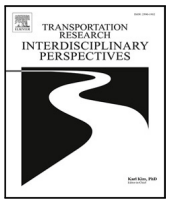
Transportation Research Interdisciplinary Perspectives 35 (2026) 101787

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.elsevier.com/locate/trip



Reasons and principles for automated vehicle decisions in ethically ambiguous everyday scenarios: The case of cyclist overtaking

Lucas Elbert Suryana [a,c,*], Simeon Calvert [a,c], Arkady Zgonnikov [c,b], Bart van Arem [a]

[a] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[b] Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands
[c] Centre for Meaningful Human Control, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:
Ethically ambiguous driving scenarios
Automated vehicles
Meaningful human control
Human reasons
Expert interviews
AV decision-making conceptual framework

ABSTRACT

Automated vehicles (AVs) consistently encounter ethically ambiguous situations in everyday driving, scenarios involving conflicting human interests and no clearly optimal course of action. While existing work often focuses on rare, high-stakes dilemmas (e.g., crash avoidance or trolley problems), routine decisions such as overtaking cyclists or navigating social interactions remain underexplored. This study addresses that gap by applying the tracking condition of Meaningful Human Control (MHC), which holds that AV behaviour should align with human reasons—the values, intentions, or expectations that justify actions. We conducted semi-structured interviews with 18 AV experts, who explained the reasons behind the considerations AV should make when planning a manoeuvre. Thirteen reason categories emerged, organised across normative, strategic, tactical, and operational levels. Using a case study on cyclist overtaking, we demonstrate how these reasons interact in practice and expose tensions in the decision-making process. Building on this analysis, we derive a reason-prioritisation principle grounded in the cyclist-overtaking scenario for AV behaviour in ethically ambiguous routine situations: prioritising vulnerable road users' safety above all, treating systemic safety and regulation as important but conditional, and permitting secondary values only when safety is not compromised. This hierarchy supports human-aligned behaviour by allowing pragmatic actions when strict legal compliance would undermine higher-priority values. Our findings offer conceptual principles intended to inform future research and design for AV decision-making in ethically challenging routine situations.

1. Introduction

Automated vehicle (AV) technology is advancing quickly, yet significant challenges remain, particularly when AVs must make decisions in ethically complex situations (Nyholm and Smids, 2016; Wang et al., 2020; Saber et al., 2024). Such situations arise when AVs must balance multiple priorities such as safety, efficiency, and compliance with societal expectations, ranging from minimising risks for all road users to resolving dilemmas involving conflicting values, interests, or trade-offs. Such conflicts often create ambiguity about the most appropriate course of action for AVs (Himmelreich, 2018; Bergmann, 2022). Addressing these dilemmas requires not only technical advancements, especially in planning and decision-making (Schwarting et al., 2018; Geisslinger et al., 2023), but also the development of clearer ethical principles.

1.1. Guidelines for AV in ethically straightforward situations

Stakeholders involved in AV development and regulation have proposed various recommendations and guidelines for how AVs should behave in safety-critical situations. Among these, the concept of roadmanship has been introduced as a guiding principle to ensure that AVs drive safely, avoid creating hazards, and respond effectively to hazards caused by others (Fraade-Blanar et al., 2018). Although less formalised than regulatory guidelines, roadmanship emphasises predictability and anticipatability in driving behaviour, similar to how a competent and careful human driver navigates traffic.

This concept aligns with the UNECE guidelines (UNECE, 2023), which present reference models of "competent and careful" drivers. These models serve as benchmarks for evaluating AV behaviour in safety-critical scenarios such as cut-ins, cut-outs, and lead-vehicle deceleration, reflecting how a skilled human driver would minimise risk. If an AV outperforms these reference models, it is considered safer than

* Corresponding author at: Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands.
E-mail address: l.e.suryana@tudelft.nl (L.E. Suryana).

https://doi.org/10.1016/j.trip.2025.101787
Received 14 June 2025; Received in revised form 3 December 2025; Accepted 8 December 2025
Available online 19 December 2025

a competent and careful human driver. Performance metrics include time headway, time-to-collision, and vehicle positioning, regardless of whether an accident is preventable (Olleja et al., 2025).

However, the safety-critical situations addressed by these models are ethically straightforward, since all parties benefit from the AV's behaviour. For example, the UNECE scenarios focus on mutually beneficial outcomes, such as avoiding collisions and reducing risk for all traffic participants. While these benchmarks offer important guidance for critical events, they do not address the full spectrum of ethically challenging situations that AVs may encounter in everyday driving.

1.2. Guidelines for AV in ethically ambiguous situations

In everyday driving, AVs frequently encounter ambiguous situations involving conflicting interests, where the optimal course of action is unclear. Routine traffic scenarios — such as approaching a crosswalk with limited visibility or making a left turn in the presence of oncoming traffic — exemplify such dilemmas (Himmelreich, 2018). These situations often require balancing safety, legal compliance, and traffic efficiency. For instance, when approaching a crosswalk with limited visibility, an AV must decide whether to prioritise slowing down to ensure pedestrian safety, at the expense of reducing traffic flow, or maintain speed to optimise mobility, potentially increasing risk.

Unlike human drivers, who make such decisions intuitively on a case-by-case basis, drawing on experience and an understanding of human behaviour, AVs must encode these trade-offs systematically across all vehicles. This raises a fundamental ethical challenge: how should AVs navigate scenarios where competing priorities conflict at a systemic level?

Some researchers have examined these challenges through the lens of the trolley problem, which explores moral judgements in life-and-death trade-offs. One approach is rooted in consequentialist ethics, where actions are evaluated based on outcomes. For example, Meder et al. (2019) uncovered nuanced ways in which individuals' moral judgements reflect consequentialist reasoning. In contrast, deontological perspectives prioritise adherence to moral principles, such as the protection of passengers. Liu and Liu (2021) found that participants favoured AVs programmed to protect their occupants at all costs. Complementing these, the MIT Moral Machine project adopted a virtue-based approach, highlighting significant cultural variation in ethical preferences for AV decision-making (Awad et al., 2018).

While these studies reveal diverse ethical perspectives, they also underscore the difficulty of developing a universal framework for AV behaviour. Critics argue, however, that the trolley problem has limited relevance to real-world driving, as AVs are unlikely to encounter such stark life-and-death scenarios during routine operation (Nyholm and Smids, 2016; Keeling, 2020). Instead, AVs more commonly face ethically ambiguous situations where the stakes are lower but the complexity is higher—making these scenarios critical for AV development and deployment (Himmelreich, 2018).

Recent recommendations from a European Commission Expert Group propose guidelines for addressing crash dilemmas through risk distribution and shared ethical principles (Bonnefon et al., 2020). Although these recommendations are valuable for ethically complex crash scenarios, they assume that crashes are unavoidable and do not fully address the challenges posed by routine, ethically ambiguous situations. This highlights the need for a more structured approach to navigating everyday ethical challenges, emphasising the importance of identifying principles that should guide AV behaviour in such contexts.

1.3. Guidelines based on the concept of meaningful human control

The concept of Meaningful Human Control (MHC) emerged in response to concerns over wrongful actions by automated weapon systems, which could create a responsibility gap (Asaro, 2012; Horowitz and Scharre, 2015). Without meaningful human control, such systems might make life-and-death decisions that result in unintended casualties or violations of international law, such as targeting civilians rather than enemy combatants while the chain of responsibility is not clear. This concern also applies to AVs, where failure to resolve ethical dilemmas, such as balancing pedestrian safety with traffic flow, could result in safety-critical failures.

Santoni de Sio and Van den Hoven (2018) subsequently developed a foundational theory defining what MHC entails. According to this theory, automated systems should be responsive to the reasons provided by human agents for their decisions, ensuring alignment with human values and intentions—a principle referred to as the tracking condition of MHC. Researchers have since proposed ways to operationalise MHC for AVs. For instance, Mecacci and Santoni de Sio (2020) explores how the concept can be applied to the AV domain, while Calvert and Mecacci (2020) builds a conceptual model for implementing MHC in the control systems of connected and autonomous vehicles (CAVs) in mixed traffic environments.

We argue that the tracking condition of MHC provides a suitable foundation for developing guidelines to support AV behaviour in ethically ambiguous routine driving scenarios. Unlike rigid ethical frameworks based on predefined rules (e.g., utilitarian or consequentialist principles), MHC emphasises understanding and incorporating the reasoning behind human decisions. In such scenarios, human drivers often know how to act, relying on their ability to interpret context, weigh competing considerations, and make judgements accordingly. This natural adaptability underscores the importance of designing AVs that can dynamically respond to human reasoning in specific contexts. By focusing on the tracking of human reasons, MHC enables AVs to align their actions with human values and intent, an essential capability in real-world driving environments where ambiguity and unpredictability are common.

1.4. Research gaps and objectives

This study addresses three main gaps in current AV ethics research:

- **Theoretical gap**: While several AV ethics frameworks, such as the Integrative Ethical Decision-Making Framework (Rhim and Urban, 2021), have been proposed to address moral dilemmas, they primarily focus on high-stakes or exceptional crash scenarios. As a result, they often overlook the ethical complexity inherent in routine driving situations. The concept of MHC, particularly its tracking condition (Santoni de Sio and Van den Hoven, 2018), offers a promising basis for addressing this gap by facilitating alignment between AV behaviour and human reasoning in everyday contexts. However, its application remains underexplored in ethically ambiguous routine scenarios. This study contributes by operationalising MHC to inform AV decision-making in such contexts.
- **Practical gap**: Although there is growing recognition that AVs must exhibit socially sensitive behaviour aligned with human values, expectations, and contextually appropriate responses (D'Amato et al., 2022; Lu et al., 2025), current frameworks lack a structured, empirically grounded set of human reasons for AVs to consider. This gap is especially evident in ethically complex, routine situations, such as overtaking a cyclist. This study addresses this gap by developing a principled framework that categorises expert-elicited reasons across normative, strategic, tactical, and operational levels of AV behaviour.
- **Methodological gap**: Existing research often relies on simulation models or moral vignettes, with limited empirical input from domain experts (Dubljević et al., 2023). While some studies incorporate expert perspectives in extreme scenarios (Milford et al., 2025), few systematically capture expert reasoning relevant to the daily ethically ambiguous conditions AVs encounter in everyday

real-world operation. This study fills that gap through qualitative interviews with 18 AV experts, using a structured analysis informed by the MHC framework.

This study focuses specifically on perspectives from experts based in Western countries and should therefore be interpreted as reflecting Western views on AV decision-making. Rather than assuming universality, we aim to contribute an empirical framework for understanding the types of human reasons that AVs should consider in ethically ambiguous, routine driving scenarios. These region-specific insights form the basis for our methodological approach, which builds on the tracking condition of MHC to address the identified research gaps. Building on this foundation, the study has two primary objectives:

- **Objective 1**: To gather the reasons provided by AV experts regarding the factors they believe AVs should consider when planning a manoeuvre.
- **Objective 2**: To derive reason-based principles for AV decision-making by analysing the ethically ambiguous routine scenario of overtaking a cyclist using expert-derived reasons from Objective 1.

To address Objective 2, we adopt a two-step approach. In the first step, we classify the reasons elicited from AV experts into groups, aiming to identify underlying principles. In the second step, a case study of an ethically ambiguous routine scenario, such as overtaking a cyclist, is presented to the experts. They are asked to provide recommendations on the manoeuvre decisions AVs should make in such situations and to explain the reasoning behind their recommendations. These reasons are then mapped back to the classifications from the first step to uncover relationships and derive expert-informed guidelines for AV behaviour.

2. Reasons that influence considerations for AV manoeuvre planning

2.1. Methods

To identify the reasons automated vehicles (AVs) should track in ethically ambiguous routine driving situations, we conducted expert interviews. This section describes the selection of experts, recruitment procedures, interview protocol, and analytical approach used to extract and categorise these reasons.

2.1.1. Expert participants

- *Selection Criteria*
  To ensure that the study reflected insights from individuals with substantive knowledge of AV systems, we employed a purposive sampling strategy, complemented by snowball sampling. Initial participants were selected through the professional networks of the authors and evaluated based on their publication records, institutional affiliations, and topic relevance, as reflected in publicly available sources such as Google Scholar profiles. This approach enabled the identification of experts with demonstrable contributions to AV-related research and development, in line with accepted practices in qualitative transportation studies. For example, Ma and Feng (2024) recruited AV professionals through LinkedIn based on their hands-on experience with automated systems, while Hilgarter and Granig (2020) employed purposive sampling in a real-world AV deployment by selecting participants immediately after they experienced an autonomous shuttle ride. Similarly, our approach aimed to ensure that participants had domain-specific expertise in AVs and were capable of contributing informed reasoning about AV decision-making.
  Subsequent participants were recruited via expert referrals following early interviews. This snowball sampling method enabled us to reach additional individuals working in specialised domains who may not have been immediately visible through conventional directories. Comparable combined strategies have been used in AV-focused qualitative studies to capture diverse, high-level perspectives from academia, industry, and government (Milford et al., 2025).
- *Recruitment*
  Participants were recruited via personalised email invitations. Each email included a brief overview of the study, highlighting its focus on understanding trade-offs in motion planning for overtaking scenarios involving automated vehicles (AVs). We clarified that although the study focused on motion planning, participation was not limited to specialists in that area; instead, we sought a broad range of perspectives from individuals involved in AV ethics, design, policy, and engineering.
  The email also outlined the interview format and logistics: semi-structured, approximately 45–60 min in duration, conducted via Zoom, and optionally recorded with participant consent. The voluntary nature of participation and the right to withdraw at any time were clearly communicated. Recruitment and study procedures were approved by the Human Research Ethics Committee (HREC) of Delft University of Technology (ID: 132530).
- Participant Profile
  The final sample consisted of 18 expert participants from seven countries, representing a range of perspectives on AV development. Of the 35 experts initially contacted — 14 from the United States, 13 from the Netherlands, four from the United Kingdom, and one each from Italy, Belgium, Israel, and Japan — 18 agreed to participate, resulting in a response rate of 51%. All participants were informed of the study objectives and provided consent prior to their involvement.

The participant profile reflects diverse institutional affiliations and technical backgrounds. As shown in Table 1, participants were drawn from academia (n = 12) and industry (n = 6), with disciplinary expertise in motion planning, human factors, ethics, behavioural science, and legal policy. Based on self-reported experience, participants had on average more than five years of direct involvement with automated vehicle development. This diversity of expertise and roles contributed to a rich and multidimensional set of perspectives on AV decision-making in motion planning contexts.

The sample size of 18 experts aligns with prior qualitative studies that employ in-depth expert interviews in the domains of automated vehicles, where 9 to 19 participants are often sufficient to achieve conceptual saturation (Dreger et al., 2020; Tabone et al., 2021; Beringhoff et al., 2022; Lee et al., 2020; Swain et al., 2023; Habibullah et al., 2024). Although our sample size was determined by expert availability rather than a predefined saturation threshold, we conducted a retrospective assessment to evaluate whether thematic saturation was likely achieved. We tracked the emergence of new reason categories across interviews and observed that all categories described in Section 2.2 were identified by the 14th interview. The final four interviews introduced no new categories, suggesting that the major themes had stabilised. This provides additional confidence in the adequacy of the sample size within the scope and aims of this study.

2.1.2. Questionnaire design

This study used a semi-structured interview protocol, operationalised through a structured questionnaire administered synchronously during interviews. The instrument was informed by the tracking condition of the Meaningful Human Control (MHC) framework. This condition holds that automated systems should respond to relevant human reasons (Santoni de Sio and Van den Hoven, 2018). In this article, we define ‘‘reasons’’ as normative reasons or factual considerations that justify particular actions, rather than motivational reasons, following the distinction outlined by Veluwenkamp (2022).

According to the MHC framework, reasons relevant to automated vehicle (AV) decision-making can be grouped into four categories: moral, strategic, tactical, and operational. Moral reasons pertain to ethical principles or social norms (e.g., fairness, harm avoidance).

Table 1
Overview of Expert Participants by Sector, Country, and Expertise.

ID	Country	Expertise	Role
Academia			
1	Netherlands	Human-AI interaction, ethics	Researcher
2	US	Technical validation, travel behaviour	Researcher
3	US	AV safety validation	Researcher
4	Netherlands	Motion planning algorithms	Researcher
5	Netherlands	Road users and infrastructure perspectives	Researcher
6	Netherlands	Ethics of AI	Researcher
7	UK	Modelling human behaviour	Researcher
8	Netherlands	Social science of behaviour	Researcher
9	UK	AV user experience	Researcher
10	Israel	Public perception and AV ethics	Researcher
11	UK	Human factors in transport	Researcher
12	Netherlands	Legal aspects of AV	Researcher
Industry			
13	UK	AV safety and assurance	Consultant
14	Netherlands	Traffic psychology	Psychologist
15	US	Software quality assurance	Engineer
16	US	Driving strategy, business development	Consultant
17	US	AV safety	Consultant
18	US	Human factors	Researcher

Strategic reasons relate to long-term planning goals (e.g., minimising travel time). Tactical reasons involve interactions with other road users (e.g., overtaking or yielding), while operational reasons concern real-time control actions (e.g., braking, steering). These categories provided the conceptual basis for the questionnaire, which consisted of five main parts:

- **Part 1 (Questions 2–4)**: Exploration factors that influence AV manoeuvre planning.
- **Part 2 (Questions 5–9)**: Evaluation of an ethically challenging real-world AV scenario involving a cyclist, asking participants to identify and assess reasons relevant to decision-making.
- **Part 3 (Questions 10–13)**: Ranking of predefined reasons, enabling participants to indicate the most appropriate decision in the given scenario.
- **Part 4 (Questions 14–19)**: Evaluation of alternative AV decisions, using time-based assessments to examine how stakeholder reasons were addressed in a revised scenario.
- **Part 5 (Questions 20–21)**: Evaluation of how AVs might interpret stakeholder intentions and manage potential conflicts.

The protocol was informally piloted with five PhD researchers working on topics related to AVs to ensure question clarity and relevance, after which wording adjustments were made prior to data collection. Building on this refined protocol, this section focuses on participants' responses to Questions 2–4, which relate to Objective 1 and are presented in Table 2. For completeness, the full set of questions is provided in Appendix Table A.6, while details of how both open- and closed-ended questions were formatted and administered are described in Section 2.1.3.

Using these questions as the starting point, we elicited expert views on the kinds of reasons AVs should respond to. Participants answered open-ended questions designed to explore what factors should be considered in AV manoeuvre planning. Their responses to Questions 2–4 formed the basis for a theory-driven qualitative coding process aimed at identifying the types of reasons referenced and mapping them to the four categories outlined in the MHC framework. The analysis procedure is detailed in Section 2.1.4.

2.1.3. Procedure

The interviewer conducted all interviews online using Microsoft Teams, with audio, video, and automated transcriptions recorded for analysis. Each session followed a predefined protocol consisting of both open-ended and closed-ended questions presented through Qualtrics (https://www.qualtrics.com). In this synchronous format, the interviewer opened the questionnaire on their own screen and shared it with participants via screen share. A direct, non-recorded link to the same questionnaire was also provided, enabling participants to reread questions and revisit previous items independently. This link also allowed them to rewatch embedded videos, which was especially helpful in cases of video lag caused by internet issues. The interviewer read each question aloud and asked participants to respond verbally. For closed-ended questions, the interviewer recorded participants' responses directly into the questionnaire, with the input visible to participants via screen share for confirmation. To prevent duplicate data, the interviewer explained that any responses submitted via the shared link would not be recorded or considered in the analysis. The interviewer also managed the structure and flow of each session. Participants were informed of this format in advance, and no concerns or discomfort were reported during or after the interviews.

This synchronous format enabled the researcher to provide immediate clarification when needed and ensured that participants responded to questions in the intended sequence. It also helped maintain consistency across interviews, as all participants saw and heard the same content in the same order at a similar pace. The researcher did not comment on or react to participants' responses and refrained from offering prompts or interpretations, intervening only when participants explicitly requested clarification. This neutral and minimal involvement allowed for observation of subtle cues, such as hesitation or clarification requests, that could enrich qualitative analysis. This method aligns with best practices for structured qualitative interviewing and has been applied in prior research (Longhurst and Johnston, 2023; Beringhoff et al., 2022; Nordhoff et al., 2023).

2.1.4. Data analysis

- Coding Framework

  We used directed content analysis to analyse expert responses, using the Meaningful Human Control (MHC) framework (Mecacci and Santoni de Sio, 2020) as the initial coding structure. This framework distinguishes reasons according to their position on a temporal scale — that is, how close or distant they are from influencing an action — and organises them into four layers. These layers were used as deductive codes to classify the reasons experts provided for expected AV manoeuvre planning. The layers are detailed as follows:

  - **Normative reasons:** Motivations grounded in moral values, legal rules, or social expectations that guide what ought to be done. These are abstract, long-term in scope, and typically shaped by institutions or broader societal expectations.
  - **Strategic reasons:** Motivations or intentions related to high-level goals and long-term plans, such as deciding where to go or what outcome to achieve. These are moderately abstract, span longer durations, and are usually attributed to the driver as planner.
  - **Tactical reasons:** Motivations or intentions that guide short-term manoeuvring decisions in response to changing circumstances. These are more concrete and informed by the immediate driving context.
  - **Operational reasons:** Immediate motivations or intentions that correspond directly to moment-by-moment physical actions. These are highly specific and implemented by the AV system or human driver in response to moment-to-moment environmental cues.

  Based on the four layers of reasons, we created a coding matrix. Interview responses to Questions 2–4 were segmented into individual statements, which were then coded according to the type of category expressed. Once all statements were categorised, we conducted an inductive thematic analysis within each category to identify more specific sub-themes.

Table 2
Interview questions relevant to objective 1.

No.	Interview question
Question 2	What should automated vehicles (AVs) consider when planning a manoeuvre? Please give one example in as much detail as possible.
Question 3	Which moral aspects do you believe AVs should consider when planning a manoeuvre?
Question 4	How might these aspects affect the manoeuvre plan?

This process enabled us to identify which layers of reasons were most frequently cited by experts and to characterise the diversity of reasons underlying what the AV should consider when planning a manoeuvre. For example, reasons in the moral layer often referred to legal compliance or fairness, whereas strategic reasons focused on acceptance and efficiency concerns. Tactical reasons addressed situational decision-making, and operational reasons emphasised vehicle control. When statements reflected more than one type of reason, cross-coding was used to preserve interpretive nuance.

These four layers of reasons were used exclusively to code expert reasons. The subsequent analysis of what experts believed the AV should do, presented later in Section 2.2, was carried out separately as an exploratory interpretive step, using behavioural levels (normative, strategic, tactical, and operational) (Calvert and Mecacci, 2020) that were not part of the coding framework. Our approach, grounded in the theoretical framework of MHC, is consistent with other AV studies employing theory-driven content analysis. For instance, Aasvik et al. (2025) used a similar method to analyse public trust in autonomous shuttles, while Suryana et al. (2025) applied the MHC framework to explore interview data in relation to the tracking and tracing conditions.

- *Qualitative Content Analysis Procedure*

  To apply the Meaningful Human Control (MHC) framework in a structured and transparent manner, we conducted a qualitative content analysis that combined theory-driven and data-driven steps. We began by segmenting interview transcripts into individual response units. Each unit was analysed to identify four key components: (1) the AV behaviour being recommended (consideration), (2) the justification for that behaviour (reason), (3) the human agent associated with the reason, and (4) the corresponding layers of reasons: normative, strategic, tactical, or operational.

  We explicitly distinguished between considerations and reasons. "Considerations" refers to the specific behaviour that the AV is expected to perform (e.g., "the AV should slow down near pedestrians"), whereas "reasons" are the human-orientated justifications for those behaviours (e.g., "to ensure the safety of vulnerable road users"). Each reason was then evaluated according to the layers of reasons. This involved examining its temporal scale and the associated human agents to whom the reason was attributed. When a reason encompassed multiple types of justification, such as combining moral fairness with strategic efficiency, it was assigned to more than one MHC category.

  Our approach recognised that reasons could be either explicitly stated in the data or logically inferred from context. Explicit reasons were identified when participants directly articulated the justification for their statements. In other cases, implicit reasons were inferred based on the surrounding narrative. This approach draws on principles of latent content analysis, in which underlying meanings are interpreted beyond the literal language used. Latent content, as defined by Graneheim and Lundman (2004), refers to the deeper meaning embedded in a text, especially important when participants allude to motivations or norms without stating them directly. Building on this, scholars such as Vaismoradi et al. (2013) and Krippendorff (2018) have emphasised how interpreting latent content can uncover implicit yet meaningful patterns within qualitative interview data.

  Following the identification and classification of reasons, responses that addressed similar topics were grouped into broader thematic categories. This step enabled us to organise the data into a set of distinct reason types, each of which was then linked to the appropriate MHC category or categories.

- *Inter-coder Reliability*

  To ensure the reliability and transparency of the coding process, we adopted a multi-stage, consensus-based approach. First, the one of the author compiled an initial list of reasons or expectations for how AVs should act, based on expert responses. This involved interpreting each expert's response and identifying distinct reasons or expectations expressed.

  Drawing on the classification framework proposed by Mecacci and Santoni de Sio (2020), we categorised each reason or expectation into four categories: moral, strategic, tactical, and operational. Following this categorisation, two authors of this paper independently coded each item to one of the four categories to ensure consistency and analytical rigour. The initial coding was done independently using the same list, allowing for a direct comparison of interpretations.

  We calculated inter-coder agreement using Cohen's kappa, based on binary coding of whether each of the four categories was applied. Agreement varied by category and coder pair. For example, the **moral** category showed substantial agreement ($\kappa = 0.77$ for Coder 2 vs Coder 1), while the **tactical** category showed moderate to substantial agreement ($\kappa = 0.62$ for Coder 1 vs Coder 3). In contrast, agreement was lower for the **strategic** ($\kappa = 0.11$–0.22) and **operational** ($\kappa = 0.00$–0.18) categories, indicating greater interpretive variability in these dimensions. This aligns with recent work showing that annotator disagreement can itself be a signal of underlying subjectivity, especially when arguments are tied to human values (Homayounirad et al., 2025). Most discrepancies arose from ambiguous phrasing in participant responses or overlapping themes across categories (e.g., a reason could plausibly be interpreted as both moral and strategic). Operational justifications were particularly prone to divergent interpretation, likely due to their context-specific nature. These differences were discussed during a follow-up meeting until full consensus was reached, and no disagreements remained unresolved.

  After reaching agreement at the category level, we collaboratively developed sub-categories within each of the four main categories, refining the framework through discussion. During this process, one co-author noted that some sub-categories could conceptually belong to more than one main category, depending on context and interpretation. These overlaps were acknowledged and addressed through further discussion, with final coding decisions made by consensus. This approach enhanced the clarity and consistency of the framework while minimising individual bias. It also reflects established best practices for investigator triangulation and consensus coding in directed content analysis (Hsieh and Shannon, 2005; Hill et al., 2005; Campbell et al., 2013).

2.2. Results

In response to Questions 2–4, most experts described hypothetical traffic situations and outlined the considerations and actions that automated vehicles (AVs) should take before executing a manoeuvre in these contexts. They also provided explanations (reasons), articulating

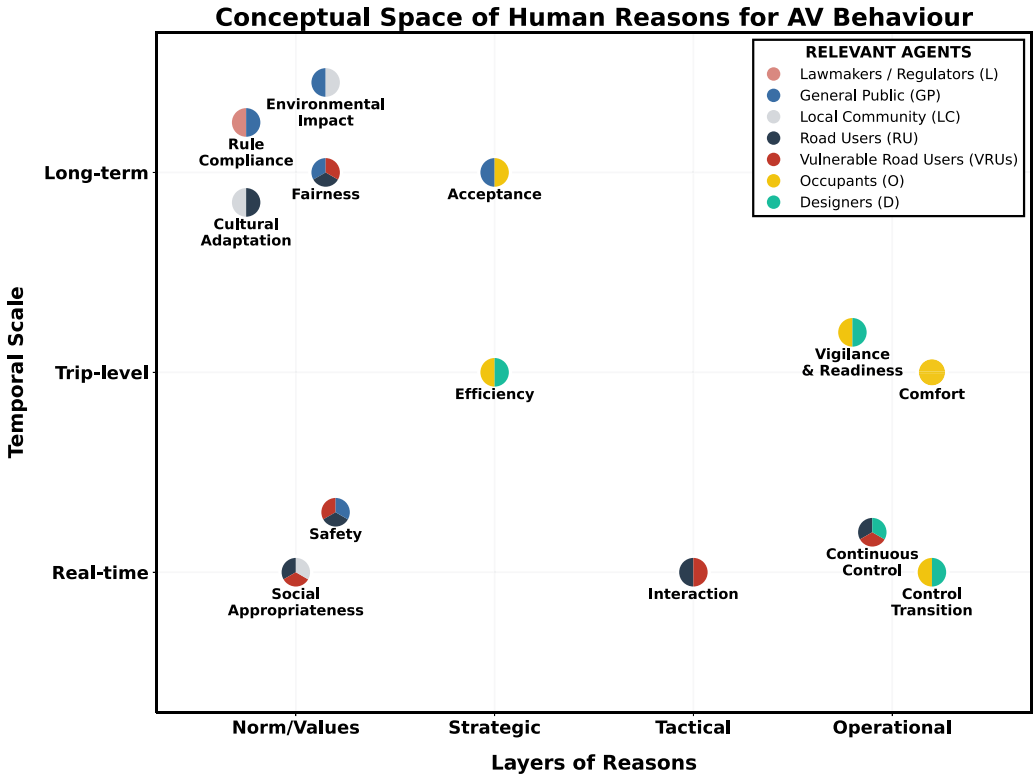


Fig. 1. Conceptual positioning of the thirteen reason categories along two dimensions: layers of reasons (norm/values, strategic, tactical, and operational) and temporal scale (real-time, trip-level, and long-term). Relevant human agents are shown using colour, indicating the individuals or groups referenced in the justification for why the AV should behave in a particular way. This representation clarifies the conceptual focus of each category.

why the considerations they mentioned were important and why they believed AVs should exhibit particular behaviours. Based on thematic analysis, we identified thirteen distinct categories of reasons that span across four layers of reasons used by experts to justify AV behaviour. These categories demonstrate that experts' expectations about what the AV should consider or do rely on diverse human reasons. Some categories may seem similar when viewed only by name, as labels necessarily compress complex reasoning into short descriptors. Examining the temporal scale and the layer to which each reason is classified helps reveal their conceptual differences. To clarify differences in layer and temporal scale, we developed a conceptual representation that positions each reason category along two dimensions.

Fig. 1 presents this conceptual space, showing how the thirteen reason categories distribute across their levels of reasons and their temporal scale, with colours indicating the relevant human agents referenced in each reason. The visual organisation highlights that distinctions among categories are more apparent when considering all dimensions together (layers of reasons, temporal scale, and the relevant human agents referenced) rather than relying solely on category labels.

Following this visual clarification, Table 3 summarises each reason category and shows how we organised the associated reasons across four behavioural levels (normative, strategic, tactical, and operational), providing a clearer structure for understanding the types of actions experts described. We now describe each of the thirteen reason categories in turn.

2.2.1. *Rule compliance*

Rule compliance refers to how the AV follows formal traffic laws and signals. Experts described this category as centred on legal duties that define how road users are expected to act by regulators and the general public.

One expert explained that the AV should follow traffic signals and road rules because these reflect shared legal norms that shape public expectations (ID 1). Another expert said the AV should obey traffic signals and signage to respect legal obligations and maintain rule-following behaviour in traffic (ID 3).

An expert also discussed how the AV should respond when dangerous rule violations occur. They suggested that the AV could warn nearby vehicles and record serious breaches, such as running a red light, so that responsibility can be assigned when needed (ID 16). This was described as supporting safety by providing clear information for those who enforce the rules.

Experts further commented on the connection between rule-following and responsible behaviour. One expert said the AV should follow road rules and act like a "good driver", meaning that rule compliance is part of demonstrating responsible conduct (ID 18). Another expert argued that the AV should follow traffic rules even when moral guidance is unclear, since rules remain a stable reference in those situations (ID 11).

A final contribution highlighted that strict compliance can reduce situations where the AV faces unclear or conflicting choices. One expert explained that following traffic law helps avoid ambiguity because the rules offer a widely accepted basis for deciding how to act (ID 12).

In summary, rule compliance was described by experts as relevant across multiple behavioural levels. At the normative level, rule compliance was described as the AV behaving in accordance with legal expectations established by lawmakers and society. At the tactical level, rule compliance appeared in discussions of visible behaviours such as obeying signals and signage, which other road users rely on to understand how the vehicle will act. At the operational level, experts emphasised that the AV must consistently execute rule-following actions, ensuring that its behaviour aligns with legal requirements in real time. This shows that rule compliance depends on how the AV connects widely accepted legal expectations with the behaviours that road users observe during interaction.

2.2.2. Social appropriateness

Social appropriateness refers to how AV behaviour is interpreted by other road users during real-time interactions. It concerns whether the AV behaves in ways that people on the road recognise as appropriate. Although these judgements occur in the moment between the AV and nearby pedestrians, cyclists, or drivers, they are shaped by broader expectations within society and local communities about how road users should behave.

Several experts explained that social appropriateness depends on how traffic rules are interpreted in everyday practice. One expert noted that AVs should evaluate their actions in light of how other road users interpret legality, since rule-following can influence whether behaviour appears safe or fair (ID 1). Another expert emphasised that certain actions, such as stopping at a red light even when there is no immediate hazard, express respect for societal expectations (ID 6). These views show that compliance with traffic rules has social meaning in addition to legal meaning, and that people often use visible rule-following to judge whether behaviour aligns with shared expectations.

Experts also pointed to interpersonal qualities of behaviour. One expert stated that AVs should reflect human social values such as politeness in their driving style (ID 7). Politeness, as used by the expert, relates to driving in ways that other road users experience as considerate and cooperative. This includes avoiding abrupt or intrusive manoeuvres that could disrupt others' movement or sense of mutual accommodation.

Another expert highlighted that appropriateness also involves accounting for local norms. They explained that AVs may need to balance broad ethical principles, such as avoiding harm, with practices that are specific to local communities (ID 8). This indicates that legitimate behaviour is partly context-dependent: an action may align with expectations in one community but not in another, and recognising this variation is important for maintaining trust.

Finally, experts noted that the appearance of behaviour matters. One emphasised that AVs should avoid actions that feel threatening to others (ID 11). This treats social appropriateness not only as a matter of rule interpretation or courtesy, but also as avoiding behaviours that could create fear or discomfort among pedestrians or other road users.

Taken together, these expert perspectives show that social appropriateness involves AV behaviour that aligns with how people expect road users to act. At the normative level, experts described socially appropriate behaviour as acting in ways that reflect the expectations of society and local communities regarding respectful and fair conduct. At the tactical level, they emphasised observable behaviours—such as courteous driving, predictable signalling, or avoiding threatening movements—that pedestrians, cyclists, and other drivers interpret in real time. This shows that social appropriateness depends on linking widely shared expectations about appropriate conduct with the moment-to-moment behavioural cues that road users rely on during interaction.

2.2.3. Environmental impact

Environmental impact concerns how the AV's behaviour can contribute to or reduce environmental harm. The public and local communities generally expect AVs to demonstrate environmentally responsible driving behaviour, including reducing emissions, supporting smooth traffic flow, avoiding unnecessary fuel consumption, and acting in ways that align with community sustainability priorities.

One expert described environmental impact as a reason for AV behaviour. They recommended that AVs should minimise traffic disruption and reduce emissions, noting that smoother movement helps avoid inefficient stopping patterns and cuts fuel consumption (ID 17). The expert also linked these actions to safety, explaining that unstable flow can increase the likelihood of unsafe interactions. The justification indicates that environmental concerns are tied to how AV behaviour influences broader conditions on the road.

These expert contributions correspond to the behavioural levels used in our analysis. Environmental impact involved several behavioural levels. At the normative level, experts described environmentally responsible behaviour as acting in ways that reflect public expectations about sustainability and reduced harm. At the strategic level, they emphasised planning AV behaviour to support steady movement and minimise unnecessary energy use. These perspectives show that environmental impact depends on how the AV links broad sustainability expectations with behavioural planning that influence emissions.

2.2.4. Fairness

Fairness concerns how the AV should treat different people and groups without bias. Experts described this category as relating to how AV behaviour can avoid discrimination, protect those who are more vulnerable, and ensure that safety and access are not distributed unequally.

One expert stated that AVs should adapt their driving behaviour to account for the vulnerability of certain road users (ID 1). The expert explained that this is important for avoiding harm to people who face higher physical risk, such as pedestrians or cyclists. The focus was on recognising differences in exposure and adjusting behaviour accordingly.

Another expert discussed fairness in relation to infrastructure. They argued that AV-only lanes and similar designs could create inequality by restricting access or shaping mobility conditions in ways that disadvantage some groups (ID 5). The expert treated the avoidance of such outcomes as part of ensuring that socio-economic status does not determine who benefits from automated systems.

An additional expert emphasised that AVs should avoid causing harm to nearby people and animals (ID 13). Although the reason refers broadly to welfare rather than to a specific social group, it was framed as a matter of protecting those who may be vulnerable during interaction with the vehicle.

Experts also linked fairness to system design. One expert noted that AVs should not blame or penalise inattentive users and should be designed to provide fallback safety even when people make mistakes (ID 9). In this view, fairness concerns responsibility allocation and the need to ensure that design limitations do not disproportionately affect certain users.

Two experts described fairness more explicitly as a matter of equal treatment. One argued that pedestrians and occupants should be treated equally regardless of wealth or identity (ID 10). Another stated that safety standards should not vary with a user's socio-economic status (ID 12). These accounts highlight fairness as a requirement that all users receive the same protection, independent of individual characteristics.

These expert accounts align with the behavioural levels used in our analysis. Fairness emerged across multiple behavioural levels. At the normative level, experts described fairness as requiring AV behaviour that provides equal consideration to all people and avoids discriminatory outcomes. At the strategic level, they emphasised behavioural decisions shaped by system and infrastructure design, influencing who benefits from the system and how access and safety features are distributed. At the tactical level, their examples highlighted real-time behavioural adjustments—such as protecting vulnerable road users or preventing harm during interaction. Overall, these perspectives show that fairness depends on recognising differences in vulnerability and ensuring that AV behaviour does not produce outcomes that advantage or disadvantage particular users based on their social or economic position.

2.2.5. Efficiency

Efficiency concerns how the AV supports effective movement through traffic while aligning with the user's travel goals. Experts described this category as relating to how the AV manages speed and flow to enable timely travel without introducing unnecessary risk or disruption. They highlighted that efficiency depends both on occupant expectations

Table 3
Expert-elicited reasons, organised by reason category (rows) and behavioural levels (columns): normative, strategic, tactical, and operational. Each cell summarises what experts believed the AV should do or consider in that category. *Not mentioned* indicates no corresponding expert response for that level. Parenthetical codes indicate the human agents associated with behaviour at that level: (L) = Lawmaker/Policymaker/Regulator (RU) = Road Users, (VRU) = Vulnerable Road Users, (D) = Designer, (O) = Occupant, (GP) = General Public, (LC) = Local Community.

Reason Category	Moral/Normative	Strategic	Tactical	Operational
Rule Compliance	- Follow traffic rules as societal duty and moral baseline[(L)] (ID 1,3,11,12)	*Not mentioned*	- Behave like a predictable "good driver"[(RU)] (ID 18)	- Alert and log red-light running[(D)] (ID 16)
Social Appropriateness	- Visible compliance signals fairness, integrity, cultural respect[(RU,GP,LC)] (ID 1,6,8)	*Not mentioned*	- Courteous, non-threatening driving to gain trust[(RU)] (ID 7,11)	*Not mentioned*
Environmental Impact	- Reduce emissions for sustainability[(GP, D)] (ID 17)	- Smooth traffic flow to cut fuel/stop-and-go[(D)] (ID 17)	*Not mentioned*	*Not mentioned*
Fairness	- Equal treatment[(GP, RU)] - Protect VRUs and animals[(GP, RU)] - Designers bear safety duty[(GP, RU)] (ID 9,10,13)	- Avoid exclusionary AV-only lanes[(D)] - Ensure safety functions for all[(D)] (ID 5,12)	- Adjust driving to shield vulnerable users[(VRU)] (ID 1)	*Not mentioned*
Efficiency	*Not mentioned*	- Balance speed with trip efficiency[(D,O)] - Honour faster-arrival goals[(D,O)] (ID 1,6)	- Avoid over-caution that disrupts flow[(RU)] (ID 16)	*Not mentioned*
Acceptance	*Not mentioned*	- Ensure passenger safety and comfort for acceptance[(O)] (ID 5)	- Drive in ways passengers find comfortable and acceptable[(O)] (ID 5)	*Not mentioned*
Cultural Adaptation	- Uphold legal standards despite unsafe local habits[(L, LC)] (ID 17)	- Adapt to region-specific traffic behaviours[(D, RU)] (ID 8)	- Adapt yielding/right-of-way to local norms[(RU)] (ID 8)	*Not mentioned*
Safety	- Adapt speed if safer[(D, GP)] (ID 17)	- Limit speed in pedestrian zones[(RU)] - Minimise manoeuvres[(RU)] - Plan for sensor-failure contexts[(RU)] (ID 4,13,15)	- Safe overtake/merge[(RU)] (ID 1) - Anticipatory VRU buffers[(VRU)] (ID 2) - Early hazard signal and brake for violator[(RU)] (ID 14,16) - Extra buffer when visibility blocked[(RU)] (ID 2)	- Detect and signal sudden obstructions early[(D)] (ID 14)
Interaction Management	- Transparent communication upholds fairness[(D, GP)] (ID 15)	- Distinguish reactive vs. goal-driven actions[(D)] (ID 13)	- Detect users, infer intent, and signal manoeuvres[(O)] (ID 3,9,15) - Predict pedestrian motion[(VRU)] (ID 4,7,8) - Cooperative merge/influence traffic[(RU)] (ID 7,18) - Decelerate to signal crossing[(RU)] (ID 11)	*Not mentioned*
Comfort	*Not mentioned*	- Integrate safety, efficiency, comfort[(D, O)] (ID 7)	- Maintain comfort to avoid overrides[(O)] (ID 5) - Smooth merge/decel[(O, RU)] (ID 13) - Avoid harsh braking and needless yielding[(O)] (ID 9)	*Not mentioned*
Continuous Control	*Not mentioned*	*Not mentioned*	*Not mentioned*	- Maintain continuous environmental monitoring[(D)] (ID 14)
Control Transition	*Not mentioned*	*Not mentioned*	*Not mentioned*	- Give clear, timely takeover warning[(D, O)] (ID 5)
Vigilance & Readiness	*Not mentioned*	*Not mentioned*	*Not mentioned*	- Do not depend on continuous driver alertness[(D, O)] - Manage engagement[(D, O)] (ID 14)

for reaching destinations reliably and on design decisions that enable smooth and stable movement within traffic.

One expert explained that AVs should balance speed and responsiveness with the need to maintain efficiency throughout the trip (ID 1). The expert noted that efficiency should not come at the expense of safety, indicating that the AV must manage its pace in a way that supports steady progress without creating hazards.

Another expert described efficiency in relation to user goals. They stated that AVs should consider the user's intention for the trip — such as wanting to reach a destination sooner — when selecting actions (ID 6). This perspective connects efficiency with the passenger's preferences and how the AV plans its route or driving behaviour to match them.

A further expert highlighted the effect of overly cautious behaviour on traffic flow. They recommended that the AV should avoid unnecessary hesitation that disrupts surrounding traffic or leads to inefficient movement patterns (ID 16). This reflects a view that efficiency includes how the AV interacts with others and maintains stability within the broader flow of travel.

These expert views correspond to the behavioural levels used in our analysis. Efficiency appeared at the strategic and tactical behavioural levels. At the strategic level, experts described behaviours related to planning and routing that balance speed with trip efficiency and support user goals, reflecting the involvement of both designers and occupants. At the tactical level, they highlighted real-time driving behaviour that avoids overly cautious actions or unnecessary hesitation that could disrupt surrounding traffic, which road users depend on for stable flow.

2.2.6. Acceptance

Acceptance concerns how the AV's behaviour is experienced by passengers and by society over longer time scales. While at first glance this may appear similar to social appropriateness, the focus here is different: this category does not address how other road users interpret the AV in real-time interactions, but rather whether people feel comfortable trusting and adopting AV technology at all.

One expert explained that AVs should balance safety with long-term user comfort and acceptance (ID 6). In this view, safety remains essential, but acceptance also depends on whether the vehicle behaves in a way that passengers find comfortable and reassuring over time. The expert's description indicates that user acceptance involves both comfort considerations and a sustained sense that the AV behaves safely across repeated experience. Because acceptance develops over time, this category concerns how passengers and society evaluate the AV's behaviour beyond any single moment in traffic.

These statements fit the behavioural levels applied in our analysis. Acceptance appeared at the strategic and tactical behavioural levels. At the strategic level, experts emphasised behaviour related to balancing comfort and safety over time so that people can rely on the system. At the tactical level, they highlighted specific driving behaviours that influence passenger comfort and confidence during use.

2.2.7. Cultural adaptation

Cultural adaptation concerns how the AV should account for location-specific behavioural expectations and informal practices in different traffic environments. Experts described this category as relating to the way norms vary across places and how these variations shape what local road users and local community expect from an AV.

One expert explained that AVs should adapt their behaviour to reflect local cultural norms and road conventions so that their actions match what people in that environment consider appropriate (ID 8). This includes modifying responses to align with expectations that differ across locations, such as how pedestrians or cyclists typically behave in that region.

The same expert illustrated how local expectations influence right-of-way decisions by comparing cyclist behaviour in the Netherlands and the United States. They explained that if behaviour were guided only by a rule such as "do not harm", a Dutch cyclist would stop whenever a pedestrian might cross their path. In practice, cyclists in the Netherlands usually continue unless the pedestrian clearly commits to entering the road. In California, however, continuing in this way could be viewed as improper or even immoral. The expert used this to explain that AVs should adjust their responses based on location-specific expectations about yielding and movement patterns (ID 8).

Another expert highlighted situations where local driver behaviour may be inconsistent with formal rules. For example, they noted that an AV should obey traffic laws even when local drivers act unpredictably in features such as roundabouts (ID 17). The expert emphasised that the AV should not reproduce informal or unsafe habits, even when these habits are common, but should still recognise them in order to navigate reliably.

In summary, experts discussed cultural adaptation in relation to different behavioural levels. At the normative level, experts described expectations grounded in local values about what counts as appropriate behaviour in a given place. At the strategic level, they referred to the need for AVs to adjust decisions to match location-specific road conventions, including differences in right-of-way expectations. At the tactical level, they discussed how the AV should respond to behaviours that occur in real time, such as informal practices at roundabouts or crossings, while still respecting legal requirements. These observations show that AVs are required to understand local practices while maintaining behaviours that remain consistent with legal and ethical expectations when local norms diverge from them.

2.2.8. Safety

Safety concerns how the AV avoids unsafe situations and reduces the likelihood of harm for both the people directly interacting with the vehicle and the wider public who depend on safe road systems. Experts described safety as relating to how the AV anticipates threats, manages uncertainty, and responds in ways that limit the possibility of collisions and support public expectations of safety.

One expert stated that AVs should execute overtaking and merging manoeuvres in ways that reduce crash likelihood (ID 1). The expert explained that safe positioning during interaction is necessary for limiting immediate risk on the road.

Another expert described how AVs should act when the presence of vulnerable road users is uncertain. They noted that the AV should perform anticipatory manoeuvres and leave buffer space when visibility is limited, so that unexpected encounters do not lead to unsafe situations (ID 2). This reasoning emphasised caution when information about the environment is incomplete.

Experts also referred to planning behaviour around pedestrians. One expert explained that manoeuvres should be planned to limit speed and create enough space for pedestrians to pass safely (ID 4). Similar points were made about planning based on the status of the AV and the position of surrounding road users so that manoeuvres are informed by the conditions of the environment (ID 13).

Other contributions focused on how the AV should detect and communicate changes in the environment. One expert stated that early detection of hazards helps the AV mitigate risk before unsafe situations develop (ID 14). Another expert described how unnecessary manoeuvres can introduce additional risk, and suggested limiting such behaviour when safe progress can still be maintained (ID 15). They also noted that when sensor reliability is compromised, the AV should assess the situation in full before responding (ID 15). These points emphasised that risk management includes adapting to changing conditions.

Some experts highlighted behaviours relevant to exceptional or unpredictable scenarios. One expert noted that the AV should detect red-light violations by others and brake when needed to avoid collision (ID 16). The same expert explained that, in uncertain situations, the AV should prioritise protecting its occupants (ID 16). Another expert stated that adapting to realistic traffic speeds, rather than relying only

on strict legal limits, can support safer movement when surrounding flow differs from posted rules (ID 17).

These observations relate directly to the behavioural levels defined in our analysis. Safety appeared across several behavioural levels. At the normative level, experts referred to expectations from the wider public that the AV should prioritise protecting both occupants and other road users when risk is present. At the strategic level, they described planning manoeuvres that account for pedestrian movement, speed limits, sensor performance, and road-user positioning, reflecting the design decisions that shape safe movement. At the tactical level, their examples focused on real-time adjustments such as leaving buffer space, adapting to occlusion, responding to red-light violators, and braking when hazards emerge. At the operational level, experts highlighted behaviours such as detecting sudden obstructions. Overall, these perspectives show that safety depends on how the AV links protective priorities with planned manoeuvres and immediate responses to unfolding events.

2.2.9. Interaction management

Interaction management refers to how the AV interprets the actions of people outside the vehicle and makes its own behaviour understandable to them. Experts consistently described this category as concerned with how the AV interprets the behaviour of pedestrians, cyclists, and other drivers, and how it makes its own behaviour understandable to them.

Several experts emphasised that the AV should detect relevant road users and infer their intentions so that it can respond in a way that avoids unsafe interaction (ID 3). This includes understanding whether others are slowing, crossing, or changing direction. The expert framed this as a necessary part of responding appropriately to surrounding behaviour.

Experts also described the importance of communication during interaction. One expert explained that the AV should convey its intended manoeuvres clearly to ensure that other road users can anticipate what it will do (ID 3). Another noted that the use of visible cues, such as cinematic indicators or clear deceleration, helps pedestrians interpret the AV's movement and judge whether they can proceed (ID 9). A similar point was made regarding hazard signals, where one expert stated that clear warnings help prevent misinterpretation during abnormal situations (ID 15).

Anticipating the movement of others was another common theme. One expert highlighted that the AV should predict pedestrian motion at both marked and unmarked crossings to prevent unsafe encounters (ID 4). Other experts provided similar examples involving the prediction of vehicle, cyclist, or scooter trajectories in more complex settings such as intersections (ID 8, ID 7). These points treat anticipation as a central part of managing shared road space.

Experts also referred to how the AV's behaviour affects the actions of others. One expert explained that the AV should influence the behaviour of surrounding road users by using cooperative or predictable manoeuvres (ID 7). Another described merging behaviour as an example, noting that the AV should cooperate with other vehicles during such manoeuvres so that its behaviour aligns with what others expect (ID 18). These examples show how interaction management involves shaping not only the AV's responses but also how others adjust their behaviour.

These perspectives reflect the behavioural levels that structure our analysis. Interaction management appeared across several behavioural levels. At the normative level, experts framed clear and transparent signalling as a behavioural responsibility that supports fairness and public trust in shared road environments. At the strategic level, they described behavioural decisions about how the AV positions itself, prepares manoeuvres, and distinguishes between reactive and goal-directed actions so that its behaviour fits the broader flow of traffic. At the tactical level, their examples highlighted real-time behaviours such as detecting nearby road users, interpreting their intentions, predicting their movement, and using signalling or cooperative manoeuvres to make the AV's actions understandable. Overall, these views show that interaction management depends on how the AV interprets others' actions, communicates its own intentions, and coordinates behaviour within dynamic shared road space.

2.2.10. Comfort

Comfort concerns how passengers perceive the AV's driving behaviour and how this perception shapes both their immediate reactions and their longer-term experience of automated travel. Although this category may appear similar to interaction management, the focus here is different: comfort concerns the effects of the AV's movement on people inside the vehicle, whereas interaction management addresses how people outside the vehicle interpret and respond to what the AV does.

These perspectives reflect the behavioural levels that structure our analysis. Several experts described comfort as closely connected to safety because discomfort can prompt unnecessary intervention. One expert explained that the AV should maintain passenger comfort to prevent overrides caused by uncertainty or confusion (ID 5). This perspective treats comfort as part of maintaining a stable relationship between the passenger and the vehicle, since discomfort may lead to actions that interfere with automated control.

Another expert noted that the AV should drive in a way that feels comfortable to passengers and intuitive to nearby road users (ID 13). In this account, comfort supports acceptance by ensuring that the vehicle's motion aligns with what passengers expect and what other road users can understand.

Experts also mentioned that comfort should be considered when planning actions. One expert described comfort as a factor the AV should address alongside safety and efficiency when determining how to act (ID 7). This reflects the idea that comfort is part of how the AV should structure its behaviour over time, not only in immediate responses.

Two examples from another expert concerned specific driving practices that influence comfort. Avoiding harsh braking was described as important for passengers on board (ID 9). The same expert noted that the AV should avoid yielding when doing so would create unnecessary disruption for the occupants (ID 9). These examples show how comfort appears in particular manoeuvres.

Taken together, these contributions describe comfort as aspects of AV behaviour that shape how passengers interpret and respond to automated driving. Comfort influences whether passengers feel secure and whether they choose to intervene. It also affects how understandable the AV's movement appears to people who interact with the vehicle.

These statements fit the behavioural levels applied in our analysis. Comfort mapped onto strategic and tactical behavioural levels. At the strategic level, they referred to comfort as something the AV should incorporate when planning how to act over longer stretches of driving. At the tactical level, their examples focused on how specific manoeuvres — such as braking or yielding — affect the passenger's immediate experience. This shows that comfort operates at several levels of behaviour, shaping how passengers respond both in individual moments and across sustained interactions with the AV.

2.2.11. Continuous control

Continuous control concerns how the AV sustains awareness of its surroundings and remains responsive during driving so that its behaviour is predictable and safe for other road users, particularly those who are vulnerable. Experts described this category as relating to the AV's ability to monitor traffic conditions continuously rather than relying only on discrete updates or isolated events, which depends on design choices that enable consistent awareness throughout the trip.

One expert stated that the AV should maintain attention to changes in the traffic environment even during routine driving (ID 14). They explained that doing so supports ongoing awareness and allows the AV

to respond when conditions shift. The reason emphasised that continuous monitoring is necessary for timely and appropriate adjustment to what happens around the vehicle.

The statement aligns with the operational task from behavioural levels. Experts described the need for ongoing monitoring and real-time responsiveness as traffic conditions change. This suggests that continuous control, as described by experts, centres on maintaining persistent situational awareness and the capacity for immediate behavioural adjustment in response to evolving traffic conditions.

2.2.12. Vigilance and readiness

Vigilance and readiness concerns how the AV manages attention and responsibility in situations where control is shared between the vehicle and the human driver. Experts described this category as relating to whether the AV should depend on the driver, who may not remain continuously alert during extended periods of automation, and to the role of system designers in determining how responsibility is allocated when attention declines.

One expert stated that the AV should avoid relying on driver alertness in long-term shared-control situations (ID 14). They explained that drivers often become inattentive when automation manages the majority of the driving task and that expecting the driver to remain vigilant under these conditions is unsafe. The reason emphasised the need for the AV to handle responsibility directly when prolonged automation reduces the likelihood of sustained human attention.

Vigilance and readiness were discussed only at the operational level, where experts described the need for the system to always remain alert, avoid depending on the driver, and maintain driver engagement. This suggests that vigilance and readiness, as described by experts, centre on the dual tasks of ensuring the driver remains engaged and prepared for operational demands, while the AV itself remains continuously attentive to the surrounding environment.

2.2.13. Control transition

Control transition concerns how responsibility shifts between the AV and the human driver, and how this transition is supported by the system's design. Experts discussed this category in relation to how the AV prepares the driver to resume manual control safely, emphasising the role of designers in determining how handover is communicated and managed.

One expert stated that the AV should provide sufficient warning before the driver takes back control (ID 5). They explained that this is needed to prevent confusion and to ensure that the driver can safely resume the task. The expert emphasised that a clear and timely transition reduces the likelihood of unsafe responses when authority over the vehicle changes.

This explanation only aligns with the operational task of behavioural levels. The expert highlighted the need for clear and timely cues that allow the driver to safely resume control. This shows that control transition centres on how the AV provides real-time support during the moment when responsibility shifts to the driver.

2.3. Discussion

To address Objective 1, we aimed to identify and structure the category of human reasons that should inform AV manoeuvre planning. Our study contributes to addressing the **practical gap** by offering a layered mapping of thirteen reason categories, organised across moral, strategic, tactical, and operational levels, and explicitly linked to the roles of relevant human agents. This structure organises the expert-derived human reasons that influence AV manoeuvre planning into a layered form of guidance on what considerations should inform AV decision-making, supporting future research and system design aligned with the tracking condition of Meaningful Human Control (MHC).

This section also responds to the methodological gap identified in the introduction by demonstrating how expert interviews and directed content analysis can reveal the structure of human reasons relevant to AV behaviour. Our approach systematically captured how experts from diverse domains interpret what matters in AV behaviour, allowing reason categories to emerge inductively while using the Meaningful Human Control framework to position them across behavioural levels. Table 3 illustrates how these reasons connect to AV behaviour across different levels, from normative expectations to operational execution, and identifies the human agents affected by those behaviours. This mapping provides a bridge between the reasons experts expressed and the types of behaviour designers may need to support in AV systems.

Multiple overlapping reasons in a single manoeuvre. A central insight from our findings is that AV manoeuvre planning rarely relies on a single type of reason. Instead, a single manoeuvre often engages multiple overlapping reasons that reflect different layers of ethical and practical concern. For example, Expert ID 17 explained that choosing not to follow traffic rules strictly in some situations can be justified for more than one reason at the same time, such as improving safety and reducing environmental impact. Additionally, reason categories themselves are not confined to one behavioural layer. Rule compliance, for instance, spans normative expectations (e.g., respecting laws), tactical execution (e.g., behaving like a predictable driver), and operational functionality (e.g., logging red-light violations). This layered nature of reasons suggests that manoeuvre planning systems must support multi-reason, multi-level responsiveness, rather than relying on rule-based execution alone.

Human proximity and agent roles. Our analysis of Fig. 1 indicates a relationship between reason type and the proximity of the human agents referenced by participants. Normative reasons were typically associated with more socially and institutionally distant agents — such as policymakers, the general public, and local communities — who shape and enforce broader ethical standards. In contrast, strategic, tactical, and operational reasons were more closely connected to agents physically proximate to AV operation and who directly interact with or design AV behaviour, such as vehicle occupants and system designers. Notably, road users and vulnerable road users appear across the full range from normative to operational levels, suggesting that participants viewed them as important agents to consider throughout all layers of reasoning. This supports and extends the proximity-based model introduced by Mecacci and Santoni de Sio (2020) and elaborated by Calvert and Mecacci (2020), which posits that meaningful human control depends on responsiveness to human reasons distributed across different layers of reasoning. Our findings provide empirical evidence for this framework and offer a structured account of how agent proximity to the AV and the type of reasons are interconnected.

Variation in behavioural level depending on task interpretation. We also found that the behavioural level at which a reason is situated can vary depending on how the AV task is interpreted. For example, time efficiency is often treated as a strategic concern, but depending on the situation, it may also appear at tactical or even operational levels. A strategic interpretation might involve planning the most efficient route, while a tactical one may involve decisions such as overtaking or avoiding hesitation that disrupts flow. Despite being motivated by the same underlying reason of efficiency, these interpretations correspond to different layers of action. This highlights the importance of distinguishing between the justification for a behaviour and the specific behavioural level at which it is operationalised.

Reason variations across levels of automation. We also observed that the relevance of certain reasons shifts with the AV's level of automation. For instance, considerations such as control transition and vigilance were particularly prominent for lower levels of automation (L2/L3), where human fallback is still required. At higher levels (L4/L5), these concerns recede, and the focus shifts towards trade-offs among values such as fairness, efficiency, and comfort — especially when AVs operate without direct human supervision. Our framework accommodates these shifts by revealing which reasons—and which

agents — are most relevant at each automation stage and behavioural level.

Interpretative flexibility in core categories. A further nuance in our data involves the diverse interpretations of rule compliance. While several experts (e.g., IDs1, 3, 11, 12) framed rule-following as a strict moral baseline, others (e.g., ID17) viewed traffic laws as flexible guidelines to be overridden when necessary to ensure fairness or safety. This reflects the contextual and scenario-sensitive nature of AV ethics: manoeuvre planning must not only track rules but also balance them against competing values such as social appropriateness and safety.

Positioning within existing literature. Our framework also offers a way to contextualise and relate existing efforts to define what AVs should consider when making driving decisions. Prior work has provided focused contributions on specific types of consideration: for example, UNECE (2023), Olleja et al. (2025) define legal and behavioural benchmarks for competent driving; Geisslinger et al. (2021) formalise ethical principles for risk-sensitive planning; Schwarting et al. (2018) model social value orientation for cooperative behaviour; and Thornton et al. (2018) apply value-sensitive design to embed stakeholder values in AV system logic. While these approaches differ in their aims and methodologies, our framework does not attempt to replace or rank them. Rather, it provides a layered structure through which these contributions can be situated — by connecting the types of reasons they represent (e.g., legality, fairness, efficiency, comfort) to specific levels of AV behaviour (e.g., moral/normative, strategic, tactical, operational). In this sense, our empirically derived categorisation can serve as a common referential model — one that helps clarify how diverse AV design goals and values interact across system layers and in relation to different human agents.

Practical design guidance and limitations. Beyond theoretical insights, our structured mapping offers practical guidance for AV developers and policy designers. For example, developers working on L2/L3 vehicles may use our findings to prioritise clear and timely takeover cues, while those building L4/L5 systems may focus on fairness–efficiency trade-offs and behaviour intelligibility in mixed traffic. This structured mapping of thirteen reason categories across behavioural levels and agent roles can help translate ethical expectations into system-level specifications—by clarifying what kinds of human concerns should be considered, at which layer of system behaviour, and by whom. Furthermore, as the questions did not solely focus on ethically challenging situations, the identified reasons are applicable not only in edge cases but also in routine and general AV driving scenarios where aligning AV behaviour with human reasons is essential.

Nonetheless, our approach has limitations. First, the scope of this study is restricted to manoeuvre-level decision-making, rather than broader systemic influences such as infrastructure, corporate strategy, or legal frameworks. Second, interpretations of the identified reasons are likely influenced by cultural and regional contexts, which may affect the generalisability of the findings. Our expert pool was primarily composed of individuals from Western countries, particularly the Netherlands, the United States, and the United Kingdom. As such, our findings should be understood as reflecting predominantly Western ethical and social norms, and not assumed to be universally applicable. Future research is needed to explore how these categories of reasons manifest in other cultural environments.

Third, we emphasise that the 13-category taxonomy is not intended as exhaustive. It reflects insights from a specific group of experts, and future research should investigate how cultural, regional, or stakeholder diversity may yield additional or context-dependent reason types. In particular, cells marked "Not mentioned" in Table 3 do not imply that no relevant reasoning exists at that level. Theoretical frameworks suggest that any reason could, in principle, be interpreted across multiple control layers. However, because our study is empirical in nature, the absence of content in certain cells reflects the limits of what was raised by experts, not a conceptual impossibility. Future research could further investigate these gaps through targeted questioning or normative modelling. Future work may also extend this approach by incorporating a broader and more diverse stakeholder base — such as regulators, insurers, and urban planners — or by developing prioritisation models to resolve conflicts among overlapping reasons.

Finally, while every effort was made to ensure conceptual clarity, we acknowledge that some reason categories, such as *social appropriateness* and *acceptance*, may overlap in practice. This reflects the interpretive nature of qualitative analysis. However, these overlaps do not affect the core findings regarding how experts prioritise reasons in ethically ambiguous situations. Consensus coding and iterative refinement were used to mitigate interpretive bias, and we encourage future research to further validate and refine the categorisation scheme using participatory or quantitative methods.

Summary of contributions. In summary, addressing Objective 1, we identified thirteen categories of human reasons relevant to AV manoeuvre planning and examined how these reasons informed expert expectations about what the AV should do across different behavioural layers, as well as which human agents were affected. Our findings highlight that even routine AV decisions can involve ethically sensitive considerations, and that manoeuvre planning must account for multiple, sometimes conflicting, human expectations. By integrating these findings with the MHC framework, our study offers a pathway towards AV systems that behave in ways aligned with human reasons.

3. Reason-based decision principles for ethically challenging AV scenarios: Cyclist overtaking as a case study

3.1. Methods

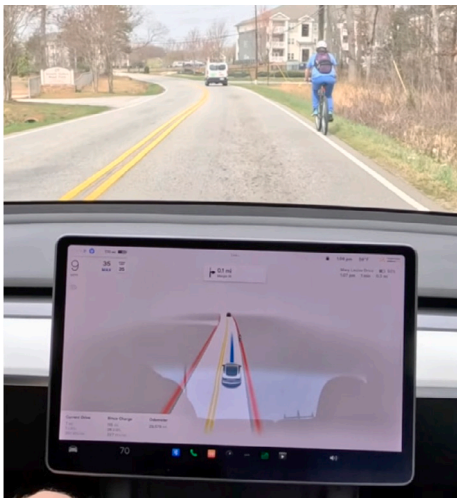
To investigate reason-based decision principles in ethically challenging cyclist overtaking situations, we analysed expert responses to Questions 5–11 of the semi-structured interview protocol administered to the same participants described in Section 2.1.1. This section outlines the scenario design, question format, and the analytical approaches used to examine both implicit and explicit prioritisation of reasons.

3.2. Scenario and questionnaire design

To explore the reason-based principles for AV decision-making in ethically challenging traffic situations, we incorporated a specific overtaking scenario in our interview protocol. This scenario involved an AV travelling behind a slow-moving cyclist on a two-way road marked with a double yellow line. This road marking typically prohibits overtaking, thereby introducing a regulatory constraint that renders the situation ethically ambiguous: the AV could either remain behind the cyclist at a reduced speed or initiate an overtaking manoeuvre by crossing the double yellow line.

After responding to Questions 2–4, which explored their reasoning about what the AV should consider when planning its manoeuvre, experts were shown a short video clip depicting the overtaking scenario (see Table 4). They then answered a series of structured follow-up questions (5–9) that asked them to predict how the AV and other traffic participants would behave, explain the reasons for these actions, identify additional influencing factors, and describe possible conflicts between different intentions. Finally, in questions 10–11, experts ranked the intentions of three predefined stakeholders in the scenario (the AV passenger, the cyclist, and the road policymaker), where "intentions" were operationalised as proxies for underlying reasons, and explained the reasoning behind their rankings. These questions were designed to elicit interpretations of AV decision-making, the factors influencing it, and how experts implicitly and explicitly prioritised competing reasons, without explicitly prompting normative judgments about what the AV should do.

Table 4
Interview questions relevant to objective 2.

No.	Interview Questions
Please watch the video below and read its description  **Video description** A passenger uses an automated vehicle (AV) for a morning commute to the office. The passenger has an important meeting and must arrive on time. If the vehicle maintains the current speed, the passenger can reach the office on time in 20 minutes. The AV is on a road with solid double yellow lines, which prohibit vehicles from crossing in both directions due to safety reasons. During the trip, the AV approaches a cyclist traveling at half of the speed of the AV. There is no safe passing zone visible from the vehicle; however, the opposite lane is currently empty.	
Question 5	If the video continues, what do you believe all traffic participants will do next?
Question 6	What are the reasons for the traffic participants performing the actions you mentioned?
Question 7	Besides those traffic participants, can you think of any other factors that might influence their decisions?
Question 8	What do you think the reasons are for the other factors you mentioned (e.g., traffic signs and the double yellow line)?
Question 9	Can you think of any situations where the intentions of the [traffic participants / other factors the experts mentioned] might conflict? Please share any examples you can think of, and let me know when these conflicts may typically occur.
Recall the scene from the previous video. There are three different people, each with their own intentions: • The automated vehicle (AV) passenger wants to pass the cyclist to get to the office on time. • The cyclist wants a safe distance from the AV for safety concerns. • The road policymaker wants both AV and cyclist to use their designated lanes, marked by solid yellow lines, for everyone's safety. Keep this in mind as you answer the rest of the questions.	
Question 10	From your perspective, whose intentions should be given the most importance? Please answer this question by ranking the individuals below, with "1" indicating the highest rank.
Question 11	Could you please explain the reasons behind the rank you provided in your previous answer?

	1	2	3
AV passenger	◯	◯	◯
Cyclist	◯	◯	◯
Road policymakers	◯	◯	◯

3.3. Expert reasoning in the cyclist overtaking scenario

3.3.1. AV behaviour justifications

To understand how experts justify their predictions about AV behaviour in overtaking scenarios, we used *Qualitative Content Analysis (QCA)* (Schreier, 2012) to analyse their responses to Questions 5–9. These questions asked experts to predict what the AV would do in a specific overtaking scenario and to justify their predictions. Following Schreier's structured approach, we developed a mixed-category coding strategy that incorporated both:

- **Concept-driven categories**, used to group expert responses by predicted AV behaviour: either overtaking the cyclist or not overtaking the cyclist.

- **Data-driven categories**, used to inductively identify the specific justifications experts provided for their predictions.

The first stage of the analysis involved classifying each expert response according to the predicted action (i.e., follow vs overtake). Within each group, expert justifications were then collected and analysed to identify recurring patterns of reasoning. These inductively derived justifications were subsequently mapped to the predefined set of thirteen reason types originally developed from the open-ended responses to Questions 2–4 (see Section 2.2).

As part of the interpretation phase, we conducted a qualitative synthesis of the coded data to explore patterns in how experts linked specific reasons to predicted AV behaviours. This involved examining which reasons frequently co-occurred, how the same reasons were interpreted differently across contexts, and instances where multiple reasons appeared within a single justification. This helped us understand how reasons relate to each other or are prioritised. The synthesis was guided by principles of thematic pattern analysis (Braun and Clarke, 2006) and constructivist grounded theory (Charmaz, 2014), with analytical interpretations discussed collaboratively between authors to enhance transparency and reduce individual bias.

To analyse reason prioritisation, we examined reason prioritisation in two complementary ways. The first approach captured how priorities emerged naturally within participants' open-ended reasoning (implicit prioritisation). The second approach asked participants to directly state their preferences in a structured ranking task (explicit prioritisation). Analysing both allowed us to compare context-driven unprompted prioritisation with deliberate stated preferences. Together, these perspectives gave us a fuller understanding of how experts weigh competing reasons.

3.3.2. Implicit reason prioritisation

Implicit analysis allowed us to capture how prioritisation emerged naturally in participants' reasoning, without being influenced by a predefined ranking task or fixed response options. This approach provided insight into how trade-offs were navigated in context and how multiple considerations interacted within the flow of open-ended discussion.

Building on the qualitative synthesis described in Section 3.3.1, we conducted an analysis of implicit prioritisation patterns evident in participants' responses. Although the questionnaire did not directly ask experts to rank reasons, the overtaking scenario was intentionally designed to present conflicting reasons within a single context, particularly through the video depiction. This design allowed us to observe how participants navigated trade-offs between different considerations when explaining their expected or preferred AV behaviour.

To systematically analyse this *implicit prioritisation*, we examined how participants justified their decision for the AV to either follow or overtake the cyclist. We focused on statements where reasons were weighed against each other in explaining that decision. For example, some participants accepted a rule violation (crossing the double yellow line) because it would reduce cyclist discomfort. Others rejected an overtake because rule compliance outweighed the driver's travel efficiency. Such comparisons allowed us to infer the relative importance of reasons. In the first example, cyclist comfort was treated as a higher priority than strict rule compliance, whereas in the second example, rule compliance was treated as a higher priority than the driver's travel efficiency. From these comparisons, we identified which reasons were positioned as primary and which were conditional or secondary, even without an explicit ranking task.

3.3.3. Explicit prioritisation

Explicit ranking provided a direct statement of participants' preferences, enabling systematic comparison across individuals and alignment checks with the priorities inferred from the implicit analysis. This approach also allowed us to quantify the relative importance assigned to each stakeholder's reason.

To complement the implicit analysis in Section 3.3.2, we included a structured ranking task in Questions 10–11 to explicitly elicit prioritisation preference. Experts were asked to rank the intentions (used as proxies for broader reasons, but phrased this way for participant clarity) of three human agents involved in the overtaking scenario: the AV passenger (who wants to reach the office on time), the cyclist (who wants a safe buffer zone), and the road policymaker (who designed the double yellow lines for public safety).

Experts were provided both a numerical ranking and a free-text explanation of their choices. This approach enabled us to collect both quantitative and qualitative data on how experts explicitly prioritised different stakeholder reasons. Rankings were aggregated to identify overall patterns, and justifications were thematically analysed based on the order of rank to uncover the themes guiding these decisions.

3.4. Results

This section presents the findings from the expert interviews. Based on their responses to Question 5–9, most experts interpreted the AV as the sole traffic participant whose actions were being evaluated. Two primary behaviours that the AV might adopt in the given scenario were identified: (1) *following the cyclist*, and (2) *overtaking the cyclist*.

In addition, some experts distinguished between what the AV is likely to do (**expected action**) and what the AV ought to do (**preferred action**). To maintain clarity, this distinction will be upheld throughout the remainder of the paper: expected action refers to what the AV is predicted to do, whereas preferred action refers to what the AV should do from the expert's normative perspective.

Fig. 2 summarises the reasons experts provided for predicting whether the AV **will** or **should** follow or overtake the cyclist. Thirteen distinct reasons are presented, grouped by action type. Blue shading represents predicted ("will") actions, while lighter red shading represents preferred ("should") actions.

In general, experts cited a more limited set of reasons for why the AV will follow the cyclist, most commonly grounded in *rule compliance* and *safety*. By contrast, overtaking was associated with a broader range of justifications, including *efficiency*, *comfort*, and *interaction management*, alongside the aforementioned safety and legal concerns.

This pattern suggests that *following the cyclist* is predominantly justified by a narrow range of risk-averse or rule-based considerations, whereas *overtaking* is viewed as a more complex decision that draws upon a wider set of overlapping reasons. We then go beyond listing reasons to examine how experts weighted competing reasons in their explanations (implicit prioritisation) and how they explicitly ranked stakeholder reasons in a structured task (explicit prioritisation). Together, these analyses provide a fuller picture of not only what reasons experts described, but also how they balanced them when reasoning about AV behaviour.

3.4.1. Reasons for the AV will follow the cyclist

From Fig. 2, panel (a), the most frequently cited reasons for why the AV is expected to follow the cyclist were *rule compliance* and *safety*. Several experts (e.g., ID2, ID12) emphasised that the AV is programmed to obey traffic laws, such as not crossing a double yellow line, making overtaking legally impermissible. Others (e.g., ID3, ID11) pointed to the importance of safety, arguing that the AV would stay behind the cyclist to avoid potential collisions or unsafe manoeuvres. In many cases, both legal and safety considerations were closely intertwined in the experts' reasoning. One expert (ID2) also mentioned driver discomfort as a potential outcome, noting that while the AV is obligated to follow the cyclist, this may result in frustration or discomfort for the human passenger. However, this was framed not as a reason for the AV's behaviour, but rather as a consequence of its strict adherence to rules and safety protocols.

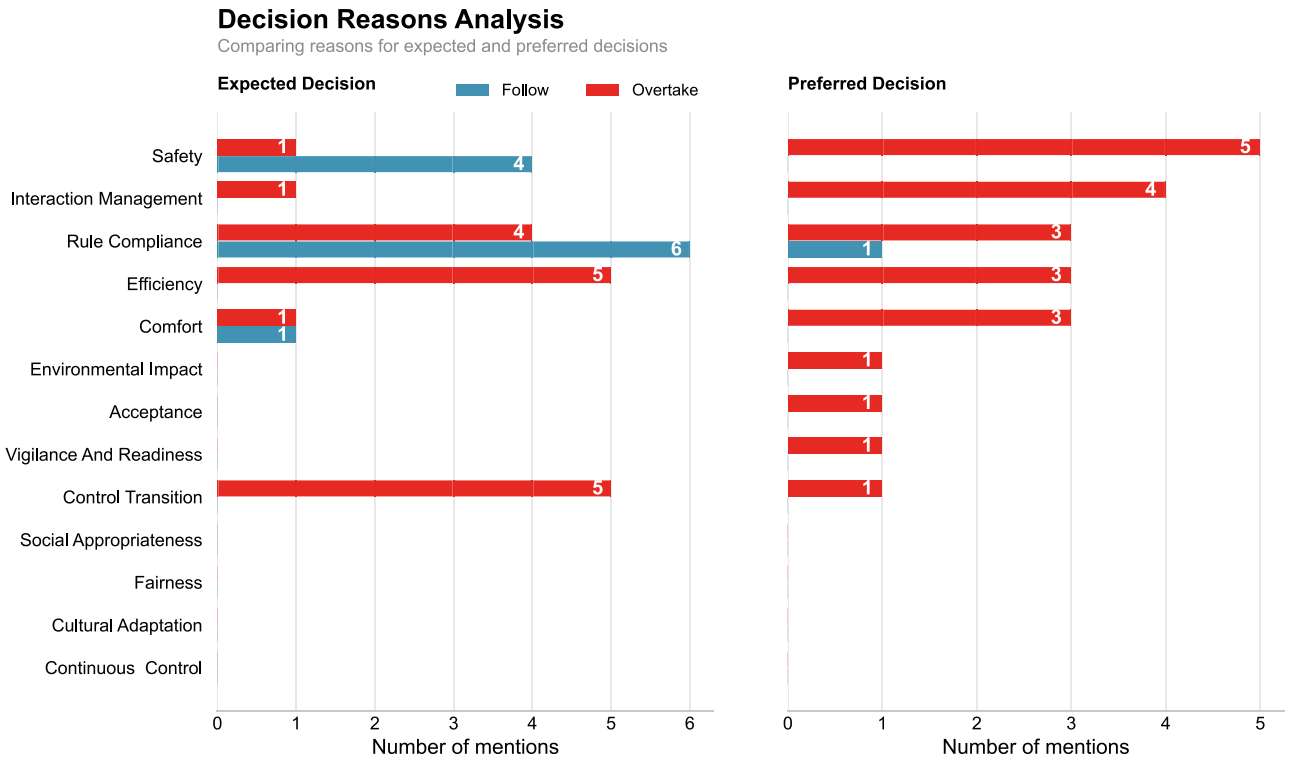


Fig. 2. Experts' reasons for predicting whether the AV will or should follow (a) or overtake (b) the cyclist. Thirteen reasons are listed in the first column, expert IDs in the second, and total counts in the third. Blue indicates predicted behaviour (will), red indicates preferred behaviour (should).

3.4.2. Reasons for the AV should follow the cyclist

Notably, only one expert (ID12) explicitly stated that the AV *should* follow the cyclist, as indicated by the lighter red shading in panel (b). This expert argued that, in addition to the fact that current AVs are programmed to follow traffic rules, they *should* continue to be programmed in accordance with those rules. Interestingly, this expert was the only one with a background in the legal aspects of AVs.

3.4.3. Reasons for the AV will overtake the cyclist

In panel (a) of the figure, the experts cited a variety of reasons for why the AV is expected to overtake the cyclist. The most frequently mentioned reasons included *efficiency*, the possibility of *control transition*, and *rule compliance*. *Safety* and *comfort* were discussed less frequently but still featured in some responses.

Several experts emphasised time efficiency as a key factor. One expert (ID15) stated that the driver would likely take over control and overtake the cyclist in order to arrive at work on time. Another expert (ID10) also mentioned the urgency of reaching a destination as a reason for manual takeover. A third expert (ID7) explained that while the AV might initially follow the cyclist, the driver would likely take over if delays occurred, particularly because they would not want to be late. Similarly, an expert (ID13) noted that overtaking would allow the AV to maintain a more efficient speed, and another (ID14) emphasised that roads are designed for higher speeds than cyclists typically travel, making it uncomfortable for drivers to go significantly slower than expected.

Rule compliance was discussed in relation to whether overtaking is legally permissible. One expert (ID13) stated that in the UK, it is legal to overtake a cyclist if they are travelling below 10 mph, and therefore expected the AV to do so. Another expert (ID6) viewed the AV's failure to overtake when the opposite lane is empty as irrational, suggesting that such behaviour would seem "stupid" to human drivers. An expert (ID7) commented that while drivers are aware of traffic rules, they often weigh the risk of breaking those rules against the need to stay on schedule. One expert (ID15) believed that the AV would not overtake because it is programmed to strictly follow traffic rules, but that the driver would override this behaviour when time becomes a priority.

The possibility of *control transition* was also commonly mentioned. Multiple experts (IDs 6, 7, 10, 14, 15) described scenarios in which the human driver would take over control to overtake the cyclist. Some (e.g., ID7, ID10) mentioned this as a temporary takeover, while others (e.g., ID14, ID15) connected it to the frustration caused by prolonged low-speed travel. One expert (ID6) stressed that such a manoeuvre would not compromise safety and therefore believed the driver would proceed with overtaking.

A few experts raised *social and comfort*-related concerns. One expert (ID13) emphasised the importance of not frustrating other road users who may be following the AV. Another expert (ID14) pointed out that drivers are not accustomed to travelling so slowly, especially on roads designed for higher speeds, and would likely feel discomfort in such situations—prompting a takeover.

Although less frequently mentioned, *safety* still featured in the discussion. One expert (ID6) stated that overtaking would be acceptable if the opposite lane were empty, implying that safety would not be compromised. Another (ID10) said the driver would overtake only if it could be done without putting anyone at risk, suggesting that manual takeovers are still constrained by a concern for *harm avoidance*.

3.4.4. Reasons for the AV should overtake the cyclist

Several experts argued that the AV *should* overtake the cyclist, even if it involves crossing a double yellow line, because failing to do so could cause confusion, frustration, and even safety issues. A common theme among the responses was that not overtaking could disrupt traffic flow, potentially leading to cascading negative effects—such as increased risk (ID1, ID17), discomfort (ID5, ID11), higher emissions (ID17), and reduced acceptance (ID11). These disruptions were framed not only as inefficiencies but also as factors that could compromise safety and overall system performance.

For some experts, *safety* did not necessarily align with strict rule-following. Instead, safety was interpreted contextually, such as maintaining adequate distance from the cyclist (ID8), supporting a steady

traffic flow (ID16, ID17), or reducing cognitive workload for drivers (ID5). This suggests that pragmatic, situational decisions — even if technically in violation of traffic regulations — can still serve safety-related objectives.

Experts also emphasised that a single justification was rarely sufficient; rather, multiple factors often combined within a single rationale. For example, an expert might state that the AV should overtake because it is both *safer* and *more comfortable* (ID5), or because it improves *traffic flow* and aligns with *human driving behaviour* (ID17). These reasons were presented as equally important, rather than hierarchically ordered.

Regarding *rule compliance*, experts acknowledged that overtaking may violate traffic rules (ID8, ID11, ID16). However, they generally agreed that strict rule adherence should not override other practical concerns such as *safety* and *traffic efficiency*. In this context, rule violations were often framed as acceptable when they led to better outcomes for all road users.

In addition, *interaction management* and *comfort* played significant roles in shaping expert expectations. Some experts noted that cyclists may feel uncomfortable or stressed when a vehicle follows too closely without overtaking (ID5, ID11, ID17), and that passengers or drivers may become frustrated by overly cautious AV behaviour (ID1, ID11). These concerns link comfort with public trust and acceptance, suggesting that AVs should behave in ways that are intelligible and relatable to human road users.

Finally, some experts (e.g., ID5) stated that in such situations, they would personally choose to overtake the cyclist by taking control of the vehicle. This reinforces the view that manual takeover may remain a practical necessity when AVs are constrained by rules that fail to account for situational flexibility.

3.4.5. Implicit prioritisation patterns

Analysis of participants' explanations revealed consistent patterns in how reasons were prioritised when predicting or prescribing AV behaviour in the overtaking scenario. Across interviews, safety consistently emerged as the primary, non-negotiable consideration. Even when participants supported overtaking, they emphasised that it should only occur if safety could be maintained. For example, some stated that overtaking was only acceptable if the opposite lane was clear (ID6) or if there was adequate distance from the cyclist (ID8, ID10).

Other reasons, such as efficiency, comfort, and interaction management, were frequently mentioned, but typically in conjunction with safety. These were often framed as benefits that could be achieved only if the manoeuvre met safety conditions. For instance, participants linked overtaking to maintaining steady traffic flow and reducing emissions (ID17), preventing frustration for following drivers (ID1, ID13), and relieving stress for cyclists (ID5, ID11, ID17).

Rule compliance was generally treated as a conditional obligation. Some participants cited it as a reason for the AV not to overtake, noting that the system would be programmed to follow the double yellow line rule (ID2, ID3, ID12). Others described it as the very reason a manual takeover might occur, since human drivers could choose to overtake when the AV, bound by traffic laws, would not (ID7, ID10, ID15). Many argued that crossing the line could still be justified when safety and traffic flow benefits outweighed strict adherence (ID6, ID7, ID13, ID16).

A smaller set of reasons, including environmental impact, driver workload reduction, and maintaining steady traffic flow, appeared less frequently and were often context-specific (ID5, ID14, ID17).

Overall, participants tended to treat safety as a primary consideration; efficiency, comfort, and human-interaction concerns were usually framed in relation to safety; and rule compliance was often conditional—sometimes cited as the reason the AV would not overtake, leading to manual takeover by the driver, and at other times set aside when safety or traffic flow benefits outweighed strict adherence (Fig. 2).

Table 5
Number of experts assigning each rank position to each stakeholder, based on their stated intentions in Question 10.

Stakeholder	1st place	2nd place	3rd place
Cyclist	12	5	0
Road policymakers	5	5	7
AV passenger	0	7	10

3.4.6. Explicit prioritisation patterns

In addition to the implicit prioritisation observed in their open-ended reasoning, experts were explicitly asked in Question 10 to rank the intentions of three stakeholders in the scenario. These stakeholders were the AV passenger, the cyclist, and the road policymakers. A rank of "1" indicated the highest priority. Table 5 summarises the aggregated rankings.

Out of 18 experts, one expert (ID02) declined to provide a ranking in Question 10, arguing that prioritising between stakeholders' intentions is inappropriate because legal and safety obligations should determine behaviour rather than preference-based trade-offs.

Thematic analysis of Question 11 justifications showed that the strong prioritisation of cyclists was grounded in their vulnerability as unprotected road users and in the moral obligation to minimise harm (for example, ID01, ID03, ID04, ID06, ID07, ID08, ID09, ID14, ID15, ID16, ID17, ID18). Several experts linked this priority to broader societal goals such as Vision Zero (ID09) and to the legal principle of protecting vulnerable road users (ID14). For these experts, safety considerations outweighed concerns for efficiency, trip time, or strict adherence to traffic regulations.

Experts who ranked road policymakers first (ID10, ID12, ID13) justified this choice by referring to the importance of systemic regulation and the rule of law in ensuring safe and predictable interactions for all road users. Some viewed policymakers as legitimate representatives of the public interest who are responsible for embedding safety into infrastructure and traffic rules (ID10, ID13). Others emphasised that the priorities of policymakers should align with the protection of vulnerable users, which indirectly supports cyclists (ID12).

The fact that no expert gave AV passengers the first rank was explained by their protected status within a vehicle and by the perception that their main concern, which is timely arrival, carries lower ethical weight compared to the safety of others (for example, ID06, ID08, ID09, ID11, ID16). When AV passengers were ranked second (for example, ID03, ID04, ID08, ID11, ID14, ID15, ID16), they were recognised as being directly involved in the situation. However, their needs were still seen as secondary to the safety of cyclists.

Overall, the explicit ranking task reinforces the patterns observed in the implicit analysis. Safety of vulnerable road users formed the primary decision criterion. Systemic safety considerations came second, and individual convenience came last.

3.5. Discussion

To address Objective 2, we applied the expert-derived reason categories developed in Section 3 to a context-specific case study involving a routine but ethically ambiguous AV scenario: overtaking a cyclist. This application contributes to **the theoretical gap** by operationalising the tracking condition of Meaningful Human Control (MHC) in everyday driving contexts, moving beyond high-stakes dilemmas to examine how AVs might align with human reasons in complex, real-world situations.

This section also speaks to **the methodological gap** by demonstrating how structured qualitative analysis — linking expert-elicited reasons to specific behavioural recommendations — can yield actionable insights. By capturing expert judgments on both expected ("will") and preferred ("should") AV behaviours, we identify how contextual

factors, value tensions, and individual reasoning strategies shape nuanced expectations for AV decision-making. These insights form the basis for deriving a reason-based prioritisation principle and for developing an empirically grounded conceptual representation that maps how such reasons emerge and interact in context.

This focus on routine, ethically ambiguous scenarios expands on prior work that has largely centred on high-stakes dilemmas, such as crash scenarios or trolley problems (Rhim and Urban, 2021; Milford et al., 2025). Whereas those studies highlight binary moral choices, our study surfaces the nuance of everyday trade-offs and expert reasoning in context-rich decisions. To support the derivation and explanation of the reason-prioritisation principle, we developed a conceptual representation of reason-based AV decision-making for the cyclist-overtaking case (Fig. 3) that illustrates how contextual factors give rise to reasons, how those reasons interact, and how prioritisation occurs in practice. This representation helps clarify the dynamics that underpin the prioritisation principle and demonstrates its applied relevance in ethically ambiguous everyday situations. While initially developed for the cyclist-overtaking scenario, its structure may also inform analyses of other routine ethically ambiguous driving contexts; however, further empirical investigation is required to examine such applicability.

Contextual Background Influencing Emerging Reasons. The reasons identified across expert responses consistently emerged from underlying contextual assumptions, including local regulations, technological capabilities, traffic situations, and individual differences. While the tracking condition in MHC theory requires that automated systems respond to human moral reasons (Santoni de Sio and Van den Hoven, 2018), existing literature does not explain how such reasons emerge from contextual circumstances. This study contributes a new insight by showing that expert reasons are shaped by situational background assumptions.

For instance, local regulations play a foundational role. One expert (ID13) stated that overtaking would be permissible under UK traffic laws. Conversely, experts ID7 and ID15 assumed stricter rules prohibiting overtaking, leading them to suggest manual takeover as a necessity. Technological capabilities and automation level also constrain or enable reasoning. Experts (IDs4, 15) noted that current AVs are programmed to avoid rule violations, thus requiring human intervention when overtaking is contextually necessary. This aligns with assumptions about Level 2 automation, where driver readiness remains essential. When experts assumed no manual override was available, they introduced alternative reasons such as *environmental impact*, *interaction management*, and *acceptance*.

Traffic situations, such as encountering a slow cyclist on a bidirectional road, influenced reasoning around safety, vigilance, and comfort. For instance, Expert ID5 emphasised that prolonged following increases driver workload, reinforcing the need for AVs to act pragmatically. Individual differences in expert values were also significant. Expert ID11 advocated strict rule-following as a matter of principle, while Expert ID 6 endorsed a consequentialist view, suggesting AVs should act based on outcomes (e.g., safety), regardless of legality.

Distinct Reasons Leading to Different Expected Behaviours. Distinct sets of reasons translate into differing AV behaviours. As shown in Fig. 2, experts cited *rule compliance* and *safety* as dominant reasons for why AVs are expected to follow the cyclist. In contrast, overtaking behaviour was associated with a wider range of reasons, including *efficiency*, *comfort*, *acceptance*, and *interaction management*.

Three mechanisms were identified as key to understanding how reasons lead to behaviours. First, reason interpretation varies by context. *Safety*, for example, was interpreted as a reason to follow the cyclist (avoiding risky manoeuvres, ID4), but also to overtake (reducing traffic disruptions, ID17). This shows that a single reason can support opposing behaviours depending on situational framing.

Second, a tension between personal and collective interests was apparent. Personal motivations (e.g., arriving on time, IDs7, 15) often

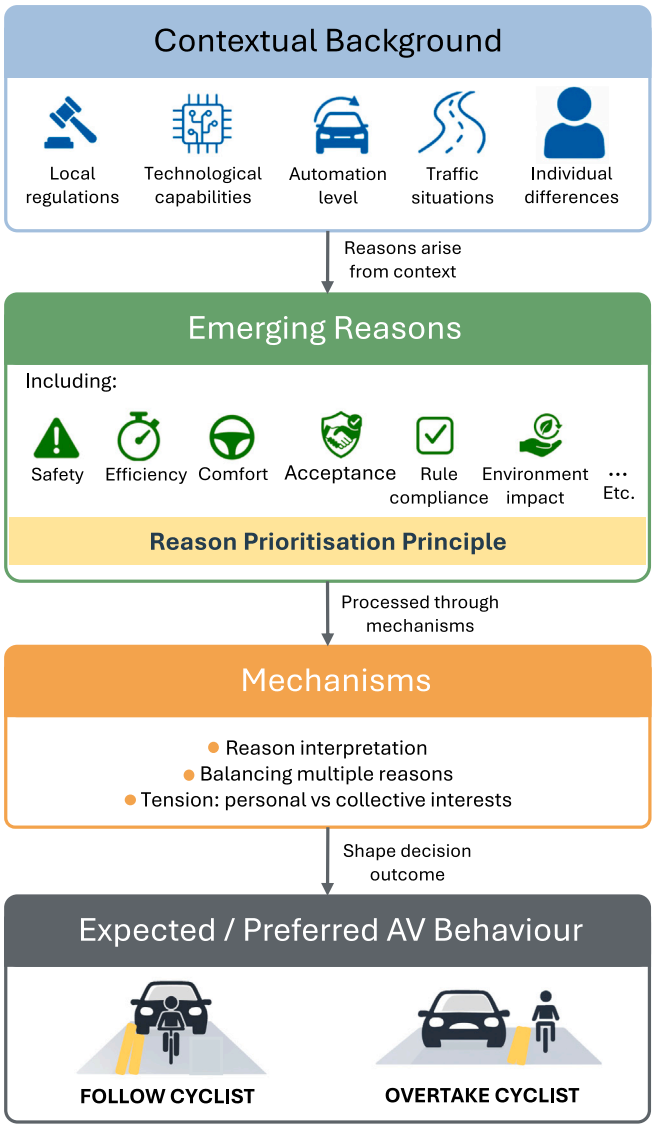
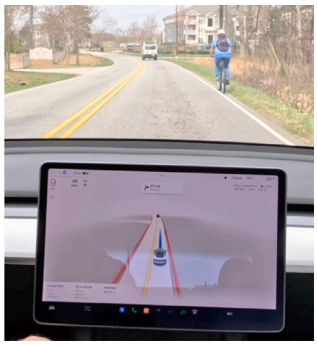


Fig. 3. Conceptual representation of reason-based AV decision-making in the cyclist-overtaking scenario. The representation illustrates how contextual background factors give rise to reasons for AV behaviour and where the derived reason-prioritisation principle operates within the AV decision-making process. These reasons, together with their prioritisation, are processed through mechanisms such as reason interpretation, tensions between personal and collective interests, and the balancing of multiple reasons. Collectively, these dynamics shape expected and preferred AV behaviour when deciding whether to follow or overtake a cyclist in an ethically ambiguous everyday situation.

shaped expected behaviour. Meanwhile, collective values (e.g., environmental sustainability, ID17; shared road safety, ID 5) influenced preferred actions. This tension reflects a broader ethical divide in AV decision-making, as noted in prior work (Bonnefon et al., 2016).

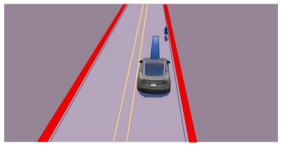
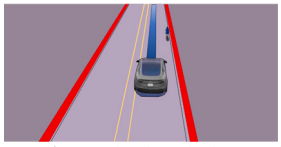
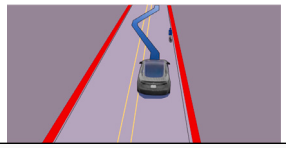
Third, balancing multiple reasons emerged frequently. Experts combined several motivations in a single rationale (e.g., ID 6 cited both safety and comfort). Notably, when rule compliance conflicted with more practical or safety-oriented reasons, the latter were often prioritised. These dynamic prioritisation mechanisms echo the tracking model proposed by Mecacci and Santoni de Sio (2020), in which AV systems are designed to respond to the most proximal reasons unless overridden by more distal values. Our findings complement this by empirically showing that human experts interpret and prioritise reasons fluidly, often allowing situational context to shape whether a reason

Table A.6
Interview questions.

Question number	Question
Q1	Where do you work, what is your position and role, what are your activities with regard to automated vehicles (AVs)?
Q2	What should automated vehicles (AVs) consider when planning a maneuver? Please give one example in much detail as possibl
Q3	Which moral aspects do you believe AVs should consider when planning a maneuver?
Q4	How might these aspects affect the manoeuvre plan?
Please watch the video below and read its description  **Video description** A passenger uses an automated vehicle (AV) for a morning commute to the office. The passenger has an important meeting and must arrive on time. If the vehicle maintains the current speed, the passenger can reach the office on time in 20 minutes. The AV is on a road with solid double yellow lines, which prohibit vehicles from crossing in both directions due to safety reasons. During the trip, the AV approaches a cyclist traveling at half of the speed of the AV. There is no safe passing zone visible from the vehicle; however, the opposite lane is currently empty.	
Q5	If the video continues, what do you believe all traffic participants will do?
Q6	What are the reasons for the [traffic participants mentioned by the experts] performing the [actions the experts mentioned]?
Q7	Besides the [traffic participants that are mentioned by the experts]'s, can you think any other factors that might influence the traffic participant decisions?
Q8	What do you think the reasons are for the [other factors that are mentioned by the experts]?
Q9	Can you think of any situations where the intentions of the [traffic participants / other factors the experts mentioned] might conflict? Please share any examples you can think of, and let me know when these conflicts may typically occur.
Recall the scene from the previous video. There are three different people, each with their own intentions: • The automated vehicle (AV) passenger wants to pass the cyclist to get to the office on time. • The cyclist wants a safe distance from the AV for safety concerns. • The road policymaker wants both AV and cyclist to use their designated lanes, marked by solid yellow lines, for everyone's safety. Keep this in mind as you answer the rest of the questions.	
Q10	From your perspective, whose intentions should be given the most importance? Please answer this question by ranking the individuals below, with '1' indicating the highest rank. **1** **2** **3** AV passenger ○ ○ ○ Cyclist ○ ○ ○ Road policymakers ○ ○ ○
Q11	Could you please explain the reasons behind the rank you provided in your previous answer?
Watch the three video scenarios below! These scenarios show three possible actions the AV might take if the previous video continues. The blue line ahead of the AV indicates the path it will follow. Imagine the AV's speed is the same in scenarios 2 and 3. **Scenario 1: AV stays behind the cyclist** In this scenario, the AV only considers the cyclist's need for a safe distance and the road rules that require the AV to stay in its lane. But it doesn't consider the AV passenger's desire to get to the office on time. **Scenario 2: AV overtakes the cyclist on its own lane** In this scenario, the AV is solely concerned with the AV passenger's goal of getting to the office on time and the road rules that insist on it staying in its lane. But it doesn't consider the cyclist's wish to ride with a sense of safety. **Scenario 3: AV overtakes the cyclist by using the opposite lane** In this scenario, the AV is focused on the AV passenger's concern about getting to the office on time and the cyclist's concern about a safe distance. But it doesn't consider the road rules that require it to stay in its own lane.	
Q12	Which of the above scenarios do you prefer? Please answer this question by ranking the scenarios, with '1' indicating the highest preference.

(*continued on next page*)

Table A.6 (*continued*).

Question number	Question
	 Scenario 1: AV stays behind the cyclist  Scenario 2: AV overtakes the cyclist on its own lane  Scenario 3: AV overtakes the cyclist by using the opposite lane
Q13	Could you please explain the reasons behind the rank you provided in your previous answer?

Take a look at the video below!

Scenario 4: AV overtakes the cyclist by crossing some part of the opposite lane
In this scenario, the AV only considers the cyclist's need for a safe distance and the road rules that require the AV to stay in its lane. But it doesn't consider the AV passenger's desire to get to the office on time. You will now assess how much you believe the AV considers the intentions of three different stakeholders across 7 moments in this scenario. The same response table below will be used to answer Questions 14, 16, and 18.

	0	10	20	30	40	50	60	70	80	90	100
Time instance 1 	○	○	○	○	○	○	○	○	○	○	○
Time instance 2 	○	○	○	○	○	○	○	○	○	○	○
Time instance 3 	○	○	○	○	○	○	○	○	○	○	○
Time instance 4 	○	○	○	○	○	○	○	○	○	○	○
Time instance 5 	○	○	○	○	○	○	○	○	○	○	○
Time instance 6 	○	○	○	○	○	○	○	○	○	○	○
Time instance 7 	○	○	○	○	○	○	○	○	○	○	○

(*continued on next page*)

Table A.6 (*continued*).

Question number	Question
Q14	Imagine you are the AV passenger in scenario 4. Please assess, for each time instance, how much you believe the AV considers your intention to arrive at the office on time. (0 = Not consider at all; 100 = Fully consider)
Q15	Please clarify why the scores you provided for each time instance are either constant or change at each time instance
Q16	Imagine you are the cyclist that is passed by the AV in scenario 4. Please assess, for each time instance, how much you believe the AV considers your intention to bike with a sense of safety. (0 = Not consider at all; 100 = Fully consider)
Q17	Please clarify why the scores you provided for each time instance are either constant or change at each time instance
Q18	Imagine you are a road policymaker, and you see the AV briefly occupying small part of the opposite lane when overtaking the cyclist in scenario 4. Please assess, for each time instance, how much you believe the AV considers your intention putting a solid double yellow lines on the road for safety reasons. (0 = Not consider at all; 100 = Fully consider)
Q19	Please clarify why the scores you provided for each time instance are either constant or change at each time instance
Q20	How can an AV understands the intention of people on the road and other stakeholders in a real traffic situation?
Q21	From your point of view, what approach should AVs use to manage possible conflicts between the intentions of people on the road and other stakeholders in real traffic situations?

like safety or legality dominates. This underscores the challenge of implementing fixed hierarchies of reasons in AV design and supports the need for flexible, context-sensitive reason-tracking mechanisms.

Implicit and Explicit Prioritisation Among Reasons. While experts cited distinct reasons for AV behaviour, a clear prioritisation pattern emerged across both implicit and explicit analyses. In their open-ended reasoning, *safety* consistently served as the primary, non-negotiable consideration. Other reasons, such as *efficiency*, *comfort*, *interaction management*, and *acceptance* appeared were frequently mentioned but typically framed in relation to safety, making them secondary. *Rule compliance* was treated as conditional obligation: important when it aligned with safety and traffic flow, but often overridden when it did not, particularly when deviations could serve other high-priority values without compromising safety.

The explicit ranking reinforced this structure. Cyclists, as the most vulnerable road users in the scenario, were ranked first by majority of experts and never ranked last. AV passengers were never ranked first, most often placed last, and seen as holding lower priority due to their protected status inside the vehicle. Road policymakers occupied a middle position, valued for their role in setting systemic safety rules, but secondary to the immediate protection of vulnerable road users.

Together, these findings indicate that experts expect AVs to prioritise the safety of the most vulnerable above all else, followed by broader public and systemic safety, and lastly the convenience of more protected individuals such as AV passengers. Secondary values like comfort, efficiency, and environmental impact were seen as important, but acceptable only when they did not conflict with the safety of vulnerable users.

This prioritisation structure aligns with previous findings in AV ethics literature, where safety is widely regarded as the paramount consideration. For instance, the Moral Machine experiment (Awad et al., 2018) and expert-based studies (Milford et al., 2025) show broad consensus that risk minimisation should guide AV behaviour. Similarly, rule compliance has been treated as a conditional obligation in earlier work, particularly when rigid adherence may compromise safety or efficiency (Ma and Feng, 2024).

However, our results extend this discussion by demonstrating that such prioritisation patterns are not limited to high-stakes dilemmas but also emerge in routine, ethically ambiguous situations. This suggests that value trade-offs are not confined to emergencies, but are an ongoing feature of everyday AV operations. Moreover, while existing frameworks often rely on normative claims or abstract models of ethical reasoning, our findings reveal how experts actually balance and contextualise competing considerations — including legality, comfort, and environmental impact — based on real-world constraints and assumptions. This contributes a more fine-grained, empirically grounded understanding of how prioritisation unfolds across different layers of AV behaviour, and how certain reasons rise or recede in relevance depending on the driving context.

Principle of Reason Prioritisation in the Overtaking Cyclist Scenario. Building on these prioritisation patterns, we derive a principle of reason prioritisation in ethically ambiguous routine situations, particularly in the overtaking cyclist scenario. The synthesis of expert reasoning suggests three core guidelines.

First, AVs must always prioritise safety of the most vulnerable road users, such as cyclists in this scenario, over the interests of more protected users, including AV passengers.

Second, legal compliance should be the default behaviour. However, when strict rule-following would conflict with the safety of vulnerable users, or when it would go against collective interests such as avoiding discomfort, confusion, or indirect risks to road users and broader public, then carefully constrained deviations may be justified-provided that safety can be fully maintained.

Third, any justified deviation from traffic rules should be aimed at serving the greater public good rather than the convenience of the AV's occupants alone.

Unlike previous models that focus on rare, high-stakes scenarios—such as the MIT Moral Machine's global crash dilemma study (Awad et al., 2018), or algorithmic approaches like Augmented Utilitarianism (AU) (Gros et al., 2025)—our principle addresses a less examined but highly relevant domain: ethically ambiguous, everyday driving situations.

The Moral Machine revealed diverse cultural preferences in life-or-death crash dilemmas, underscoring the challenge of creating globally acceptable AV ethics. However, its emphasis on binary, extreme scenarios limits its applicability to nuanced real-world contexts. Similarly, AU advances ethical reasoning by incorporating diverse moral theories into adaptable goal functions grounded in empirical data. It uses attributes like harm, fairness, and legality, refined through participatory methods, to compute ethical decisions dynamically. In their work, Gros et al. (2025) designed AU to address both critical and non-critical contexts. However, the scenario used to illustrate and evaluate AU, such as brake-failure dilemmas or unusual narrow-road positioning with vulnerable pedestrians, are still relatively rare compared to the more commonplace, low-stakes situations AVs regularly encounter in everyday driving situations such as overtaking a slow cyclist.

In contrast, our principle complements these by providing a practice-orientated structure for routine AV behaviour, grounded in expert judgement rather than abstract moral theory or crowdsourced moral preferences. It emphasises prioritising the safety of the most vulnerable, allowing pragmatic flexibility when safety is not a risk, and applying a public-good focus when making legal exceptions. This structure

captures the complex value trade-offs that arise in cyclist overtaking scenarios in everyday AV operation. While these decisions may not involve stark life-or-death choices, they can meaningfully contribute to the development of AVs that are ethically designed to build public trust and acceptance.

Further, in addition to its relevance to everyday rather than exceptional moral dilemmas, our prioritisation principle has potential relevance to computational implementation. Prior work has shown that human-provided reasons can be operationalised by quantifying them through human-factors research and embedding them, with associated weights, into a trajectory evaluation framework to handle ethically challenging routine driving scenarios (Suryana et al., 2025b). This quantification and weighting structure provide a possible pathway for integrating our reason categories and prioritisation principle into future evaluation stages of trajectory planning.

Recommendations and Implications. The recommendations presented here apply directly to overtaking situations and may be applicable to similar ethically ambiguous driving scenarios—contexts in which AVs must navigate tensions between rule adherence, safety, comfort, and public expectations. Our findings on contextual reason prioritisation framework suggest that AV systems should incorporate flexible, context-aware decision logic. Developers should embed mechanisms to interpret reasons dynamically and prioritise them based on real-time conditions. Policymakers should consider enabling AVs to operate within regulated bounds that allow principled flexibility—particularly in routine scenarios where rigid rule-following may be counterproductive. Designers should also consider how human-like behaviour and acceptance intersect with safety and efficiency. Expert ID 11 highlighted that users are more likely to trust AVs that behave like human drivers, provided that safety is preserved. This has implications for user-centred AV design and for the development of regulatory frameworks that accommodate safe and socially acceptable deviations.

Limitations and Future Directions. This framework, while grounded in rich expert input, has limitations. Expert reasoning may reflect regional or cultural biases, and its generalisability across different AV scenarios — such as urban versus rural environments, or contexts with varying cultural norms — remains untested.

A further limitation concerns the interpretative nature of the implicit prioritisation analysis. Although the study included an explicit prioritisation task in which experts directly ranked the importance of reasons, the implicit prioritisation structure was inferred through qualitative interpretation of expert explanations. While independent coding and consensus resolution helped mitigate potential bias, these processes reduce rather than eliminate subjectivity. Future research could incorporate larger-scale empirical validation to examine how the prioritisation structure generalises beyond expert accounts.

Additionally, future research could examine how human reasons evolve over time in real or simulated driving contexts, and how AVs might adapt their decision-making while maintaining the safety of vulnerable road users and enabling practical action in situations where strict rule compliance conflicts with other considerations. While prior work (Suryana et al., 2025b) demonstrates the feasibility of operationalising human-provided reasons by quantifying them within a trajectory-evaluation framework, this work addresses the evaluation of candidate trajectories rather than full real-time control. The integration of the reason-prioritisation structure proposed in the present study into such computational frameworks has not yet been implemented or validated. Future work should develop and test this integration in both simulation and controlled real-world scenarios to assess practical effectiveness and feasibility.

4. Conclusion

This study derives a reason-prioritisation principle from expert reasoning about cyclist overtaking in ethically ambiguous routine driving situations. Grounded in the tracking condition of Meaningful Human Control (MHC), the principle supports an expert-derived framework that maps how human reasons influence AV decisions. Through qualitative interviews with AV experts, we identified thirteen categories of reasons that influence manoeuvre planning, structured across normative, strategic, tactical, and operational levels of AV behaviour, and linked to the roles of relevant human agents.

The findings show that AV decisions often involve multiple overlapping reasons, with *safety* consistently regarded as the primary concern. Other reasons, such as *efficiency*, *comfort*, and *acceptance*, were frequently mentioned alongside safety but rarely overrode it. *Rule compliance* was treated as a conditional obligation and often deprioritised when it conflicted with more context-sensitive goals. These prioritisation patterns a set of empirically grounded principles that upholds safety while permitting carefully constrained deviations from legal rules when justified by practical or ethical considerations.

By mapping expert reasons into a layered structure and positioning them within a conceptual representation of reason-based AV decision-making, this case-specific model offers guidance that complements existing high-stakes ethical approaches and may inform future research on AV behaviour in dynamic, real-world scenarios. Future work should evaluate the broader applicability of this representation across diverse cultural contexts, automation levels, and driving situations.

CRediT authorship contribution statement

Lucas Elbert Suryana: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Simeon Calvert:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Arkady Zgonnikov:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Bart van Arem:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lucas Elbert Suryana reports financial support was provided by the Indonesia Endowment Fund for Education. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the use of OpenAI ChatGPT to enhance the clarity and conciseness of, and to proofread, the manuscript; the original text was entirely authored by the researchers. This work was supported by the Indonesia Endowment Fund for Education (LPDP) under Grant No. 0006552/TRA/D/19/lpdp2021.

Appendix A. Questionnaire

See Table A.6.

Data availability

Data will be made available on request.

References

Aasvik, O., Hagenzieker, M., Ulleberg, P., 2025. I trust norway – investigating acceptance of shared autonomous shuttles using open and closed questions in short-form street interviews. Transp. Res. Interdiscip. Perspect. 31, 101414. http://dx.doi.org/10.1016/j.trip.2025.101414.

Asaro, P., 2012. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. Int. Rev. Red Cross 94, 687–709.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., Rahwan, I., 2018. The moral machine experiment. Nat. 563, 59–64. http://dx.doi.org/10.1038/S41586-018-0637-6.

Bergmann, L.T., 2022. Ethical issues in automated driving—opportunities, dangers, and obligations. User Exp. Des. Era Autom. Driv. 99–121. http://dx.doi.org/10.1007/978-3-030-77726-5-5.

Beringhoff, F., Greenyer, J., Roesener, C., Tichy, M., 2022. Thirty-one challenges in testing automated vehicles: Interviews with experts from industry and research. In: 2022 IEEE Intelligent Vehicles Symposium IV. IEEE, pp. 360–366. http://dx.doi.org/10.1109/IV51971.2022.9827097.

Bonnefon, J.-F., Černy, J., Devillier, N., Johansson, V., Kovacikova, T., Martens, M., Mladenovic, M., Palade, P., Reed, N., et al., 2020. Ethics of connected and automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility. Placeholder J. http://dx.doi.org/10.2777/035239.

Bonnefon, J.-F., Shariff, A., Rahwan, I., 2016. The social dilemma of autonomous vehicles. Sci. 352, 1573–1576. http://dx.doi.org/10.1126/science.aaf2654.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3, 77–101. http://dx.doi.org/10.1191/1478088706qp063oa.

Calvert, S.C., Mecacci, G., 2020. A conceptual control system description of cooperative and automated driving in mixed urban traffic with meaningful human control for design and evaluation. IEEE Open J. Intell. Transp. Syst. 1, 147–158. http://dx.doi.org/10.1109/OJITS.2020.3021461.

Campbell, J.L., Quincy, C., Osserman, J., Pedersen, O.K., 2013. Coding in-depth semistructured interviews: problems of unitization and intercoder reliability and agreement. Sociol. Methods Res. 42, 294–320. http://dx.doi.org/10.1177/0049124113500475.

Charmaz, K., 2014. Constructing Grounded Theory. SAGE publications Ltd.

D'Amato, A., Dancel, S., Pilutti, J., Tellis, L., Frascaroli, E., Gerdes, J., 2022. Exceptional driving principles for autonomous vehicles. JL Mobil. 1, https://repository.law.umich.edu/jlm/vol2022/iss1/2/.

Dreger, F.A., de Winter, J.C., Happee, R., 2020. How do drivers merge heavy goods vehicles onto freeways? a semi-structured interview unveiling needs for communication and support. Cogn. Technol. Work. 22, 825–842. http://dx.doi.org/10.1007/S10111-019-00601-3.

Dubljević, V., Douglas, S., Milojevich, J., Ajmeri, N., Bauer, W.A., List, G., Singh, M.P., 2023. Moral and social ramifications of autonomous vehicles: a qualitative study of the perceptions of professional drivers. Behav. Inf. Technol. 42, 1271–1278. http://dx.doi.org/10.1080/0144929X.2022.2070078.

Fraade-Blanar, L., Blumenthal, M.S., Anderson, J.M., Kalra, N., 2018. Measuring Automated Vehicle Safety: Forging a Framework. RAND Corporation, Santa Monica, Calif, https://Www.Rand.Org/T/RR2662. Library of Congress Cataloging-in-Publication Data Available.

Geisslinger, M., Poszler, F., Betz, J., Lütge, C., Lienkamp, M., 2021. Autonomous driving ethics: From trolley problem to ethics of risk. Philos. Technol. 34, 1033–1055. http://dx.doi.org/10.1007/S13347-021-00449-4.

Geisslinger, M., Poszler, F., Lienkamp, M., 2023. An ethical trajectory planning algorithm for autonomous vehicles. Nat. Mach. Intell. 5, 137–144. http://dx.doi.org/10.1038/S42256-022-00607-Z.

Graneheim, U., Lundman, B., 2004. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ. Today 24, 105–112. http://dx.doi.org/10.1016/J.Nedt.2003.10.001.

Gros, C., Werkhoven, P., Kester, L., Martens, M., 2025. A methodology for ethical decision-making in automated vehicles. AI SOCIETY 1–12. http://dx.doi.org/10.1007/S00146-025-02370-2.

Habibullah, K.M., Heyn, H.-M., Gay, G., Horkoff, J., Knauss, E., Borg, M., Knauss, A., Sivencrona, H., Li, P.J., 2024. Requirements and software engineering for automotive perception systems: an interview study. Requir. Eng. 29, 25–48. http://dx.doi.org/10.1007/S00766-023-00410-1.

Hilgarter, K., Granig, P., 2020. Public perception of autonomous vehicles: A qualitative study based on interviews after riding an autonomous shuttle. Transp. Res. Part F: Traffic Psychol. Behav. 72, 226–243. http://dx.doi.org/10.1016/J.Trf.2020.05.012.

Hill, C.E., Thompson, B.J., Williams, E.N., 2005. Consensual qualitative research: An update. J. Couns. Psychol. vol. 52, 196–205. http://dx.doi.org/10.1037/0022-0167.52.2.196.

Himmelreich, J., 2018. Never mind the trolley: The ethics of autonomous vehicles in mundane situations. Ethical Theory Moral Pr. 21, 669–684. http://dx.doi.org/10.1007/S10677-018-9896-4.

Homayounirad, Amir, Liscio, Enrico, Wang, Tong, Jonker, Catholijn M., Siebert, Luciano Cavalcante, 2025. Will annotators disagree? identifying subjectivity in value-laden arguments. In: Findings of the Association for Computational Linguistics: EMNLP 2025. Association for Computational Linguistics, pp. 15237–15252. http://dx.doi.org/10.18653/v1/2025.findings-emnlp.824.

Horowitz, M.C., Scharre, P., 2015. Meaningful Human Control in Weapon Systems: A Primer. Center for a New American Security, Working Paper.

Hsieh, H.-F., Shannon, S.E., 2005. Three approaches to qualitative content analysis. Qual. Health Res. 15, 1277–1288. http://dx.doi.org/10.1177/1049732305276687.

Keeling, G., 2020. Why trolley problems matter for the ethics of automated vehicles. Sci. Eng. Ethics 26, 293–307. http://dx.doi.org/10.1007/S11948-019-00096-1.

Krippendorff, K., 2018. Content Analysis: An Introduction To Its Methodology, fourth ed. SAGE Publications, http://dx.doi.org/10.4135/9781071878781.

Lee, S.C., Nadri, C., Sanghavi, H., Jeon, M., 2020. Exploring user needs and design requirements in fully automated vehicles. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–9. http://dx.doi.org/10.1145/3334480.3382881.

Liu, P., Liu, J., 2021. Selfish or utilitarian automated vehicles? deontological evaluation and public acceptance. Int. J. Human–Computer Interact. 37, 1231–1242. http://dx.doi.org/10.1080/10447318.2021.1876357.

Longhurst, R., Johnston, L., 2023. 10 semi-structured interviews and focus groups. Key Methods Geogr. 168.

Lu, H., Zhu, M., Lu, C., Feng, S., Wang, X., Wang, Y., Yang, H., 2025. Empowering safer socially sensitive autonomous vehicles using human-plausible cognitive encoding. In: Proceedings of the National Academy of Sciences, vol. 122, http://dx.doi.org/10.1073/pnas.2401626122.

Ma, J., Feng, X., 2024. Analysing the effects of scenario-based explanations on automated vehicle hmis from objective and subjective perspectives. Sustain. 16, 63. http://dx.doi.org/10.3390/Su16010063.

Mecacci, G., Santoni de Sio, F., 2020. Meaningful human control as reason-responsiveness the case of dual-mode vehicles. Ethics Inf. Technol. 22, 103–115. http://dx.doi.org/10.1007/S10676-019-09519-W.

Meder, B., Fleischhut, N., Krumnau, N.-C., Waldmann, M.R., 2019. How should autonomous cars drive? a preference for defaults in moral judgments under risk and uncertainty. Risk Anal. 39, 295–314. http://dx.doi.org/10.1111/Risa.13178.

Milford, S.R., Malgir, B.Z., Elger, B.S., Shaw, D.M., 2025. All things equal: ethical principles governing why autonomous vehicle experts change or retain their opinions in trolley problems—a qualitative study. Front. Robot. AI 12, http://dx.doi.org/10.3389/frobt.2025.1544272.

Nordhoff, S., Lee, J.D., Calvert, S.C., Berge, S., Hagenzieker, M., Happee, R., 2023. (mis-) use of standard autopilot and full self-driving (fsd) beta: results from interviews with users of tesla's fsd beta. Front. Psychol. 14, 1101520. http://dx.doi.org/10.3389/Fpsyg.2023.1101520.

Nyholm, S., Smids, J., 2016. The ethics of accident-algorithms for self-driving cars: An applied trolley problem? Ethical Theory Moral Pr. 19, 1275–1289. http://dx.doi.org/10.1007/S10677-016-9745-2.

Olleja, P., Markkula, G., Bärgman, J., 2025. Validation of human benchmark models for automated driving system approval: How competent and careful are they really? Accid. Anal. Prev. 213, 107922. http://dx.doi.org/10.1016/J.Aap.2025.107922.

Rhim, H.J., Urban, J.M., 2021. A deeper look at moral dilemmas in autonomous driving: The integrative ethical decision-making framework. Front. Robot. AI 8, 632394. http://dx.doi.org/10.3389/Frobt.2021.632394.

Saber, E.M., Kostidis, S.-C., Politis, I., 2024. Ethical Dilemmas in Autonomous Driving: Philosophical, Social, and Public Policy Implications. Springer, pp. 7–20. http://dx.doi.org/10.1007/978-3-031-55044-7-2.

Schreier, M., 2012. Qualitative Content Analysis in Practice. Sage, http://dx.doi.org/10.4135/9781529682571.

Schwarting, W., Alonso-Mora, J., Rus, D., 2018. Planning and decision-making for autonomous vehicles. Annu. Rev. Control. Robot. Auton. Syst. 1, 187–210. http://dx.doi.org/10.1146/Annurev-Control-060117-105157.

Santoni de Sio, F., Van den Hoven, J., 2018. Meaningful human control over autonomous systems: A philosophical account. Front. Robot. AI 5, 15. http://dx.doi.org/10.3389/frobt.2018.00015.

Suryana, L.E., Nordhoff, S., Calvert, S., Zgonnikov, A., van Arem, B., 2025. Meaningful human control of partially automated driving systems: insights from interviews with tesla users. Transp. Res. Part F: Traffic Psychol. Behav. 113, 213–236. http://dx.doi.org/10.1016/J.Trf.2025.04.026.

Suryana, L.E., Rahmani, S., Calvert, S.C., Zgonnikov, A., van Arem, B., 2025b. A framework for human-reason-aligned trajectory evaluation in automated vehicles. arXiv preprint arXiv:2507.23324.

Swain, R., Truelove, V., Rakotonirainy, A., Kaye, S.-A., 2023. A comparison of the views of experts and the public on automated vehicles technologies and societal implications. Technol. Soc. 74, 102288. http://dx.doi.org/10.1016/J.Techsoc.2023.102288.

Tabone, W., Winter, J.De., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P.A., et al., 2021. Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. Transp. Res. Interdiscip. Perspect. 9, 100293. http://dx.doi.org/10.1016/J.Trip.2020.100293.

Thornton, S.M., Lewis, F.E., Zhang, V., Kochenderfer, M.J., Gerdes, J.C., 2018. Value sensitive design for autonomous vehicle motion planning. In: 2018 IEEE Intelligent Vehicles Symposium IV. IEEE, pp. 1157–1162. http://dx.doi.org/10.1109/IVS.2018.8500441.

UNECE, 2023. UN regulation no. 157: Uniform provisions concerning the approval of vehicles with regard to automated lane keeping systems. https://unece.org/transport/documents/2023/03/standards/un-regulation-no-157-amend4. (Accessed 07 December 2024).

Vaismoradi, M., Turunen, H., Bondas, T., 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs. Health Sci. 15, 398–405. http://dx.doi.org/10.1111/Nhs.12048.

Veluwenkamp, H., 2022. Reasons for meaningful human control. Ethics Inf. Technol. 24, 51. http://dx.doi.org/10.1007/S10676-022-09673-8.

Wang, H., Huang, Y., Khajepour, A., Cao, D., Lv, C., 2020. Ethical decision-making platform in autonomous vehicles with lexicographic optimization based model predictive controller. IEEE Trans. Veh. Technol. 69, 8164–8175. http://dx.doi.org/10.1109/MITS.2019.2953556.