CHALLENGES FOR THE EVALUATION OF AUTOMATED DRIVING SYSTEMS USING CURRENT ADAS AND ACTIVE SAFETY TEST TRACK PROTOCOLS

Scott Schnelle
Kristofer D. Kusano
Francesca Favaro
Guy Sier
Trent Victor
Waymo, LLC
United States

Paper Number 23-0329

ABSTRACT

A number of public safety stakeholders have advocated for the application of traditional consumer-focused testing protocols (e.g., NCAP programs) for the evaluation of safety for Automated Driving Systems (ADSs). Even though test protocols only exist for some ADAS and active safety technologies (i.e. SAE Level 0, Level 1, Level 2), proposals for expansion and adaptation of these tests for ADS have been brought forth within the industry community. To gain practical insight into the types of challenges and limitations arising from the application of these existing test protocols to ADSs, the Waymo Driver^{TM}, a SAE Level 4 ADS, was the subject of a testing campaign that leveraged several of the most difficult currently available ADAS and active safety test procedures. The main challenge discovered was that while these ADAS and active safety tests are aimed at evaluating the systems' collision avoidance behavior, most of these tests were unable to be evaluated as designed due to the increased capabilities of the Waymo Driver that prevented the vehicle from even entering into a conflict to begin with. Difficulties encountered included creating the type of occlusions envisioned in some test protocols due to the location and performance of the Waymo Driver's sensor suite and insufficient information in the test procedure regarding the roadway and map information. For example, in the occluded vulnerable road user (VRU) scenarios, the Waymo Driver could sense the test target prior to it starting to move and could proactively slow down, resulting in the desired collision avoidance interaction in the scenario not being tested. To make the test conditions representative of the intended collision avoidance interactions in the test procedure, either extra vehicles and/or different vehicle types were used as the occluding vehicles (e.g., large trucks). Similarly for the car-to-car test, a larger obstructing lead vehicle was used for the cut-out test so the Waymo Driver could not see over the lead vehicle. Also, without specifying additional details for the roadway that were not in the original test procedure, the Waymo Driver would proactively slow down due to the presence of parked cars or other roadway features on the test track, such as intersections. Beyond these required modifications to enable the interactions described in the test procedure, additional optional modifications were made to the test to increase the difficulty of the test. For example, in the NCAP cut-in test, the distance at which the vehicle was cutting in was reduced from 7.5m to 3m to try to elicit collision avoidance behavior. For all the test runs, including those run to specification and those with modifications, the Waymo Driver was able to avoid collisions which would have resulted in the highest rating for this evaluation.

Our conclusion is that existing ADAS and active safety test protocols cannot be applied as-is for an ADS such as the Waymo Driver. The highlighted challenges, ranging from the need to heighten the difficulty of the proposed scenarios to the under-specification of certain aspects of the test protocol, result in ambiguous requirements for both the test developers, the test facilities, and the test site administrators. This further indicates that Level 0-2 systems need to be separately considered from Level 4 ADS, such as the Waymo Driver. Furthermore, the results of this testing calls into question the feasibility and utility of adapting ADAS and active safety test for ADSs.

Keywords: Automated Driving System, Test Track Protocol, Physical Testing

INTRODUCTION

The safety evaluation for the development and deployment of Automated Driving Systems (ADSs) is a challenging endeavor. The operational design domain (ODD) of an ADS can include complex environments with numerous potentially hazardous situations. Additionally, unlike advanced driver assistance systems (ADAS) and active safety systems such as automatic emergency braking (AEB), an ADS like the Waymo Driver™ is responsible for the entire dynamic driving task (DDT) which makes verifying and validating the safety of an ADS a considerable undertaking. One methodology proposed in industry, standards and literature [1 Wimmer - 4 ISO] and used as a one[1] of the Waymo safety methodologies [5 Webb - 6 Kusano] for safety evaluation is scenario-based testing. Scenario-based testing can leverage a combination of virtual, closed-course, and real-world driving to enable a complimentary and comprehensive assessment of an ADSs safety performance within scenarios that it may encounter in its ODD. There is significant research ongoing around identifying what types and how many scenarios to test for ADS safety [7 Ding - 9 Riedmaier] and how representative these scenarios and their distributions need to be to that encountered in the real world [6 Kusano]. Regulatory and ratings agencies are also developing scenario-based closed-course and simulation tests for ADAS and ADS evaluations [10 Euro NCAP - 13 UNECE]. A specific subset of these scenario-based evaluations is a closed-course consumer ratings test for ADAS and active safety systems, like those in [14 Euro NCAP - 17 NHTSA]. These scenario-based tests target frequent and/or severe crashes that current ADAS and active safety technologies have the potential to mitigate or avoid that can be tested with current test track and test tool capabilities. Euro NCAP echoes these goals and objectives in the introduction to their testing procedures: for example, in the introduction of the Car-to-Car test procedure [14 Euro NCAP]: *"Car-to-Car rear impacts are one of the most frequent accidents happening on the roads [...] While injury severities are usually low, these accidents are very frequent and represent over a quarter of all crashes."*, and in their Car-to-Vulnerable Road User (VRU) test procedure [16 Euro NCAP]: *"car-to-VRU impacts are one of the most frequent accidents [...] These types of accidents with vulnerable road users usually coincide with severe injuries and leave the driver with very little reaction time to apply the brakes."* Furthermore, based on the existing precedence and availability of ADAS and active safety evaluation, one possibility is to investigate the adoption and/or adaptability of existing procedures as a starting point to generate scenarios for ADS (i.e., L3-L4-L5) performance evaluation for consumer information. Therefore, the focus of this paper will be on these closed-course consumer ratings tests for ADAS and active safety systems and challenges that arise from applying them to an ADS, along with whether this approach is feasible or even useful to achieve the goal of safety evaluation and consumer ratings to garner public trust. To enable this evaluation, an ADS, the Waymo Driver, was the subject of a testing campaign leveraging several existing ADAS and active safety test procedures.

The rest of the paper is organized as follows: the Methodology section will give an overview of each type of test, the Tests' Execution and Results section reviews test results and required modifications for the specific tests to enable assessment of our ADS, the discussion section presents overarching challenges and limitations regarding the feasibility and utility of closed-course testing for ADS safety assessment and consumer ratings. Finally, conclusions about the role closed-course testing plays in ADS evaluation for consumer information and potential alternatives are presented.

METHODOLOGY

This section gives a brief overview of the type of ADAS and active safety consumer rating tests that were selected for the testing campaign. The selection process started with a review of existing test procedures of which specific tests that leveraged scenarios that are part of the known unsafe/hazardous situations that ADSs encounter frequently were prioritized. From here, a final selection was made based on the anticipated difficulty of the test, along with potential difficulties of adapting the test to an ADSs based on the Waymo Driver's ODD and design combined with the selected test track's capabilities and tools, with more information on the selection process provided below.

---

[1] Scenario-based testing is only one methodology in various methods proposed in [6 Webb] for the holistic safety assurance of the Waymo Driver.

The selected tests come from the Euro NCAP AEB Car-to-Car [14 Euro NCAP], Euro NCAP AEB Car-to-VRU [16 Euro NCAP], Euro NCAP Highway Assist Systems [15 Euro NCAP] test protocols and NHTSA's draft Traffic Jam Assist System Confirmation Test [18 NHTSA]. The goal of these protocols is to test a specific ADAS or active safety systems function in collision avoidance scenarios. These test procedures start with a test overview which discusses specifics for testing ADAS and active safety systems, such as specifying behavior of the vehicle under test (VUT), driver behavior, and pre-test behavior. For example, from the Euro NCAP Front Turn Across Path test procedure [14 Euro NCAP], a specific path for the VUT is specified as shown in Figure 5. This already raises concerns since these requirements do not apply to the Waymo Driver since the ADS is responsible for the entire DDT, including trajectory planning, lateral and longitudinal control, and Object and Event Detection and Response (OEDR). Therefore, specifying a path for an ADS not only isn't possible but also would test the ADSs ability to follow directions instead of evaluating its capabilities as they were designed for the given ODD and scenario. Instead, the test procedures were adapted so only the inputs to the scenario within the ADSs ODD can be adjusted to try to elicit the desired interaction between the ADSs and the other safety relevant entities in the scene similar to the method described in [19 NHTSA].

After an overview of the test in the test procedure, specifications for the roadway that the test will take place on is provided. This specification consists of requirements on road surface (e.g., smooth, no holes, solid paved surface, flat, <1% slope, mu >0.9) roadway markings (e.g., lane line color, style, reflectivity, width, lane width) along with what can and can't be in the vicinity of the vehicle as it is tested (e.g., only specific abnormalities within 3m to either side of test path and 30 m ahead of VUT.) These requirements are specific to the ADAS or active safety system that is being tested and focus on the inputs required to test that system in the target ODD, namely what is required to activate and maintain functionality of the system under test (SUT). They do not provide additional information relevant to the HD map that the Waymo Driver leverages, which will be discussed in more detail later.

In addition to certain roadway features pertinent to the system under test and specific test, requirements for the types of targets that are used during testing are provided to ensure the tests are repeatable and reproducible. Since these ADAS and active safety tests are addressing safety critical scenarios that can have near-miss and collision interactions, some tests require the use of surrogate targets. These targets are designed to look realistic to various sensors (radar, lidar, camera) and be strikable without damaging the VUT or the target. The surrogate test targets used in this testing campaign included a child mannequin, an adult mannequin, an adult mannequin on a bicycle, and a surrogate vehicle referred to as the Global Vehicle Target (GVT).

After the specifications for the test and requirements for the roadway and targets are provided, details regarding the specific scenario that is being tested are provided. A summary of these scenarios for each test used in this report are provided below for both the Car-to-VRU tests and the Car-to-Car tests.

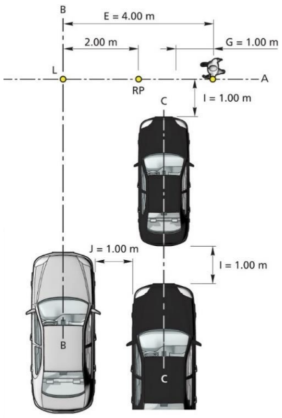
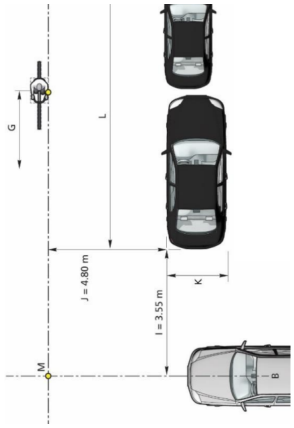
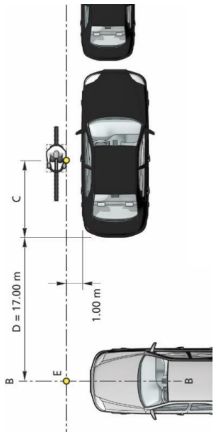
Car-to-VRU Tests

The scenarios in Table 1 contain an interaction between the VUT and a VRU from the Euro NCAP test protocol for AEB VRU systems[2] [16 Euro NCAP]. These scenarios are frequently encountered in the Waymo Driver's ODD and have the potential for severe injuries. These specific test protocols were selected since they were similar to NHTSA's Pedestrian Automatic Emergency Brake System Confirmation Tests [20 NHTSA], which were still in draft at the time of testing and did not contain bicyclists interactions. The selected scenarios are: Car-to-Pedestrian Nearside Child (CPNC), Car-to-Bicyclist Nearside Adult Occluded (CBNAO), Car-to-Bicyclist Nearside Adult (CBNA) and Car-to-Bicyclist Farside Adult (CBFA). A description of each scenario along with a birds-eye-view of the scenario can be found in Table 1.

---

[2] April 2021, Version 3.0.4

Table 1: Overview of selected Euro NCAP AEB VRU Tests [16 Euro NCAP]

<table>
<tr><th>CPNC</th><th>CBNAO</th><th>CBNA</th><th>CBFA</th></tr>
<tr><td>“a collision in which a vehicle travels forwards towards a child pedestrian crossing its path running from behind and obstruction from the nearside and the frontal structure of the vehicle strikes the pedestrian at 50% of the vehicle's width when no braking action is applied.”</td><td>“a collision in which a vehicle travels forwards towards a bicyclist crossing its path cycling from the nearside from behind an obstruction and the frontal structure of the vehicle strikes the bicyclist at 50% of the vehicle's width when no braking action is applied.”</td><td>“ a collision in which a vehicle travels forwards towards a bicyclist crossing its path cycling from the nearside and the frontal structure of the vehicle strikes the bicyclist when no braking action is applied.”</td><td>“a collision in which a vehicle travels forwards towards a bicyclist crossing its path cycling from the farside and the frontal structure of the vehicle strikes the bicyclist at 50% of the vehicle's width when no braking action is applied.”</td></tr>
<tr><td>
</td><td>
</td><td>
</td><td>
</td></tr>
</table>

Car-to-Car Tests

The following scenarios contain an interaction between the VUT and another vehicle. These scenarios are frequently encountered in the Waymo Driver’s ODD. These specific test protocols come from both Euro NCAP (Test Protocol for AEB Car-to-Car systems[3] [14 Euro NCAP] and Highway Assist Systems Test & Assessment Protocol[4] [15 Euro NCAP]) and NHTSA (draft Traffic Jam Assist test procedure[5] [18 NHTSA]) test programs. Test procedures from both of these testing programs were selected due to unique scenarios between test programs and variations in similar scenarios that may affect scenario difficulty.

NHTSA Lead Vehicle Lane Change with Braking (LVLCB)

The lead vehicle lane change with braking tests (LVLCB), commonly referred to as a cut-in scenario, comes from the NHTSA draft Traffic Jam Assist System Confirmation Test [18 NHTSA]. As stated in the draft test procedure, the object of the test is to “evaluate the TJA system’s ability to detect and respond to a moving POV that brakes during or after performing a lane change into a space between the SV and SOV.”

---

[3] April 2021, Version 3.0.3

[4] September 2020, Version 1.0

[5] July 2019, Working Draft

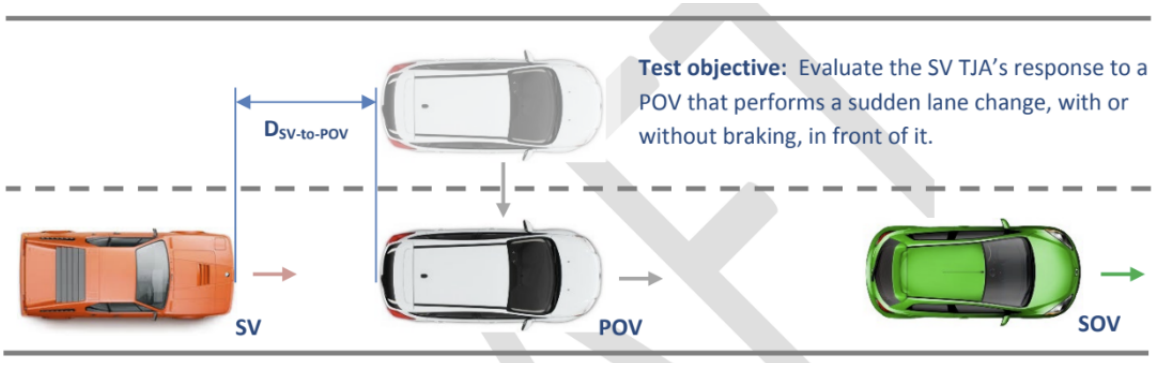


Figure 1: NHTSA TJA LVLCB scenario from [18 NHTSA]

<u>Euro NCAP Cut-in</u>
Similar to the NHTSA LVLCB test, the Euro NCAP cut-in test [15 Euro NCAP] scenario consists of “The GVT in the adjacent lane will perform a full lane change (3.5m lateral displacement) into the lane of the VUT. The indicated TTC is defined as the TTC at the point in time that the GVT has finished the lane change manoeuvre, where the rear centre of the GVT is in the middle of the VUT driving lane. “

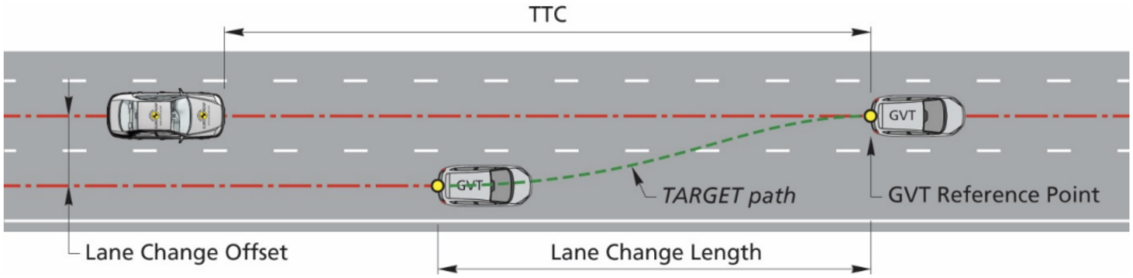


Figure 2: Euro NCAP Cut-in scenario from [15 Euro NCAP]

<u>Euro NCAP Cut-out</u>
The Euro NCAP cut-out test [15 Euro NCAP] scenario consists of “The vehicle cutting out will perform a full lane change (3.5m lateral displacement) into the adjacent lane to avoid the stationary GVT. The indicated TTC is defined as the TTC of the lead vehicle to the GVT when the lead vehicle will start the lane change.”

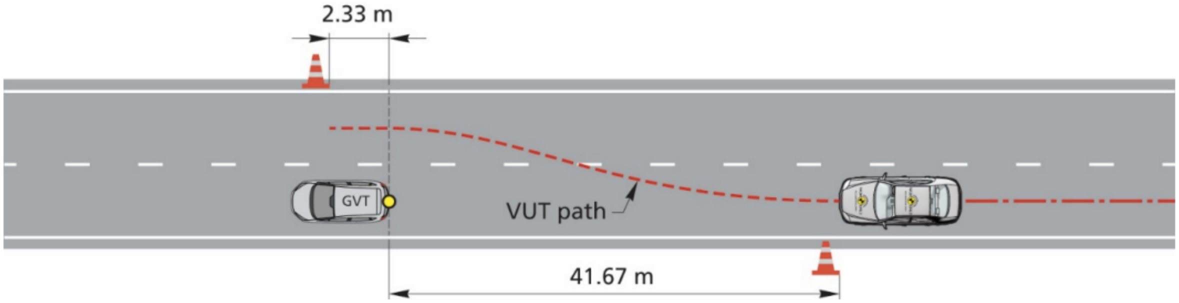


Figure 3: Euro NCAP Cut-out scenario from [15 Euro NCAP]

<u>NHTSA Lead Vehicle Deceleration, Accelerates, then Decelerates (LVDAD)</u>
The NHTSA Lead Vehicle Deceleration, Accelerates, then Decelerates (LVDAD) [18 NHTSA] test’s objective is to “evaluate the TJA system’s ability to detect and respond to a POV that moderately brakes to a stop, pauses, accelerates back to its initial speed, then brakes aggressively to a stop ahead of the SV”

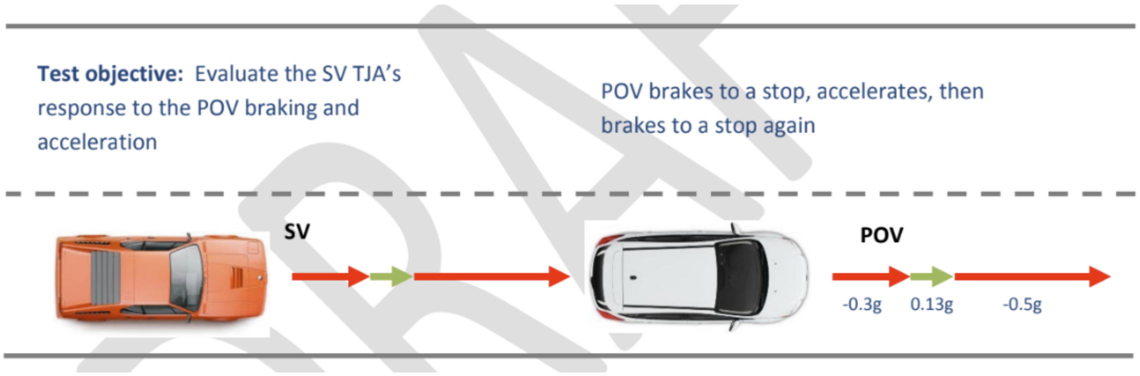


Figure 4: NHTSA TJA LVDAD scenario from [18 NHTSA]

Euro NCAP Front Turn Across Path (FTAP)

The final test that was selected was the Euro NCAP Car-to-car Front turn across path (FTAP) [14 Euro NCAP]. The test consists of “a collision in which a vehicle turns across the path of an oncoming vehicle traveling at constant speed, and the frontal structure of the vehicle strikes the front structure of the other.”

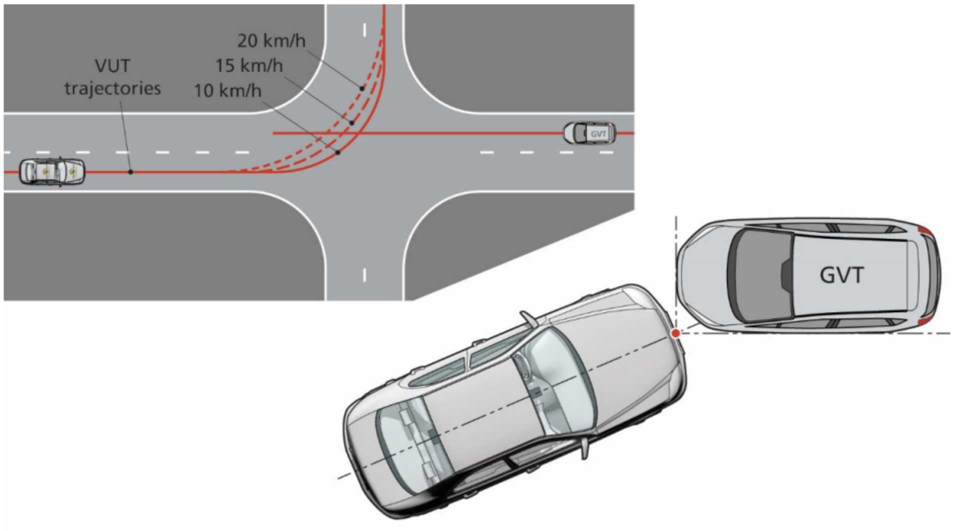


Figure 5: Euro NCAP FTAP scenario from [15 Euro NCAP]

Excluded Tests

As previously mentioned, not all of the tests from the Euro NCAP and NHTSA test procedures were selected. Priority was given to the tests that had a higher anticipated difficulty based on the Waymo Driver’s capability and ODD. Some of the excluded tests, like the Car-to-Pedestrian Longitudinal Adult (CPLA) from Euro NCAP’s AEB VRU Tests [16 Euro NCAP] were performed once and deprioritized since the Waymo Driver was able to easily detect and respond the the test target and there were no new implementation challenges presented by these scenarios. Additionally, a few tests from Euro NCAP’s AEB VRU Tests Annex B: Testing At Low Ambient Lighting Conditions [16 Euro NCAP] were executed, but these were also deprioritized since there were no observed performance differences from the same daylight tests due to the Waymo Driver’s sensor and perception capability, mainly lidar performance in low ambient lighting.

TESTS' EXECUTION AND RESULTS

Each of the above tests were first attempted with the specifications provided in the test procedure. However, due to the proactive safe driving capability of the Waymo Driver which prevented the scenarios from turning critical, modifications were needed to force the Waymo Driver into the collision avoidance maneuvers as originally intended in the ADAS and active safety test procedures. The two main types of modifications were: 1) Additions and alterations of the test procedure due to ***under-specification of the protocol*** for application to the Waymo Driver and 2) Additions and alterations to ***increase the difficulty*** of the protocol to ensure alignment with the original intent of the protocol. With these modifications, the Waymo Driver was able to be evaluated in the intended interactions described in the test procedures. For each of these tests, including those originally run to the test specifications and those that were modified, the Waymo Driver was able to avoid contact with the test target. The rest of the section provides further details and rationale for each of the modifications made.

NOTE: The birds-eye view visualization provided in the figures below is a simplification of the Waymo Driver's perception system for illustrative purposes and does not represent the full extent of objects tracked by the system nor show sensor data.

<table>
<tr><td colspan="2">Protocol: Euro NCAP CPNC</td></tr>
<tr><td colspan="2">Implementation Challenges</td></tr>
<tr><td colspan="2">● The Waymo Driver is able to detect the VRU test target to the right of the two parked vehicles specified in the test procedure prior to it becoming occluded (this starting position is required to meet the stated VRU speed per procedure). This is shown in Figure 6. This results in the Waymo Driver tracking the pedestrian and proactively slowing down before the intended reveal.
● The Waymo Driver would slow down below the required testing speed due to the presence of parked cars to the side and/or to unclear classification of those vehicles as parked and/or stationary in the active lane. Depending on the exact configuration of the testing site, the Waymo Driver would also slow down if it knew of or expected an upcoming intersection as shown in Figure 6 (the test was executed at various locations within the test facility)</td></tr>
<tr><td colspan="2">Modifications</td></tr>
<tr><td>Protocol under-specification</td><td>● Need to specify marking that separates the active lane for VUT and the lane that stationary vehicles are parked in to avoid spurious ADS slow-downs for uncertain classification of side vehicles
● Need to specify road graph beyond the active site of the pedestrian crossing to avoid spurious ADS slow-downs for intersections
Adopted modifications: the HD map was altered by removing the perpendicular intersection leg and stop sign to enable the Waymo Driver to maintain the required test speed. Additionally, all signs on the test track were covered as shown in Figure 7.</td></tr>
<tr><td>Difficulty</td><td>● Need to ensure enhanced perception capabilities are accounted for in the test
Adopted modifications: two additional obstructing vehicles were placed to the right of the original vehicles as shown in Figure 7 along with objects placed under the vehicles to block the Waymo Driver from detecting and tracking the child pedestrian</td></tr>
</table>

Figure 6: Euro NCAP CPNC scenario before modifications. Image source: Waymo.

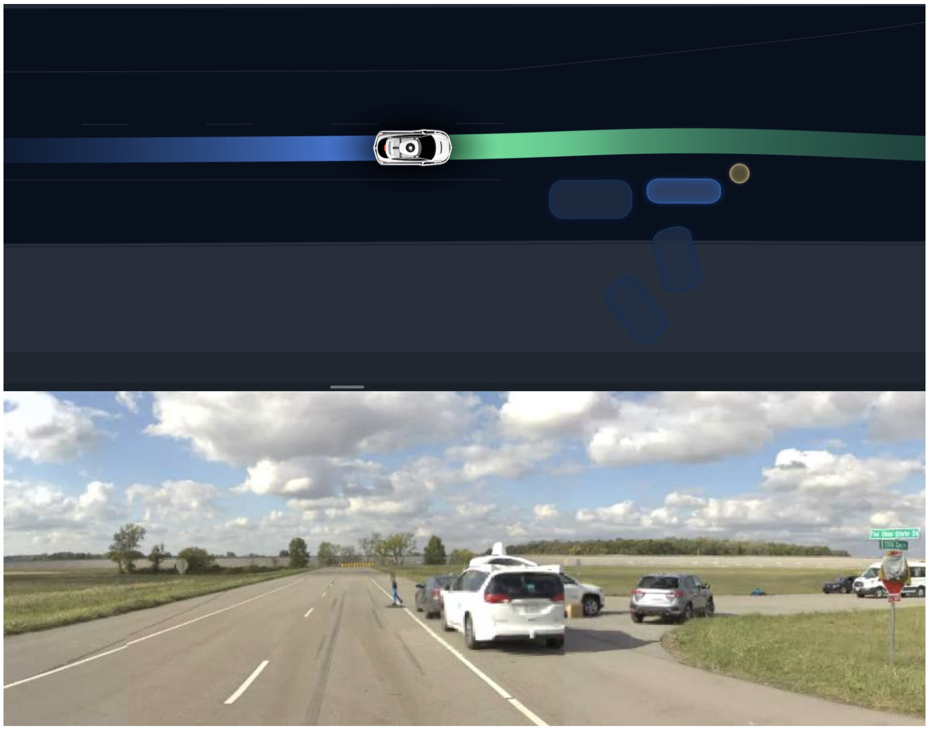


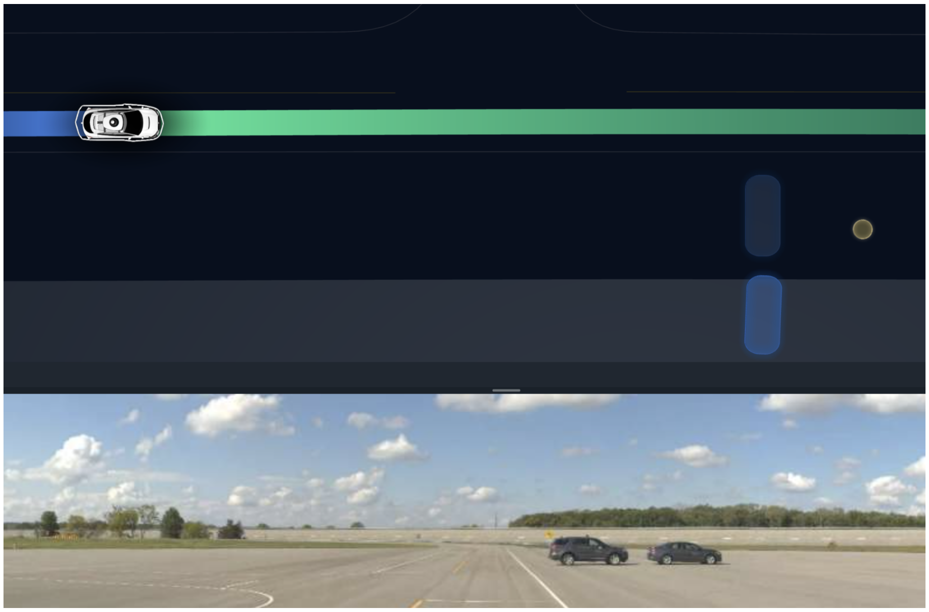
Figure 7: Euro NCAP CPNC scenario after modifications. Image source: Waymo.

Protocol: Euro NCAP CBNAO

Implementation Challenges

- The Waymo Driver is able to detect the VRU test target between and overtop of the two parked vehicles specified in the test procedure due to the location and performance of its sensors (the starting position of the test target is required to meet the stated VRU speed per procedure) as shown in Figure 8. Additionally, the Waymo Driver could detect the test target via sensor returns from under the parked obstructions. These both contribute to the Waymo Driver tracking the cyclist and proactively slowing down before the intended reveal.
- The Waymo Driver would slow down below the required testing speed due to the presence of perpendicularly parked cars to the side and/or to unclear classification of those vehicles as parked and/or stationary in an active lane/unmarked intersection/alley. Depending on the exact configuration of the testing site, the Waymo Driver would also slow down if it knew of or expected an upcoming intersection (the test was executed at various locations within the test facility).

Modifications

<table>
<tr><td>Protocol under-specification</td><td>● Need to specify marking that separates the active lane for VUT and the lane that stationary vehicles are parked in to avoid spurious ADS slow-downs for uncertain classification of side vehicles
● Need to specify road graph beyond the active site of the pedestrian crossing and/or specify straight road ahead to avoid spurious slow-downs for intersections
Adopted modifications: the HD map was altered to enable the Waymo Driver into maintaining the test required speed</td></tr>
<tr><td>Difficulty</td><td>● Need to ensure enhanced perception capabilities are accounted for in the test
Adopted modifications: An additional taller obstructing vehicle[6] was placed in front of and in between the original vehicles as shown in Figure 9 along with objects placed under the vehicles to block the Waymo Driver from detecting and tracking the cyclist</td></tr>
<tr><td colspan="2">
Figure 8: Euro NCAP CBNAO scenario before modifications. Image source: Waymo.</td></tr>
</table>

---

[6] Another possible obstruction is a temporary wall such as [21 4activesystems]

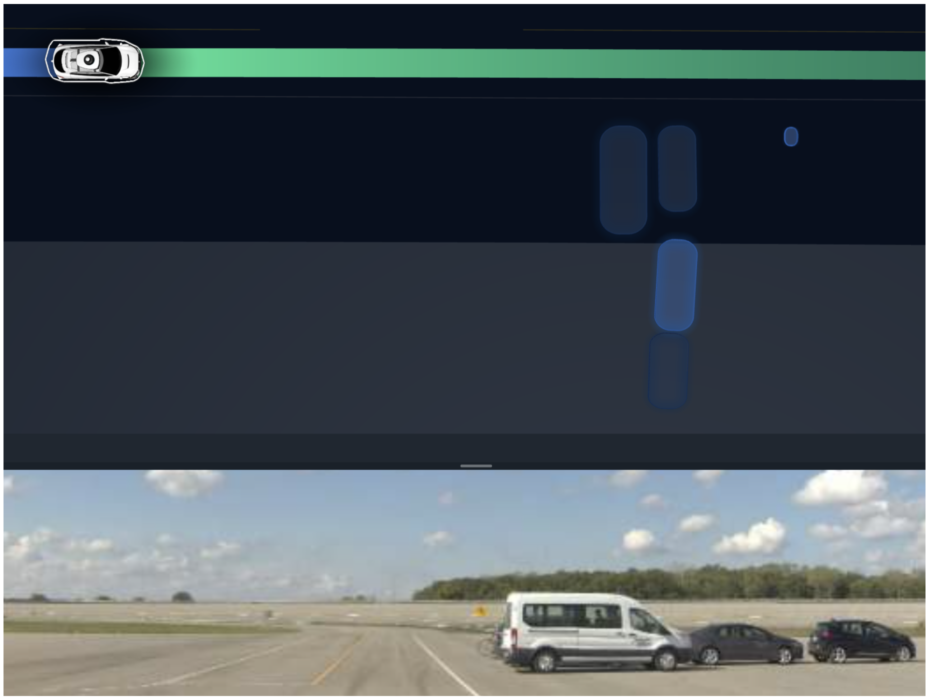


Figure 9: Euro NCAP CBNAO scenario after modifications. Image source: Waymo.

Protocol: Euro NCAP Cut-out	
Implementation Challenges	
The Waymo Driver is able to detect the stopped vehicle ahead in the lane over the vehicle it was following as shown in Figure 10. This results in the Waymo Driver tracking the stopped vehicle and proactively slowing down before the intended reveal.	
Modifications	
Protocol under-specification	● Need to specify markings for all lane lines to avoid mismatches in prediction of actors actions ● Need to specify road graph beyond and in adjacent lanes of the stopped vehicle to avoid spurious ADS slow-downs **Adopted modifications:** HD map was updated with lane line information to enable testing.
Difficulty	● Need to ensure enhanced perception capabilities are accounted for **Adopted modifications**: A taller lead vehicle was used as shown in Figure 11 to block the Waymo Driver from detecting and tracking the stopped vehicle overtop of the leading vehicle.

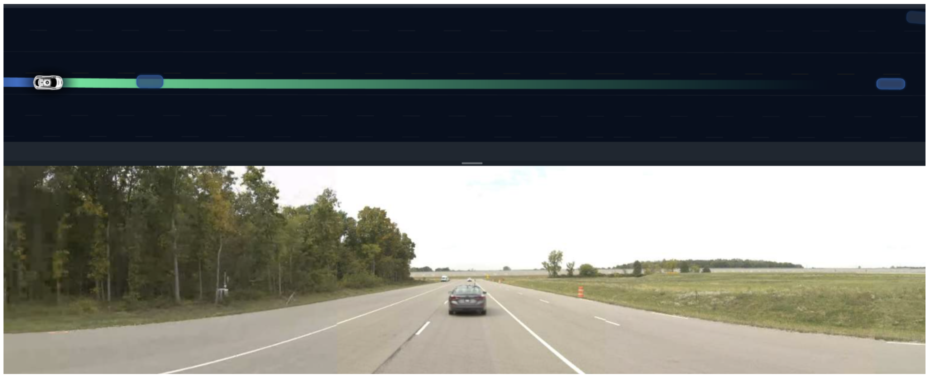


Figure 10: Euro NCAP Cut-out scenario before modifications. Image source: Waymo.

Figure 11a: Euro NCAP Cut-out scenario after modifications. Image source: Waymo.

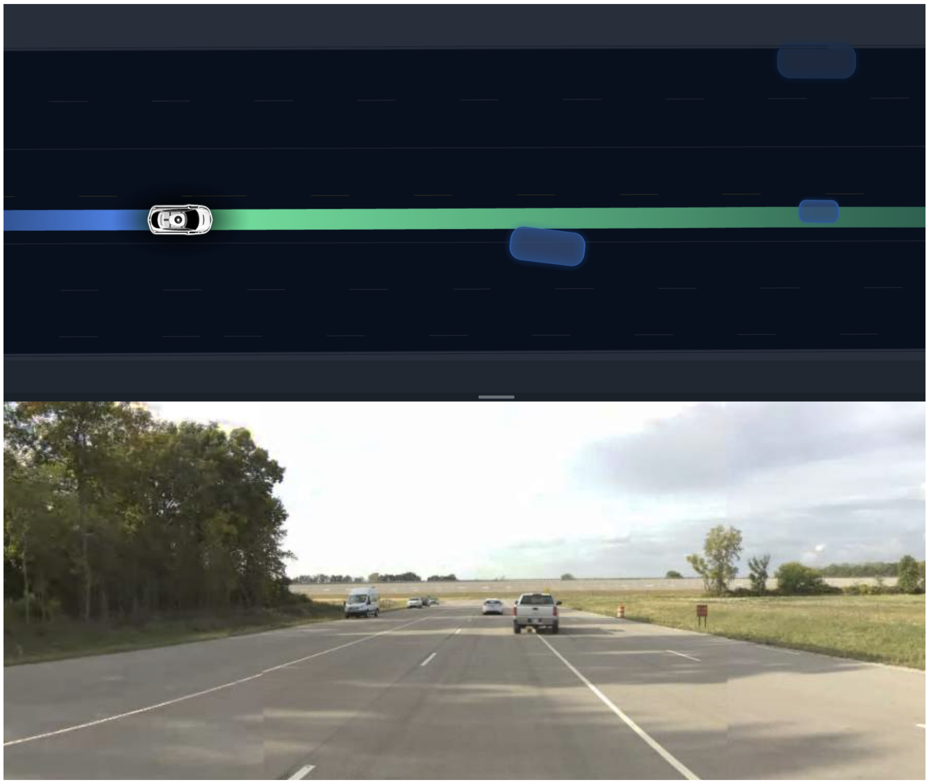


Figure 11b: Euro NCAP Cut-out scenario after modifications[7]. Image source: Waymo.

Protocol: NHTSA LVLCB (Cut-in)	
Implementation Challenges	
The tests were run as originally specified in the test procedure, including the two cut-in distances, 10.7m and 7.5m, and two decelerations, 0.3g and 0.5g, after the lane change. The Waymo Driver's ability to detect, track and predict the cutting-in vehicles behavior resulted in low decelerations and large proximities between the two vehicles as shown in Figure 12.	
Modifications	
Protocol under-specification	● Need to specify and add additional capabilities for the test target to signal, or not, intent of lane change **Adopted modifications**: None, test were ran with the GVT
Difficulty	● Need to ensure enhanced perception and prediction capabilities are accounted for in the test **Adopted modifications**: Two closer cut-in distances, 6m and 3m were used[8].

---

[7] The top visualization is for illustrative purposes and does not represent the full extent of sensor data or objects tracked by the system

[8] Deceleration was not increased above 0.5g due physical limitations of the GST which has a max specified deceleration of 0.8g. 0.5g was found to be the maximum practical deceleration for repeatable GST movements and without damaging the GST wheels.

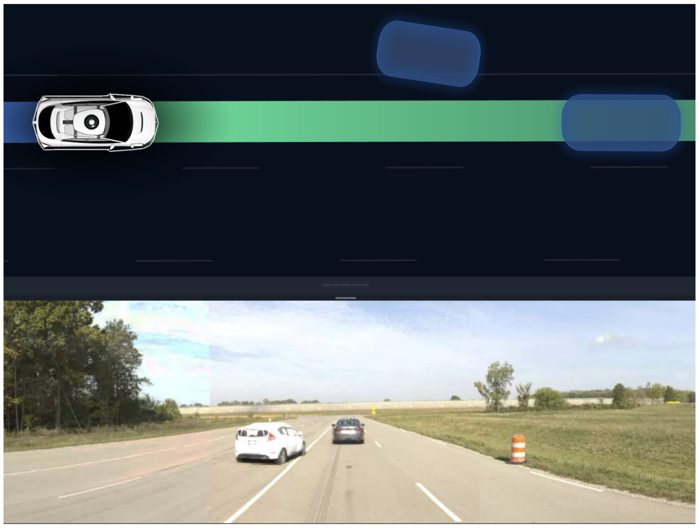


Figure 12: Euro NCAP Cut-in (7.5m) scenario before modifications. Image source: Waymo.

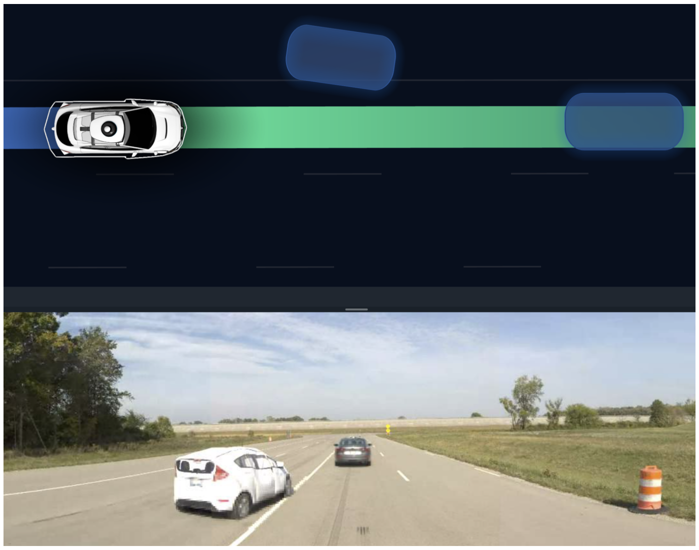


Figure 13: Euro NCAP Cut-in (6m) scenario after modifications. Image source: Waymo.

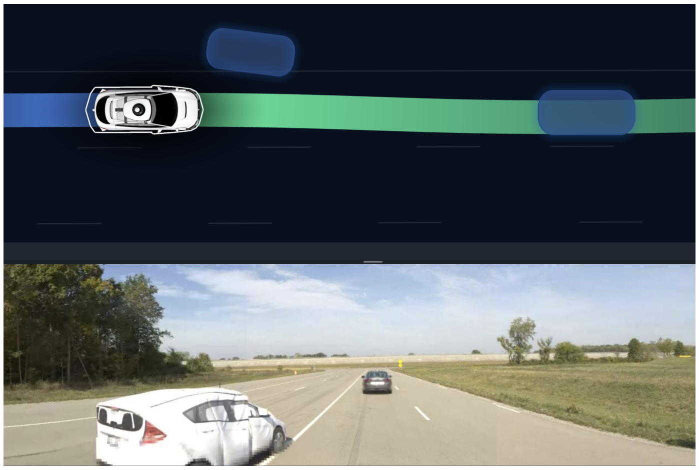


Figure 14: Euro NCAP Cut-in (3m) scenario after modifications. Image source: Waymo.

A test matrix of all of the tests performed, including those with modifications, is shown in Table 2. As previously mentioned, all of these tests, including those at various speeds and those that were modified, were passed by the Waymo Driver in that no contact was made with the test target.

Table 2: Test Matrix

Test	Speeds (kph)		Modifications (beyond HD map)	Pass/Fail
Euro NCAP CPNC	10, 30, 40, 50, 60		Extra obstructions	Pass
Euro NCAP CBNAO	32, 50, 60		Taller extra obstruction	Pass
Euro NCAP CBNA	60			Pass
Euro NCAP CBFA	60			Pass
NHTSA TJA LVLCB	24.1	w/0.3g decel, w/0.5g decel		Pass
	40	w/0.3g decel, w/0.5g decel	6m and 3m cut in distance	Pass
Euro NCAP Cut-In	50, 80			Pass
NHTSA TJA LVDAD	24.1, 37			Pass
Euro NCAP LTAP	15, 20			Pass
Euro NCAP Cut-Out	50, 70		Tall lead vehicle	Pass

DISCUSSION

Beyond the specific, practical challenges mentioned, two additional issues of *feasibility* and *utility* arise when considering adapting or adopting these ADAS and active safety tests that are designed for a specific driving function in a specific collision scenario to the Waymo Driver which is required to perform the entire DDT without immediate human intervention for all scenarios in its ODD. The rest of this section elaborates on these two issues.

Feasibility

Unlike ADAS and active safety systems, which act on fewer sources of input information (e.g., surrounding vehicle movement, lane markings) with less capable sensing, prediction, and planning capabilities, the Waymo Driver takes the holistic, long-horizon road scene into account when planning its behavior. Therefore, additional considerations need to be given regarding the specification of the road scene, environment, surrounding safety critical objects, and timing aspects for these tests. Identifying which of these factors are pertinent for assessing the safety of the ADS without proprietary ODD, performance, and design information and varying those factors is difficult to do and would result in an untenable set of potential scenarios for closed-course testing, commonly referred to as the “parameter explosion” problem (IEEE 2846-2022, Annex A) [22] . Additionally, those pertinent factors may vary widely depending on the intended ODD and capabilities of the system [6 Kusano], making it difficult to design a robust closed-course test program that could be reused across many different ADS as is done for ADAS and active safety systems today. Moreover, the behavior of the ADS may be affected by prior information collected during field testing (e.g., likelihood of parked vehicles, likelihood of pedestrians crossing) and a closed-course test on a synthetic map with surrogate actors may not be indicative of actual on-road performance. Furthermore, this reliance on map data introduces the risk of tampering with the test results through the creation of maps. Even without malicious intent, the creation of a special map for a test track may inadvertently create an unfair or unrepresentative evaluation of the ADS’s capabilities.

Another practical issue pertains to the physical limitations of test facilities, mainly that a test track needs to have the desired ODD features for a specific ADS. These limitations coupled with limitations of test equipment (e.g. top speeds, decelerations, lateral accelerations, etc.) significantly reduces the type of scenarios, especially high-severity car-to-car scenarios like those in [23 Scanlon], that can be executed on a closed-course. Closely related to this test facility limitations issue are implementation considerations that need to be taken into account for various scenarios on a closed-course. For example, if you want to test varying aspects of the ODD related to road features, such as speed, each run of an individual protocol requires the generation of ad-hoc metadata and an accompanying map configuration along with making sure that any speed limit signs on the test track are covered or updated so it is not conflicting with the test procedure. This implies that the test execution team needs to manage numerous different map and road configurations that need to be updated on the vehicles and test facilities. This results in an additional test burden and the necessity to set up appropriate test management practices that adds complexity beyond that in these already complex testing campaigns, as stated in [24 Manahan] "...we acknowledge the test burden associated with the use of complex and highly synchronized track-based efforts, and why NHTSA must determine the role of simulation in its research efforts". Beyond these complexity challenges, consideration also needs to be given to make sure that a given test is applicable for a specific ADS's ODD and functions, not to mention that its ODD and functions can quickly change with software updates. This is already being seen in some more recent ADAS vehicles where the test results significantly depend on the current software version, which can be remotely updated, disrupting testing or negating previous tests performed on a different software [25 Bauchwitz - 26 Cummings]. Another issue related to the physical limitation of the test facilities and equipment that was briefly discussed earlier is the representativeness of these idealized tests to scenarios encountered in on-road driving. These test protocols take place on facilities that are designed to be flexible and general purpose with surrogate targets, therefore it is hard to evaluate how representative these idealized tests are to scenarios and behaviors encountered in the real world where conditions can be less than ideal. This is discussed in greater detail in Waymo's Collision Avoidance Testing paper [6 Kusano] and these stated limitations further reinforce the need to follow a scenario-based testing program like that outlined in [6 Kusano].

A final practical feasibility consideration is with the iterative process of trying to characterize the SUT to get the scenario timing correct. Traditionally, ADAS and active safety tests have been evaluated in a black box manner, where the timing in the scenarios are calibrated based on repeated test track runs. Due to the behavior variation challenges that an ADS may have as mentioned above, using the planner information generated by an ADS to trigger events (e.g., the start movement of the pedestrian) would both lead to more efficient testing and more repeatable results. Without this information, getting the choreography of the scenario correct is challenging due to the increased sensing and planning capabilities of an ADS and the increased input dimension an ADS may act on. Therefore, trying to characterize timing and coordination of actors in the tests from repeated trials may be more difficult.

Utility

Closely related to the issue of the feasibility of closed-course testing to inform consumers of ADS safety performance is the issue of utility of the conclusions that can be drawn from this type of testing. One of the main goals of these consumer-based tests is to inform the public on the capabilities of a given system and to be able to compare the performance of different manufacturers in frequent and/or severe crashes that current technologies have the potential to mitigate or avoid that can be tested with current test track and test tool capabilities. This comparison amongst ADS manufacturers will be much harder and less meaningful since many details about true capabilities of a system and its dependency on specific ODD features may be proprietary. Additionally, these ADAS and active safety tests procedures are scoped in their design to push the current state of the art (SOTA) so consumers can see differentiation amongst systems, i.e. the tests can't be too easy so everyone passes nor too difficult that no one passes. This is in contrast to other scenario-based testing methods, such as the one outlined in [6 Kusano], that aim to be representative of real-world crashes, including their frequency and severity distributions, rather than selecting tests targeted for differentiation of current SOTA systems. This difficulty of trying to design tests to differentiate performance amongst ADAS and active safety systems becomes almost impossible for ADSs given 1) the current SOTA in test tools and test facilities as previously mentioned and 2) the increased capabilities and performance of ADSs.

To provide insight into this challenge of designing scenarios for an ADS that aim to address the aforementioned goal of these consumer-ratings, specifically targeting severe crashes, the maximum injury potential (maxIP) [27 Kusano] for the baseline scenarios and the scenarios that were modified is shown in Figure 15 and 16 below. The maxIP is a metric that describes the worst-case outcome that is insensitive to avoidance maneuvers of the VUT. It propagates counterfactual trajectories based on actors capabilities, sizes of actors, and their inertial properties. Any potential collisions are input into a collision and injury model which outputs the maximum probability of injury severity using the Abbreviated Injury Scale (AIS). An injury with a score of 3 or higher on the AIS is classified as clinically seriously injured (MAIS3+). Figure 15 shows that for the car-to-car test scenarios, including the modified scenarios (ncap_cut_out_tall_lead_vehicle, nhtsa_cut_in_0.5g_3m_headway, and nhtsa_cut_in_0.5g_6m_headway in Figure 15), none of the scenarios have high probability for severe injuries mainly due to the test equipment and facility limitations. This limits the ability of these consumer ratings tests to meet their stated goal of targeting severe crashes. In contrast, the VRU interaction scenarios do have a higher potential for severe injuries if no intervention is taken by the VUT, as shown for the VRU cyclist tests in Figure 16. This figure also shows that the implemented modifications (cbnao_tall_obstruction and cbnao_tall_and_extra_obstruction in Figure 16) increased the maxIP for the scenario. In the unmodified scenarios, the Waymo Driver slowed down much earlier than the intended reveal, decreasing the probability for any severe injury interactions within the maxIP counterfactual trajectory propagation look ahead window. This again limits the stated goal of targeting severe crashes for these consumer ratings tests.

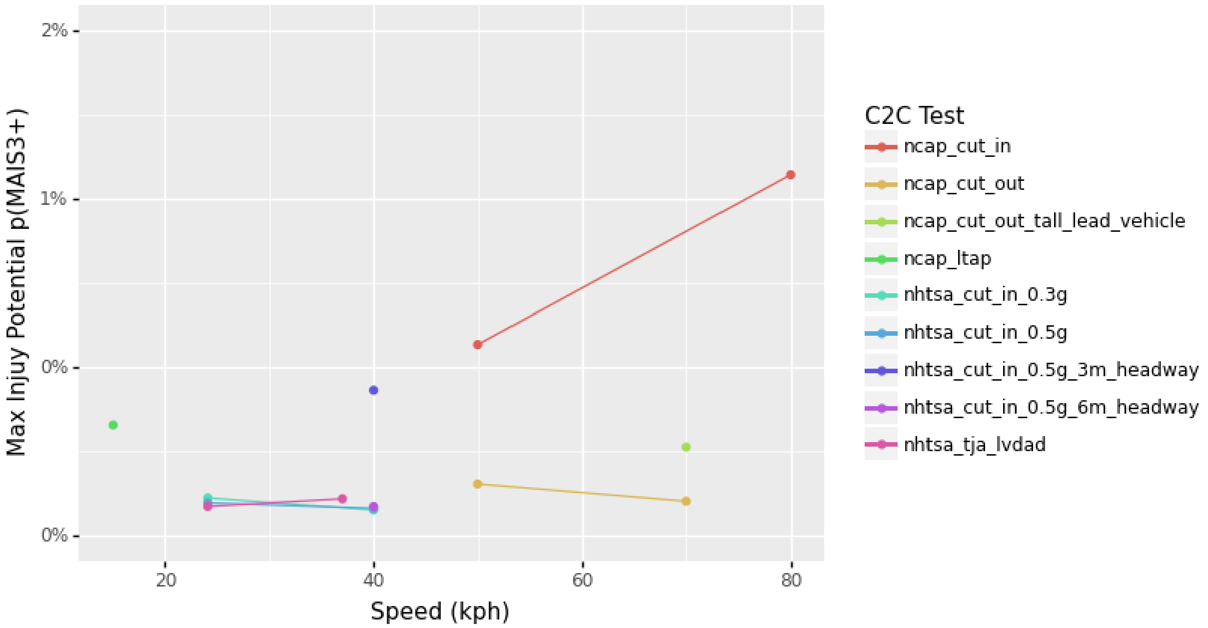


Figure 15: C2C maxIP

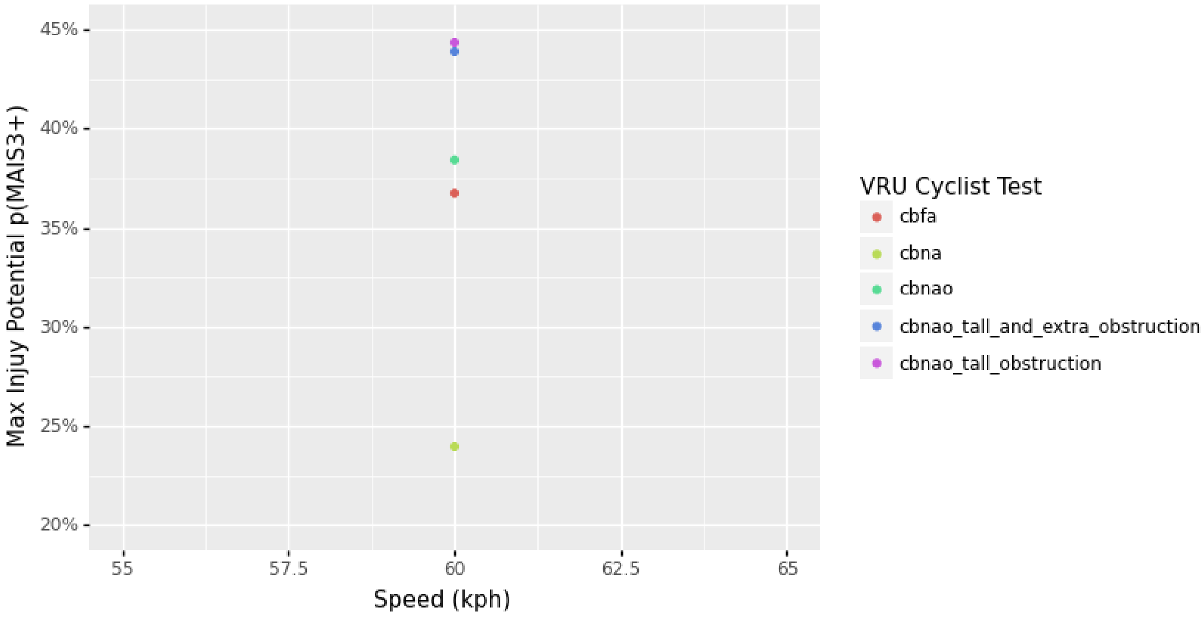


Figure 16: VRU Cyclist maxIP

CONCLUSION

While closed-course consumer-based ADAS and active safety ratings tests could be informative for demonstrating safety of an ADSs to the public, they are far from sufficient. The Waymo Driver was able to pass all of the selected ADAS and active safety tests, with required modifications, which is a logical first step for demonstrating safety given the availability of these tests. However, due to the stated challenges, issues, and limitations, procedures developed for Level 2 systems turned out to be less challenging and constraining for a Level 4 ADS, such as the Waymo Driver, further proving the point that these types of technological solutions need to be separately considered. The role of traditional consumer-testing and the intuitiveness of these protocols has a place in informing the selection of a few tests and engender comparison across platforms, but cannot be the basis for consumer safety information for ADS (neither was it ever the intent for these test procedures to do so). Additionally, it is important to note that claiming to pass these test procedures unmodified without further consideration to necessary changes to enable the intent of the original test procedure may result in false confidence in the safety of an Level 4 vehicle.

Therefore, to address the practical challenges raised and the issues relating to the feasibility and utility of consumer-focused safety testing, an alternative scenario-based safety evaluation process is proposed in [6 Kusano] that focuses on simulation and on-road testing to provide the public with information regarding safety assurance of an ADSs in its intended ODD. A staple in the behavioral evaluation of the Waymo Driver comes from the intensive pressure-testing of collision avoidance responses of our ADS in thousands of scenario variations. This principled and systematic scenario-based testing program is built upon the creation of vast libraries of situations that the Waymo Driver may be exposed to. This type of assessment can provide consumers with a more holistic safety evaluation of an ADS.

REFERENCES

[1] Wimmer, P., Düring, M., Chajmowicz, H., Granum, F., King, J., Kolk, H., ... & Wagner, M. (2019). Toward harmonizing prospective effectiveness assessment for road safety: Comparing tools in standard test case simulations. *Traffic injury prevention*, *20*(sup1), S139-S145.

[2] International Organization for Standardization. (2022). *Road Vehicles - safety of the intended functionality* (ISO 21448:2022).

[3] International Organization for Standardization. (2022). *Road vehicles - Test Scenarios for Automated Driving Systems - Scenario-based safety evaluation framework for Automated Driving Systems* (ISO 34502:2022).

[4] International Organization for Standardization. (2021*). Intelligent transport systems — Low-speed automated driving (LSAD) systems for predefined routes — Performance requirements, system requirements and performance test procedures* (ISO 22737:2021)

[5] Webb, N., Smith, D., Ludwick, C., Victor, T., Hommes, Q., Favaro, F., ... & Daniel, T. (2020). *Waymo's safety methodologies and safety readiness determinations.* arXiv preprint arXiv:2011.00054.

[6] Kusano, K., Beatty, K., Schnelle, S., Favarò, F., Crary, C., Victor, T. (2022). *Collision Avoidance Testing of the Waymo Automated Driving System*. arXiv preprint arXiv:2212.08148.

[7] Ding, W., Xu, C., Lin, H., Li, B., & Zhao, D. (2022). *A Survey on Safety-critical Scenario Generation from Methodological Perspective*. arXiv preprint arXiv:2202.02215.

[8] Zhang, X., Tao, J., Tan, K., Törngren, M., Sánchez, J. M. G., Ramli, M. R., ... & Felbinger, H. (2021). *Finding critical scenarios for automated driving systems: A systematic literature review*. arXiv preprint arXiv:2110.08664.

[9] Riedmaier, S., Ponn, T., Ludwig, D., Schick, B., & Diermeyer, F. (2020). *Survey on scenario-based safety assessment of automated vehicles*. IEEE access, 8, 87456-87477.

[10] EUROPEAN NEW CAR ASSESSMENT PROGRAMME (Euro NCAP) *What's New?* Accessed on November 21, 2022 from: https://www.euroncap.com/en/for-engineers/whats-new/

[11] National Highway Traffic Safety Administration [Docket No. NHTSA-2021-0002] (2021) *New Car Assessment Program Request for comments (RFC).* Accessed on November 21, 2022 from https://www.nhtsa.gov/sites/nhtsa.gov/files/2022-03/NCAP-ADAS-RFC-03-03-2022-web.pdf

[12] IIHS (2022) *Test protocols and technical information* Accessed on November 21, 2022 from https://www.iihs.org/ratings/about-our-tests/test-protocols-and-technical-information

[13] UNECE (2021) *UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS)* Revision 3 https://unece.org/sites/default/files/2021-03/R157e.pdf

[14] EUROPEAN NEW CAR ASSESSMENT PROGRAMME (Euro NCAP) (2021) *TEST PROTOCOL – AEB Car-to-Car systems* Version 3.0.3 https://cdn.euroncap.com/media/62794/euro-ncap-aeb-c2c-test-protocol-v303.pdf

[15] EUROPEAN NEW CAR ASSESSMENT PROGRAMME (Euro NCAP) (2020) *Assisted Driving Highway Assist Systems Test & Assessment Protocol* Version 1.0 https://cdn.euroncap.com/media/58813/euro-ncap-ad-test-and-assessment-protocol-v10.pdf

[16] EUROPEAN NEW CAR ASSESSMENT PROGRAMME (Euro NCAP) (2021) *TEST PROTOCOL – AEB VRU systems* Version 3.0.4 https://cdn.euroncap.com/media/62795/euro-ncap-aeb-vru-test-protocol-v304.pdf

[17] National Highway Traffic Safety Administration (2019) *Advanced Driver Assistance Systems Draft Research Test Procedures* Docket No. NHTSA-2019-0102 https://www.federalregister.gov/documents/2019/11/21/2019-25217/advanced-driver-assistance-systems-draft-research-test-procedures#footnote-6-p64405

[18] National Highway Traffic Safety Administration (2019, July). *Traffic jam assist system confirmation test (working draft)*. Washington, DC: National Highway Traffic Safety Administration. https://www.regulations.gov/document/NHTSA-2019-0102-0002

[19] National Highway Traffic Safety Administration (2021, May). *An Approach for the Selection and Description of Elements Used to Define Driving Scenarios.* Report Number : DOT HS 813 073: https://rosap.ntl.bts.gov/view/dot/55465

[20] National Highway Traffic Safety Administration (2019, September). *Pedestrian automatic emergency brake system confirmation test (working draft)*. Washington, DC: National Highway Traffic Safety Administration. https://www.regulations.gov/document/NHTSA-2019-0102-0005

[21] 4activesystems (2022) *4activeEQ-Wall* Accessed on November 21, 2022 from: https://www.4activesystems.at/4activeeq-wall

[22] IEEE Vehicular Technology Society (2022). *IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems.* (IEEE Std 2846-2022).

[23] Scanlon, J. M., Kusano, K. D., Daniel, T., Alderson, C., Ogle, A., & Victor, T. (2021). *Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain*. Accident Analysis & Prevention, 163, 106454.

[24] Manahan, T., Forkenbrock, G. (2020). *Exploring the Practical Limits of Multi-Actor Test Track Evaluations*. 2020 SAE Government and Industry Meeting. Accessed on November 20, 2022 from https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/exploring_the_practical_limits_of_multi-actor_test_track_evaluations_tag.pdf

[25] Bauchwitz, B., & Cummings, M. (2022). *Effects of Individual Vehicle Differences on Advanced Driver-Assist System Takeover Alert Behavior*. Transportation Research Record, 03611981211068362.

[26] Cummings, M. L., & Bauchwitz, B. (2021). *Safety Implications of Variability in Autonomous Driving Assist Alerting*. IEEE Transactions on Intelligent Transportation Systems.

[27] Kusano, K., & Victor, T. (2022). *Methodology for determining maximum injury potential for automated driving system evaluation.* Traffic injury prevention, 1-4. DOI: 10.1080/15389588.2022.2125231