showed that aspects of personality, driving style and attitudes towards the presence of AVs on the road had a significant effect on moral disengagement. Although scales like the MDSI could be potentially used to identify individuals with a higher likelihood to be morally disengaged, a better practice could be the elimination of such type of behaviour during the training stages, as discussed previously.

Except for individual characteristics related to aggressive driving behaviour, attitudes against sharing the road with AVs also had a significant positive impact on moral disengagement. If these attitudes are considered as a proxy of acceptance of sharing the road, efforts need to focus on approaches to improve general acceptance of the technology. Strömberg et al. (2021) suggested that factors defining attitudes and acceptance towards AVs could have a significant

role on elicited anger and human drivers' behaviour during interactions with AVs. Hence, although the development of a driving culture promoting the harmonious interactions of human drivers with AVs may be very important, the main prerequisite to achieve the latter will ultimately be the reliable performance of AVs, which will eventually enable the general acceptance of the technology and consequently acceptance of sharing the roads. To that end, factors related to general acceptance may not only need to be considered in the context of using the AVs, but also for improving interactions between the latter with the human drivers. Perceived safety is a main factor of behavioural intention to use AVs (Montoro et al., 2019). Ensuring their safe performance may also lead to less aggressive and more cooperative behaviour of human drivers. Towards that direction, it is necessary that authorities will develop a framework of trust and address a broad range of issues related to the presence of AVs as legal aspects (Adnan et al., 2018) or data privacy (Alonso et al., 2021).

5.3 Limitations of the study and future research

The results presented in the current study come with a series of limitations. First, all measures were based on self-report surveys and hence may suffer from bias related to misinterpretation of the questions or the social desirability effect. Moreover, the sample refers to a worldwide panel, which may not share the same attitudes or behaviour with respect to the driving task, due to cultural differences. Also, collecting the data online might have further reduced the non-representativeness of the data, since the sample may be biased towards respondents familiar to the use of the internet technology. The latter is also reflected in the age distribution of sample, as more respondents were below 25 years old. Regarding the personality and driver style items, longer versions of the surveys could be considered in the future, as tools that are more representative. In addition, the moral disengagement scale concerned hypothetical situations that referred to vehicles that are not currently a part of the road transportation systems. Lack of familiarity and experience with respect to these situations may have affected participants' responses. Also, it should be mentioned that given that the AVMDS was an author-developed survey, the items included might not be the most representative and further validation might be required. Finally, the SEM model considered aspects of personality, driving style and attitudes. On top of these individual characteristics, some additional could affect moral disengagement and might need to be considered in future work.

The results suggested that attitudes towards sharing the road could influence moral disengagement. As discussed in Section 5.2, these can further extend to general acceptance of AVs hence, a more rigorous investigation is needed. The acceptance formation of AVs is a complex framework that includes aspects of trust (Liu et al., 2019), technological anxiety (Keszey, 2020), legal liability and data privacy issues, but also expected benefits such as reduced driving demands, greater connectivity/accessibility or energy consumption savings (Useche et al., 2021). As general acceptance and intention to use AVs may also shape human drivers' acceptance to share the road with them, the role of the aforementioned factors need to be investigated in future research. The relation to psychological models such as the Technology Acceptance Model (TAM) (Davis et al., 1989) or the Unified Theory of Acceptance and Use of Technology model (UTAUT) (Venkatesh et al., 2003) is also worth of investigation. Moreover, in the current paper, the impact of personality on moral disengagement was approximated using facets of the FFM, as these have been previously used in research related to aggressive driving. However, in future research, a more comprehensive framework of personality traits could be included. For instance, in relevance to the issue of behavioural intention to use AVs, Keszey (2020), highlighted that personality traits such as hedonic or utilitarian motivation could have an important role. Also, other personality traits such as

(empathy, locus of control or moral identity, to name a few) have been previously linked to moral disengagement (Detert et al., 2008). The role and relevance of these traits with human driver behaviour, could be the focus of future studies. Finally, despite the relationship between factors related to aggressive driving (as personality) and moral disengagement, the extent of the actual aggressive behaviour is still uncertain. More research is required towards this direction for instance in the form of naturalistic or simulator studies.

5.4 Conclusion

In conclusion, the findings of the paper supported the construct validity of a moral disengagement scale that was developed in the context of AVs. The latter was related to personality, driving style and attitudes towards sharing the road with AVs. As familiarity and experience with the latter increases, the behaviour of human drivers may change. However, the approach of moral disengagement could be used to identify potential cases of poor driving behaviour, especially during the early stages, as the results indicated that human drivers might disengage their self-regulations mechanisms during their interactions with AVs. Further research is necessary towards this direction, for a holistic understanding of the issue and its potential consequences.

Acknowledgement

This research is supported by the PAsCAL project, funded by the European Union's Horizon 2020 research and innovation programme (grant agreement No 815098, https://www.pascal-project.eu/), aimed at enhancing driver behaviour and public acceptance of connected and automated vehicles. The online survey was approved by the University of Leeds Ethics Committee (reference no. LTTRAN-141).

References

Abduh, & Abdul Razak. (2012). Customers' attitude towards diminishing partnership home financing in Islamic banking. *American Journal of Applied Science, 9*(4), 593-599.

Ackermann, Beggiato, Schubert, & Krems. (2019). An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? *Applied ergonomics, 75*, 272-282.

Adnan, Md Nordin, bin Bahruddin, & Ali. (2018). How trust can drive forward the user acceptance to the technology? In-vehicle technology for autonomous vehicle. *Transportation Research Part A: Policy and Practice, 118*, 819-836. doi:10.1016/j.tra.2018.10.019

Ajzen. (1991). The theory of planned behavior. Orgnizational Behavior and Human Decision Processes, 50, 179–211. In.

Akbari, Kamran, Heydari, Motevalian, Tabrizi, Asadi-Shekari, Sullman, & research. (2019). Meta-analysis of the correlation between personality characteristics and risky driving behaviors. *11*(2), 107.

Alonso, Faus, Esteban, & Useche. (2021). Is There a Predisposition towards the Use of New Technologies within the Traffic Field of Emerging Countries? The Case of the Dominican Republic. *Electronics, 10*(10). doi:10.3390/electronics10101208

Alonso Raposo, Grosso, Després, Fernández Macías, Galassi, Krasenbrink, Krause, Levati, Mourtzouchou, & Saveyn. (2018). An analysis of possible socio-economic effects of a Cooperative, Connected and Automated Mobility (CCAM) in Europe.

Bailey, Lennon, & Watson. (2016). Getting mad may not mean getting even: The influence of drivers' ethical ideologies on driving anger and related behaviour. *Transportation*

Research Part F: Traffic Psychology and Behaviour, 36, 104-116. doi:10.1016/j.trf.2015.11.004
Bandura. (1986). Social foundations of thought and action. Englewood Cliffs, NJ7 Prentice-Hall. In: Inc.
Bandura. (1990). Selective activation and disengagement of moral control. *Journal of social issues, 46*(1), 27-46.
Bandura. (2002). Selective Moral Disengagement in the Exercise of Moral Agency. *Journal of Moral Education, 31*(2), 101-119. doi:10.1080/0305724022014322
Bandura. (2014). Moral disengagement in the perpetration of inhumanities. In *Perspectives on Evil and Violence* (pp. 193-209): Psychology Press.
Bandura, Barbaranelli, Caprara, & Pastorelli. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71*(2), 364.
Benson, McLaughlin, & Giles. (2015). The factors underlying the decision to text while driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 35*, 85-100. doi:10.1016/j.trf.2015.10.013
Bianchi, & Summala. (2002). Moral judgment and drivers' behavior among Brazilian students. *Psychological reports, 91*(3), 759-766.
Brooks. (2017). The big problem with self-driving cars is people. *IEEE spectrum: technology, engineering, and science News, 27*.
Campbell, & Kumar. (2012). Moral reasoning on the ground. *Ethics, 122*(2), 273-312.
Charness, Yoon, Souders, Stothart, & Yehnert. (2018). Predictors of Attitudes Toward Autonomous Vehicles: The Roles of Age, Gender, Prior Knowledge, and Personality. *Front Psychol, 9*, 2589. doi:10.3389/fpsyg.2018.02589
Chorlton, Conner, & Jamson. (2012). Identifying the psychological determinants of risky riding: an application of an extended Theory of Planned Behaviour. *Accid Anal Prev, 49*, 142-153. doi:10.1016/j.aap.2011.07.003
Chraif, Aniţei, Burtăverde, & Mihăilă. (2016). The link between personality, aggressive driving, and risky driving outcomes–testing a theoretical model. *19*(6), 780-797.
Cleary, Lennon, & Swann. (2016). *Should we be aiming to engage drivers more with others on-road? Driving moral disengagement and self-reported driving aggression.* Paper presented at the Proceedings of the 26th Canadian Multidisciplinary Road Safety Conference.
Conner, Lawton, Parker, Chorlton, Manstead, & Stradling. (2007). Application of the theory of planned behaviour to the prediction of objectively assessed breaking of posted speed limits. *Br J Psychol, 98*(Pt 3), 429-453. doi:10.1348/000712606X133597
Connor. (2016). First self-driving cars will be unmarked so that other drivers don't try to bully them. *The Guardian*.
Cuadrado-Gordillo, & Fernandez-Antelo. (2019). Analysis of Moral Disengagement as a Modulating Factor in Adolescents' Perception of Cyberbullying. *Front Psychol, 10*, 1222. doi:10.3389/fpsyg.2019.01222
Dahlen, Edwards, Tubre, Zyphur, & Warren. (2012). Taking a look behind the wheel: an investigation into the personality predictors of aggressive driving. *Accid Anal Prev, 45*, 1-9. doi:10.1016/j.aap.2011.11.012
Davis, Bagozzi, & Warshaw. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science, 35*(8), 982-1003.
Deffenbacher, Oetting, & Lynch. (1994). Development of a driving anger scale. *74*(1), 83-91.
Delbosc, Naznin, Haslam, & Haworth. (2019). Dehumanization of cyclists predicts self-reported aggressive behaviour toward them: A pilot study. *Transportation Research Part F: Traffic Psychology and Behaviour, 62*, 681-689. doi:10.1016/j.trf.2019.03.005

Della Vedova, Loscalzo, Giannini, & Matthey. (2022). An exploratory and confirmatory factor analysis study of the EPDS in postnatal Italian-speaking women. *J Reprod Infant Psychol, 40*(2), 168-180. doi:10.1080/02646838.2020.1822993

Denny. (2000). Getting Behind the Wheel With Aggressive Drivers. *Journal of the Northwest Communication Association, 29*.

Detert, Treviño, & Sweitzer. (2008). Moral disengagement in ethical decision making: a study of antecedents and outcomes. *93*(2), 374.

Du, Shen, Chang, & Ma. (2018). The exceptionists of Chinese roads: The effect of road situations and ethical positions on driver aggression. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*, 719-729. doi:10.1016/j.trf.2018.07.008

Dula, Geller, & Chumney. (2011). A Social-Cognitive Model of Driver Aggression: Taking Situations and Individual Differences into Account. *Current Psychology, 30*(4), 324-334. doi:10.1007/s12144-011-9120-3

Elliott, & Thomson. (2010). The social cognitive determinants of offending drivers' speeding behaviour. *Accid Anal Prev, 42*(6), 1595-1605. doi:10.1016/j.aap.2010.03.018

Fagnant, & Kockelman. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice, 77*, 167-181. doi:10.1016/j.tra.2015.04.003

Fernandez-Antelo, & Cuadrado-Gordillo. (2019). Moral Disengagement as an Explanatory Factor of the Polyivictimization of Bullying and Cyberbullying. *Int J Environ Res Public Health, 16*(13). doi:10.3390/ijerph16132414

Fornell, & Larcker. (1981). Evaluating structural equation models with unobservable variables and measurement error. *18*(1), 39-50.

Gallais, Gagnon, Forgues, Cote, & Laberge. (2017). Further evidence for the reliability and validity of the Fatigue and Daytime Sleepiness Scale. *J Neurol Sci, 375*, 23-26. doi:10.1016/j.jns.2017.01.032

Galovski, & Blanchard. (2004). Road rage: a domain for psychological intervention? *Aggression and Violent Behavior, 9*(2), 105-127. doi:10.1016/s1359-1789(02)00118-0

Gauld, Lewis, White, Fleiter, & Watson. (2017). Smartphone use while driving: What factors predict young drivers' intentions to initiate, read, and respond to social interactive technology? *Computers in Human Behavior, 76*, 174-183. doi:10.1016/j.chb.2017.07.023

Ge, Qu, Jiang, Du, Sun, & Zhang. (2014). The effect of stress and personality on dangerous driving behavior among Chinese drivers. *Accident Analysis & Prevention, 73*, 34-40.

Georgiou, Charalambous, & Stavrinides. (2020). Mindfulness, impulsivity, and moral disengagement as parameters of bullying and victimization at school. *Aggress Behav, 46*(1), 107-115. doi:10.1002/ab.21876

Gini. (2006). Bullying as a social process: The role of group membership in students' perception of inter-group aggression at school. *Journal of school psychology, 44*(1), 51-65.

Gini. (2007). Who is blameworthy?: Social identity and inter-group bullying. *School Psychology International, 28*(1), 77-89.

Gini, Pozzoli, & Hymel. (2014). Moral disengagement among children and youth: a meta-analytic review of links to aggressive behavior. *Aggress Behav, 40*(1), 56-68. doi:10.1002/ab.21502

Golbabaei, Yigitcanlar, Paz, & Bunker. (2020). Individual Predictors of Autonomous Vehicle Public Acceptance and Intention to Use: A Systematic Review of the Literature. *Journal of Open Innovation: Technology, Market, and Complexity, 6*(4). doi:10.3390/joitmc6040106

Goldberg. (1990). An alternative" description of personality": the big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216.
Goldberg. (1993). The structure of phenotypic personality traits. *American psychologist, 48*(1), 26.
Graham, Haidt, Koleva, Motyl, Iyer, Wojcik, & Ditto. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. 47, pp. 55-130): Elsevier.
Herrero-Fernandez. (2021). Do people drive as they live, or are they transformed when they drive? A comparison of driving styles and living styles. *Accid Anal Prev, 161*, 106342. doi:10.1016/j.aap.2021.106342
Hoadley. (2018). Road Vehicle Automation and cities and regions.
Holman, & Havârneanu. (2015). The Romanian version of the multidimensional driving style inventory: Psychometric properties and cultural specificities. *Transportation Research Part F: Traffic Psychology and Behaviour, 35*, 45-59.
Huang, Cheng, Sun, & Chou. (2021). The Effects of Perceived Identity Threat and Realistic Threat on the Negative Attitudes and Usage Intentions Toward Hotel Service Robots: The Moderating Effect of the Robot's Anthropomorphism. *International Journal of Social Robotics, 13*(7), 1599-1611. doi:10.1007/s12369-021-00752-2
Hwang, Li, Stough, Lee, & Turnbull. (2020). People with disabilities' perceptions of autonomous vehicles as a viable transportation option to improve mobility: An exploratory study using mixed methods. *International journal of sustainable transportation*, 1-19. doi:10.1080/15568318.2020.1833115
IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp
Jing, Du, Chen, Shi, Zhan, & Xie. (2021). Factors that influence parents' intentions of using autonomous vehicles to transport children to and from school. *Accid Anal Prev, 152*, 105991. doi:10.1016/j.aap.2021.105991
Jing, Xu, Chen, Shi, & Zhan. (2020). The Determinants behind the Acceptance of Autonomous Vehicles: A Systematic Review. *Sustainability, 12*(5). doi:10.3390/su12051719
Jobst, Bader, & Moshagen. (2021). A Tutorial on Assessing Statistical Power and Determining Sample Size for Structural Equation Models. *Psychological Methods*. doi:https://doi.org/10.1037/met0000423
Jonah. (1997). Sensation seeking and risky driving: a review and synthesis of the literature. *Accident Analysis Prevention, 29*(5), 651-665.
Jones, Woodman, Barlow, & Roberts. (2017). The darker side of personality: Narcissism predicts moral disengagement and antisocial behavior in sport. *The Sport Psychologist, 31*(2), 109-116.
Jovanović, Lipovac, Stanojević, & Stanojević. (2011). The effects of personality traits on driving-related anger and aggressive behaviour in traffic among Serbian drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(1), 43-53.
Kapoor, Balaji, Maity, & Jain. (2021). Why consumers exaggerate in online reviews? Moral disengagement and dark personality traits. *Journal of Retailing and Consumer Services, 60*, 102496.
Keijsers, & Bartneck. (2018). *Mindless Robots get Bullied*. Paper presented at the Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction.
Keijsers, Kazmi, Eyssel, & Bartneck. (2021). Teaching robots a lesson: determinants of robot punishment. *International Journal of Social Robotics, 13*(1), 41-54.
Keszey. (2020). Behavioural intention to use autonomous vehicles: Systematic review and empirical extension. *Transportation Research Part C: Emerging Technologies, 119*. doi:10.1016/j.trc.2020.102732

Kim. (2018). The role of legal and moral norms to regulate the behavior of texting while driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 52*, 21-31. doi:10.1016/j.trf.2017.11.004

Kohlberg. (1969). Stage and sequence: The cognitive-developmental approach to socialization. *Handbook of socialization theory and research, 347*, 480.

Konstabel, Lönnqvist, Walkowitz, Konstabel, & Verkasalo. (2012). The 'Short Five'(S5): Measuring personality traits using comprehensive single items. *European Journal of Personality, 26*(1), 13-29.

Kraus, Scholz, & Baumann. (2020). What's Driving Me? Exploration and Validation of a Hierarchical Personality Model for Trust in Automated Driving. *Human Factors*, 0018720820922653.

Kuilman, Jansen, Middel, Mulder, & Roodbol. (2019). Moral reasoning explained by personality traits and moral disengagement: A study among Dutch nurse practitioners and physician assistants. *J Adv Nurs, 75*(6), 1252-1262. doi:10.1111/jan.13939

Kuldas, Foody, & Norman. (2021). Does Ethnicity of Victims and Bullies Really Matter? Suggestions for Further Research on Intra-Ethnic Bullying/Victimisation. *International Journal of Bullying Prevention*, 1-12.

Lam. (2012). Impact of competitiveness on salespeople's commitment and performance. *Journal of Business Research, 65*(9), 1328-1334. doi:10.1016/j.jbusres.2011.10.026

Lee, Momen, & LaFreniere. (2021). Attributions of social interactions: Driving among self-driving vs. conventional vehicles. *Technology in Society, 66*. doi:10.1016/j.techsoc.2021.101631

Lee, Segal, Kimberlin, Smith, & Weiler. (2014). Reliability and validity for the measurement of moral disengagement in pharmacists. *Res Social Adm Pharm, 10*(2), 297-312. doi:10.1016/j.sapharm.2013.06.006

Lennon, & Watson. (2011). "Teaching them a lesson?" A qualitative exploration of underlying motivations for driver aggression. *Accid Anal Prev, 43*(6), 2200-2208. doi:10.1016/j.aap.2011.06.015

Liu, Du, Wang, & Da Young. (2020). Ready to bully automated vehicles on public roads? *Accid Anal Prev, 137*, 105457. doi:10.1016/j.aap.2020.105457

Liu, & Xu. (2020). Public attitude toward self-driving vehicles on public roads: Direct experience changed ambivalent people to be more positive. *Technological Forecasting and Social Change, 151*. doi:10.1016/j.techfore.2019.119827

Liu, Xu, & Zhao. (2019). Road tests of self-driving vehicles: Affective and cognitive pathways in acceptance formation. *Transportation Research Part A: Policy and Practice, 124*, 354-369. doi:10.1016/j.tra.2019.04.004

Long, & Ruosong. (2019). Reliability and validity of the Multidimensional Driving Style Inventory in Chinese drivers. *Traffic injury prevention, 20*(2), 152-157.

Louis, Esses, & Lalonde. (2013). National identification, perceived threat, and dehumanization as antecedents of negative attitudes toward immigrants in A ustralia and C anada. *Journal of Applied Social Psychology, 43*, E156-E165.

Luo, & Bussey. (2019). The selectivity of moral disengagement in defenders of cyberbullying: Contextual moral disengagement. *Computers in Human Behavior, 93*, 318-325. doi:10.1016/j.chb.2018.12.038

Machin, & Sankey. (2008). Relationships between young drivers' personality characteristics, risk perceptions, and driving behaviour. *Accid Anal Prev, 40*(2), 541-547. doi:10.1016/j.aap.2007.08.010

Madigan, Nordhoff, Fox, Amini, Louw, Wilbrink, Schieben, & Merat. (2019). Understanding interactions between Automated Road Transport Systems and other road users: A video

analysis. *Transportation Research Part F: Traffic Psychology and Behaviour, 66*, 196-213.
Merriman, Plant, Revell, & Stanton. (2021). Challenges for automated vehicle driver training: A thematic analysis from manual and automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 76*, 238-268. doi:10.1016/j.trf.2020.10.011
Michael. (2020). The invisible car: the cultural purification of road rage. In *Car cultures* (pp. 59-80): Routledge.
Mirnig, Gärtner, Meschtscherjakov, & Tscheligi. (2020). *Blinded by novelty: a reflection on participant curiosity and novelty in automated vehicle studies based on experiences from the field.* Paper presented at the Proceedings of the Conference on Mensch und Computer.
Mitchell. (2016). Human drivers will bully robot cars, says CEO of Mercedes-Benz USA. *Los Angeles Times*.
Moan, & Rise. (2011). Predicting intentions not to "drink and drive" using an extended version of the theory of planned behaviour. *Accident Analysis & Prevention, 43*(4), 1378-1384.
Montoro, Useche, Alonso, Lijarcio, Bosó-Seguí, & Martí-Belda. (2019). Perceived safety and attributed value as predictors of the intention to use autonomous vehicles: A national study with Spanish drivers. *Safety Science, 120*, 865-876. doi:10.1016/j.ssci.2019.07.041
Moore, Currano, Shanks, & Sirkin. (2020). *Defense against the dark cars: Design principles for griefing of autonomous vehicles.* Paper presented at the Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction.
Moshagen, & Erdfelder. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 54-60.
Nair, & Bhat. (2021). Sharing the road with autonomous vehicles: Perceived safety and regulatory preferences. *Transportation Research Part C: Emerging Technologies, 122*. doi:10.1016/j.trc.2020.102885
Newman, Le, North-Samardzic, & Cohen. (2019). Moral Disengagement at Work: A Review and Research Agenda. *Journal of Business Ethics, 167*(3), 535-570. doi:10.1007/s10551-019-04173-0
Ogunfowora, Nguyen, Steel, & Hwang. (2021). A meta-analytic investigation of the antecedents, theoretical correlates, and consequences of moral disengagement at work. *Journal of Applied Psychology*.
Ojala, & Nesdale. (2004). Bullying and social identity: The effects of group norms and distinctiveness threat on attitudes towards bullying. *British journal of developmental psychology, 22*(1), 19-35.
Oltedal, & Rundmo. (2006). The effects of personality and gender on risky driving behaviour and accident involvement. *Safety Science, 44*(7), 621-628.
Otto, Finley, McMahill, & Arpin. (2021). *Guidance on Messaging to Avoid Psychological Reactance and Address Moral Disengagement*. Retrieved from
Padilla, Doncel, Gugliotta, & Castro. (2018). Which drivers are at risk? Factors that determine the profile of the reoffender driver. *Accident Analysis & Prevention, 119*, 237-247.
Page, Jones, & King. (2013). *Moral judgements in driving situations and their implications.* Paper presented at the Proceedings of the 2013 Australasian Road Safety Research, Policing and Education Conference.
Parker, Manstead, & Stradling. (1995). Extending the theory of planned behaviour: The role of personal norm. *British journal of social psychology, 34*(2), 127-138.
Parkin, Clark, Clayton, Ricci, & Parkhurst. (2018). Autonomous vehicle interactions in the urban street environment: a research agenda. *Proceedings of the Institution of Civil Engineers - Municipal Engineer, 171*(1), 15-25. doi:10.1680/jmuen.16.00062

Pavetich, & Stathi. (2021). Investigating antecedents of Islamophobia: The role of perceived control over terrorism, threat,
meta-dehumanization
, and dehumanization. *Journal of Community & Applied Social Psychology, 31*(4), 369-382. doi:10.1002/casp.2512
Pettigrew, Fritschi, Norman, & health. (2018). The potential implications of autonomous vehicles in and around the workplace. *15*(9), 1876.
Rahman, Dey, Das, & Sherfinski. (2021). Sharing the road with autonomous vehicles: A qualitative analysis of the perceptions of pedestrians and bicyclists. *Transportation Research Part F: Traffic Psychology and Behaviour, 78*, 433-445. doi:10.1016/j.trf.2021.03.008
Redshaw. (2001). Changing driving behaviour—a cultural approach. *Australian Journal of Social Issues, 36*(4), 315-331.
Redshaw. (2006). Safer driving through reflective thinking.
Rengifo, & Laham. (2022). Big Five personality predictors of moral disengagement: A comprehensive aspect-level approach. *Personality and Individual Differences, 184*, 111176.
Rest. (1986). Moral development: Advances in research and theory.
Reynolds, & Ceranic. (2007). The effects of moral judgment and moral identity on moral behavior: an empirical examination of the moral individual. *J Appl Psychol, 92*(6), 1610-1624. doi:10.1037/0021-9010.92.6.1610
Riendeau, Stinchcombe, Weaver, & Bédard. (2018). Personality factors are associated with simulated driving outcomes across the driving lifespan. *Transportation Research Part F: Traffic Psychology and Behaviour, 54*, 171-187.
Rosseel. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software, 48*(2), 1-36.
Sachdeva, Singh, & Medin. (2011). Culture and the quest for universal principles in moral reasoning. *Int J Psychol, 46*(3), 161-176. doi:10.1080/00207594.2011.568486
Saffron. (1982). *Driver instruction: some future research directions.* Paper presented at the Driver Training-Steering a Course for the Future, Seminar, 1982, Melbourne, Australia.
Saidon, Galbreath, & Whiteley. (2010). *Antecedents of moral disengagement: Preliminary empirical study in Malaysia.* Paper presented at the Proceedings of the 24th annual australian and new zealand academy of management conference.
Schumacker, & Lomax. (2004). *A beginner's guide to structural equation modeling*: psychology press.
Shen, Ge, Qu, Sun, & Zhang. (2018). The different effects of personality on prosocial and aggressive driving behaviour in a Chinese sample. *Transportation Research Part F: Traffic Psychology and Behaviour, 56*, 268-279. doi:10.1016/j.trf.2018.04.019
Solbraa Bay. (2016). *Innovation adoption in robotics: consumer intentions to use autonomous vehicles.*
South, & Wood. (2006). Bullying in prisons: the importance of perceived social status, prisonization, and moral disengagement. *Aggressive Behavior, 32*(5), 490-501. doi:10.1002/ab.20149
Strömberg, Ramos, Karlsson, Johansson, Ekman, Bligård, & Bergstad. (2021). A future without drivers? Comparing users', urban planners' and developers' assumptions, hopes, and concerns about autonomous vehicles. *European Transport Research Review, 13*(1), 1-12.
Sutton. (2010). *The effects of video game exposure, unmitigated agency, and moral disengagement on risky driving behavior in adolescents and young adults*: The University of Alabama.

Swann, Lennon, & Cleary. (2017). Development and preliminary validation of a scale of driving moral disengagement as a tool in the exploration of driving aggression. *Transportation Research Part F: Traffic Psychology and Behaviour, 46*, 124-136. doi:10.1016/j.trf.2017.01.011

Tabone, de Winter, Ackermann, Bärgman, Baumann, Deb, Emmenegger, Habibovic, Hagenzieker, & Hancock. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives, 9*, 100293.

Taubman-Ben-Ari, Mikulincer, & Gillath. (2004). The multidimensional driving style inventory—scale construct and validation. *Accident Analysis & Prevention, 36*(3), 323-332.

Taubman-Ben-Ari, & Skvirsky. (2016). The multidimensional driving style inventory a decade later: Review of the literature and re-evaluation of the scale. *Accident Analysis & Prevention, 93*, 179-188.

Taubman-Ben-Ari, & Yehiel. (2012). Driving styles and their associations with personality and motivation. *Accident Analysis & Prevention, 45*, 416-422.

Taylor. (2009). *Animals and ethics*: broadview Press.

Team. (2013). R: A language and environment for statistical computing.

Thornberg, & Jungert. (2013). Bystander behavior in bullying situations: basic moral sensitivity, moral disengagement and defender self-efficacy. *J Adolesc, 36*(3), 475-483. doi:10.1016/j.adolescence.2013.02.003

Thornberg, & Jungert. (2014). School bullying and the mechanisms of moral disengagement. *Aggress Behav, 40*(2), 99-108. doi:10.1002/ab.21509

Trende, Unni, Weber, Rieger, & Luedtke. (2019). *An investigation into human-autonomous vs. human-human vehicle interaction in time-critical situations*. Paper presented at the Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments.

TRL. (2017). *GATEway Project Report PPR807: Driver responses to encountering automated vehicles in an urban environment*. Retrieved from https://gateway-project.org.uk/wp-content/uploads/2017/02/D4.6_Driver-responses-to-encountering-automated-vehicles-in-an-urban-environment_PPR807.pdf

Turner, Layton, & Simons. (1975). Naturalistic studies of aggressive behavior: aggressive stimuli, victim visibility, and horn honking. *Journal of Personality and Social Psychology, 31*(6), 1098.

Ulleberg, & Rundmo. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. *Safety Science, 41*(5), 427-443.

Useche, Peñaranda-Ortega, Gonzalez-Marin, & Llamazares. (2021). Assessing the Effect of Drivers' Gender on Their Intention to Use Fully Automated Vehicles. *Applied Sciences, 12*(1). doi:10.3390/app12010103

van den Berg, Kroesen, & Chorus. (2020). Does morality predict aggressive driving? A conceptual analysis and exploratory empirical investigation. *Transportation Research Part F: Traffic Psychology and Behaviour, 74*, 259-271. doi:10.1016/j.trf.2020.08.017

van Griethuijsen, van Eijck, Haste, den Brok, Skinner, Mansour, Savran Gencer, & BouJaoude. (2014). Global Patterns in Students' Views of Science and Interest in Science. *Research in Science Education, 45*(4), 581-603. doi:10.1007/s11165-014-9438-6

van Huysduynen, Terken, & Eggen. (2018). The relation between self-reported driving style and driving behaviour. A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 56*, 245-255.

Veldscholten. (2015). *Moral reasoning in traffic: About the possible relations between moral reasoning and traffic safety*. University of Twente,

Venkatesh, Morris, Davis, & Davis. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Viki, Osgood, & Phillips. (2013). Dehumanization and self-reported proclivity to torture prisoners of war. *Journal of Experimental Social Psychology, 49*(3), 325-328. doi:10.1016/j.jesp.2012.11.006

Wang, Lei, Liu, & Hu. (2016). Moderating effects of moral reasoning and gender on the relation between moral disengagement and cyberbullying in adolescents. *Personality and Individual Differences, 98*, 244-249. doi:10.1016/j.paid.2016.04.056

Wang, Wong, Li, & Yuen. (2020). This is not me! Technology-identity concerns in consumers' acceptance of autonomous vehicle technology. *Transportation Research Part F: Traffic Psychology and Behaviour, 74*, 345-360. doi:10.1016/j.trf.2020.06.005

Waytz, Heafner, & Epley. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113-117. doi:10.1016/j.jesp.2014.01.005

West, Taylor, & Wu. (2012). Model fit and model selection in structural equation modeling. *Handbook of structural equation modeling, 1*, 209-231.

Wong. (2018). Rage against the machine: self-driving cars attacked by angry Californians. *The Guardian, 6*.

Yang, Du, Qu, Gong, & Sun. (2013). Effects of personality on risky driving behavior and accident involvement for Chinese drivers. *Traffic Inj Prev, 14*(6), 565-571. doi:10.1080/15389588.2012.748903

Yang, Qin, Chen, Ji, Zhang, & Ma. (2019). *Empirical Study on the Impact of Short Video Content Marketing on Consumer's Purchasing Intention based on the Integrated Model of TRA and ELM.* Paper presented at the 1st International Symposium on Economic Development and Management Innovation (EDMI 2019).

Yogeeswaran, Złotowski, Livingstone, Bartneck, Sumioka, & Ishiguro. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *5*(2), 29-47.

Young, & Monroe. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology, 85*. doi:10.1016/j.jesp.2019.103870

Yuen, Ma, Wang, & Lee. (2021). The role of trust in influencing consumers' adoption of automated vehicles: an application of the health belief model. *15*(11), 837-849.

Zhou, Zheng, & Gao. (2018). The relationship between the big five and cyberbullying among college students: the mediating effect of moral disengagement. *Current Psychology, 38*(5), 1162-1173. doi:10.1007/s12144-018-0005-6

Złotowski, Yogeeswaran, & Bartneck. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International journal of human-computer studies, 100*, 48-54. doi:10.1016/j.ijhcs.2016.12.008

Zuckerman. (1994). *Behavioral expressions and biosocial bases of sensation seeking*: Cambridge university press.

Appendix

Table A.1 The AV moral disengagement scale

Item code	MD mechanism	Item	Comment
MD1	Moral justification	Being aggressive towards AVs or "testing" their behaviour is justifiable, if this will lead to its improvement by the manufacturer	
MD6	Moral justification	It is fine to behave aggressively towards AVs to highlight potential problems in their behaviour	
MD8	Moral justification	It is fine to violate some traffic or priority rules when interacting with AVs, to help society understand issues with their behaviour in such occasions	
MD9	Euphemistic labelling	It is justifiable to "test" the behaviour of AVs, as this would be a fun part of the process to better understand their capabilities	
MD11	Euphemistic labelling	It is not bad to take advantage of the cautious behaviour of an AV, as this would also help other human drivers to understand the range of their capabilities	
MD16	Euphemistic labelling	It would be fine to behave aggressively towards AVs out of curiosity to observe their behaviour	
MD5	Advantageous comparison	It is fine to drive aggressively against AVs, considering that other people have physically attacked (vandalised) them	Excluded in the final model
MD10	Advantageous comparison	It is not a big problem to behave aggressively towards AVs, thinking that other people behave aggressively towards human drivers	
MD17	Advantageous comparison	Taking advantage of the overly safe behaviour of AVs to move faster through traffic is not a great issue, considering that other people are running red lights or drink and drive	
MD3	Displacement of responsibility	If someone is running late to arrive at the workplace in time, it is fine to take advantage of the safe behaviour of the AVs around and weave through traffic	Excluded in the final model
MD7	Displacement of responsibility	Drivers cannot be blamed for driving competitively against AVs, if their friend pressured them into it	
MD18	Displacement of responsibility	If a driver is facing issues in personal life, it is justifiable to be aggressive towards AV that behaves overly safe	
MD4	Diffusion of responsibility	If the other drivers are violating some traffic priority rules when interacting with AVs, then it is fine if I would do so too	
MD12	Diffusion of responsibility	It is ok to behave aggressively towards AVs, if the other drivers are doing the same as well	
MD19	Distortion of consequences	It is fine to tailgate an AV to push it move faster, as this is not a very dangerous behaviour	
MD20	Distortion of consequences	It is fine to take risks when encountering an AV, as it is a machine supposed to always behave safely	
MD23	Distortion of consequences	It is not bad to take advantage of the cautious behaviour of an AV to move faster, if you do it in a safe way	
MD2	Attribution of blame	If human drivers take advantage of the overly cautions behaviour of AVs, it is probably because there is something wrong with it	
MD14	Attribution of blame	If AVs are overly cautious, then it is fine if drivers violate some traffic priority rules when interacting with them	
MD22	Attribution of blame	It is normal for other drivers to behave aggressively towards AVs, if these behave extremely cautiously	
MD13	Dehumanisation	AVs are just machines and it is normal if they are not treated as human drivers	
MD15	Dehumanisation	If AVs do not behave as humans, it is justifiable for human drivers to be assertive towards them	Excluded in the final model
MD21	Dehumanisation	It is fine to violate some traffic priority rules when interacting with AVs, to move faster, as human needs are more important	
MD24	Dehumanisation	There is no problem if human drivers behave aggressively towards an AV, as it is a machine that cannot react	

Table A.2 Items related to the presence or sharing the road with AVs

Item code	Item	References
AT1	AVs should never be on streets without a human on board	Nair and Bhat (2021), Rahman et al. (2021)
AT2	AVs should not be allowed in certain areas (e.g. around schools)	Nair and Bhat (2021), Rahman et al. (2021)
AT3	I would feel nervous to drive among unoccupied AVs without passengers inside	(Charness et al., 2018)
AT4	As much as infrastructure allows, AVs should be separated from the rest of the traffic	Nair and Bhat (2021)
AT5	The behaviour of AVs should always be courteous to other cars and road users even if this means that they will be reducing speed or braking often	Author-developed

Table A.3 Items related to perceived threats by AVs

Item code	Item
T1	I am concerned that equipment failures of AVs will cause more accidents than currently
T2	I am concerned that the computer systems of AVs may be hacked.
T3	I am concerned about data privacy issues related to AVs
T4	I am worried that many people will lose their jobs because of AVs
T5	Increase in AV services could have a negative impact on active mobility and public health
T6	Recent advances in AV technology are challenging the very essence of what it means to be human driver with a personal style.
T7	The increased presence of AVs could force human drivers to drive like robots diminishing personal styles
T8	The increased presence of AVs will change informal traffic rules and communication among drivers
T9	In the long run, the increased presence of AVs may lead to ban of human driving

Table A.4 Personality items

Item code	Item
A1	I am caring and attentive when it comes to other people.
A2	When someone needs assistance, I dis-continue my activities to help
A3	I do not want to deal with other people's problems
A4	I am considered a selfish and egotistical person
E1	I crave new experiences and excitement.
E2	I do not like to take risks
E3	I like to test myself in unknown situations
E4	I am not looking for excitement or adventures.
N1	I am often nervous, fearful, and anxious and I worry that something might go wrong
N2	I am a calm person who does not worry much about what may go wrong
N3	I am easily offended and I often feel angry and bitter; even small details may upset me
N4	I am a well composed person and it is difficult to upset or anger me

Table A.5 Normlessness items

Item code	Item
NL1	It is all right to do anything you want as long as you keep out of trouble
NL2	It is OK to get round laws and rules as long as you don't break them directly
NL3	If something works, it is less important whether it is right or wrong
NL4	Some things can be wrong to do even though it is legal to do it

Table A.6 MDSI items

Item code	Driving style	Item
AS1	Angry	I often blow my horn or 'flash' the car in front as a way of expressing my frustration.
AS2		I often swear at other drivers
AS3		When someone does something on the road that annoys me, I flash them with the high beams
AnS1	Anxious	I feel nervous while driving
AnS2		On a clear freeway, I usually drive at or a little below the speed limit
AnS3		Driving usually makes me feel frustrated
AnS4		I feel distressed while driving
AnS5		It worries me when driving in bad weather
CS1	Cautious	I drive cautiously
CS2		I am always ready to react to unexpected manoeuvres by other drivers
CS3		I tend to drive cautiously
SS1	High-velocity	I often purposely tailgate other drivers
SS2		When I am in a traffic jam and the lane next to mine starts to move, I try to move into that lane as soon as possible
SS3		When a traffic light turns green and the car in front of me doesn't get going, I just wait for a while until it moves
SS4		In a traffic jam, I think about ways to get through the traffic faster
SS5		When a traffic light turns green and the car in front of me doesn't get going immediately, I try to urge the driver to move on
PS1	Patient	At an intersection where I have to give right-of-way to oncoming traffic, I simply wait patiently for cross-traffic to pass
PS2		I base my behaviour on the motto "better safe than sorry"
RS1	Risky	I usually enjoy the sensation of driving on the limit (dangerously)
RS2		I like to take risks while driving
RS3		I usually enjoy the excitement of dangerous driving

Table A.7 Factor loadings of threat items (Rotated solution – Varimax rotation)

	Component	
	1	2
T1		0.750
T2		0.795
T3		0.763
T4	0.599	0.432
T5	0.554	0.477
T6	0.710	
T7	0.779	
T8	0.668	
T9	0.638	

Table A.8 Measurement model results

paths			Estimate	p-value
SS1	←	High-velocity	0.461	0.000
SS2	←	High-velocity	0.886	0.000
SS3	←	High-velocity	-0.624	0.000
SS4	←	High-velocity	0.884	0.000
SS5	←	High-velocity	0.980	0.000
RS1	←	Risky	1.058	0.000
RS2	←	Risky	0.765	0.000
RS3	←	Risky	1.127	0.000
A1	←	Altruism	0.902	0.000
A2	←	Altruism	0.652	0.000
A3	←	Altruism	-0.800	0.000
A4	←	Altruism	-0.884	0.000
NL1	←	Normlessness	0.257	0.005
NL2	←	Normlessness	0.583	0.000
NL3	←	Normlessness	0.803	0.000
AT1	←	Attitudes	0.831	0.000
AT2	←	Attitudes	0.727	0.000
AT3	←	Attitudes	0.874	0.000
AT4	←	Attitudes	0.703	0.000
AT5	←	Attitudes	0.361	0.000
T1	←	Threat realistic	0.509	0.000
T2	←	Threat realistic	0.457	0.000
T3	←	Threat realistic	0.445	0.000
T4	←	Threat identity	0.721	0.000
T5	←	Threat identity	0.662	0.000
T6	←	Threat identity	0.681	0.000
T7	←	Threat identity	0.801	0.000
T8	←	Threat identity	0.415	0.000
T9	←	Threat identity	0.521	0.000
Threat realistic	←	Threat	2.078	0.000
Threat identity	←	Threat	1.301	0.000
MD6	←	CR	0.385	0.000
MD8	←	CR	0.383	0.000
MD1	←	CR	0.282	0.000
MD11	←	CR	0.414	0.000
MD16	←	CR	0.423	0.000
MD9	←	CR	0.45	0.000
MD17	←	CR	0.358	0.000
MD10	←	CR	0.407	0.000
MD18	←	OM	0.255	0.000
MD7	←	OM	0.262	0.000
MD4	←	OM	0.272	0.000
MD12	←	OM	0.263	0.000
MD19	←	DC	0.292	0.000
MD20	←	DC	0.344	0.000
MD23	←	DC	0.304	0.000
MD2	←	BD	0.126	0.000
MD14	←	BD	0.22	0.000
MD22	←	BD	0.224	0.000
MD13	←	BD	0.176	0.000
MD21	←	BD	0.218	0.000
MD24	←	BD	0.229	0.000
CR	←	MD	1.435	0.000
OM	←	MD	2.161	0.000
DC	←	MD	1.701	0.000
BD	←	MD	2.625	0.000

Table A.9 Model covariances

Variables		Estimate	p-value
High-velocity	Risky	0.447	0.000
Altruism	Normlessness	-0.326	0.000
NL1	NL2	0.555	0.000