ARTICLE POST-PRINT

TRANSPORTATION RESEARCH RECORD

Driver Takeover Performance and Monitoring Behaviour with Driving Automation at System-Limit versus System-Malfunction Failures

Chelsea A. DeGuzman
Department of Mechanical and Industrial Engineering
University of Toronto, Toronto, Canada, M5S 3G8

Samantha A. Hopkins
Department of Mechanical and Industrial Engineering
University of Toronto, Toronto, Canada, M5S 3G8

Birsen Donmez
Department of Mechanical and Industrial Engineering
University of Toronto, Toronto, Canada, M5S 3G8

ABSTRACT

Today's vehicles are becoming highly automated, however, if the automation fails, drivers must take over control of the vehicle. Automation may fail due to known system limits (system-limit failure) or due to malfunctions that are unforeseen by system designers (system-malfunction failure). The aim of this research is to quantify the differences between how these two failure types influence driver takeover performance and monitoring behaviours. In a simulator with SAE Level 2 driving automation, 18 drivers experienced both a system-limit and system-malfunction failure while engaging in a secondary task. Results show that drivers put their hands on the wheel 0.62 seconds sooner and took over 0.51 seconds faster for the system-limit failure compared to the system-malfunction failure. Eye tracking data revealed that the percent of time looking at the secondary task display was 12.7% lower and the percent of time looking at the roadway was 11.2% higher before the system-limit failure compared to before the system-malfunction failure. Given that takeover performance and monitoring behavior differ significantly based on failure type, a distinction should be made in the literature between system-limit and system-malfunction failures, and comparisons between previous studies using these failures should not be done without considering this distinction. Furthermore, as Level 2 vehicles are currently available to consumers, efforts should be focused on supporting drivers' mental models of automated systems, so that drivers are able to successfully predict system-limit failures.

Keywords: Automated vehicle, driving simulator, automation failure, SAE Level 2

INTRODUCTION

With automated vehicles available for public purchase, research on how drivers interact with driving automation is becoming increasingly important. Based on the SAE taxonomy of driving automation (*1*), the highest level of driving automation currently available for purchase is Level 2 (e.g., Tesla Autopilot, Cadillac Super Cruise). In SAE Level 2 (L2) driving automation, the automation can control longitudinal and lateral movement of the vehicle through the use of adaptive cruise control (ACC) and lane keeping (LK) systems, respectively, but the driver is still responsible for monitoring the environment and the automation at all times to determine when it is necessary to resume manual control (*1*). ACC works similar to cruise control, but has the added benefit of being able to maintain a headway gap between the ego-vehicle and the lead vehicle. LK controls the lateral motion of the vehicle by maintaining control of the vehicle's steering. Current L2 systems are not perfect, with many limitations listed in the owner's manuals. For example, missing or worn lane markings and limited visibility due to poor weather (e.g., snow, fog) are listed as circumstances that can result in the automation failing to properly control the vehicle's lateral and/or longitudinal movement (e.g., *2*, *3*) and that the drivers must take over control whenever the automation fails.

The media has reported numerous instances of L2 systems failing and drivers unable to takeover in time to avoid a collision. For example, a highly publicized crash involved an Uber vehicle being controlled by the automated system that mistakenly classified a pedestrian first as an unknown object, then a vehicle, then a bicycle, before finally determining 1.3 seconds before impact that an emergency braking maneouvre was necessary (*4*). There have also been several instances of Tesla vehicles with Autopilot engaged colliding with stopped emergency vehicles, a well-known limitation of ACC systems (*5*). In general, failures of these automated driving systems can be due to known system limits (e.g., Tesla manual states that the system may have problems recognizing stationary vehicles), but they can also be due to system malfunctions unforeseen by system designers (e.g., an algorithmic error such as the one in the Uber crash; many argued that the Uber crash should have been an easy scenario for automation to handle given the current know-how). If drivers are informed about known system limitations then they can predict related automation failures based on their knowledge of the automated system and cues in the environment (i.e., the driver could look at the road and see that there is a situation that the automation would not be able to handle). However, for system malfunctions that are unforeseen even by their own designers, the failure would happen with no indicators that a driver can use to prepare for it.

Researchers generally study driver behavior when encountering either system-limit failures or system-malfunction failures, with very limited research including both failure types to compare how driver behavior differs between them. Helldin et al. (*6*) examined driver takeover performance and monitoring when encountering a system-limit failure associated with weather conditions. The authors used scenarios with heavy snow, which participants were told could impact the automation's performance. However, the purpose of their study was to compare a baseline condition (L2 driving automation with no display) to a display condition in which participants were given continuous information about the automation's uncertainty regarding its ability to drive based on the weather. While in the Helldin et al. (*6*) study, the participants could understand why the automation might have failed, other studies use system-malfunction failures, where drivers are presented with automation failures that occur for seemingly no reason. In one study with a lead vehicle braking, ACC either did not brake or did not brake at full capacity, with no observable reason for these failures (*7*). The authors were investigating how having both ACC and LK engaged (i.e., L2 driving automation) versus only ACC engaged would impact driver reaction to automation failures. Beller et al. (*8*) used scenarios in which the ACC would brake when it was not necessary or fail to brake when braking was required seemingly for no reason (i.e., a malfunction failure). Similar to Helldin et al. (*6*), this study had two groups, one that received no uncertainty information and another that would be notified via a display of the automation's uncertainty. Beller et al. (*8*) also included the automation's performance (reliable versus unreliable) as a within-subjects factor to examine the relationship between uncertainty information and automation reliability on driver performance. Although these studies are informative, they do not allow a comparison of how driver performance differs between failure types.

To the best of our knowledge, there is only one study investigating the impact of these different failure types on takeover performance and monitoring behavior. Dogan et al. (*9*) told their participants that the automation would fail when travelling over 50 km/h, and then exposed them to automation failures at 50 km/h (system-limit) and 30 km/h (system-malfunction). Drivers looked more at the speedometer when they were approaching a speed of 50 km/h, presumably because they were anticipating the failure. While this effect on monitoring was significant, there was no effect of failure type on how quickly participants took over control from the automation, which was likely due to the fact that they were alerted of the need to take over for both failure types via a takeover request. Seppelt and Lee (*10*) found that a better understanding of driving automation's limitations was associated with quicker responses to system-limit failures. These results would suggest that drivers might takeover quicker when they can predict a failure. Given that drivers can predict system-limit failures but not system-malfunction failures, we would thus expect a faster takeover for system-limit failures. While the Seppelt and Lee (*10*) study can provide insights into the differences between system-limit and system-malfunction failures, it did not explicitly aim to compare these two failure types as they only used system-limit failures.

Different system failure types should be treated differently in the literature, and studies that use system-limit failures should not be compared directly to those that use system-malfunction failures. It seems obvious that drivers will react differently when experiencing these different failure types, as they can anticipate and prepare for a system-limit failure, while this is not possible for a system-malfunction failure. However, there is only one study to date that we know of that investigates the extent of this difference, and this study focused on cues within the vehicle (e.g., the speedometer) as opposed to those in the driving environment, and used takeover requests, so drivers were alerted of the need to take over (*1*). Thus, we conducted a driving simulator study to quantify the differences between L2 driving automation system-limit and system-malfunction failures in terms of driver takeover performance and monitoring, specifically for system-limit failures that are predictable based on cues in the driving environment and when drivers must decide on their own when to take over. A recent survey found that in a sample of Tesla Autopilot users, 62% of participants experienced at least one "unexpected or unusual" event while using Autopilot (*11*). Most of the reported events were due to failures of the lane detection, which may have been predictable had the drivers known the limits of the lane detection system. Therefore, in this study, we focused on failures with the LK system. It was hypothesized that for system-limit failures, drivers would have improved monitoring (i.e., looking more on road and less at a secondary task) and in turn better takeover performance (e.g., take over sooner, smaller steering movements).

METHODS

The experiment reported here is part of a larger study investigating the effects of informational displays that alert drivers when a takeover is required during system-limit and system-malfunction failures. This paper only reports the data of participants who experienced a baseline condition (i.e., no informational display) first. That is, the participants who are included in the current analysis did not experience display conditions before completing their baseline drives. There were two baseline drives, one with system-limit failures and one with system-malfunction failures; the order of the drives was counterbalanced across participants. Our analysis reported in this paper focuses on the first failure event experienced by participants in each drive (i.e., the first system-limit and first system-malfunction event). Therefore, the independent variable of this analysis is within-subjects and has two levels: whether the failure experienced was due to a system limitation or a system malfunction. All procedures were approved by the University of Toronto's Research Ethics Board (protocol # 00035693).

Participants

18 participants (10 males, 8 females) were included in this preliminary analysis. Participants were between the ages of 25 and 30 (Mean = 27.1, Standard Deviation = 2.2), had a valid full Canadian driver's license or equivalent for at least 2 years, drove at least several times a month, and had normal or corrected-to-normal vision. Participants were selected based on their responses to a screening questionnaire, with the intent of recruiting participants who were not prone to simulator sickness and who had no previous experience driving with ACC or LK systems. Participants were compensated at a rate of $14/hour plus an $8 bonus. Participants were told that the bonus was based on secondary task performance to incentivize secondary task engagement; however, each participant received the full bonus.

Apparatus

The study was conducted using a National Advanced Driving Simulator (NADS) MiniSim driving simulator. This fixed-base simulator has three 42" widescreen displays, creating a 130° horizontal and 24° vertical field of view at a 48" viewing distance. The simulated driving experiment was developed using the MiniSim Software Suite. The road network was created using a roadmap provided by MiniSim, and the driving scenarios were created using the Interactive Scenario Authoring Tool. The simulator collects driving measures at 60 Hz.

A secondary task (described further in the Procedures section) was displayed on a Microsoft Surface Pro 2 tablet positioned to the right of the dashboard where it would not be visually obstructed by the steering wheel and a head-mounted Dikablis Glasses 3 eye tracker was used to collect gaze data (**Figure 1**). Seven areas of interest (AOIs) were created: Left Screen, Front Screen, Right Screen, Rear View Mirror, Secondary Task, Dashboard, and Steering Wheel. However, results are only reported for the front screen (roadway) and secondary task as it was found that participants looked mainly in these two areas: as reported in more detail later, 93% of the time before the system-limit failure and 94% before the system-malfunction failure.

FIGURE 1: Simulator set up with Surface Pro 2 tablet (1) for the secondary task, and a Dikablis eyetracker (2)

Procedures

Upon arrival, participants were given an introduction to the experiment and completed an informed consent process. Participants then performed a practice drive to acquaint themselves to manual driving in the simulator. In order to guarantee that all participants had the same knowledge of automated driving prior to the experimental drives, they were trained on how to use the ACC and LK systems as well as on the limits of these systems to give them the knowledge to recognize system-limit failure events. Participants were also told that "seemingly random issues" may occur due to sensor errors or algorithmic failures so that they were aware that system-malfunction failures could also occur. First, participants were trained on how to use the ACC system, which involved the experimenter going through a series of powerpoint slides explaining how ACC worked and what the limitations of the system were. Then there was a short quiz in which participants were shown a picture of a road scene (e.g., with poor weather conditions), and asked if the automation would work in that scenario. For any incorrect answer, the experimenter would remind the participant of the limitations. After the quiz, participants were shown how to turn the ACC on and off in the simulator. Once the ACC training was complete, the same training was completed for the LK system. All participants then performed a six-minute training drive where they experienced a system-malfunction failure and a system-limit failure (the same as the ones in experimental drives). Participants then completed the two baseline drives; each drive lasted roughly six minutes.

Using the Automation

The driving automation consisted of ACC and LK systems. Participants were responsible for monitoring the automation and taking over when they felt it was necessary. Participants were told that they had to drive safely throughout the drive and that they were free to disengage the automation when they felt that it was necessary, however, they were told to keep the automation on as much as possible. The ACC and LK systems used in this experiment are similar to the systems detailed in previous research (e.g., *12*, *13*). Each of the systems were set to be on at the beginning of the experiment by the experimenter, and the drivers could then engage or disengage the automated systems. To disengage the ACC, participants could use either the ACC cancel button on the steering wheel or the brake pedal. To disengage the LK, participants could either use the LK on/off button or turn the steering wheel. Participants had to move the steering wheel more than 5° to disengage the LK system, in order to prevent the system from being turned off accidentally.

Driving Scenario

Each of the drives occurred on a rural two-lane highway, where the road was 12 feet across (3.66 m), and had double yellow lines separating the opposing lanes of traffic. Surrounding traffic was light (9-11 cars per minute) and the posted speed limit was 50 mph. Participants were instructed to follow a lead vehicle throughout the drive, which maintained a gap of 3 seconds.

Secondary Task

A self-paced visual-manual secondary task adapted from Donmez, Boyle, and Lee (*14*) was provided in all drives. This task mimicked searching for a song on an infotainment system and has been shown to degrade driving performance in several simulator studies (e.g., *14–16*). Specifically, it was a word matching task presented on the Surface Pro 2 tablet. Out of a list of 10 closely related phrases, participants needed to select one correct phrase that matched the target phrase "Discover Project Missions". A phrase qualified as a match if any of these three conditions were met: "Discover" was in the first position, "Project" was in the second position, or "Missions" was in the third position. Thus "Discover Missions Project" is a match because it has "Discover" first, whereas "Project Discover Misguide" is not a match because none of the target words are in the correct place. Only two options were displayed on the screen at a time, and to scroll through the options, participants could tap the up and down arrows with their fingers (see **Figure 2**). Participants entered their choice by pressing the submit button and received feedback on whether their choice was correct or incorrect. The task was available throughout the drive and participants could choose when to engage in the task.

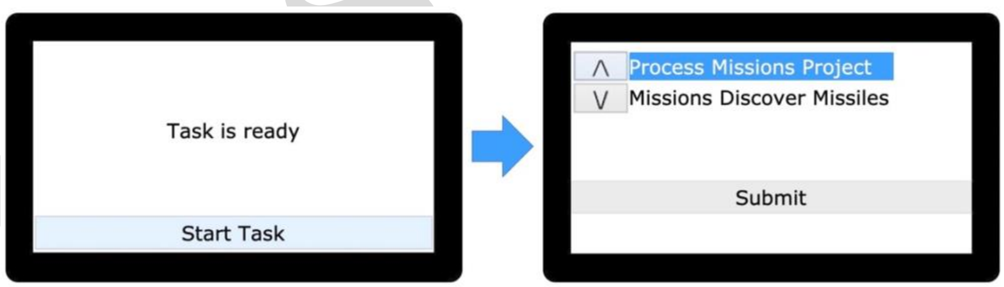

FIGURE 2: Secondary task, adapted from (*14*)

Participants were told that they could receive up to an $8 bonus based on their secondary task performance (specifically that they would receive an additional $0.20 for each correct answer, and they would lose $0.40 for each incorrect answer). However, participants were also informed that their driving performance would be rated, and poor performance would result in deductions from their bonus. Thus, they were instructed to drive safely, but had an incentive to perform the secondary task.

Automation Failure Events

The failures were in the LK system and were lane departure events as motivated at the end of the Introduction.

System-Limit Failure

The system-limit failure occurred at an intersection, nearly 2230 m from the start of the drive. Rather than continuing straight to follow the lead vehicle, the ego-vehicle followed the solid white line marking the edge of the pavement and entered the right-turn lane (**Figure 3**). As participants were taught about the automation limits prior to the experiment, it was assumed that they knew that when the LK was engaged, the ego-vehicle may not continue straight through the intersection, but instead would follow the road edge line.

Features that indicated the upcoming intersection included a road sign 184m prior to the intersection, and a street light and several buildings surrounding the intersection. As the rest of the drive consisted of a rural landscape, these were clear indicators of an upcoming intersection, and thus a potential failure event. If a participant took over control from the automation prior to reaching the beginning of the right-turn lane, the participant would simply drive the vehicle through the intersection. If the participant took over control from the automation after the start of the failure event, then the participant would have to steer the vehicle back into the center lane to follow the lead vehicle.

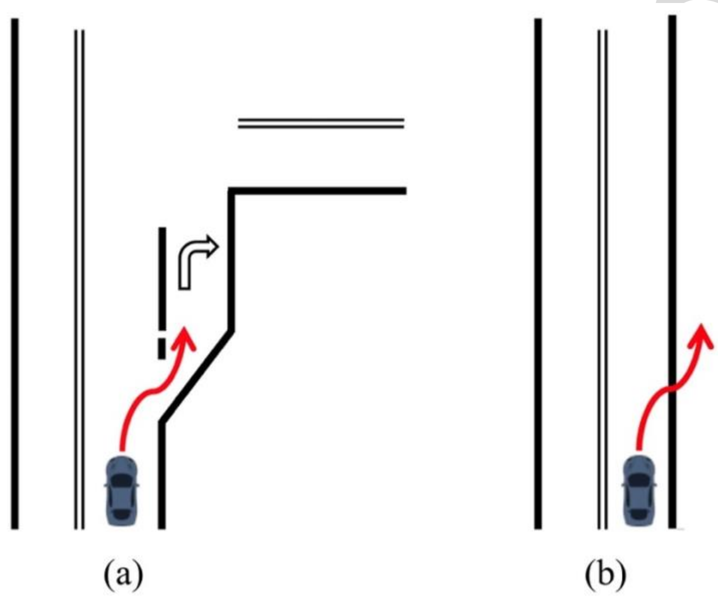

FIGURE 3: (a) Path of vehicle at the system-limit failure. (b) Path of the vehicle at the system-malfunction failure.

System-Malfunction Failure

The manner in which the vehicle failed during the system-malfunction failure (**Figure 3**) was redesigned several times during development until it was visually determined to have the same level of rapidness, swerve at a similar angle, and require a similar level of steering, as the system-limit failure. When the system-malfunction failure occurred, the ego-vehicle veered to the right, without any indicators that the vehicle would fail. The malfunction was similar to the system-limit condition in geometry; the initial movement was a slight swerve out of the ego-lane onto the shoulder (similar to swerving into the right-turn lane for the system-limit failure), but then the vehicle continued to veer right until the participant took over. However, all participants took over, thus nobody experienced the continued veer off-road. This failure occurred at nearly 2168 m from the start of the drive, on a straight section of road. If a participant took over control of the vehicle prior to the failure event, the failure would not be triggered (same as with the system-limit failure), and the participant would continue to drive. If the participant took over control of the vehicle after the start of the failure event, the participant would have to steer the vehicle back into the lane and continue driving. If a participant did not take over at all, the car would continue to go off road and eventually hit the rumble strip, thus giving the participant an auditory cue that they went off road.

Dependent Variable and Statistical Models

Dependent Variables

Takeover performance was assessed using the following measures that were adopted from previous research (e.g., *12*, *17*):

- Hands-on-wheel time (seconds): When, relative to the failure, the driver put their hands on the wheel to take over control. Participants were told to drive with their hands off the wheel when the LK was engaged, so this was measured by identifying the first movement of the steering wheel.
- Takeover time (seconds): When, relative to the failure, the driver disengaged the LK. Participants had to turn the steering wheel 5° to disengage the LK, so they could put their hands on the wheel without immediately disengaging the LK.
- Standard deviation of steering wheel angle (degrees): Standard deviation of the steering wheel angle for the 20 s after participants took control of the vehicle during a failure event.
- Steering wheel angle range (degrees): The range that the steering wheel was moved by the driver during the 20 s after participants took control of the vehicle during a failure event.

Monitoring was measured during the period starting 15.5 s before the failure until takeover (i.e., LK was disengaged by the participant), which will be referred to as the time before takeover. The starting point of 15.5 s before the failure was chosen because this was approximately when participants had a first indication that an intersection was coming up (the "intersection ahead" sign became legible) in drives with a system-limit failure. The equivalent time period was used for system-malfunction failures as a comparison to investigate whether participants would look more at the roadway when there were cues in the environment to allow them to anticipate the upcoming failure. Because participants took over at different times, the time before takeover differed for each person. The following three measures were extracted for both AOIs (i.e., front screen and secondary task) during the time before takeover:

- Rate of glances (per minute): Rate of glances on an AOI (number of glances/time).
- Percent of time: Percent of time before takeover looking at an AOI.
- Average glance duration (ms): average length of glances to each AOI.

Rate of glances and percent of time were used to assess driver monitoring behavior, while the average glance duration was used to further explore the relationship between rate of glances and percent of time. Glances were defined based on ISO Standards (*18*, *19*). While the combination of glance data on the front screen and secondary task can be used as a measure of the driver's monitoring behavior, glances towards the secondary task were also used as a visual measure of secondary task engagement.

Statistical Models

All measures were analyzed through mixed linear models using the SAS MIXED procedure, with failure type introduced as a fixed factor and participant as a random factor. Order (and its interaction with failure type) was initially included in all models to check for learning effects. Neither the main effect nor the interaction were significant in any models so they were removed. A total of three observations where one participant did not have the lane keeping on for at least 20 s prior to a failure event and where two participants did not follow the experimenter instructions, were removed, leaving 33 observations for the takeover performance analysis. Due to errors in equipment setup, eye tracking data was missing for an additional three observations, resulting in 30 total observations for the monitoring analysis. As the homogeneity of variance assumption was not met for any of the measures, an unstructured variance-covariance matrix was used for all analyses. For average glance duration to the secondary task, a log-transform was used to correct for non-normality.

RESULTS

Takeover Performance

The results for hands-on-wheel time, $F(1, 22.6) = 8.16$, $p < 0.01$, and takeover time were significant, $F(1, 22.8) = 7.38$, $p < 0.05$. Both hands-on-wheel time and takeover time were significantly lower for the system-limit failure compared to system-malfunction failure (see **Table 1**). Participants put their hands on the wheel 0.62 s sooner and disengaged the lane keeping 0.51 s sooner at the system-limit failure compared to the system-malfunction failure. However, only two participants took over before the system-limit failure event; as expected, none of the participants took over before the system-malfunction failure event (see **Figure 4**). The results for standard deviation of steering wheel angle, $F(1, 22.5) = 0.02$, $p = 0.89$, and steering wheel angle range, $F(1, 23.3) = 0.10$, $p = 0.76$, were not significant.

TABLE 1: Difference (Δ) of Least Squares Means (LSM), System-Limit Minus System-Malfunction, for Hands-on-Wheel Time and Takeover Time Measures

Measure	System-Limit		System-Malfunction		Δ LSM	Standard Error	df	*t*-*value*	*p*-*value*	95% Confidence Interval	
	LSM	Min., Max.	LSM	Min., Max.						Lower	Upper
Hands-on-Wheel Time (s)	0.58	-3.83, 1.10	1.20	0.65, 2.38	-0.62	0.22	22.6	-2.86	.009**	-1.06	-0.17
Takeover Time (s)	0.78	-2.48, 1.32	1.29	0.77, 2.50	-0.51	0.19	22.8	-2.72	.01*	-0.90	-0.12
Standard Deviation of Steering Wheel Angle (degrees)	4.21	1.62, 7.09	4.18	2.90, 5.70	0.04	0.27	22.5	0.14	0.89	-0.53	0.60
Steering Wheel Angle Range (degrees)	26.67	7.42, 65.75	26.02	18.07, 45.67	0.65	2.11	23.3	0.31	0.76	-3.72	5.02

* $p < .05$, ** $p < .01$

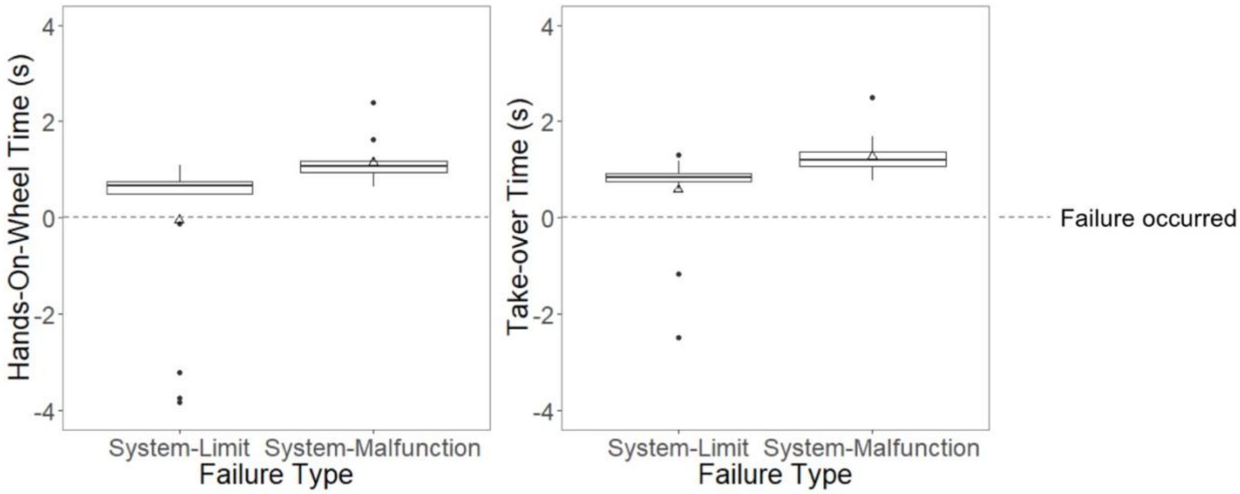

FIGURE 4: Box-plots for hands-on-wheel time and takeover time by failure type. Negative numbers indicate that participants put their hands on the wheel or took over before the failure. Boxplots in this figure and later show the minimum, first quartile, median, third quartile, maximum, and potential outliers; triangles indicate the mean.

Monitoring Performance and Visual Engagement in Secondary Task

Participants looked mainly at the front screen and secondary task in the time before takeover (93% of the time before the system-limit failure and 94% before the system-malfunction failure), thus statistical analysis was conducted only on these two areas. There was a significant effect of failure type for rate of glances to the secondary task, $F(1, 15) = 5.09$, $p < .05$, and percent of time spent looking at the secondary task, $F(1, 13.6) = 6.68$, $p < .05$. At the system-limit failure compared to the system-malfunction failure, both the rate of glances towards (by 5.10/minute) and percent of time looking at the secondary task were significantly lower (by 12.74%; **Table 2**), indicating that participants engaged less in the secondary task before the system-limit failure. The effect of failure type on rate of glaces to the front screen was not significant, $F(1, 13.9) = 1.94$, $p = .19$. However, there was a significant effect for percent of time looking at the front screen, $F(1, 14.1) = 5.30$, $p < .05$, which was significantly higher (by 11.23%) for system-limit failures compared to system-malfunction failures (**Table 2**).

TABLE 2: Difference of Least Squares Means (System-Limit Minus System-Malfunction) for Rate (/min) of Glances and Percent of Time Measures

Area	**Measure**	**Estimate**	**Standard error**	**df**	***t*-value**	***p*-value**	**95% Confidence Interval**	
							Lower	**Upper**
Secondary Task	Rate of Glances	-5.10	2.26	15	-2.26	.04*	-9.91	-0.28
	% of Time	-12.74	4.93	13.6	-2.58	.02*	-23.33	-2.14
Front Screen	Rate of Glances	-3.38	2.43	13.9	-1.39	.19	-8.58	1.83
	% of Time	11.23	4.88	14.1	2.30	.04*	0.78	21.68

Note. Negative estimates indicate lower values before the system-limit failure compared to the system-malfunction failure. Positive estimates indicate higher values before the system-limit failure compared to the system-malfunction failure. * $p < .05$

Figure 5 shows that both the rate of glances to the secondary task and percent of time looking at the secondary task were, on average, lower before the system-limit failure. However, the higher percent of time looking at the front screen was not associated with a higher rate of glances to the front screen.

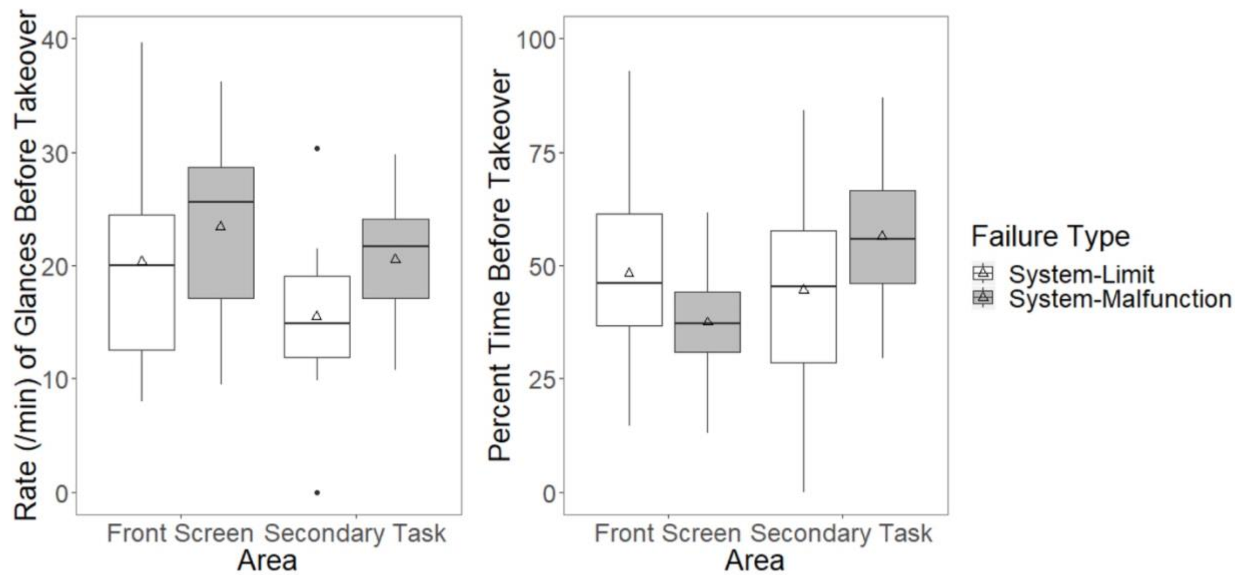

FIGURE 5: Rate (/min) of glances and percent of time data for the front screen and secondary task, by failure type.

A higher percent of time looking at the front screen without an increase in the rate of glances indicated that glances to the front screen were longer for system-limit failures, thus we analyzed average glance duration for the front screen. In our initial analysis (with all data points), there was a clear outlier (studentized residual > 4; **Figure 6**), and the model was only marginally significant, $F(1, 15.1) = 3.97$, $p = .06$. This data point appeared to be one conservative driver who engaged less with the secondary task than the other participants. Given that this participant displayed markedly different behavior from the others, the outlier was removed. The subsequent analysis was significant, $F(1, 14.7) = 5.88$, $p < .05$, and average glance duration to the front screen was significantly longer before the system-limit failure than the system-malfunction failure, indicating that participants made longer glances to the roadway when approaching the location of the system-limit failure. This result explained why we found a significantly higher percent of time looking at the roadway without a change in the rate of glances to the roadway. In other words, people made a similar number of glances to the roadway regardless of the failure type, but when the failure was predictable, these glances were longer (**Figure 6**). There was no difference in average glance duration to the secondary task, $F(1, 13.6) = 0.02$, $p = .89$, which was consistent with our findings that that the lower rate of glances before system-limit failures was associated with a lower percent of time looking at the secondary task.

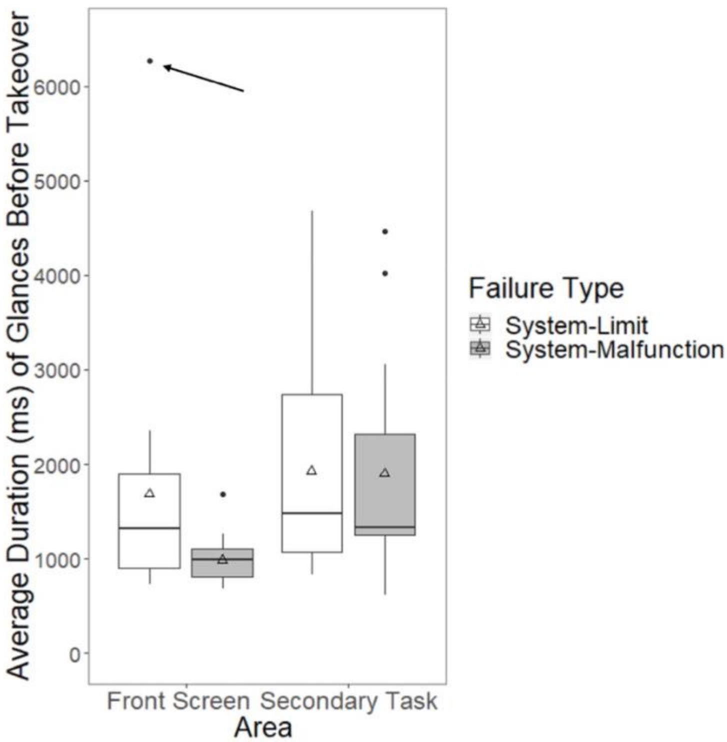

FIGURE 6: Average glance duration before takeover for the front screen and secondary task, by failure type. The arrow indicates the outlier removed from the analysis.

DISCUSSION

When system-malfunction failures occur, drivers are unable to anticipate the failure and thus it makes sense that their behavior would not change prior to such an event. Therefore, the purpose of this study was not to identify that differences in takeover performance and monitoring exist between failure types, but to quantify these differences to assess the impact of failure type on takeover performance and monitoring. An additional objective was to bring these differences to the attention of the research community in order to encourage resarchers to more consciously think about the type of system failures that they implement in their studies.

Takeover Performance

Results show that participants put their hands on the wheel 0.62 seconds sooner and took over from the automation 0.51 seconds sooner at the system-limit failure compared to the system-malfunction failure. At a speed of 50 mph (the speed limit in this study), participants would have an additional 37.4 ft to respond to the failure, which can make have a significant effect on whether they can avoid a critical situation. These results confirm our expectations that participants were better able to prepare for system-limit failures and quantify the improvements in takeover performance for system-limit failures compared to system-malfunction failures. Previous research comparing system-limit to system-malfunction failures did not find a significant difference in takeover performance (*9*). However, takeover requests were used to alert drivers of the need to take over, meaning drivers did not have to decide when they needed to take over as they did in the current experiment. Our findings suggest that when drivers must decide independently when to resume manual control of the vehicle (as in L2 driving automation), being able to predict an upcoming failure is associated with faster takeover times (i.e., drivers take over sooner). These results are consistent with previous work in manual driving showing that driver reaction time is quicker to expected events (e.g., *20*) and work in automated driving showing that a higher perceived predictability of the automation is associated with faster takeover times (*10*).

However, while participants prepared for the system-limit failure sooner than the system-malfunction failure, the data shows that only two participants took over control from the automation prior to the system-limit failure. While this result is unexpected, it likely is due to the lack of criticality in the failure event. Dogan et al. (*21*) found that a failure with higher criticality (i.e., a sudden obstacle in the ego-lane) resulted in a faster takeover time compared to one with lower criticality (i.e., missing lane markings). The system-limit failure in this study would cause the vehicle to veer off-road to the right, which was a grassy area, with no other vehicles around, causing no immediate danger to the driver. Additionally, as participants had only experienced one prior system-limit failure (in the training drive), they may have wanted to wait until the failure event started to see if the vehicle would consistently fail at intersections.

Surprisingly, while participants put their hands on the wheel sooner and took over sooner at the system-limit failure, there were no significant differences in the other two takeover performance variables (standard deviation of steering and steering wheel angle range). This could be due to sample size limitations or a result of participants not taking over until after the failure occurred. Had they taken over prior to the vehicle swerving out of the lane for the system-limit failure, they would likely have had smaller steering movements, resulting in a larger difference between failure types.

Monitoring Performance and Visual Engagement in Secondary Task

Prior to the failure, the percent of time participants spent looking at the front screen was 11.2% higher for the system-limit failure compared to the system-malfunction failure, and the time they spent looking at the secondary task was 12.7% lower. In addition, glance rate towards the secondary task was lower by 5.10/minute before the system-limit failure. These results suggest that participants noticed the upcoming intersection and reduced their visual engagement in the secondary task and changed their monitoring behavior (looking more at the roadway and less at the secondary task) in anticipation of the system-limit failure. While we used a system-limit failure that drivers could predict based on cues in the roadway, these findings are consistent with previous research showing that drivers looked more at relevant in-vehicle cues (i.e., speed on the speedometer) when approaching a system-limit failure (*9*). Further, they support the idea that participants were more aware of the upcoming system-limit failure than the system-malfunction failure, which was expected given that by definition, participants should not be able to anticipate a system-malfunction failure. However, participants would need to identify the intersection ahead and understand that it might lead to an automation failure before they could change their monitoring behavior accordingly. Thus, these results also confirm that after training on the limitations of the LK, participants were able to predict the system-limit failure based on cues in the environment.

There was no significant change in rate of glances to the front screen before the system-limit failure compared to the system-malfunction failure, even though the percent of time looking at the front screen was higher. It was confirmed through further analysis that this was due to participants increasing the average length of each glance to the front screen before the system-limit failure.

Future Work

One area for future research is to investigate different types of system-limit failures. We used only one system-limit failure (an intersection), but it would be worthwhile to see whether the same pattern of results is found across various types of system-limit failures (e.g., poor weather, stopped vehicle in ego-lane). Future studies could also vary the criticality of the failure in order to make the takeover more urgent, to see if drivers will take over before the start of the failure to prevent it from occurring. As previously mentioned, even though our results indicate that drivers were able to predict the system-limit failure, most of the participants waited until the failure occurred before taking over. While Dogan et al. (*21*) found that more critical failures had faster takeover times, drivers in their study were told when they needed to takeover via takeover requests. More research is needed to see if a similar pattern will be observed when drivers cannot rely on takeover requests and must decide on their own when to take over. Similarly, we used only one type of secondary task, a mainly visual search task, but there are other types of tasks that drivers may engage in. Future work should investigate how monitoring behaviour and takeover time may differ depending on the type of secondary task available.

The results of this study also support the importance of improving drivers' mental models of driving automation systems. The system-limit failures in this experiment were predictable because participants were trained to know the specific situations in which the automation might fail. However, without training, these failures could be unpredictable to drivers, and thus they may be slower to take over and pay less attention to the roadway before a system-limit failure occurs. Thus, there could be three failure types: 1) system-limit failures that drivers can predict because they are aware of the limitations, 2) system-limit failures that may be unpredictable because drivers are untrained, and 3) system-malfunction failures, which are always unpredictable. In this study, we compared type 1 to type 3. Future studies could compare all three failure types to see how type 2 failures differ from the others in terms of monitoring and takeover performance.

Using L2 driving automation is essentially a vigilance task, as drivers are required to monitor the environment and the automation to identify any situations that would require them to take over control of the vehicle. It is well established in the human-automation interaction literature that humans are not good at vigilance tasks (e.g., *22*), and research in automated driving showed that when drivers were required to monitor the roadway for an extended period of time (40 minute drive), there was a significant decrease in their ability to detect situations that may lead to an automation failure and slower reaction time as the drive progressed (*23*). In our study, participants completed relatively short drives (5-6 minutes), but it would be important for future work to investigate how monitoring behavior and takeover time are affected over longer drives.

Finally, the data analyzed here came from a larger study about the impact of in-vehicle informational displays (i.e., takeover request versus reliability display) on driver takeover performance and monitoring. The next step will be to investigate the impact of failure type when these different informational displays are present.

CONCLUSIONS

Given that takeover performance and monitoring behavior differ significantly based on failure type, a distinction should be made in the literature between system-limit and system-malfunction failures, and comparisons between previous studies using these failures should not be done without considering this disctinction. Furthermore, as L2 vehicles are currently available to consumers, efforts should be focused on supporting drivers' mental models of automated systems. Drivers are usually not trained on how to use driving automation, with one study reporting that approximately 70% of drivers were unaware of the limitations of the ACC in their vehicle (*24*). Thus, current methods that drivers use to learn about driving automation (e.g., reading the owner's manual, trial-and-error) may not be sufficient. In fact, even when drivers are able to remember system limitations after reading the manual, limitations that they do not continue to experience are forgotten over time (*25*). Our findings suggest that if drivers are aware of the automation's limitations, they will be able to anticipate system-limit failures and subsequently may pay more attention to relevant aspects of the driving environment in preparation of taking over, and thus take over sooner, compared to if these failures are unpredictable to them. Therefore, developing a method to calibrate drivers' mental models so that they are aware of system limitations has the potential to improve the safety of the driver interaction with automated vehicles.

ACKNOWLEDGMENTS

Some of these findings have been published in a short extended abstract in the 2019 Proceedings of the Human Factors and Ergonomics Society Annual Meeting. The research was funded by the Natural Sciences and Engineering Reserch Council of Canada.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: SH, BD; data collection: SH; analysis and interpretation of results: CD, SH, BD; draft manuscript preparation: CD, SH, BD. All authors reviewed the results and approved the final version of the manuscript. The authors do not have any conflicts of interest to declare.

REFERENCES

1. SAE International. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. SAE Standard J3016_201806. 2018.
2. Tesla. *Model S Owner's Manual*. 2018.
3. Cadillac. *CT6 Super Cruise Convenience & Personalization Guide*. 2018.
4. National Transportation Safety Board. *Preliminary Report Highway: HWY18MH010*. 2018.
5. Matousek, M. A Tesla Model S Crashed into a Parked Police Car While Autopilot Was Activated. *Business Insider*. https://www.businessinsider.com/tesla-model-s-with-autopilot-on-crashes-into-police-car-2018-5. Accessed Jul. 31, 2018.
6. Helldin, T., G. Falkman, M. Riveiro, and S. Davidsson. Presenting System Uncertainty in Automotive UIs for Supporting Trust Calibration in Autonomous Driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Eindhoven, The Netherlands, 2013, pp. 210-217.
7. Strand, N., J. Nilsson, I. C. M. Karlsson, and L. Nilsson. Semi-Automated versus Highly Automated Driving in Critical Situations Caused by Automation Failures. *Transportation Research Part F: Psychology and Behaviour*, Vol. 27, 2014, pp. 218–228.
8. Beller, J., M. Heesen, and M. Vollrath. Improving the Driver-Automation Interaction: An Approach Using Automation Uncertainty. *Human Factors*, Vol. 55, No. 6, 2013, pp. 1130–1141.
9. Dogan, E., M. C. Rahal, R. Deborne, P. Delhomme, A. Kemeny, and J. Perrin. Transition of Control in a Partially Automated Vehicle: Effects of Anticipation and Non-Driving-Related Task Involvement. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 46, 2017, pp. 205–215.
10. Seppelt, B. D., and J. D. Lee. Keeping the Driver in the Loop: Dynamic Feedback to Support Appropriate Use of Imperfect Vehicle Control Automation. *International Journal of Human-Computer Studies*, Vol. 125, 2019, pp. 66–80.
11. Dikmen, M., and C. M. Burns. Autonomous Driving in the Real World: Experiences with Tesla Autopilot and Summon. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Ann Arbor, MI, 2016, pp. 225-228.
12. Shen, S., and D. M. Neyens. Assessing Drivers' Response during Automated Driver Support System Failures with Non-Driving Tasks. *Journal of Safety Research*, Vol. 61, 2017, pp. 149–155.
13. Zeeb, K., A. Buchner, and M. Schrauf. Is Take-over Time All That Matters? The Impact of Visual-Cognitive Load on Driver Take-over Quality after Conditionally Automated Driving. *Accident Analysis and Prevention*, Vol. 92, 2016, pp. 230–239.
14. Donmez, B., L. N. Boyle, and J. D. Lee. Safety Implications of Providing Real-Time Feedback to Distracted Drivers. *Accident Analysis and Prevention*, Vol. 39, No. 3, 2007, pp. 581–590.
15. Chen, H. Y. W., L. Hoekstra-Atwood, and B. Donmez. Voluntary- and Involuntary-Distraction Engagement: An Exploratory Study of Individual Differences. *Human Factors*, Vol. 60, No. 4, 2018, pp. 575–588.
16. Merrikhpour, M., and B. Donmez. Designing Feedback to Mitigate Teen Distracted Driving: A Social Norms Approach. *Accident Analysis and Prevention*, Vol. 104, 2017, pp. 185–194.

17. Knappe, G., A. Keinath, K. Bengler, and C. Meinecke. Driving Simulator as an Evaluation Tool - Assessment of the Influence of Field of View and Secondary Tasks on Lane Keeping and Steering Performance. In *Proceedings of the 20th International Technical Conference on the Enhanced Safety of Vehicles*, Lyon, France, 2007, pp. 1-11.
18. ISO. *Road Vehicles — Measurement of Driver Visual Behaviour with Respect to Transport Information and Control Systems — Part 1: Definitions and Parameters*. ISO/FDIS 15007-1:2013(E). 2013.
19. ISO. *Road Vehicles — Measurement of Driver Visual Behaviour with Respect to Transport Information and Control Systems — Part 2: Equipment and Procedures*. ISO/TS 15007-2:2013(E). 2013.
20. Green, M. "How Long Does It Take to Stop?" Methodological Analysis of Driver Perception-Brake Times. *Transportation Human Factors*, Vol. 2, No. 3, 2000, pp. 195–216.
21. Dogan, E., V. Honnêt, S. Masfrand, and A. Guillaume. Effects of Non-Driving-Related Tasks on Takeover Performance in Different Takeover Situations in Conditionally Automated Driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 62, 2019, pp. 494–504.
22. Molloy, R., and R. Parasuraman. Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects. *Human Factors*, Vol. 38, No. 2, 1996, pp. 311–322.
23. Greenlee, E. T., P. R. DeLucia, and D. C. Newton. Driver Vigilance in Automated Vehicles: Hazard Detection Failures Are a Matter of Time. *Human Factors*, Vol. 60, No. 4, 2018, pp. 465–476.
24. Jenness, J. W., N. D. Lerner, S. Mazor, J. S. Osberg, and B. C. Tefft. *Use of Advanced In-Vehicle Technology by Young and Older Early Adopters: Survey Results on Adaptive Cruise Control Systems*. Report No. DOT HS 810-917. U.S. Department of Transportation, NHTSA, 2008.
25. Beggiato, M., M. Pereira, T. Petzoldt, and J. Krems. Learning and Development of Trust, Acceptance and the Mental Model of ACC. A Longitudinal on-Road Study. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 35, 2015, pp. 75–84.