Bears in our midst: familiarity with Level 2 driving automation and situational awareness of on-road events

Main research paper

March 2021

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

Amy Benedick
Doreen De Leonardis
Rick Huey
Westat

Preprint March 2021. This paper has not been peer reviewed. Please do not copy or cite without the author's permission.

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 study presents a novel paradigm to evaluate driver situational awareness (SA) when using Level 2 (L2) driving automation. An oversized pink teddy bear was mounted to the back of a study vehicle that overtook participants three times while they drove another study vehicle, a 2019 Mercedes-Benz C300 equipped with a L2 system, for approximately 1 hour. The L2 system was turned on or off for the drive, depending on the assigned condition, and participants varied in their familiarity with L2 systems. Post-drive surveys measured SA about the bear and road by asking participants to recall the bear and the number of bear presentations as well as landmarks along the route. Cameras recorded participant eye glance behavior. Results show that the driving automation support only gave participants familiar with L2 systems an advantage for greater bear SA. Unfamiliar participants were at a disadvantage when assisted by the L2 system, having poorer bear SA compared with unfamiliar participants who drove with the system off. Better bear SA corresponded with better landmark recall and wider on-road gaze dispersion. Our findings support the effectiveness of this paradigm to measure a driver's SA of the road objectively and unobtrusively when using a L2 system under real-world conditions.

Keywords: attention; inattentional blindness; adaptive cruise control; lane centering; driver assistance; experience.

Introduction

Driver assistance systems are growing in popularity and vary considerably in how they support the driving task. Some of these technologies are designed to prevent crashes, such as forward collision warning and automatic emergency braking (Cicchino, 2017). Other systems are marketed more for driver convenience and comfort by providing sustained support for certain aspects of the driving task, such as Level 2 driving automation (SAE International, 2018). Level 2 driving automation is a partially automated system that simultaneously combines adaptive cruise control (ACC), which controls the vehicle speed and distance to a vehicle in front, and lane centering, which controls the vehicle's steering to keep it within the lane for extended periods.

Although Level 2 systems are designed to make driving less effortful, they are not capable of autonomous driving and are limited in where and how they can operate. This means that the driver is responsible for the vehicle's behavior and must monitor what the vehicle is doing and what is happening on the road. Despite reducing the physical workload of driving, this supervisory role ironically adds to the cognitive load of the driver's task (Norman, 1989), and it can become difficult for a driver to maintain the necessary vigilance over time (Endsley, 2017b; Manzey, Reichenbach, & Onnasch, 2012). One reason why reduced vigilance can be especially dangerous when using Level 2 driving automation is that current production systems can suddenly behave in ways that are unexpected to the driver (American Automobile Association [AAA], 2020; Insurance Institute for Highway Safety [IIHS], 2018), meaning the driver must be ready and able to intervene at a moment's notice. Unexpected vehicle behavior includes when it drifts out of the lane because the system cannot detect the lane lines, or when the system is tracking a lead vehicle that suddenly changes lanes to reveal a slower or stationary vehicle ahead that the system is unable to detect.

It is unclear whether Level 2 driving automation affects all drivers the same way. Situational awareness (SA) refers to a driver's ability to perceive and respond to objects and events in the environment as well as the ability to respond to how those objects and events relate to the driver's goals (Endsley, 1996). Driving experience automatizes cognitive process and behavioral routines to free cognitive resources that can be allocated elsewhere (Trick, Enns, Mills, & Vavrik, 2004). As a result, experienced drivers tend to have more efficient and effective visual information processing than novice drivers (Mueller & Trick, 2013), and a similar effect might occur as drivers gain experience with Level 2 systems. The ability to safely operate these systems requires an accurate understanding about how they operate and how drivers should use them (Seppelt & Lee, 2017, 2019), and familiarity with ACC has been shown to improve a driver's ability to intervene when the system encounters situations it cannot handle (Larsson, Kircher, & Hultgren, 2014).

Drivers who are not familiar with Level 2 systems might experience greater cognitive demand when using the technology than those who are familiar with these systems, and cognitive load tends to narrow visual scanning patterns (Wang, Reimer, Dobres, & Mehler, 2014). The cognitive demand involved with using driving automation might lead unfamiliar drivers to have fewer cognitive resources available to monitor and scan the roadway because they might devote more resources to supervising the vehicle's lane-keeping and headway-maintenance behavior. Having fewer resources available might limit a driver's awareness of driving-relevant information, such as other road users who represent latent hazards. In contrast, the driving automation's assistance should free cognitive resources among familiar drivers and lead to more dispersed visual scanning behavior and greater on-road SA.

On-road testing is necessary to understand how drivers interact with Level 2 systems in production vehicles in everyday situations because these systems perform differently on public roads than on closed-course testing (AAA, 2020; IIHS, 2018). Public roads, however, are challenging environments in which to evaluate SA. While there are numerous laboratory SA paradigms (e.g., Endsley & Garland, 2000), they tend to not transfer well to on-road testing because they often involve either priming the driver to look for stimuli of interest that will be recalled after the drive or interrupting the drive to deliver the recall task in real time. Self-report of one's own state can be an unreliable indicator of actual behavior (Schmidt et al., 2009) and likely would not accurately capture how situationally aware one is when using a Level 2 system. The manipulation used in this study was modeled after Simons and Charbis (1999), where participants watched videos of people playing basketball during which a person in a gorilla suit or with an umbrella walked through the game. SA was evaluated in terms of observers' inattentional blindness to those surprise events when asked about whether they had seen anything odd when viewing the videos.

The surprise events used in the current study were salient and relevant to the driving task. Each was delivered consistently across participants during the drive and did not interfere with the driving task. They involved an oversized pink teddy bear in a high-visibility jacket mounted to the rear of a study vehicle that was driven by a researcher. The researcher overtook the participant three times at predetermined locations during a 1-hour drive. At the end of the drive, participants were asked whether they had seen something odd and how many times, thereby providing objective measures of SA through the degree of inattentional blindness to the pink teddy bear and the number of times it appeared. This study used a small sample and a 2019 Mercedes-Benz C300 that was equipped with a Level 2 system to present a proof of concept for the application of this paradigm. Experimental groups were established according to participant familiarity with Level 2 driving automation and whether they drove with the study vehicle's Level 2 system on or off.

Method

Sample

There were 31 participants and three experimental groups that varied according to familiarity with a Level 2 system and driving condition (the study vehicle with the Level 2 system on or off). The unfamiliar or inexperienced users with the Level 2 system off, herein called the inexp-L2-off group (n = 10, 5 male and 5 female), were on average 42 years old (SD = 7.7), had 24 years of driving experience (SD = 8.0), and drove 6 days a week (SD = 1.0). Unfamiliar users with the Level 2 system on (inexp-L2-on) (n = 10, 5 male and 5 female) were on average 47 years old (SD = 14.4), had 31 years of driving experience (SD = 14.6), and drove 6 days a week (SD = 0.9). Familiar users with the Level 2 system on (exp-L2-on) (n = 11, 6 male and 5 female) were on average 43 years old (SD = 7.8), had 24 years of driving experience (SD = 8.7), and drove 6 days a week (SD = 1.4).

We determined familiarity with Level 2 driving automation based on whether participants owned or had regular access to a vehicle with the system. Those who reported owning or having regular access to such vehicles were interviewed in detail to ensure that their vehicles were equipped with both ACC and lane centering and that they regularly used both systems when driving. Experienced Level 2 system users reported using the ACC and lane centering systems on average 4 days a week (SD = 2.0 and 2.2, respectively). Video collected from every participant's drive in this study was reviewed to ensure those in the Level 2 system-on conditions used the study vehicle's system for the majority of the drive and those in the Level 2 system-off condition never turned the study vehicle's system on.

Using advertisements on social media, Craigslist, and Westat's intranet site open to friends and family of staff, participants were recruited from the general population. They provided informed consent and were paid between $100 and $300 for participating (payment was increased in the latter half of the study to encourage recruitment). A total of 43 participants had been recruited to participate, but 12 were removed because of the following reasons: the driver frequently turned off the Level 2 system when they were in the Level 2 system-on condition (n = 5), the bear reveals were unacceptable or unable to be performed due to traffic conditions or participant behavior (n = 2), the participant saw the bear before the start of the drive (n = 1), the bear reveal ran over into a baseline epoch (n = 2) (for information on baseline epochs, see *Video epochs* under *Materials*), traffic congestion encountered during the drive prevented bear presentations (n = 1), and inclement weather began during the drive (n = 1). The study was approved by Westat's Institutional Review Board (IRB).

Materials

Vehicles. The study vehicle was a 2019 Mercedes-Benz C300 equipped with a Level 2 system. It was instrumented with four cameras (Figure 1) that recorded (1) the driver's face; (2) the forward roadway; (3) the instrument cluster, to ensure the driver had the appropriate systems enabled or disabled; and (4) the driver and front passenger seat. Video from the four cameras was collected through a mosaic box and stored in a high-definition (HD) video recorder. In order to present the bear at predetermined locations, the C300 and the other study vehicle, herein called the bear vehicle, had computers and cellular modems to collect and share GPS information.

Figure 1. Still frames depicting the camera angles used to capture the driver's face, forward roadway, instrument cluster, and front seats. Top right frame shows the bear vehicle during a reveal from the participant's point of view.

The primary subject of recall for the situational awareness measurement was the oversized pink teddy bear wearing a high-visibility jacket that was mounted to the rear of the bear vehicle (see Figure 2). The driver of the bear vehicle monitored the participant's movement along the route to appropriately time the bear reveals via a large touchscreen tablet that was mounted in the cockpit.

Figure 2. Teddy bear stimulus

Video epochs. The GPS and video data were integrated to allow for video segments to be extracted at fixed locations throughout the drive. Those video segments are herein called baseline epochs, as they collected baseline driver behavior outside of the bear-reveal presentations. Twenty-three 30-second baseline epochs were extracted from each participant's drive. After every drive, the video footage was reviewed to ensure the vehicle systems were used as instructed according to the participant's experimental condition.

An in-house app integrated the GPS information from both study vehicles to provide prompts to the driver of the bear vehicle through a tablet about the participant's location, including real-time proximity to the bear vehicle and the next presentation location, as well as alerts for when the bear reveals should occur. Using the app, the bear vehicle's driver identified where each bear reveal began and when it ended. The bear-reveal portions of the drive resulted in three epochs that were distinct from the baseline epochs, and the two types of epochs never overlapped. Every bear-reveal epoch was reviewed after the drive to ensure the presentation met study inclusion criteria, as discussed below.

Surveys. In-person surveys were delivered after the drive. SA of the bear was measured by first asking participants if they had seen anything odd about any of the vehicles in front of them during their drive. Response accuracy and detail varied across participants. If participants provided vague or incorrect responses for the bear-identification question (incorrect responses included participants saying they had not seen anything odd), they were prompted by the researcher to recall whether they had seen anything odd about the appearance of the back of any vehicles they might have encountered. If participants identified the bear or a toy on a vehicle, they were considered to have correctly recalled the bear. Participants who correctly identified the stimulus were then asked how many times they had seen it during their drive. Only participants who answered that they had seen it three times were considered to have correctly recalled the number of presentations.

Road environment SA was measured by asking participants whether they had seen certain places or signs along the study route. The survey asked about ten landmarks (e.g., weigh stations, fairgrounds, golf course), five of which the participants encountered along their drives and five that were not present on the route. Participants were also asked how familiar they were with the route using a 5-point Likert scale (responses ranged from a value of 1 = not at all, 2 = slightly, 3 = somewhat, 4 = moderately, and 5 = extremely familiar).

Procedure

Participants met the researcher at Urbana Park and Ride off Interstate 270 in Maryland and provided written informed consent. They were informed about the recording equipment in the vehicle and signed a video and photo release. The actual 1-hour experimental drive took place on Interstate 70 in Maryland, and the speed limit was 70 mph. Interstate 270 was selected for the practice drive because it was similar in road geometry to I-70, and closer in proximity to the meeting location, which reduced the overall length of the study session. The practice drive enabled the participants to familiarize themselves with the vehicle and its various systems.

Orientations. Once in the driver's seat of the study vehicle, every participant was given verbal and video orientations to the functionality and operation of the vehicle's infotainment system and Bluetooth. Participants were shown how to use the infotainment system and the Bluetooth connection for their smartphones and were encouraged to connect them, but they were instructed not to modify any other vehicle settings. If participants were in an experimental condition with the system on, the researcher gave detailed verbal and video orientations to the vehicle's Level 2 system functionality and how to operate it. Only participants in the system-on conditions were shown and permitted to modify the Level 2 system settings with respect to adjusting the ACC set speed and headway distance. All other system settings were kept constant across participants. Participants in the system-on conditions were instructed to use the Level 2 system as often as possible while on the interstate, which was within the intended design domain of the C300's Level 2 system, as long as they felt it was safe to do so.

Before the start of a practice drive, participants were instructed to look to different areas of the cabin to establish glance areas of interest (AOI) to assist with eye glance coding for analysis. These locations included the left and right side mirrors, the rearview mirror, straight ahead, the instrument cluster, and the center stack where the infotainment screen was located. Participants in the system-on conditions were informed that the Level 2 system was not a fully autonomous or self-driving system, and therefore, they were always responsible for the vehicle's behavior, even when the system was on. All participants were instructed to drive as they normally would while obeying traffic laws. The vehicle's GPS system with preprogrammed waypoints guided participants along the route, and they were also given a sheet with a map and route instructions as a backup.

Bear presentations. In order for the bear vehicle to perform the necessary overtaking maneuvers to present the bear, participants were instructed to drive in the middle lane when on three-lane portions of the interstate or to drive in the right lane when on the two-lane segments. The bear vehicle would perform the overtaking maneuvers at three predetermined locations, whereby it would overtake the participant on the left and remain in view, two to three vehicle lengths ahead, for approximately 30 seconds, after which it would fall back behind the participant out of view until the next presentation or the end of the study. If traffic conditions or participant behavior prevented the bear vehicle from falling back in the same lane, it moved in front of the participant to fall back in the other free lane (this happened in seven out of 93 trials, and never more than once per participant). Bear reveals were only considered to be successful if the presentation allowed for the bear to be directly visible to the participant for approximately 30 seconds. The presentation was determined to be a highly salient and long enough event for any driver who was paying attention to the road to detect the bear. Only participants who experienced three successful bear reveals were included in the final sample ($N = 31$).

Practice and experimental drives. Drives were conducted during clear weather and when traffic was light and free flowing. Every participant completed a 15-minute practice drive before the 1-hour experimental drive. During the practice drive, the participant drove with the researcher in the front passenger seat, who answered any questions and helped to ensure safe and appropriate use of the vehicle systems. Participants in the system-on conditions were encouraged to use the Level 2 system and to modify its set speed and headway distance to familiarize themselves with the controls. The participant then dropped the researcher off at the park-and-ride to begin the experimental drive alone, at which point the researcher notified the bear vehicle's driver to start following the participant for the bear reveals.

The researcher administered the surveys after participants returned from the drive, followed by the study debrief. For response consistency and completeness, the researcher guided participants through the surveys and entered the participant responses for each online survey using REDCap (Harris et al., 2009; Harris et al., 2019) through a tablet that was connected to a portable Wi-Fi unit.

Analysis

SA surveys. As this was a case study to establish a proof of concept for the paradigm with on-road testing, the sample was small, which limited statistical testing. SA of the bear was measured by the number of participants who correctly recalled the bear and, of those who correctly recalled the bear, how many correctly recalled the number of presentations as a function of experimental group. SA of the road environment through the landmark recall survey involved tallying responses to provide the percentage-correct score. The data were explored according to bear SA accuracy by separating the sample into three groups: participants who correctly identified the bear and recalled the number of presentations (i.e., all correct); those who correctly identified the bear but did not correctly recall the number of presentations (i.e., partially correct); and those who did not correctly identify the bear, and therefore, were never asked to recall the number of presentations (i.e., none correct).

Video coding. Eye glance behavior was coded from the video data using Morae Manager (TechSmith, Okemos, MI, USA). One coder coded eye glance frame by frame for every epoch for the whole sample. For six participants, all the bear-reveal and baseline epochs were reviewed by a second coder. For the remaining sample, the second coder reviewed all three bear-reveal epochs and seven out of 23 randomly selected baseline epochs per participant. Coding disagreements resulted in another round of coding for the affected epochs by the first and second coders, independently. A third coder reviewed the whole dataset for the final sample.

Gaze patterns. One of this study's hypotheses was that gaze dispersion should be related to SA. The glance data were explored with respect to the bear situational data by grouping participants based on their bear-presentation recall accuracy. Bear-reveal epochs varied in duration across trials and participants (m = 35.7 seconds, SD = 8.0, min = 20.4, max = 62.0), limiting the ability to investigate the average frequency and total duration of glances to AOIs among the bear-reveal epochs; however, this was not an issue for the baseline epochs as they were all 30 seconds. Gaze behavior was examined in terms of the average percentage of time spent looking to an AOI per epoch for bear-reveal and baseline epochs, separately. Average glance frequency and the total duration of glances to an AOI per epoch were investigated for the baseline epoch data only.

On-road AOIs were defined as the forward center, forward periphery, side, and to the rearview mirror. The location of the rearview mirror served as the separation between the forward center and periphery AOIs. On-road center glances were defined as glances in the area from the rearview mirror (but not including the rearview mirror) to the driver's left forward field of view while looking ahead. On-road periphery glances were characterized as glances in the area from the rearview mirror (but not including the rearview mirror) to the driver's right forward field of view while looking ahead. Side glances were defined as glances when the driver looked to the extreme left or right, including over-the-shoulder checks and glances to the left and right side mirrors. The analysis combined forward periphery and side glances. Off-road AOIs were the instrument cluster, center stack, and "off-road other" (i.e., center console, smartphone, rear passenger seat, ceiling, door controls, map, front passenger seat, food, and driver's watch).

Results

Situational awareness

Bear. There was an interaction between familiarity with Level 2 driving automation and system activity, as shown in Figure 3. Inexperienced Level 2 users were at a disadvantage when using the Level 2 system as compared with inexperienced users who drove with the system turned off. Experienced users with the Level 2 system on seemed to have an advantage over both inexperienced groups. Specifically, almost all of the exp-L2-on participants correctly identified the bear, followed by inexp-L2-off participants and least of all inexp-L2-on participants, respectively. Of those who correctly identified the bear, more exp-L2-on participants correctly recalled the number of times the bear had been presented than either of the inexperienced groups.

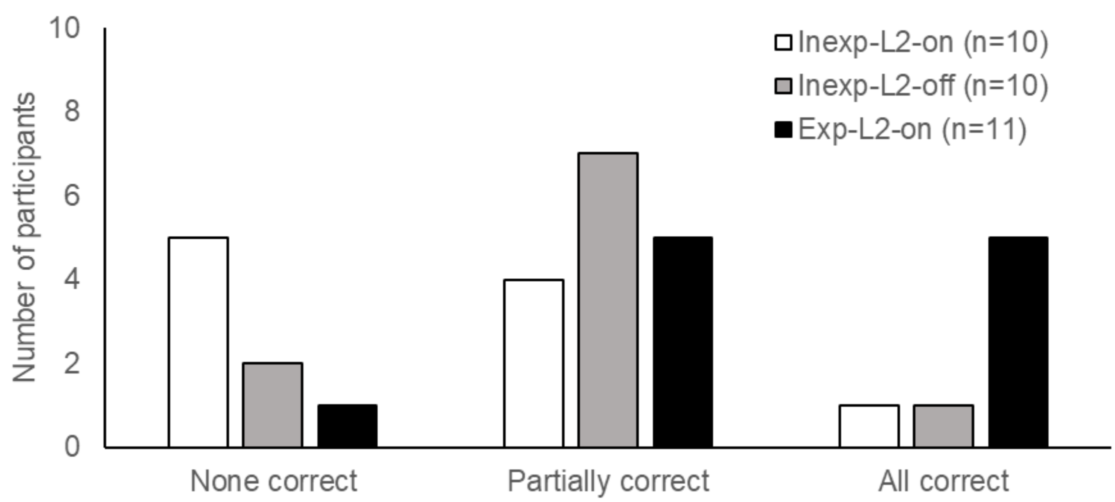

Figure 3. Number of participants who were correct on all (bear recall and number of presentations), some (correct bear recall but incorrect number of presentations; partially correct), and none of the bear SA questions (incorrect bear recall) as a function of experience with Level 2 driving automation and system activity.

Landmarks. Both groups with the Level 2 system on scored, on average, higher on landmark recall (inexp-L2-on *m* score = 75%, *SD* = 19; exp-L2-on *m* score = 73%, *SD* = 11) than the group with the system off (inexp-L2-off *m* score = 61%, *SD* = 12); however, these data should be interpreted with caution due to a confound in route familiarity between groups. Twice as many inexp-L2-on participants (80%) reported being moderately or extremely familiar with the route (Likert scores of 4 and 5, respectively) than those participants in the inexp-L2-off (40%) and exp-L2-on groups (45%). Even so, the landmark recall data aligned with the bear SA data. Those participants who responded correctly to all the bear SA questions had higher landmark recall scores (*n* = 7, *m* = 76%, *SD* = 15) than those who were only partially correct (*n* = 16, *m* = 70%, *SD* = 16). Participants with no correct bear SA responses had on average the lowest landmark recall scores (*n* = 8, *m* = 64%, *SD* = 12). Moreover, participants who demonstrated greater SA of the bear were not more familiar with the route than those who showed less SA. Participants who answered all the bear SA questions correctly reported an average value of 3.6 (i.e., between somewhat and moderately familiar with the route; *SD* = 1.3), those who were only partially correct reported an average value of 3.4 (*SD* = 1.3), and those who were not at all correct reported an average value of 4 (i.e., moderately familiar with the route; *SD* = 0.9).

Glance behavior

Bear-reveal epochs. As shown in Table 1, participants who correctly answered all the bear SA questions, on average, showed the widest on-road gaze dispersion with the greatest percentage of time spent looking to the forward periphery and side as well as to the rearview mirror per bear-reveal epoch. Partially correct participants spent more time looking at the forward periphery and side and a similar amount of time looking to the rearview mirror compared with participants who failed to identify the bear. Participants who were not at all correct spent the greatest percentage of time, on average, looking at the center stack and instrument cluster but the least amount of time looking elsewhere as captured in the “other” category per epoch.

Table 1. Mean percentage of time spent looking to areas of interest per bear-reveal epoch and baseline epoch, separately, as a function of bear identification and presentation number recall accuracy

		None correct	Partially correct	All correct
		M % (*SD*)	*M* % (*SD*)	*M* % (*SD*)
Bear-reveal epochs	Forward periphery + side	5.7 (7.5)	7.4 (3.3)	12.7 (6.3)
	Forward center	76.5 (7.4)	79.6 (10.4)	69.7 (13.2)
	Rearview mirror	2.4 (3.1)	2.5 (2.0)	3.3 (3.8)
	Instrument cluster	7.0 (5.8)	5.0 (3.2)	6.0 (3.9)
	Center stack	8.2 (7.4)	5.4 (8.6)	5.3 (5.1)
	Other	0.2 (0.6)	0.2 (0.4)	2.9 (7.2)
Baseline epochs	Forward periphery + side	3.5 (1.7)	5.4 (2.7)	6.2 (3.2)
	Forward center	76.9 (6.2)	73.5 (7.4)	71.2 (7.6)
	Rearview mirror	2.6 (1.2)	4.1 (2.8)	4.3 (1.8)
	Instrument cluster	7.8 (3.2)	6.9 (2.9)	9.6 (4.0)
	Center stack	7.6 (3.6)	9.6 (3.1)	7.7 (3.0)
	Other	1.6 (2.3)	0.6 (1.2)	0.9 (1.0)

Note. All correct: correct bear recall as well as the number of presentations; partially correct: correct bear recall but incorrect recall of the number of presentations; none correct: incorrect bear recall.

Baseline epochs. Participants displayed similar, albeit weaker, on-road gaze dispersion patterns in the baseline epochs: SA of the bear corresponded with a greater percentage of time looking at the forward periphery and side per epoch. The same pattern was observed in the average frequency and total duration of glances per baseline epoch, as shown in Table 2. Moreover, regardless of presentation number recall accuracy, participants who correctly identified the bear spent more time looking at the rearview mirror than incorrect participants. Poorer SA of the bear corresponded with longer total duration of glances (Table 2) and a greater percentage of time looking to the forward center (Table 1).

Table 2. Mean frequency and total duration of glances (seconds) per baseline epoch as a function of bear identification and recall accuracy

		None correct	Partially correct	All correct
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
Frequency of glances per epoch	Forward periphery + side	1.2 (0.5)	1.8 (0.9)	2.1 (1.4)
	Forward center	8.0 (1.9)	9.0 (3.3)	9.0 (2.6)
	Rearview mirror	1.1 (0.5)	1.8 (1.5)	1.8 (0.8)
	Instrument cluster	2.8 (1.3)	2.5 (1.3)	3.1 (0.9)
	Center stack	2.1 (1.1)	2.7 (1.0)	2.0 (0.8)
	Other	0.4 (0.7)	0.2 (0.3)	0.2 (0.1)
Total duration of glances per epoch	Forward periphery + side	1.1 (0.5)	1.6 (0.8)	1.9 (1.0)
	Forward center	23.1 (1.9)	22.1 (2.2)	21.4 (2.3)
	Rearview mirror	0.8 (0.4)	1.2 (0.8)	1.3 (0.5)
	Instrument cluster	2.3 (0.9)	2.1 (0.9)	2.9 (1.2)
	Center stack	2.3 (1.1)	2.9 (0.9)	2.3 (0.9)
	Other	0.5 (0.7)	0.2 (0.3)	0.3 (0.3)

Note. All correct: correct bear recall as well as the number of presentations; partially correct: correct bear recall but incorrect recall of number of presentations; none correct: incorrect bear recall.

Discussion

The findings from this study support the utility of this paradigm for objectively evaluating on-road SA when using Level 2 driving automation under real-world conditions in a manner that does not interfere with the drive. Our data suggest that Level 2 driving automation has the potential to improve a driver's SA once he or she is familiar with the technology, although it does not guarantee it. Unfamiliar drivers, however, appear to have even more difficulty maintaining SA when using the system than when driving without it. On average, participants who were familiar with Level 2 systems showed the highest degree of SA about the bear when using the system, unfamiliar participants who drove with the system off had moderate SA, and unfamiliar participants who drove with the system on demonstrated the lowest SA. We also found support for the bear manipulation as a way to measure a driver's SA of the road through our sample's performance on the landmark recall task, as participants who had greater

SA of the bear tended to likewise be more accurate in their recall of the landmarks they had encountered along the route. Landmark recall might be a useful measure of SA in future research, as long as consideration is given to a driver's familiarity with the route.

As this study represents a proof of concept for the paradigm, our examination into the data was limited by the small sample size. Nevertheless, our sample was comprised of drivers with similar levels of general driving experience (years licensed) and driving exposure (number of days per week that they drive); therefore, the effect of familiarity on SA observed in the current study appears to be specific to their experience with Level 2 driving automation. If mechanisms of controlled versus automatized processing (Trick, Enns, Mills, & Vavrik, 2004) help to explain the effect of familiarity with the driving automation on SA, familiar drivers may have more cognitive resources available when supported by the driving automation, hence the better SA of the bear among familiar drivers. Conversely, a lack of familiarity may increase cognitive demand when using the system, which would help to explain the poorer SA of the bear among unfamiliar drivers. As Level 2 driving automation gradually proliferates the registered vehicle fleet, there will be a growing number of new users who could be at risk for decreased SA when learning how to use these systems.

The relationship between cognitive demand when driving and driving experience (Patten, Kicher, Östlund, Nilsson, & Svenson, 2006) also corresponds with narrower visual scanning strategies (Underwood, 2007) and inattentional blindness to driving relevant information (Kass, Cole, & Stanny, 2007; Mueller & Trick, 2013). Drivers in the current study who demonstrated better bear SA spent a greater percentage of time looking to the forward periphery, the side, and the rearview mirror, which indicates more active, dispersed scanning strategies. This was especially true in the bear-reveal epochs, because in those epochs a highly conspicuous road user was consistently presented to every participant in the visual forward periphery and side (i.e., during the overtake maneuver), meaning that anyone who was paying attention to the road would have spent time looking at those areas of interest. In comparison, participants with poorer bear SA spent a larger percentage of time looking at the forward center. Gaze concentration under higher cognitive load corresponds with poorer hazard perception (Savage, Potter, & Tatler, 2013); however, while drivers tend to concentrate their gaze in the center of the roadway when cognitive demand increases (Harbluk, Noy, Trbovich, & Eizenman, 2007), they also tend to do so when mind wandering (He, Becic, Lee, & McCarley, 2011). Therefore, further study is necessary to determine the nature of the relationship between SA and gaze dispersion in the context of using Level 2 systems.

Despite our data showing a benefit of Level 2 driving automation familiarity to enhance SA when using the system, familiarity with this technology can also have drawbacks. During a 6-month case study where Endsley (2017a) drove a Tesla Model S equipped with Autopilot, she reported having more cognitive resources available and greater SA of the road when using the Level 2 system, but she also described the fairly reliable system as an “enabler” of poor behind-the-wheel behavior. Secondary task activity appears to be more frequent among drivers who are familiar with ACC when using either it or a Level 2 system (Naujoks, Purucker, & Neukum, 2016), and distraction can lead to a greater impairment of SA when using driving automation compared with when driving manually without assistance (de Winter et al, Happee, Martens, & Stanton, 2014). Driver distraction and improper use of Level 2 driving automation have already been implicated in serious crashes (National Transportation Safety Board, 2017, 2019, 2020). It remains imperative that robust safeguards be implemented into Level 2 systems to keep drivers in the loop (Mueller, Reagan, & Cicchino, 2021). By keeping driver disengagement to a minimum, driver monitoring and management systems might be able to facilitate the benefits of familiarity with the technology to improve a driver's SA when supported by Level 2 driving automation.

Conclusions

This study demonstrates the application of a novel method for evaluating situational awareness over a single drive when using Level 2 driving automation in production vehicles. Our finding of an effect of driver familiarity with the technology on inattentional blindness to the road demonstrates the need for future research to account for its influence on driver interactions with Level 2 systems.

Acknowledgements

This study was supported by the Insurance Institute for Highway Safety. Thank you to Michael Gill, Jeremy Walrath, Jazzmyne Sangster, Elisha Lubar, Chavez Lee, and Matthew Airola for their help with pilot testing, data collection, data reduction, and video coding. We are grateful to Eric Teoh for his help with data analysis and to David Zuby, Lana Trick, Bobbie Seppelt, Ensar Becic, and Trent Victor for their advice on study design and data interpretation. Inspiration for the bear is thanks to Jeremy Clarkson from the BBC's Top Gear.

References

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

Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis & Prevention, 99,* 142–152. doi:10.1016/j.aap.2016.11.009

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 Research Part F, 27,* 196–217. doi:10.1016/j.trf.2014.06.016

Endsley, M. R. (1996). *Automation and situation awareness*. In R. Parasuraman& M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 163–181). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. (2017a). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making, 11*, 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., & Garland D.(Eds.) (2000) *Situation awareness analysis and measurement.* Mahwah, NJ: Lawrence Erlbaum Associates.

Harris, P., Taylor, R., Minor, B., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software partners. *Journal of Biomedical Information, 95,* 103208. doi:10.1016/j.jbi.2019.103208

Harris, P, Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Information, 42*(2), 377–381. doi:10.1016/j.jbi.2008.08.010

Harbluk, J., Noy, Y., Trbovich, P., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention, 39*, 372–379. doi:10.1016/j.aap.2006.08.013

He, J., Becic, E., Lee, Y. C., & McCarley, J. S. (2011). Mind wandering behind the wheel: Performance and oculomotor correlates. *Human Factors,* 53, 13–21. doi:10.1177/0018720810391530

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.

Kass, S., Cole, K., & Stanny, C. (2007). Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F, 10,* 321–329. doi:10.1016/j.trf.2006.12.002

Larsson, A. F., Kircher, K., & Hultgren, J. A. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transportation Research Part F, 27*, 229–237. doi:10.1016/j.trf.2014.05.008

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

Mueller, A. S., Reagan, I. J., & Cicchino, J. B. (2021). Addressing driver disengagement and proper system use: human factors recommendations for Level 2 driving automation design. *Journal of Cognitive Engineering and Decision Making*. doi:10.1177/1555343420983126

Mueller, A. S. & Trick, L. M.. (2013). Effect of Driving Experience on Change Detection Based on Target Relevance and Size. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* Bolton Landing, New York. Iowa City, IA: Public Policy Center, University of Iowa, 2013: 341–347.

National Transportation Safety Board. (2017). *Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston, Florida, May 7, 2016* (Highway Accident Report NTSB/HAR-17/02). Washington, DC.

National Transportation Safety Board. (2019). *Collision between car operating with partial driving automation and truck-tractor semitrailer: Delray Beach, Florida, March 1, 2019* (Highway Accident Brief NTSB/HAB-20/01). Washington, DC.

National Transportation Safety Board. (2020). *Collision between a sport utility vehicle operating with partial driving automation and a crash attenuator: Mountain View, California, March 23, 2018* (Highway Accident Report NTSB/HAR-20/01). Washington, DC.

Naujoks, F., Purucker, C., & Neukum, A. (2016). Secondary task engagement and vehicle automation—Comparing the effects of different automation levels in an on-road experiment. *Transportation Research Part F, 38*, 67–82.

Norman, D. A. (1989). The “problem” of automation: Inappropriate feedback and interaction, not “overautomation” (ICS Report 8904). *Institute for Cognitive Science*, University of California at San Diego, La Jolla, CA.

Patten, C., Kircher, A., Östlund, J., Nilsson, L, & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis & Prevention, 38*, 887–894. doi:10.1016/j.aap.2006.02.014

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/

Savage, S., Potter, D., & Tatler, B. (2013). Does preoccupation impair hazard perception? A simultaneous EEG and eye tracking study. *Tranportation Research Part F, 17*, 52–62. doi:10.1016/j.trf.2012.10.002

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

Seppelt, B. D., & Lee, J. D. (2017). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies, 65,* 192–205. doi:10.1016/j.ijhcs.2006.10.001

Seppelt, B. D., & Lee, J. D. (2019). Keeping the driver in the loop: Dynamic feedback to support appropriate use of imperfect vehicle control automation. *International Journal of Human-Computer Studies, 125*, 66–80. doi:10.1016/j.ijhcs.2018.12.009

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

Trick, L. M., Enns, J. T., Mills, J., & Vavrik, J. (2004). Paying attention behind the wheel: A framework for studying the role of attention in driving. *Theoretical Issues in Ergonomics Science, 5*, 385–424. doi:10.1080/1463922042331298938

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics, 50*, 1235–1249. doi:10.1080/00140130701318707

Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing driver's gaze concentration under increase cognitive demand. *Transportation Research Part F, 26*, 227–237. doi:10.1016/j.trf.2014.08.003