People Attribute Purpose to Autonomous Vehicles When Explaining Their Behavior

Balint Gyevnar[a,*], Stephanie Droop[a], Tadeg Quillien[a], Shay B. Cohen[a], Neil R. Bramley[b], Christopher G. Lucas[a] and Stefano V. Albrecht[a]

[a]School of Informatics; University of Edinburgh
[b]School of Philosophy, Psychology and Language Sciences; University of Edinburgh

Abstract. Cognitive science can help us understand which explanations people might expect, and in which format they frame these explanations, whether causal, counterfactual, or teleological (i.e., purpose-oriented). Understanding the relevance of these concepts is crucial for building good explainable AI (XAI) which offers recourse and actionability. Focusing on autonomous driving, a complex decision-making domain, we report empirical data from two surveys on (i) how people explain the behavior of autonomous vehicles in 14 unique scenarios ($N_1 = 54$), and (ii) how they perceive these explanations in terms of complexity, quality, and trustworthiness ($N_2 = 356$). Participants deemed teleological explanations significantly better quality than counterfactual ones, with perceived teleology being the best predictor of perceived quality and trustworthiness. Neither the perceived teleology nor the quality were affected by whether the car was an autonomous vehicle or driven by a person. This indicates that people use teleology to evaluate information about not just other people but also autonomous vehicles. Taken together, our findings highlight the importance of explanations that are framed in terms of purpose rather than just, as is standard in XAI, the causal mechanisms involved. We release the 14 scenarios and more than 1,300 elicited explanations publicly as the Human Explanations for Autonomous Driving Decisions (HEADD) dataset.

1 Introduction

The field of XAI has attracted much multi-disciplinary attention in recent years. There is a shift from viewing XAI as a sterile scalpel for dissecting AI models towards using XAI to coordinate knowledge both between experts and non-expert stakeholders on one hand, and, in a more expansive near-future vision, between natural and artificial agents on the other [13, 39]. Cross-disciplinary work in XAI draws on, for example, social sciences [35], human-computer interaction [9, 36], psychology [49, 52, 10], philosophy [54], and natural language processing [45], fueling the increasing emphasis on *human-centred XAI*. However, a majority of human-centred XAI is still based on theoretical arguments that have seen limited empirical testing.

This limits the reach and potential of these methods despite promising to accessibly explain AI systems. For example, consider the topic of causality, which is well established in theory as a cornerstone of useful XAI systems [4, 35]. Much practical work has gone into constructing causal and counterfactual explanations [47, 14]. While evaluation with humans is used to validate the generated explanations, this is by no means a widespread practice [20]. Furthermore, many user studies of XAI focus on evaluating the effectiveness of explanations only in terms of various quantitative metrics (for example, perceived trust [53] and reliability [57]), but are not designed to assess whether people would give these explanations in the first place.

This means that research in XAI provides methodological recommendations for creating causal explanations but offers few insights into whether people would actually explain AI systems according to these suggestions. Accordingly, we also have a limited understanding of the way people interpret different kinds of causal explanations. For example, Miller [35] recommends using contrastive explanations of the form "Why P instead of Q" based on an extensive review of philosophical and social literature. However, this raises important questions related to the contents of P and Q. For instance, they could refer to a system's intrinsic goals or to the causal mechanisms that affect its output. In this paper, we suggest that getting traction on these questions will involve a combination of integrating theoretical insights from cognitive science and conducting targeted empirical studies of how people generate and interpret explanations in context.

Research in cognitive science reveals that generating and interpreting causal explanations involves sophisticated computations and inferences [28, 44, 24, 38]. In particular, humans often adopt an *intentional stance* [7] when they explain the behavior of a system, ascribing goals, beliefs, and intentions to the system. These explanations are inherently *teleological*; they explain an agent's decision in terms of the purpose of that decision. In contrast, XAI models generate *mechanistic explanations* that appeal to the causal, usually mathematical, logic and external conditions (e.g., input data) involved in leading to the decision [46]. An explanation of the inherent purpose or goal of the decision is often lacking.

Furthermore, to understand how people interpret causal explanations, it is also important to assess whether they tend to give mechanistic or teleological explanations, even when the agent is not a person but a machine. We also need to understand whether people's preferences for teleological and mechanistic explanations are at odds with explanations produced by applications of different theories of causation. This improved understanding would allow us to base the design of causal explanations on empirically validated principles.

In this paper, we discuss relevant research in cognitive science on causality, counterfactuals, and teleology as they relate to explanation. We then report the results of two surveys with human participants recruited through the online crowd-sourcing platform Prolific ($N_1 = 54$; $N_2 = 356$). Rather than taking a controlled toy environment detached

* Corresponding Author. Email: balint.gyevnar@ed.ac.uk

from the idiosyncrasies of the real world, we base our experiments on AI decision-making for autonomous vehicles (AV), a popular domain of application for XAI [25]. This setup has the advantage of (i) eliciting explanations from real people rather than generating them ourselves; (ii) having realistic situations where the ground truth is still relatively accessible to the explainer; (iii) allowing us to explore various scenarios while keeping constant the overall context.

In the first survey, participants were asked to watch 7 short driving scenarios sampled from a total of 14 scenarios with multiple interacting vehicles. They were then instructed to explain in their own words the behavior of a selected ego vehicle along different explanatory modes (descriptive, mechanistic, teleological, and counterfactual). In the second survey, a different set of participants evaluated these explanations along various dimensions, such as the number of perceived causes, perceived complexity, quality, and trustworthiness.

We find that people preferred teleological and mechanistic concepts to counterfactual explanations. They were also just as likely to refer to the mental states of autonomous vehicles as of human drivers. In addition, perceived teleology was consistently ranked as the best predictor of explanation quality and trustworthiness. Based on these results, we recommend the field of XAI consider the use of different explanatory modes as an important axis of analysis, especially focusing on the role and effect of teleology. To summarise, our contributions are:

- Discussions of cognitive science on causality, counterfactuals, and teleology as they relate to explanation, highlighting the role of different **explanatory modes** (i.e., teleological, mechanistic);
- We curate and release a **novel dataset** of human-elicited and evaluated explanations for autonomous driving, called the Human Explanations for Autonomous Driving Decisions (HEADD) dataset.[1]
- Two user studies providing evidence that **teleology is preferred** by people when explaining an agent's decision, regardless of whether the agent is perceived as human or machine.

2 Foundations of Explanation

Explanation has a close relationship with causality [35] and, although there are nuances in the details of how each is formalised [15, 17, 27], it is broadly accepted that explaining a phenomenon often involves asserting its cause. In turn, causality has a close relationship with counterfactuals [27, 26, 44]. The counterfactual theory of causation, prominent in philosophy and psychology, holds that the meaning of 'C caused E' is (roughly), that if C had not happened then E would not have happened [27, 51, 40].

Although the term *causal* can broadly be used for the whole class including counterfactuals, it is nonetheless useful to distinguish a narrower meaning of causal explanations from counterfactual explanations. Counterfactual explanations explicitly highlight ways that things could have turned out differently (e.g., "If I had done x, then y would have happened"), whereas causal explanations as we use the term refer explicitly to a mechanism (e.g., "y happened because x happened"). Empirically, when people give causal explanations they tend to focus on strong causes that co-vary with an outcome, for example, 'a drunk driver caused the crash' [44]. When constructing counterfactuals, they tend to focus on controllable conditions that could have altered the outcome, for example 'the crash would not have happened if the protagonist had driven home a different way' [34].

Recent research has studied whether counterfactual or causal explanations of an AI system are more effective. Empirical studies have found that users who are given a counterfactual explanation of a decision made by an autonomous system report more satisfaction with that explanation than users who are given a causal explanation [5, 50]. Counterfactual explanations are also more effective at improving the user's ability to predict the behavior of the system. For example, Celar and Byrne ([5]) showed participants the decisions made by an algorithm designed to determine whether someone's blood alcohol content (BAC) is above or below the legal limit for driving. The decisions were accompanied by a counterfactual explanation ('if the person had drunk 3 units of alcohol, they would be below the limit'), a causal explanation ('drinking 5 units of alcohol caused the person to be above the limit'), or no explanation. The fact the counterfactual explanations were rated as more satisfying than the causal explanations [5] is intriguing when placed next to computational models of counterfactual reasoning which suggest people simulate several different counterfactual worlds when they generate a causal explanation [32, 42, 44, 8].

However, the research that suggests an advantage for counterfactual explanations has focused on counterfactuals that typically highlight one possible alternative state of the world and on explanations generated by very simple algorithms. For example, determining whether someone is under the legal BAC limit can be done by applying simple rules. The question arises as to whether this advantage of counterfactuals generalizes to more complex settings. In particular, if the system is sufficiently complex, like a self-driving car, people might take an *intentional stance* toward that system, conceiving of it as an agent.

2.1 Explanatory modes

The human mind entertains different types of causal explanations [2, 12, 19]. For example, we can think of someone's actions in mechanistic terms ('she turned the handle with her hand'), or in terms of the person's goals and desires ('she opened the door to let her friends in'). Explainable AI most often takes the former mechanistic stance, as the design of XAI methods is usually targeted at tracing the causal chain from input to output in terms of mathematical manipulations and algorithmic mechanisms. In contrast, the latter example corresponds to taking an *intentional stance*, whereby we can conveniently characterise and therefore predict an agent's behavior by attributing to them mental states such as beliefs, desires and intentions [7]. The intentional stance reliably emerges very early in development [11], and its computational underpinnings are beginning to be mapped out by cognitive scientists [33, 3, 43].

Explanations that use the intentional stance are *teleological*: they explain something in terms of the purpose it serves. For example, saying that Mary opened the fridge *in order* to get some milk is a teleological explanation because it explains Mary's action in terms of its purpose. Teleological explanations are intuitive to the human mind, even outside the domain of psychological reasoning. They are readily produced and endorsed by children [21, 29]. Adults sometimes endorse teleological explanations even for inanimate processes, e.g. when under time pressure [22, 23]. Teleological explanations are generally useful because they identify causes that are *robust* to changes in background circumstances: for example, my intention to drive home would have caused me to get home even if my usual route was closed, because I would then have taken a different route [30, 28, 31, 7].

Evidence suggests that people can adopt the intentional stance toward artificial systems [41, 6]. Whenever this is the case, teleological explanations of autonomous system decisions might be particularly effective, because they are consistent with the way the user intuitively represents the system [56]. In our study, we test whether this is the case for autonomous driving which involves explaining the coupled

[1] HEADD is available at https://datashare.ed.ac.uk/handle/10283/8714 and our code for analysis at https://github.com/Stephaniedroop/AV_Explanations.

decision-making of multiple agents with a mixture of both human and artificial agents. Will people prefer explanations of a self-driving car that are framed in teleological terms, or will they prefer mechanistic explanations in terms of causal or counterfactual terms? In the next section, we explain the relevance of teleology to the contrast between counterfactual and causal explanations.

2.2 Teleology and counterfactuals

There is an interesting potential tension between counterfactual explanations and teleology. One recipe for generating counterfactual explanations is to take inspiration from the counterfactual theory of causation and produce a counterfactual of the form 'if C had not happened, then E would not have happened' where C is the cause of outcome E. Consider for example a scenario where a self-driving car stops because pedestrians are crossing the road. A counterfactual explanation obeying the standard template would be 'if the pedestrians were not crossing the road, the car would not have stopped'. This explanation effectively highlights the material cause of the car's behavior, but does not have teleological content.

To generate a counterfactual explanation of the car's behavior in teleological terms, we might instead say 'if the car had not stopped, it would have run over the pedestrians'. This counterfactual implicitly highlights the *reason* for the car's behavior: the car stopped because if it had not, a bad consequence would have followed. Note that this kind of counterfactual has a different structure than the standard 'if ¬C then ¬E' template: Instead of altering the cause (the pedestrians crossing), we alter the effect (the car stopping). Teleological explanations are still implicitly causal: the car stopped because it computed that not stopping would have worse consequences than not stopping. Nonetheless, the complexity of teleological explanations might mean that they will be difficult to express in terms of more standard counterfactuals.

These considerations suggest the following prediction: if participants intuitively conceive of self-driving cars as agents, and apply the intentional stance toward them, they might not be satisfied by counterfactual explanations of their behavior, especially if these counterfactuals are of the form 'if ¬C then ¬E'.

2.3 The present study

We curate a dataset of human-generated explanations of the decisions of autonomous vehicles, as well as evaluations of these explanations by a different set of participants. We anticipate that this dataset can shed light on a variety of questions regarding both explanation generation and interpretation. Below, we focus on our main predictions as they relate to the issues we reviewed in this section.

Our key experimental manipulation was the *explanation prompt*, i.e. the type of explanations that participants were asked to generate.[2] Our **Counterfactual** explanation prompt requested participants to 'describe changes to the scenario so that the blue vehicle takes different actions'. That is, it requested a counterfactual of the type 'if ¬C then ¬E', which is difficult to interpret in teleological terms (see above). Our **Mechanistic** explanation prompt asked participants to 'explain how the blue car was influenced in the scenario to take these actions'. This request elicits a causal explanation and does not specifically target teleological features of the situation (although it does not preclude them). Finally, our **Teleological** explanation prompt requested participants to 'explain why the blue car took these actions over different actions to reach its goal', foregrounding the agent's intentions, and emphasizing that the relevant counterfactuals are ones where the agent acts differently.

Suppose there is a robust preference for counterfactual explanation for artificial systems. In that case, we expect that explanations generated in response to the Counterfactual prompt should be rated as better than explanations generated in response to the other two prompts (replicating the results in [5, 50]). In contrast, if the behavior of a complex agent (such as a self-driving car) activates an intentional stance, then the explanations generated in response to the Teleological (and possibly Mechanistic) prompt should be rated as better than the Counterfactual explanations. Additionally, if participants adopt an intentional stance, we predict that explanations containing more teleological features should be seen as more satisfying.

3 Survey Methodology

Here, we describe the overall study procedure for our experiment including the survey methodology, the independent measurement variables, the form of the collected data, and summary statistics of the data. We performed two surveys as part of the user study. In the first survey, we elicited natural language explanations from participants about the behavior of an autonomous vehicle in various driving scenarios. The scenarios were shown in top-down animated videos (example snapshots shown in Figure 1) and participants were asked to write in four explanatory modes that included descriptive, mechanistic, teleological, and counterfactual explanations. In the second survey, we took these explanations and asked another set of participants to evaluate them according to their causal content and subjective quality.

For both surveys, we used the online crowd-sourcing platform Prolific to recruit participants. We recruited from the USA, as the surveys were in English and the video recordings used right-handed traffic. We filtered for participants whose first language was English. Participants were paid a pro-rated fee of £11 per hour and the study was approved by the ethics committee of the authors' institution. The ethics approval and a consent form were shown to the participants before allowing them to complete the surveys.

3.1 Survey 1: eliciting explanations

A total of 54 participants (25 male and 29 female) filled out the first survey with a median duration of completion of 25 minutes and 37 seconds. The participants' ages ranged between 19 to 73 years, with a median of 36 years. The majority of participants had some form of tertiary education (49 people) with the largest group having a Bachelor's degree (19 people). We also asked participants about their driving skills. Most participants reported having a valid driver's license (48 people) and the large majority of participants had been driving for at least 2 years at the time of taking the survey (44 people).

3.1.1 Survey design

After having consented to take the survey, participants were shown 7 driving scenarios picked randomly with equal chance out of a collection of 14 scenarios.[3] At the start of each scenario, participants were shown a short (5 to 15-second-long) top-down animated video recorded in the software RoadRunner 2023a by MathWorks. Example snapshots from such videos are shown in Figure 1. They were also explicitly told what the goal of the blue vehicle was. Participants were then asked to answer the following four questions in their own words:

[2] Note, we capitalise the type of explanation prompt as an experimental manipulation to differentiate it from the underlying explanatory mode.

[3] For the complete description of scenarios, refer to the appendix and the metadata descriptions in the HEADD dataset.

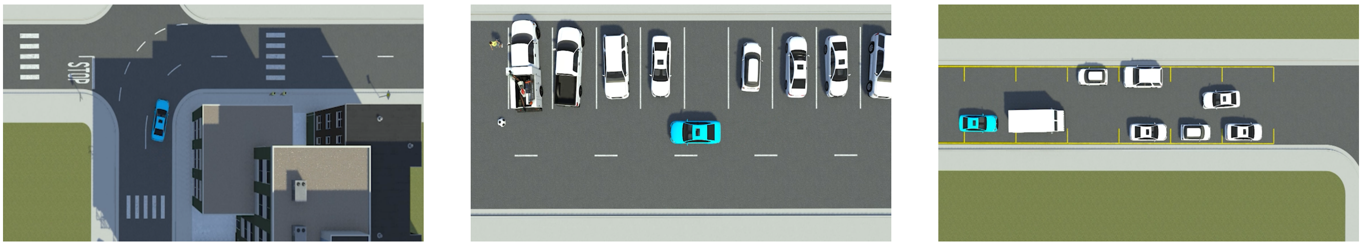

Figure 1. Three example scenarios from HEADD. Participants were always asked to explain the behavior of the blue vehicle. (**Left**; #8). The blue car slows down before turning right, as its view is blocked by a building. Once the view is clear, the blue car notices pedestrians at the crossing and stops. (**Mid**; #11). The blue car is passing a row of parked cars when it perceives a ball rolling onto the road. It sharply breaks, as a child emerges from behind a lorry. (**Right**; #12) The blue car waits behind a lorry obscuring its vision of the road. It keeps on waiting as the lorry passes between the parked cars, to avoid other oncoming vehicles.

1. *Descriptive*: describe the actions of the blue car;
2. *Teleological*: explain why the blue car took these actions over different actions to reach its goal;
3. *Mechanistic*: explain how the blue car was influenced in the scenario to take these actions;
4. *Counterfactual*: describe changes to the scenario so that the blue vehicle takes different actions.

Participants were able to re-watch the video clip as many times as they wished, at any point during this phase. At the end of the survey, participants were asked to answer questions regarding their driving habits (driving license, frequency, annual covered distance) and their demographics (age, education level, gender). We collected 1,308 free-text explanations across all four explanatory modes.

3.1.2 Independent variables

We used three independent variables to vary the experimental setup in a between-subjects design:

- *Scenario*: which 7 scenarios were selected for the participant. Scenarios were picked to ensure an equal coverage for all 14 scenarios;
- *AV*: for half of all participants, we showed "self-driving car" in place of "vehicle". This was done to measure the effects of participants knowing whether the blue car was controlled by a machine;
- *AV explanation*: for half of all participants who were shown "self-driving car" in the previous variable, we also showed a high-level explanation of how the AV works. This was done to understand whether knowing how the AV worked affects people's explanations.

Our within-subject independent variable was the *explanation prompt*, i.e. the requested explanatory mode for the explanation (descriptive, teleological, mechanistic or counterfactual; see above).

3.2 Survey 2: evaluating explanations

A total of 356 participants (176 male, 177 female, and 3 other) filled out the second survey with a median duration of completion of 23 minutes and 7 seconds. Participants' age ranged between 19 to 83 years, with a median of 38 years. The majority of participants had some form of tertiary education (320 people) with the largest group having a Bachelor's degree (150 people). Most participants reported having a valid driver's license (330 people) and the majority of participants had been driving for at least 2 years at the time of the survey (329 people). The participants of Survey 1 and 2 were different.

3.2.1 Survey design

After consenting to take the survey, participants were guided through a simple driving scenario that explained to them the core concepts required to evaluate the explanations from Survey 1. This consisted of teaching them the various explanatory modes (teleological, mechanistic, contrastive) and the definition of a cause. We excluded descriptive explanations because they do not describe causal relationships.

Following the teaching phase, we sampled at random one scenario from the same 14 scenarios as in Survey 1, and let the participant watch the video clip of the scenario. Before seeing any explanations, participants were asked to answer the following two questions to understand the factors behind whether and what sort of explanations participants wanted for the selected scenario:

- *Explanatory need*: how much an explanation would help better understand the causes behind the blue self-driving car's actions (5-point Likert scale);
- *Curiosity*: why would the participant ask for an explanation (multiple choice among 5 options regarding the decision-making of the blue vehicle targeting correctness, predictability, alternative action and outcome, and surprise).

For the selected scenario, we randomly picked 13 explanations written by participants of Survey 1. The participants were not explicitly told the goal of the blue vehicle. For each selected explanation, they were asked to answer the following questions, with parentheses containing the data type from the question:

- *Explanatory mode*: how much the displayed explanation targets the teleological or mechanistic explanatory mode. Participants rated on three 5-point Likert scales the extent to which each of the following aspects of the scenario was targeted by the explanation: the goal(s), desire(s), or intention(s) of the blue car (variable name: *Teleology*); the actions of the other traffic participants (*MechanisticAgent*); the road layout or traffic laws (*MechanisticLayout*);
- *Causes*: how many causes were mentioned in the explanation as perceived by the participant (non-negative integer);
- *Contrastive*: whether the explanation was contrastive (boolean);
- *Preferences*: participant ratings of the explanation according to completeness, sufficiency, trustworthiness, satisfaction (5-point Likert scale per attribute).

Finally, participants were asked the same driving experience- and demographics-related questions as in Survey 1. As such, the only independent variables that were manipulated in Survey 2 were which scenario and which explanations were picked for a given participant. No other manipulations were performed as the goal of Survey 2 was to provide rich evaluations of the explanations from Survey 1.

Table 1. Highest-quality explanations (as rated by Survey 2 participants) for each explanation type, for scenario #8 shown in Figure 1.

Explanation type	Example human-generated explanation
Teleological	'The blue car took those actions as it is the safest, slowing down to make the turn successfully and also stopping to allow pedestrians to cross instead of betting on the fact that the pedestrians are paying attention to the road.'
Mechanistic	'The blue car was influenced by the two pedestrians waiting to cross as it slowed to a complete stop allowing them to cross. The right turn was also 90 degrees which required the car to slow in order to make a successful turn.'
Counterfactual	'If there were no pedestrians in the scenario, then the car could have just immediately sped up to the speed limit instead of stopping in front of the crosswalk.'

Each explanation from Survey 1 received between 5 to 7 independent evaluations, yielding a total of 4,963 evaluations.

3.2.2 Linguistic processing

We also investigated how linguistic complexity correlates with the perceived qualities of explanations, so we performed processing steps to the explanations to provide additional linguistic data to analyse.

First, each explanation was processed using the Spacy NLP library, which, importantly for us, tokenizes and lemmatizes each explanation, while also providing the dependency parse trees for them. Note, that an explanation may be composed of multiple sentences, in which case, the parsing was performed per sentence.

Second, for each explanation, we extracted standard measures of complexity: the number of alphanumeric characters, tokens, unique lemmas, and sentences. We have calculated this both across the entire explanation and normalised by the number of sentences. We also found the average dependency separation between tokens in a sentence, to encode the distance between two dependent tokens.

While we performed our analyses for each measure of complexity, we found that our result did not change significantly depending on which measure we picked, therefore, here we only report results using the number of tokens (words) in the sentence.

4 Results

Table 1 shows an example participant-generated explanation for each explanation prompt. As a manipulation check, we find that explanations generated in response to a Teleological prompt do exhibit more teleological features than explanations generated in response to other prompts, $p < .001$. We also find that they are less likely to mention the actions of other agents, $p < .001$. In contrast, the type of explanation prompt has no effect on the tendency of explanations to mention aspects of the road layout or traffic laws, $p = .44$.

Figure 2 displays the zero-order correlation matrix among judgments made by evaluators. Because ratings of Satisfyingness, Completeness and Sufficient Detail were highly correlated with each other (all $r > .8$, all $p < .001$), we created a composite 'Quality' variable by averaging them. This variable will be the main target in our analyses.

4.1 Mechanistic and Teleological prompts lead to more satisfying explanations than Counterfactual

On average, explanations generated in response to a Mechanistic or Teleological prompt (in Survey 1) were perceived as better (by participants in Survey 2) than explanations generated for a Counterfactual prompt; see Figure 3. This effect was statistically significant, as assessed in a linear mixed model with random slopes at the scenario level, and random intercepts at the scenario, explanation and participant levels: relative to Counterfactual explanations, both Mechanistic explanations ($\beta = .10$, [95% CI: .05, .14]) and Teleological explanations ($\beta = .09$, [95% CI: .05, .13]) elicited higher Quality.

In contrast, there was only weak evidence for an effect of explanation type on Trustworthiness. Using a similar linear mixed modelling approach as above, we find that relative to Counterfactual explanations, Mechanistic ($\beta = .04$, [95% CI: .01, .08]) and Teleological explanations ($\beta = .04$, [95% CI: -.01, .09]) are perceived as only slightly more trustworthy, if at all.

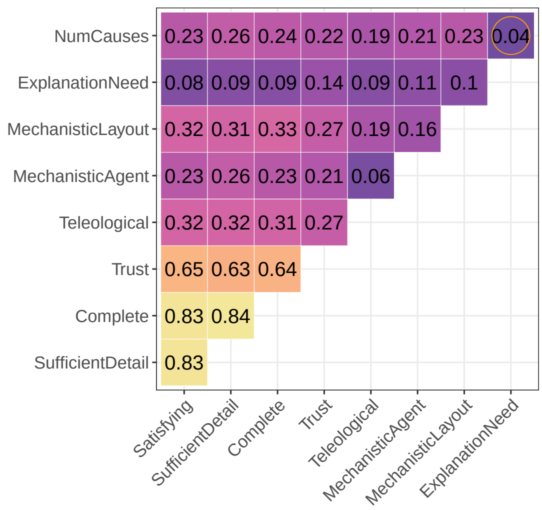

Figure 2. Zero-order correlation between ratings. Correlation coefficients circled in orange are non-significant after Bonferroni correction.

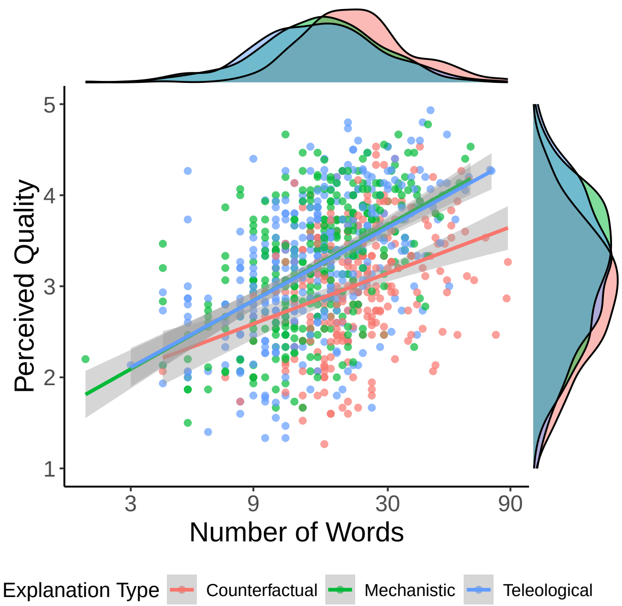

Figure 3. Perceived explanation quality as a function of the number of words in the explanation, and the explanation prompt. Each dot corresponds to the average quality rating of one explanation (computed by averaging the ratings of about 5 participants). Lines are linear fits with 95% CIs.

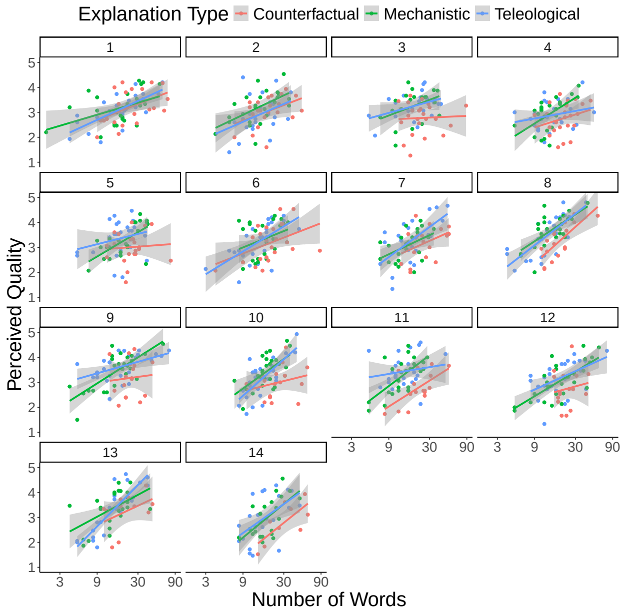

Figure 4. Perceived explanation quality as a function of number of words and explanation prompt, for each of the 14 scenarios. Each dot corresponds to the average quality rating of one explanation (computed by averaging the ratings of about 5 participants). Lines are linear fits with 95% CIs.

All our linguistic measures of explanation complexity had a positive effect on perceived Quality, all $ps < .001$. Figure 3 shows for example that longer explanations (as indexed by number of words) are rated as better. Interestingly, there was an interaction between number of words and explanation type: number of words had a larger effect on perceived quality for Mechanistic and Teleological explanations relative to Counterfactual explanations, as shown by a linear mixed model with random intercepts at the scenario, explanation and participant levels, interaction effect: $p < .001$. Intuitively, short explanations tend to be low-effort explanations that are perceived as bad regardless of the original explanation prompt, but participants who put some effort into their explanations generated better explanations in response to the Teleological and Mechanistic prompts.

Figure 4 shows that the patterns discussed above are relatively robust across scenarios. Removing the scenario-level random slopes from the linear mixed model we used to test the effect of explanation type did not decrease model fit (full model, AIC = 15070, without random slopes, AIC = 15060). On the other hand, the effect of linguistic complexity appears to vary slightly depending on the scenario: removing the scenario-level random slopes from a linear mixed model predicting perceived Quality from number of words results in a slightly lower model fit (full model, AIC = 14900, without random slopes, AIC = 14908, $p = .003$).

4.2 *Teleological features are the main predictor of perceived quality and trustworthiness*

Participants rated explanations along various features: for example, whether an explanation mentioned the agent's goals, how many causes it described, etc. We ran linear mixed models to assess how well these features predicted participants' judgments of the Quality and Trustworthiness of explanations. Figure 5 shows the standardized coefficients from two linear mixed models respectively predicting participants' judgments of Quality and Trustworthiness; with random

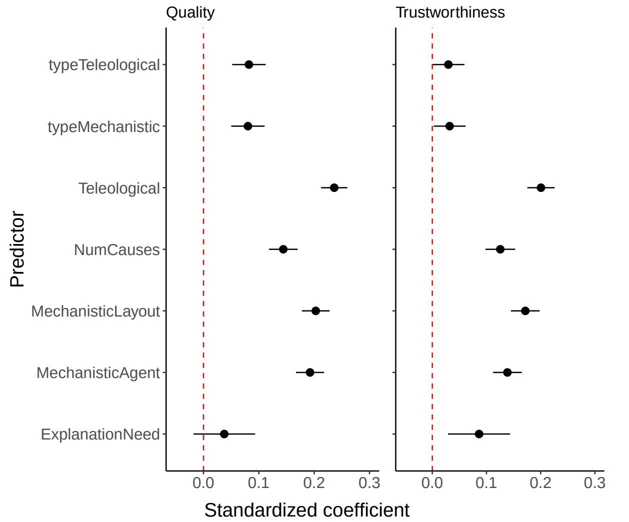

Figure 5. Standardized coefficients from linear mixed models predicting perceived Quality (left) and perceived Trustworthiness (right). The first two predictors (type) represent the experimental manipulation (the advantage of the Teleological and Mechanical prompts relative to the Counterfactual prompt, which is taken as a baseline), while the other predictors represent the effect of perceived features of explanations. Error bars represent 95% CIs.

intercepts at the scenario, explanation, and participant levels.

Overall, perceived Teleology was the best predictor of perceived Quality and Trustworthiness: i.e. the explanations that participants judged as mentioning the goals, desires or intentions of the agent were also perceived as better and more trustworthy.[4] The number of causes mentioned by an explanation, the extent to which the explanation mentioned the actions of other agents, and the extent to which it mentioned road layout or traffic laws, also reliably predicted both perceived Quality and Trustworthiness.

Importantly, perceived Teleology (how much an explanation mentioned the agent's desires, goals and intentions, as judged by participants in Survey 2) and the Teleology prompt (whether participants in Survey 1 were explicitly instructed to write Teleological explanations) had independent effects on participants' Quality judgments: Each variable has a significant effect when controlling for the other (see Figure 5). Even for explanations generated in response to a Counterfactual or Mechanistic prompt, those that mentioned more teleological features were judged as better and more trustworthy. We did not find a difference in the effect of perceived Teleology across explanation types: adding random slopes at the explanation type level did not improve the fit of linear mixed models predicting perceived Quality ($p = .62$) or perceived Trustworthiness ($p = .27$).

4.3 *Neither perceived teleology nor quality ratings are affected by autonomous vs. human driver status*

We manipulated across conditions whether participants were told the blue car was an autonomous vehicle or was driven by a human driver (independent variable 'AV' in Survey 1). We then ran linear

[4] To more formally establish that Teleology is the best predictor of perceived Quality, we computed the AICs of linear mixed models where we removed either Teleology, MechanisticLayout or MechanisticAgent as predictors. The model without Teleology had a substantially worse fit (AIC=14262) than the models without MechanisticLayout (AIC=14139) and without MechanisticAgent (AIC=14108). A similar approach yields the same result for perceived Trustworthiness (model without Teleology, AIC = 14712, without MechanisticLayout, AIC=14633, without MechanisticAgent, AIC=14577).

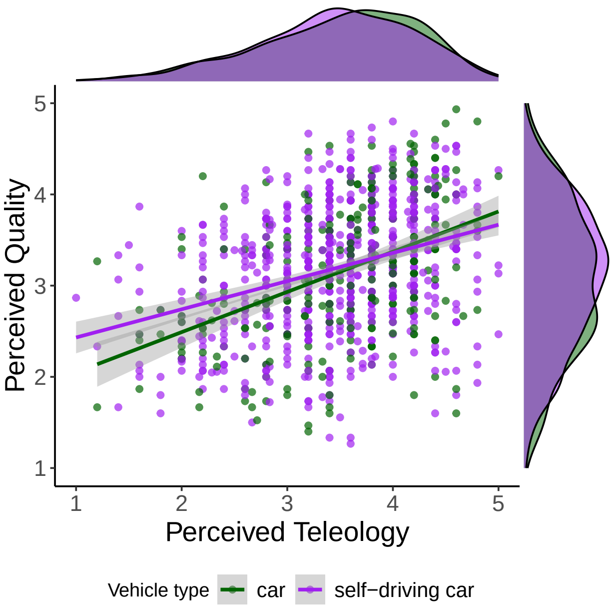

Figure 6. Perceived Quality as a function of Perceived Teleology, and whether the blue vehicle was identified as an autonomous vehicle. Each point represents one explanation. Lines are linear fits with 95% confidence intervals.

mixed models as above to assess whether the fact the person who generated the explanation in Survey 1 was explaining the actions of a human or an autonomous vehicle had any effect on either Quality ratings or perceived Teleology (how much an explanation mentioned the agent's desires, goals and intentions, as judged by participants in Survey 2). We found no significant difference in either ratings between the two conditions: including AV as a predictor variable contributed no significant improvement in the fit of the linear mixed models predicting perceived Quality ($p = .44$) or perceived degree of Teleology ($p = .4$). Furthermore, there was no improvement in model fit from including the interaction between AV and perceived Teleology ($p = .94$), indicating that the effect of perceived Teleology on perceived Quality is the same regardless of whether participants think the car is an autonomous vehicle; see Figure 6.

5 Discussion

In this paper, we introduced a rich dataset of human evaluations of explanations. These explanations were themselves human-generated, and targeted the behavior of autonomous vehicles in short video clips. We hope this dataset will be a valuable resource to help researchers better understand how laypeople generate and interpret explanations.

Our main result is that when people explain the behavior of self-driving vehicles, they often take an intentional stance, conceiving the vehicle as an agent with goals and beliefs. Specifically, we find that:

1. Explanations in response to a Teleological prompt are judged as more satisfying than explanations generated in response to a Counterfactual prompt, which discourages teleological content;
2. Explanations that are perceived as having teleological content are judged as more satisfying—in fact, perceived teleology is the most important predictor of explanation satisfaction;
3. Whether people are explaining the behavior of human drivers or autonomous vehicles has no effect on the perceived quality of the explanations or the perceived degree of teleology.

The latter result suggests that people have no qualms about referring to autonomous vehicles as having beliefs, desires and intentions. The intentional stance is not just levelled at people but can be a convenient abstraction to help us quickly conceptualise and refer to the outcome of any complex system. This is evident in the way we talk, as witnessed by utterances like 'the car doesn't want to start today', or 'my laptop won't talk to the printer'. Even if people may not actually attribute mental content to machines, they still find it convenient to reason as if machines had mental states [6, 41, 56].

Participants' preference for teleological explanations highlights the usefulness of concepts from cognitive science for XAI. Cognitive scientists emphasize the fact that most of our knowledge is organized around domain-specific intuitive theories [18, 12]. Explanations that do not conform to the intuitive theory within which people intuitively understand a system might not be the most effective ones. In the domain of autonomous vehicles, explanations that resonate with people's intuitive psychology (i.e. that take the intentional stance) are likely to be more effective. Notably, when people adopt the intentional stance, they might not favor simple counterfactual-based explanations, in contrast to previous findings in simpler contexts [5, 50]. This also provides support to the design of decision-making systems that explicitly utilise a goal-oriented model (e.g., [1, 16]), as the decisions of these systems would be more amenable to human understanding.

In addition, our study contributes to a nascent literature in psychology that investigates "naturally occurring" explanations [55, 48]. While the psychological literature on explanations traditionally uses well-controlled stimuli, asking participants to evaluate a handful of experimenter-generated explanations, recent studies have asked participants to evaluate explanations collected from online forums [55] or collected from a crowd-sourcing platform [48]. While these studies have focused on explanations that target general facts (e.g. 'Why does thunder make noise?'), we contribute to this literature by exploring how people explain specific events (e.g. 'Why did the car stop in this particular situation?'). More generally, our video stimuli depict scenes that are sufficiently rich to be interesting, but also simple enough that explainers can plausibly identify the reason for the agent's behavior. We also replicate some results from these previous studies, finding for example that more complex explanations (as indexed by the number of causes they mention) are more satisfying.

The present study has some limitations. Following previous studies, we collected *subjective* measures of explanation quality [55, 48], such as how satisfying or trustworthy the listener considers the explanation. However, subjective measures can diverge from more objective measures (such as how much the explanation improves the listener's ability to predict the system) in subtle ways [5]. It remains an open question to what extent teleological explanations help the user better predict the behavior of an autonomous vehicle, or better infer the details of what happened (cf. [24, 37, 38]). Future research should also investigate the extent to which explanation preference varies across different contexts, beyond the limited range of scenarios considered here. It seems plausible that explanatory preferences might vary in function of many features of a situation, for example, whether an agent's goal is easy to infer, or how much the agent can see. Finally, despite our dichotomous framing, good natural explanations may contain elements of all the modalities discussed.

References

[1] S. V. Albrecht, C. Brewitt, J. Wilhelm, B. Gyevnar, F. Eiras, M. Dobre, and S. Ramamoorthy. Interpretable Goal-based Prediction and Planning for Autonomous Driving. In *IEEE International Conference on Robotics and Automation (ICRA)*, Mar. 2021.

[2] Aristotle. *Metaphysics*, volume 1. Harvard University Press Cambridge, MA, 1933.
[3] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4):0064, 2017.
[4] S. Beckers. Causal Explanations and XAI. In *Proceedings of the First Conference on Causal Learning and Reasoning*, pages 90–109. PMLR, June 2022.
[5] L. Celar and R. M. Byrne. How people reason with counterfactual and causal explanations for artificial intelligence decisions in familiar and unfamiliar domains. *Memory & Cognition*, pages 1–16, 2023.
[6] H. H. Clark and K. Fischer. Social robots as depictions of social agents. *Behavioral and Brain Sciences*, 46:e21, 2023.
[7] D. C. Dennett. *The intentional stance*. MIT Press, 1987.
[8] S. Droop and N. R. Bramley. Extending counterfactual reasoning models to capture unconstrained social explanations. In *ICML Workshop on Counterfactuals in Minds and Machines*, number 1, 2023.
[9] U. Ehsan and M. O. Riedl. Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach. In C. Stephanidis, M. Kurosu, H. Degen, and L. Reinerman-Jones, editors, *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, Lecture Notes in Computer Science, pages 449–466, Cham, 2020. Springer International Publishing.
[10] M. Franklin, E. Awad, and D. Lagnado. Blaming automated vehicles in difficult situations. *Iscience*, 24(4), 2021.
[11] G. Gergely, Z. Nádasdy, G. Csibra, and S. Bíró. Taking the intentional stance at 12 months of age. *Cognition*, 56(2):165–193, 1995.
[12] T. Gerstenberg and J. B. Tenenbaum. Intuitive theories. 2017.
[13] B. Gyevnar, N. Ferguson, and B. Schafer. Bridging the transparency gap: What can explainable ai learn from the ai act? In *26th European Conference on Artificial Intelligence*, pages 964–971. IOS Press, 2023.
[14] B. Gyevnar, C. Wang, C. G. Lucas, S. B. Cohen, and S. V. Albrecht. Causal explanations for sequential decision-making in multi-agent systems. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, 2024.
[15] J. Y. Halpern. *Actual causality*. MIT Press, 2016.
[16] J. P. Hanna, A. Rahman, E. Fosong, F. Eiras, M. Dobre, J. Redford, S. Ramamoorthy, and S. V. Albrecht. Interpretable Goal Recognition in the Presence of Occluded Factors for Autonomous Vehicles. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7044–7051, Sept. 2021.
[17] D. J. Hilton and L. M. John. The course of events: counterfactuals, causal sequences, and explanation. In *The psychology of counterfactual thinking*, pages 56–72. Routledge, 2007.
[18] L. A. Hirschfeld and S. A. Gelman. *Mapping the mind: Domain specificity in cognition and culture*. Cambridge University Press, 1994.
[19] S. Joo, S. R. Yousif, and F. C. Keil. Understanding "why:" how implicit questions shape explanation preferences. *Cognitive Science*, 46(2): e13091, 2022.
[20] M. T. Keane, E. M. Kenny, E. Delaney, and B. Smyth. If Only We Had Better Counterfactual Explanations: Five Key Deficits to Rectify in the Evaluation of Counterfactual XAI Techniques. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 4466–4474, Montreal, Canada, Aug. 2021. International Joint Conferences on Artificial Intelligence Organization.
[21] D. Kelemen. Why are rocks pointy? children's preference for teleological explanations of th natural world. *Developmental psychology*, 35(6):1440, 1999.
[22] D. Kelemen and E. Rosset. The human function compunction: Teleological explanation in adults. *Cognition*, 111(1):138–143, 2009.
[23] D. Kelemen, J. Rottman, and R. Seston. Professional physical scientists display tenacious teleological tendencies: purpose-based reasoning as a cognitive default. *Journal of experimental psychology: General*, 142(4): 1074, 2013.
[24] L. Kirfel, T. Icard, and T. Gerstenberg. Inference from explanation. *Journal of Experimental Psychology: General*, 151(7):1481, 2022.
[25] A. Kuznietsov, B. Gyevnar, C. Wang, S. Peters, and S. V. Albrecht. Explainable AI for Safe and Trustworthy Autonomous Driving: A Systematic Review, Feb. 2024.
[26] D. A. Lagnado, T. Gerstenberg, and R. Zultan. Causal responsibility and counterfactuals. *Cognitive science*, 37(6):1036–1073, 2013.
[27] D. Lewis. Causation as influence. *The Journal of Philosophy*, 97(4): 182–197, 2004.
[28] T. Lombrozo. The structure and function of explanations. *Trends in cognitive sciences*, 10(10):464–470, 2006.
[29] T. Lombrozo. Explanatory preferences shape learning and inference. *Trends in cognitive sciences*, 20(10):748–759, 2016.
[30] T. Lombrozo and S. Carey. Functional explanation and the function of explanation. *Cognition*, 99(2):167–204, 2006.
[31] T. Lombrozo and D. Wilkenfeld. Mechanistic versus functional understanding. *Varieties of understanding: New perspectives from philosophy, psychology, and theology*, 209, 2019.
[32] C. G. Lucas and C. Kemp. An improved probabilistic account of counterfactual reasoning. *Psychological review*, 122(4):700, 2015.
[33] C. G. Lucas, T. L. Griffiths, F. Xu, C. Fawcett, A. Gopnik, T. Kushnir, L. Markson, and J. Hu. The child as econometrician: A rational model of preference understanding in children. *PloS one*, 9(3):e92160, 2014.
[34] D. R. Mandel and D. R. Lehman. Counterfactual thinking and ascriptions of cause and preventability. *Journal of personality and social psychology*, 71(3):450, 1996.
[35] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.
[36] T. Miller. Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven Decision Support using Evaluative AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, pages 333–342, New York, NY, USA, June 2023. Association for Computing Machinery.
[37] A. Nam, C. Hughes, T. Icard, and T. Gerstenberg. Show and tell: Learning causal structures from observations and explanations. In *Proceedings of the cognitive science society*. 2023.
[38] N. Navarre, C. Konuk, N. Bramley, and S. Mascarenhas. Functional rule inference from causal selection explanations. In *Proceedings of the cognitive science society*. 2024.
[39] A. Páez. The Pragmatic Turn in Explainable Artificial Intelligence (XAI). *Minds and Machines*, 29(3):441–459, Sept. 2019.
[40] J. Pearl. *Causality*. Cambridge university press, 2000.
[41] J. Perez-Osorio and A. Wykowska. Adopting the intentional stance toward natural and artificial agents. *Philosophical Psychology*, 33(3): 369–395, 2020.
[42] T. Quillien and M. Barlev. Causal judgment in the wild: evidence from the 2020 us presidential election. *Cognitive Science*, 2022.
[43] T. Quillien and T. C. German. A simple definition of 'intentionally'. *Cognition*, 214:104806, 2021.
[44] T. Quillien and C. G. Lucas. Counterfactuals and the logic of causal selection. *Psychological Review*, 2023.
[45] D. Slack, S. Krishna, H. Lakkaraju, and S. Singh. Explaining machine learning models with interactive natural language conversations using TalkToModel. *Nature Machine Intelligence*, 5(8):873–883, Aug. 2023.
[46] T. Speith. A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, pages 2239–2250, New York, NY, USA, June 2022. Association for Computing Machinery.
[47] I. Stepin, J. M. Alonso, A. Catala, and M. Pereira-Fariña. A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence. *IEEE Access*, 9:11974–12001, 2021.
[48] J. Sulik, J. van Paridon, and G. Lupyan. Explanations in the wild. *Cognition*, 237:105464, 2023.
[49] J. E. T. Taylor and G. W. Taylor. Artificial cognition: How experimental psychology can help generate explainable artificial intelligence. *Psychonomic Bulletin & Review*, 28(2):454–475, Apr. 2021.
[50] G. Warren, R. M. Byrne, and M. T. Keane. Categorical and continuous features in counterfactual explanations of ai systems. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 171–187, 2023.
[51] J. Woodward. *Making things happen: A theory of causal explanation*. Oxford university press, 2003.
[52] S. C.-H. Yang, N. E. T. Folke, and P. Shafto. A Psychological Theory of Explainability. In *Proceedings of the 39th International Conference on Machine Learning*, pages 25007–25021. PMLR, June 2022.
[53] M. Yin, J. Wortman Vaughan, and H. Wallach. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, New York, NY, USA, May 2019. Association for Computing Machinery.
[54] C. Zednik. Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy & Technology*, 34(2): 265–288, June 2021.
[55] J. C. Zemla, S. Sloman, C. Bechlivanidis, and D. A. Lagnado. Evaluating everyday explanations. *Psychonomic bulletin & review*, 24:1488–1500, 2017.
[56] J. Zerilli. Explaining machine learning decisions. *Philosophy of Science*, 89(1):1–19, 2022.
[57] Y. Zhang, Q. V. Liao, and R. K. E. Bellamy. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, pages 295–305, New York, NY,

USA, Jan. 2020. Association for Computing Machinery.

A Scenarios

Table 2 shows a summary table of all 14 scenarios with descriptions. The following figures (Figures 7 to 20) contain example images of each scenario, giving both a schematic overview of the scenario and a snapshot from the video that was used in the actual surveys.

Table 2. Summary table of all scenarios used for our surveys. The columns (**E**)efficiency, (**C**)omfort, and (**O**)cclusion specify whether the scenario was created with the intent to prompt an explanation related to the efficiency or comfort of the driving action, or occluded elements in the environment, respectively.

ID	Name	E	C	O	Description
1	Cutting between cars	Y	Y	N	The ego vehicle needs to exit the highway through the off-ramp, but it is blocked by two other vehicles in the outer lane. One of the other vehicles notices the ego's intention to exit and slows down to allow the ego to pass between the two other cars. The ego then exits the highway.
2	Early right turn	Y	N	N	The ego approaches a four-way crossroads from the south. At the same time a vehicle from the east approaches maintaining a high speed. In addition, from the west another vehicle approaches which comes to a rolling stop at the crossroads. This indicates to the ego that the vehicle from the west intends to turn left at the junction, and it is stopped to give way to the vehicle from the east. As the goal of the ego vehicle is to turn right, it can use this realisation to reach its destination sooner.
3	Merge into waiting line	Y	N	N	The ego approaches a T-junction. The vehicle on its right is coming to a rolling stop. Initially, this may be because the other vehicle could be turning left and is stopping to give way to the vehicle coming in the opposite direction. However, as the vehicle from the opposite direction passes, the ego predicts that the vehicle on its right is stopping to allow the ego to merge behind a waiting line of cars on the main road.
4	Enter roundabout early	Y	N	N	As the ego approaches a roundabout, it recognises the intention of another vehicle already in the roundabout to exit at the next roundabout exit, therefore, the ego does not need to stop completely at the roundabout entry and can enter the roundabout faster. The ego recognises the intentions of the other vehicle because it changes lanes as the ego arrives at the roundabout and it also begins to slow down gradually for the turn.
5	Take over parking	Y	N	N	The ego sees another vehicle slow down near a line of parked cars. It correctly predicts that the other vehicle intends to park. Since there is no one else on the oncoming lane, the ego can take over the parking vehicle to reach its goal faster which is to reach the end of the road.
6	Vehicle cut in front	Y	N	N	The vehicle in front of the ego cuts in front of the ego vehicle and begins to slow down. This is indicative of its intention to turn right at the T-junction. As the goal of the ego vehicle is to head straight it decides to change lanes to the left so that it may continue on towards its goal unimpeded.
7	Cautious on-ramp	N	N	Y	The ego vehicle approaches a highway through an onramp where the view of the highway is obstructed by bushes. To avoid a possible collision, it slows down before merging onto the highway. The ego vehicle then observes another vehicle on the highway, it gives way, and then merges onto the highway to continue ahead.
8	Occluded crossing	N	N	Y	The ego vehicle enters an urban area with a junction and pedestrian crossings. Even though it cannot see the continuation of the main road curving to the right as the view is blocked by the buildings, the ego still slows down to make sure that it can safely stop at the pedestrian crossing. Once the building is out of sight, the ego vehicle notices pedestrians at the crossing and stops according to highway rules. It then continues onwards towards its goal.
9	Occluded roundabout	N	N	Y	The ego vehicle's goal is to enter the roundabout. Even though it cannot see another vehicle in the roundabout due to the bushes in the middle, it stops at the entry because it sees another vehicle on its left being already stopped. This is indicative that the other vehicle is giving way to an unseen third vehicle in the roundabout. To execute its maneuvers safely, the ego therefore decides to give way as well and wait until both of the other vehicles have passed.
10	Parked car pulling out	N	N	Y	The ego arrives at a line of parked cars and decreases its speed slightly. It does this in order to pass the line of parked cars more safely in case an unattentive driver exits their parking spot while the ego is passing by. When this happens, the ego can break quickly thus avoiding an accident.
11	Ball rolling on road	N	N	Y	The ego notices a ball drift out from behind a truck in front of it. Thus, it decides to break quickly as it predicts the ball to be followed by a person and this action avoids a potentially fatal crash.
12	Truck narrow passage	N	N	Y	On a narrow road with cars parked on each side, the ego is following a truck ahead of itself. When the truck slows down to a stop, the ego vehicle predicts that there is an occluded third vehicle oncoming on the road, therefore it waits behind the truck. While the truck passes the parked vehicles, the ego still cannot see whether there are any other oncoming vehicles on the road. So it waits until the truck is completely clear of sight and confirms that passage is safe before continuing towards its goal.
13	Very curved road	N	Y	N	The ego approaches a road section with very high curvature and rapidly decelerates to avoid swerving out of the road despite putting a large amount of jerk on the passenger which may be uncomfortable.
14	Occluded off-ramp	N	Y	Y	The ego approaches an offramp it needs to exit through to reach its goal, however, a truck is blocking its way for changing lanes. It therefore decelerates rapidly to allow the truck to pass and to be able to confirm that there are no occluded vehicles on the outer lane, after which it can safely change lanes and off-ramp.

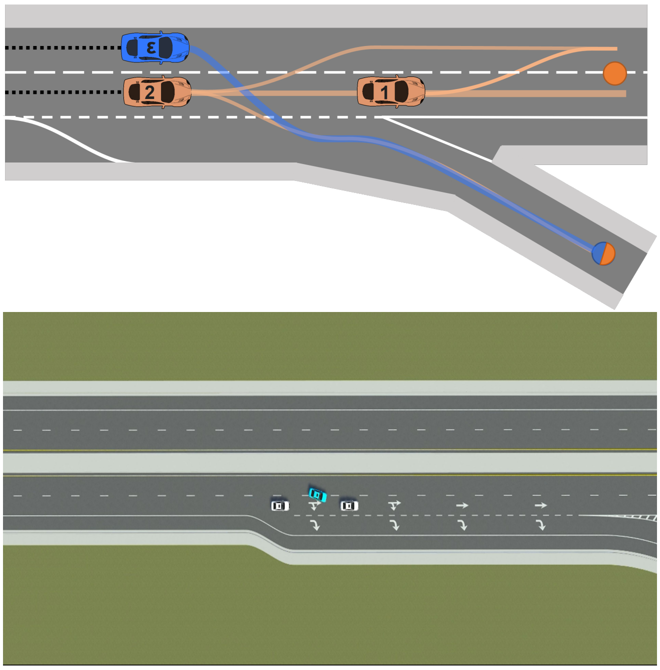

Figure 7. Scenario 1 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

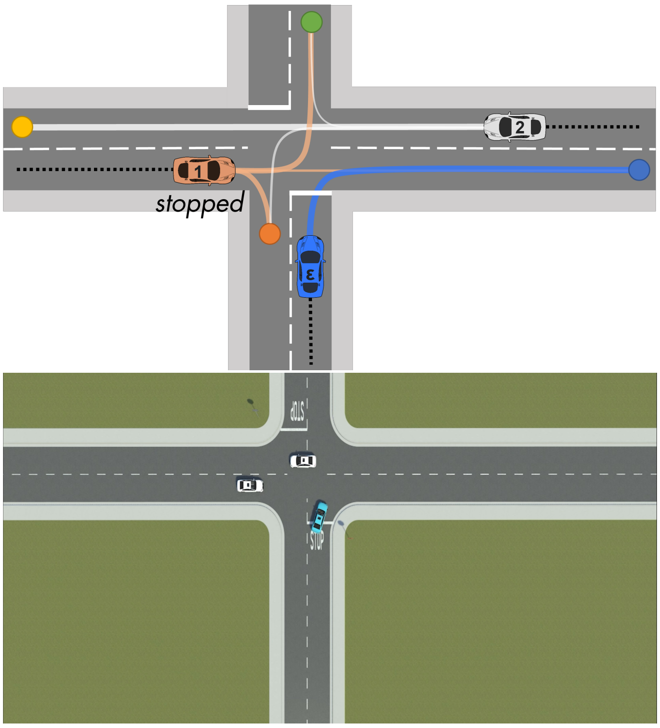

Figure 8. Scenario 2 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

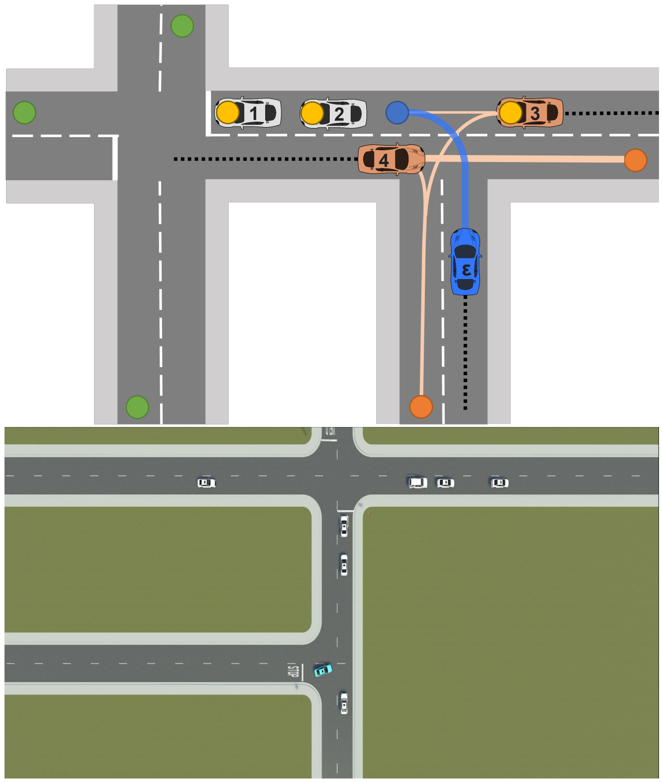

Figure 9. Scenario 3 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

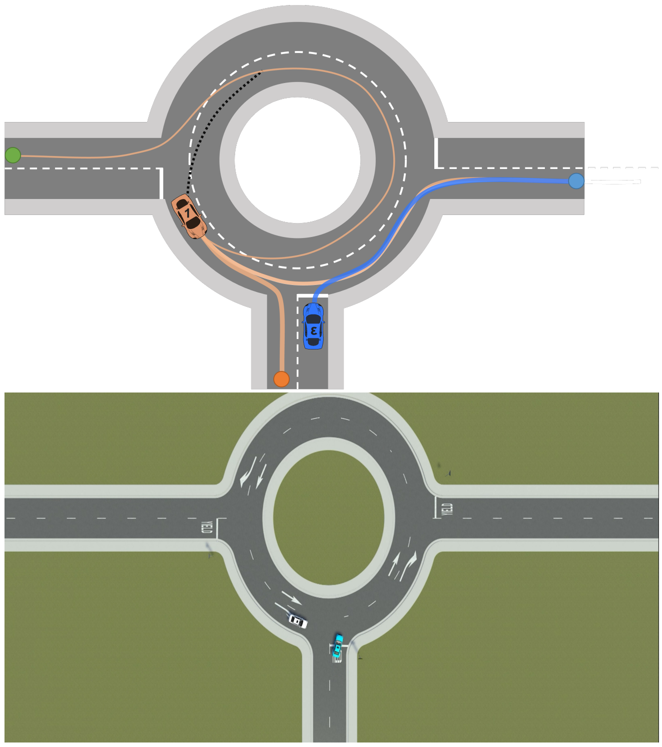

Figure 10. Scenario 4 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

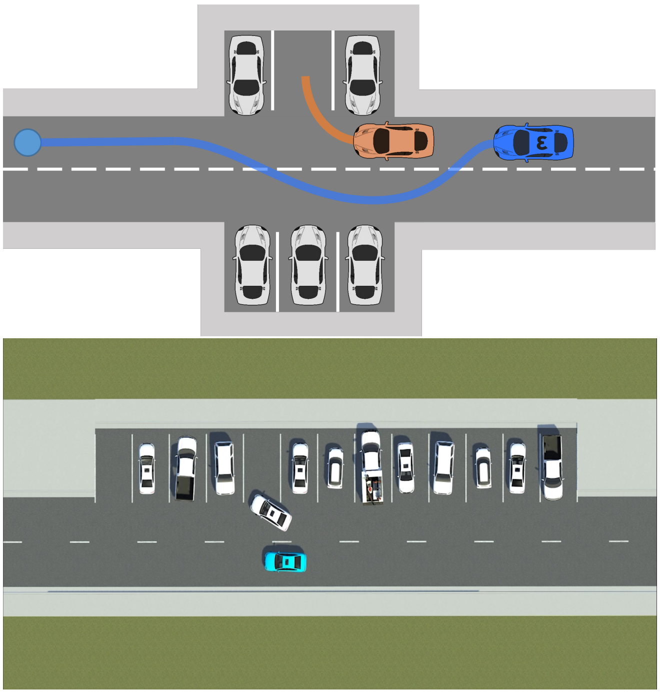

Figure 11. Scenario 5 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

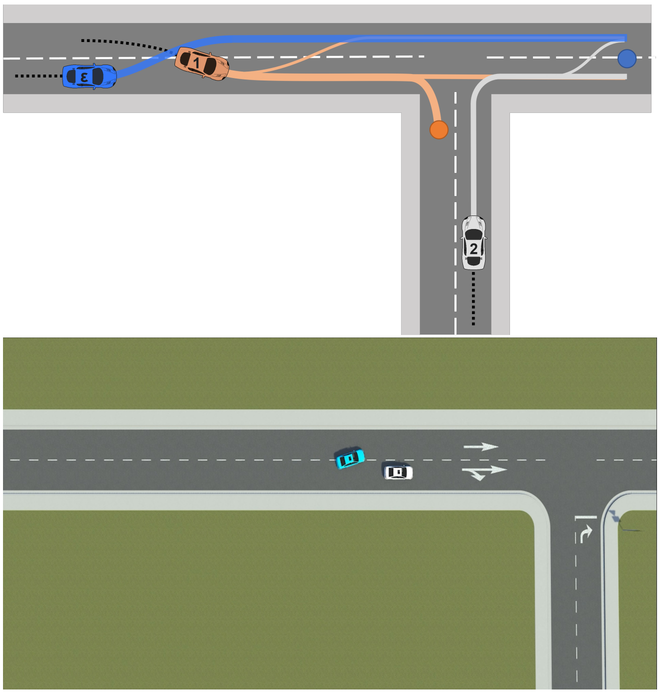

Figure 12. Scenario 6 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

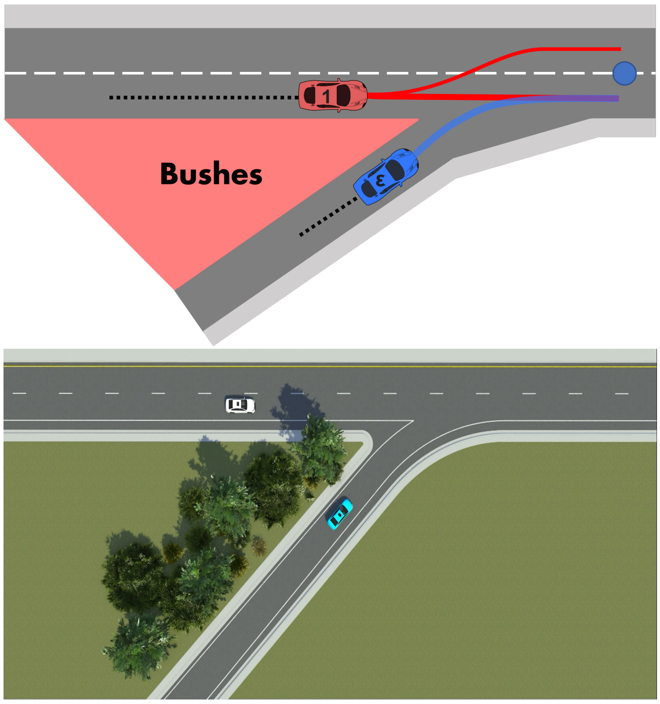

Figure 13. Scenario 7 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

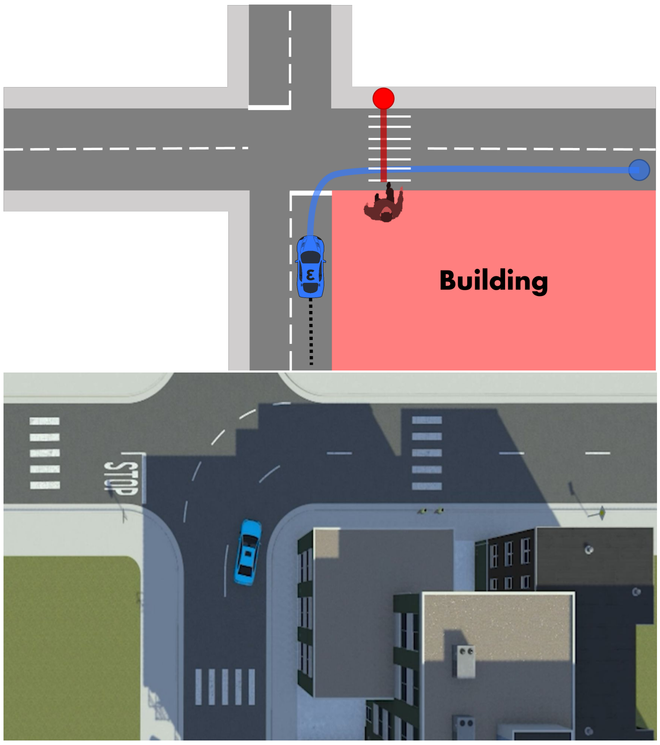

Figure 14. Scenario 8 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

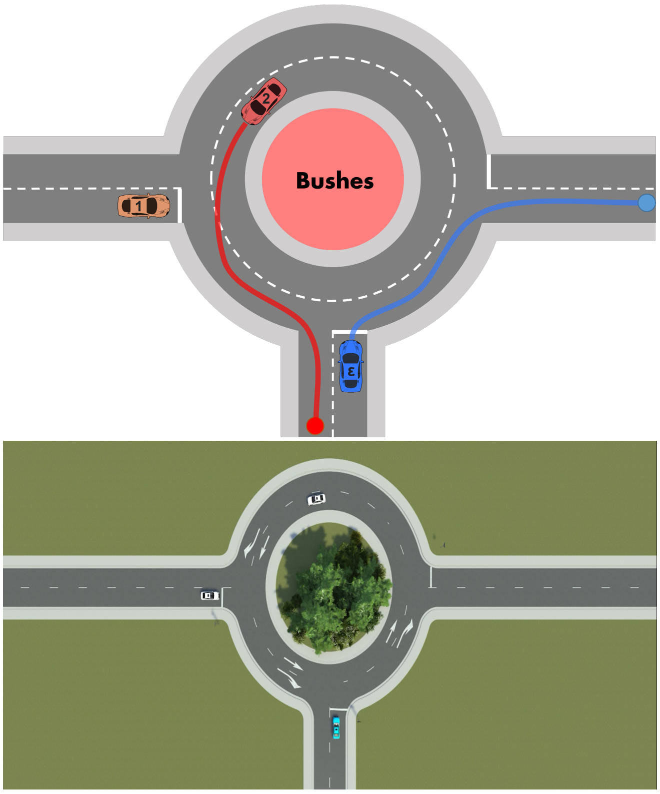

Figure 15. Scenario 9 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

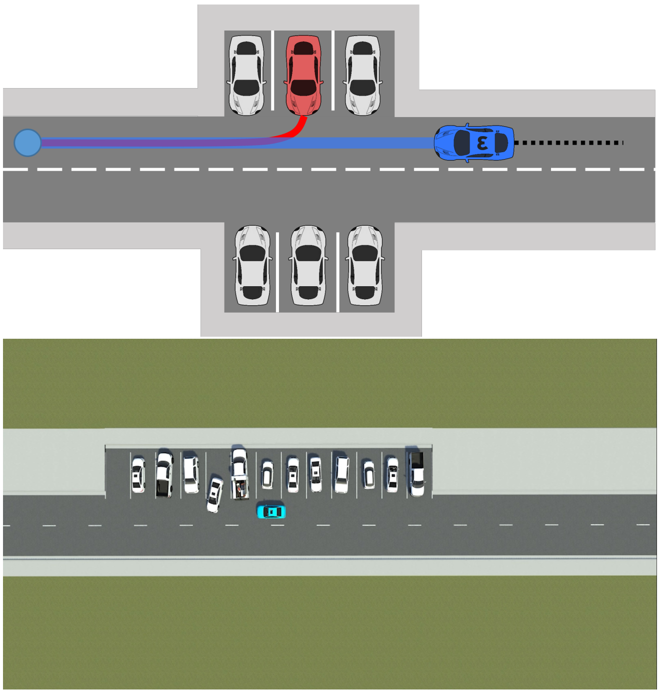

Figure 16. Scenario 10 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

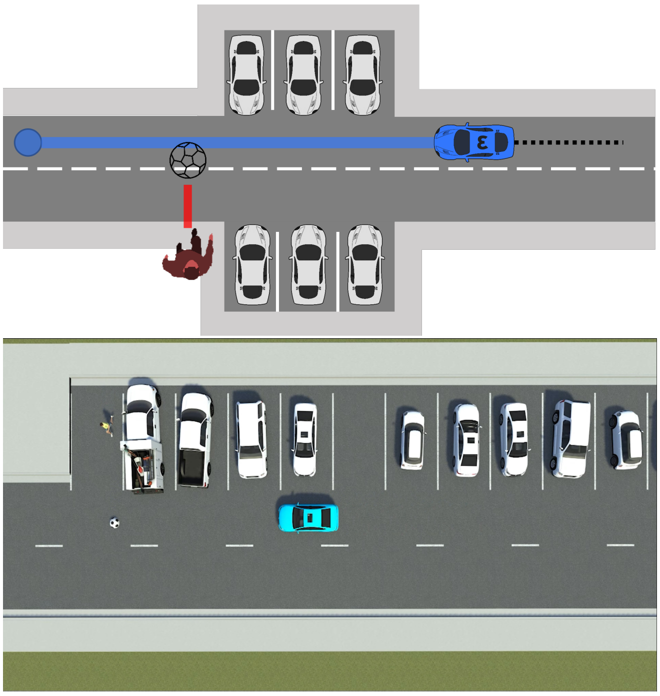

Figure 17. Scenario 11 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

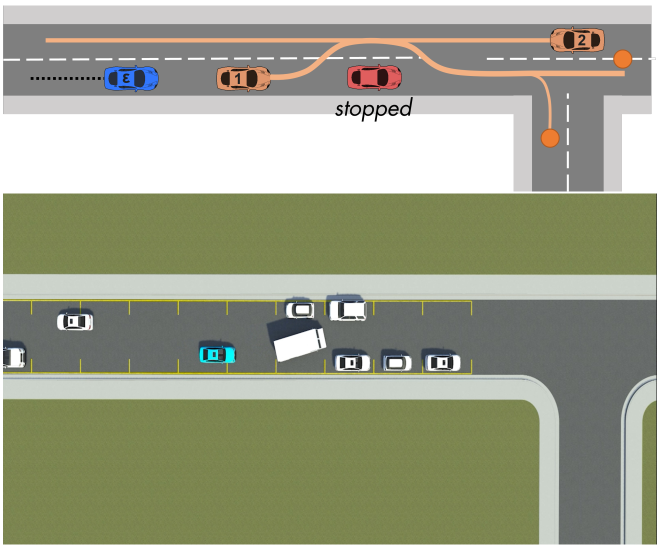

Figure 18. Scenario 12 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

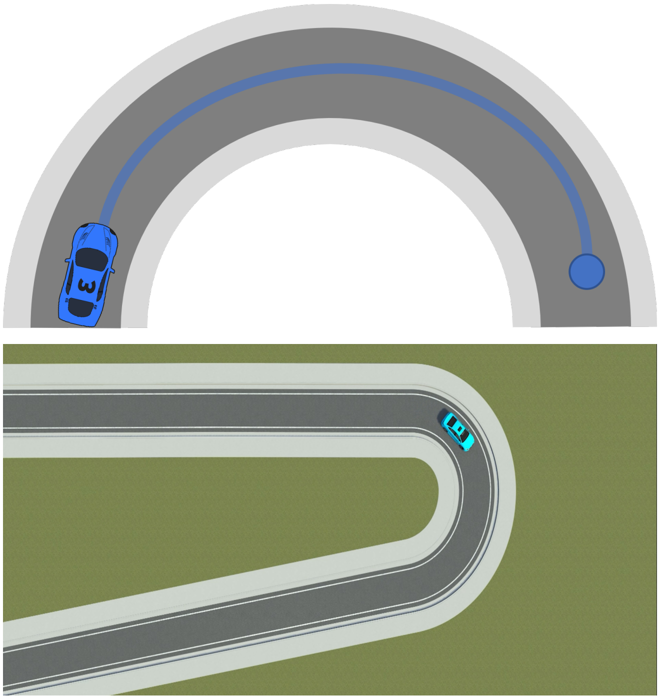

Figure 19. Scenario 13 – (**Top**) Schematic overview; (**Bot**) Video snapshot.

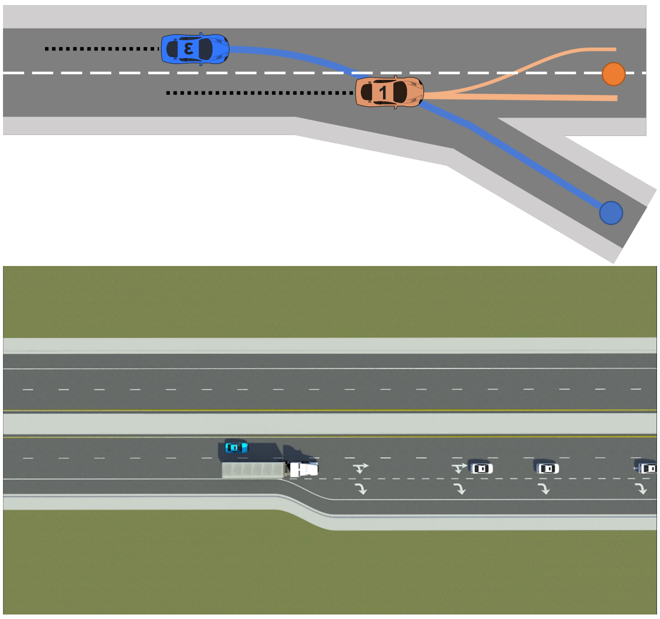

Figure 20. Scenario 14 – (**Top**) Schematic overview; (**Bot**) Video snapshot.