Teddy bears and driving automation: an on-road paradigm to evaluate situational awareness

Supplemental research paper*

March 2021

Alexandra S. Mueller
Jessica B. Cicchino
Insurance Institute for Highway Safety

Amy Benedick
Doreen De Leonardis
Rick Huey
Westat

*This paper is a supplement to the IIHS research study *Bears in our midst: familiarity with Level 2 driving automation and situational awareness of on-road events*

Insurance Institute for Highway Safety
4121 Wilson Boulevard, 6th floor
Arlington, VA 22203
researchpapers@iihs.org
+1 703 247 1500

iihs.org

Contents

Abstract

This paper discusses the lessons learned from the development of a novel paradigm in an on-road observation study that explored human factors issues when using Level 2 driving automation. There is the risk that drivers may become disengaged from the driving task when using these partially automated systems, but current production Level 2 systems vary between manufacturers in terms of their functional performance and driver management strategies. On-road methodologies are needed to objectively evaluate the degree of disengagement drivers may experience when behind the wheel of these types of vehicles. This paper presents a proof of concept of an experiment that involved an approximately 1-hour drive along a predetermined route in a 2019 Mercedes-Benz C300 equipped with a Level 2 system. The paradigm was designed to assess drivers' inattentional blindness to three driving-related surprise events, thereby providing a measure of situational awareness. Using GPS data from the participant's vehicle and another study vehicle, the other vehicle overtook the participant at three predetermined locations. The "surprise" nature of each event was an oversized pink teddy bear wearing a high-visibility jacket that was mounted to the back of the other study vehicle. Situational awareness was evaluated in a post-drive survey in which participants had to recall the bear and how many times they had seen it on the road. The aim of this paper is to help inform future research using on-road driving methods, particularly as they relate to the use of Level 2 driving automation, by using insights gained from this study's experimental procedure, vehicle equipment, and data collection protocols.

Introduction

On-road observation driving studies are time and labor intensive, but they have certain advantages over laboratory studies that make them invaluable for understanding how people behave in the real world. Laboratory setups often lack the fidelity for a truly immersive, realistic experience. The artificiality, particularly as it relates to the driving context, risks participants not taking experimental tasks seriously, meaning they might not respond the way they would on the road in a real vehicle where their behavior would have actual consequences. In addition, advanced driver assistance systems are becoming more widely available today in production vehicles, and many of these systems behave differently on the road compared to controlled test track or laboratory conditions. Level 2 driving automation (SAE International, 2018) integrates speed and headway support from the adaptive cruise control system and continuous active steering assistance from the lane centering system. The dynamic and uncontrolled road infrastructure and traffic behavior in the real world often lead Level 2 systems to struggle with conditions that are not usually present in closed-course testing (American Automobile Association, 2020; Insurance Institute for Highway Safety [IIHS], 2018). As a result, they frequently behave in ways that are unexpected to the driver, and it is necessary to test these systems under real conditions to understand how drivers interact with them.

Level 2 systems are known as partial automation because the driver remains fully responsible for the vehicle’s behavior when the system is on. A well-documented issue with using any partially automated system is the difficulty the operator faces when trying to pay attention to the environment and what the automation is doing for extended periods (Endsley, 2017b; Manzey, Reichenbach, & Onnasch, 2012). In other words, although Level 2 systems are designed to make the driving task easier, they introduce a new level of difficulty by adding another element that the driver has to monitor while simultaneously requiring less physical involvement in controlling the vehicle. This creates a perfect storm for drivers to become cognitively disengaged from the driving task over time, which may reduce their situational awareness of what is happening on the road, especially if the driving automation performs fairly reliably in the operating conditions for which it was designed (Carsten, Lai, Barnard, Jamson, & Merat, 2012; Hergeth, Lorenz, Vilimek, & Krems, 2016). In addition, there is the risk of drivers overtrusting these systems and being more likely to engage in secondary activity when using them (de Winter, Happee, Martens, & Stanton, 2014; Reimer et al., 2016), which slows responses to situations that require driver intervention (Merat, Jamson, Lai, & Carsten, 2012). The danger of reduced situational awareness under these conditions is that, even with relatively reliable systems, Level 2 driving automation will still abruptly encounter conditions it is unable to handle and, as a result, the driver must be ready and able to rapidly intervene.

There is a need for on-road testing to utilize methodologies that evaluate a driver's situational awareness when using production Level 2 systems. Although self-report is useful for understanding driver state, it is subjective and not necessarily reflective of driving behavior (Schmidt et al., 2009). Some paradigms designed to assess situational awareness, such as the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 2000), do not translate well or easily to on-road testing. Many of those methodologies require interrupting the drive to deliver the experimental task or involve priming the driver ahead of the drive to look for events of interest; consequently, such paradigms are primarily designed for laboratory settings, such as driving simulators, where the environment can be highly controlled. Although Endsley (2017a) performed a modified version of the SAGAT while operating a Tesla Autopilot system, she is a human factors subject matter expert and was the sole participant in her case study, which limits the generalizability of the findings.

Study Design

Simons and Charbis (1999) conducted the famous "gorilla" study to assess situational awareness through an observer's inattentional blindness to a person dressed in a gorilla suit who walked through a group of people playing basketball while the observer watched how many passes one of the teams made. Their methodology for capturing an observer's awareness of a salient event while he or she performed another task inspired the design of the current study. We developed an ecologically relevant manipulation that could be delivered in a controlled manner across multiple participants as they drove a production vehicle in an on-road study. A post-drive recall survey was used to measure situational awareness. It was important to not prime participants ahead of the drive in any way that would make them look for the subject of the post-drive recall survey. The stimulus had to be relevant to driving, but also unusual in appearance to stand out to the driver enough to be able to recall after the drive. We chose an oversized pink teddy bear wearing a high-visibility (i.e., bright yellow) jacket as the subject of recall. The bear was mounted at the rear of another study vehicle, herein called the bear vehicle, as shown in Figure 1.

Figure 1. Teddy bear stimulus used for situational awareness evaluation

In order for the teddy bear stimulus to be presented consistently across trials and participants, the bear vehicle monitored and coordinated with the participant vehicle's GPS output to perform the overtaking maneuvers. The bear was presented for approximately 30 seconds in front of the participant, after which the bear vehicle reduced speed and fell out of sight until the next presentation or end of the drive. This presentation occurred three times at predefined locations during the drive to provide an added level of detail in the situational awareness recall assessment. After the drive, using a survey that was administered by the researcher to ensure response consistency and completeness, participants were first asked if they had seen anything odd about any vehicles in front of them during their drive. If the participant said yes, they were asked to describe what they had seen; if they said no, the researcher prompted them to recall any odd appearance of the back of any vehicles they saw. If they described a toy on a vehicle, their response was recorded as correct. The detail and accuracy of responses varied considerably. Incorrect or vague responses were followed by the researcher prompting for further details, such as the location of the stimulus on the bear vehicle (i.e., on the lift gate), what the stimulus looked like (i.e., pink, teddy bear), if it was wearing anything (i.e., high-visibility jacket), or what maneuver the vehicle did (i.e., overtake). Participants who described a bear or toy on a vehicle were then asked how many times they had seen it.

The complexity of this paradigm required thorough pilot testing of the equipment and procedures as well as research staff training. As Level 2 systems vary in functional performance and driver management strategies between manufacturers, a single production vehicle equipped with Level 2 driving automation was used to determine the efficacy of this new paradigm: a 2019 Mercedes-Benz C300. The vehicle was instrumented with video cameras to capture the participant's behavior during the drive. Participants were informed about the monitoring equipment and most appeared to ignore the devices after only a few minutes into the drive.

Insights for Future Research

Route selection

Participants drove alone on a predetermined route. The researcher conducted a 15-minute practice drive from a staging area near the experimental route to allow the participant to become familiar with the vehicle and Level 2 system. The participant received thorough verbal and video instruction about the vehicle, infotainment system, and Level 2 system prior to the practice drive. Once the practice drive was complete, the participant drove back to the staging area, dropped off the researcher, and began the experimental session where they would drive alone. The vehicle's navigation system guided participants along the route. Participants were given verbal instructions about the route prior to the drive and a physical map with turn-by-turn directions as a backup.

The drive lasted approximately 1 hour on a limited-access road, which is within the Level 2 driving automation's operational design domain. Functional testing has shown that these systems often struggle with sharp curves and hills (IIHS, 2018); therefore, the road selected for the drive was primarily straight with only gentle curves that the system could reliably handle. The road was primarily tree-lined with few buildings or signage along the route. The entire route was one speed limit (70 mph). Speed limit was an important consideration because the bear vehicle often had to exceed the speed limit to overtake the participant vehicle and present the bear stimulus. Experimental sessions were conducted when the traffic was lighter and free flowing to be as consistent as possible across participants and also to increase the likelihood that the bear vehicle could present the bear at predefined locations. The consistent traffic levels and speeds afforded by the route and time of day selection also reduced the need for participants to adjust settings and to deactivate and reactivate the Level 2 system. This lowered the chance of drivers selecting incorrect settings and also increased opportunity for driver disengagement by having the driving automation support active for as long as possible.

Vehicle settings

Many production vehicles have various driver assistance systems and sub-features that may not be desirable for an on-road study. For example, as the purpose of this study was to explore driver disengagement, we disabled the driver attention assist system that detects driver drowsiness. We also disabled speed-limit sign recognition, which automatically adjusts adaptive cruise control's (ACC) set speed to the speed limit whenever a new one is detected, because we wanted the system to be as predictable and consistent as possible. For this reason, we also kept the vehicle in sport mode to ensure ACC had rapid responses to keep up with traffic flow and avoid participant frustration. Although participants were given explicit instructions not to alter the instrument cluster display or change the vehicle settings, it was impossible to control this during the drives. Videos from every drive were reviewed to ensure the vehicle systems were used as intended.

Secondary activity was also of interest, and so participants were required to bring their personal smartphones to the study. Their smartphones were paired with the vehicle through Bluetooth. Apple CarPlay and Android Auto features, despite being widely available in newer production vehicles, were disabled due to the mixed success of connecting and using different smartphones with those systems during the piloting phase of the study. In an attempt to encourage participants to behave as they normally would when driving, they were instructed on how to use their smartphones through the vehicle system before each drive. Participants could use their smartphones through the hands-free system or otherwise, depending on their driving habits.

Participant instruction and bear reveals

Although the aim of the study was to determine whether and how drivers become disengaged from the driving task when they use Level 2 systems, participants were informed at the beginning of the study that the driving automation was not fully autonomous and they were responsible for the vehicle's behavior at all times. Moreover, while they were encouraged to drive as they normally would, they were told to obey traffic laws and the speed limit. This degree of instruction was necessary as one pilot participant, who was familiar with Level 2 systems, had his hands off the wheel for the majority of the drive, even though the system was not designed to permit hands-off-wheel behavior. Despite explicit instructions to obey the speed limit, speeding was observed among many participants, which sometimes made it difficult or impossible for the bear vehicle to safely present the bear.

Although participants were instructed at the beginning of the study to stay in the center lane when the road had three lanes or the rightmost lane when the road had two lanes, the challenges of on-road testing required flexibility in the maneuvering protocols for the bear reveals. To remain in the final sample, during each reveal, the bear was required to be fully visible two to three vehicle lengths in front of the participant for approximately 30 seconds to ensure the participant could not overlook it if he or she was paying attention to the road.

Alternative protocols had to be established for situations where the bear vehicle was unable to fall back after overtaking the participant from the left lane. If during the bear reveal the participant moved into the left lane behind the bear vehicle or a non-study vehicle followed the bear vehicle too closely to prevent a safe deceleration maneuver in the left lane, the bear vehicle merged into the far right lane to fall back behind the participant. If the participant was driving in the furthest left lane before the bear reveal, the bear vehicle had to perform the overtaking and deceleration maneuvers in the right lane closest to the participant. Given the complexity of testing conditions and the need for consistency across trials, videos were reviewed after every drive to ensure the bear reveals met study inclusion criteria.

Performing overtaking and fall back maneuvers occasionally elicited aggressive behavior from non-study drivers, such as honking, flashing lights, and overt gestures. Having an oversized pink teddy bear mounted at the back, at times, produced tailgating behavior from other drivers, some of whom were trying to take photographs of the bear. In addition, sometimes another vehicle would cut in between the bear vehicle and the participant and the surrounding traffic prevented a fall-back maneuver in any of the other lanes. Aggressive behavior toward the bear vehicle, participant speeding behavior, and uncooperative traffic conditions typically resulted in unusable trials, leading to the affected participants being removed from the sample because three complete bear reveals were required per participant.

Vehicle equipment and data

Coordinating the bear reveals required the study vehicles be able to share data in real time. They were equipped with computers that collected GPS data, cellular modems to share data, roof-mounted antennae to improve connectivity, and dedicated external power sources were used for all recording devices, including cameras. Essential to the study’s objective was capturing what participants were doing behind the wheel. Four cameras were set up in the C300 to capture (1) driver eye glances and face; (2) the vehicle cockpit, including the driver and passenger seats (the camera was mounted on the front passenger-seat headrest); (3) the vehicle instrument cluster, to monitor the Level 2 system activity; and (4) the forward roadway. Video data were combined through a mosaic box into a high-definition (HD) video recorder, which were then copied to a secure location to avoid data loss after every drive. A pragmatic note about managing the equipment over the course of the study is that all equipment, especially the cameras, could move or overheat. This meant using the A/C climate control and paying attention to sun exposure when the vehicle was parked, as well as performing camera-angle checks before every drive.

Integrating the video and GPS data allowed GPS locations to be used to identify twenty-three 30-second segments (epochs) of baseline participant behavior in the video stream. Baseline refers to the fact that those epochs did not overlap with the bear-reveal periods, which were treated separately. Using GPS locations ensured better consistency across participants than using time markers based on duration into the drive.

The bear vehicle's driver needed a safe way to efficiently monitor the participant’s location and be alerted when and where to make the bear reveals. A large touchscreen tablet was mounted in the bear vehicle’s cockpit, through which the driver was able to track the participant’s performance using an in-house developed application. Using the GPS information from both vehicles, the app provided alerts to the driver about the participant vehicle’s progress along the route, its distance from the bear vehicle, and ideal points for the bear-reveal initiation based on proximity of the bear vehicle to the predetermined locations and the participant vehicle. The locations for the bear reveals were selected to have cushions of space and time between surrounding baseline epochs to avoid overlap; however, the precise location and duration of every bear-reveal epoch were specific to each participant because of traffic conditions and participant behavior. The app allowed the bear vehicle's driver to identify when in the video and

GPS data streams the bear reveals occurred through a simple button press to mark the beginning and end of the presentations in real time.

Conclusions

Situational awareness is a challenging construct to measure, and various paradigms have been developed to understand it in different settings. Our method allows for an objective assessment of a driver's situational awareness that can be manipulated with the assistance of a Level 2 driving automation system in a manner that does not interfere with the driving task. While the observed behavior is representative of what occurs on the road every day, it does not capture how that behavior might change over time as drivers become more familiar with the system. Therefore, this paradigm offers a middle ground between naturalistic studies that follow behavior over time and laboratory studies with limited generalizability and helps to give researchers a better understanding of the real-world benefits and limitations of vehicle technologies, such as Level 2 driving automation.

Acknowledgements

This work was supported by the Insurance Institute for Highway Safety. We would like to thank Michael Gill, Jeremy Walrath, Jazzmyne Sangster, Elisha Lubar, Chavez Lee, and Matthew Airola for their roles in pilot testing, data collection, and video coding. Thank you to Eric Teoh for helping with data analysis. We are grateful to David Zuby, Lana Trick, Bobbie Seppelt, and Trent Victor for their insights into study design, materials, and analyses. Jeremy Clarkson from the BBC's Top Gear provided inspiration for the bear.

References

American Automobile Association, Inc. (2020). *Evaluation of active driving assistance systems*. Washington, DC.

Carsten, O., Lai, F., Barnard, Y., Jamson, H., & Merat, N. (2012). Control task substitution in semi-automated driving: Does it matter what aspects are automated? *Human Factors, 54*(5)*,* 747–761. doi:10.1177/0018720812460246

de Winter, J., Happee, R., Martens, M., & Stanton, N. (2014). Effects of adaptive cruise control and highly automated driving on workload and situational awareness: A review of empirical evidence. *Transportation Part F, 27,* 196–217. doi:10.1016/j.trf.2014.06.016

Endsley, M. (2017a). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making, 11*(3)*,* 225–238. doi:10.1177/1555343417695197

Endsley, M. (2017b). From here to autonomy: Lessons learned from human-automation research. *Human factors, 59,* 5–27. doi:10.1177/0018720816681350

Endsley, M. R. & Garland, D. J. (Eds.). (2000) Situation awareness analysis and measurement. Mahwah, NJ: Lawrence Erlbaum Associates. doi:10.1201/b12461

Hergeth, S., Lorenz, L., Wilimek, R., & Krems, J. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human Factors, 58,* 509–519. doi:10.1177/0018720815625744

Insurance Institute for Highway Safety. (2018). Road, track tests to help IIHS craft ratings program for driver assistance features. *Status Report, 53*(4), 3–6.

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making, 6,* 57–87. doi:10.1177/1555343411433844

Merat, N., Jamson, A., Lai, F., & Carston, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors, 54,* 762–771. doi:10.1177/0018720812442087

Reimer, B., Pettinato, A., Fridman, L., Lee, J., Mehler, B., Seppelt, B., Park, J., & Iagnemma, K. (2016). *Behavioral impact of drivers' roles in automated driving.* Automotive'UI 16: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicle Applications, Ann Arbor, MI, USA, 217–224. doi:10.1145/3003715.3005411

SAE International. (2018). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles* (SAE Standard J3016, Report No. J3016-201806). Warrendale, PA. Retrieved from https://www.sae.org/standards/content/j3016_201806/

Schmidt, E., Schrauf, M., Simon, M., Frizsche, M., Buchner, A., & Kincses, W. (2009). Drivers' misjudgment of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention, 41,* 1087–1093. doi:10.1016/j.aap.2009.06.007

Simons, D., & Charbis, C. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28,* 1059–1074. doi:10.1068/p281059