PREPRINT VERSION. THIS ARTICLE HAS BEEN SUBMITTED TO A JOURNAL.

Finding windows of opportunity: how drivers adapt to partial automation safeguards over time

September 2024

Alexandra S. Mueller
Insurance Institute for Highway Safety

Pnina Gershon
Samantha H. Haus
MIT Center for Transportation & Logistics, AgeLab

Jessica B. Cicchino
Insurance Institute for Highway Safety

Bruce Mehler
Bryan Reimer
MIT Center for Transportation & Logistics, AgeLab

IIHS

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

iihs.org

Contents

Abstract

Introduction: Most partially automated systems have safeguards to counteract driver disengagement, but little is known about how they affect driver behavior over time. This naturalistic observation study investigated how the behavior of 14 drivers who had no partial automation experience evolved over a month of exposure to the Tesla Autopilot system in a model year 2020 Model 3.

Method: Behavior was analyzed leading up to, during, and immediately after attention reminders and emergency-slowdown-leading-to-lockout events.

Results: We found that drivers learn to internalize safeguard sequences and discover windows of opportunity to do non-driving-related activities. People learned to respond quicker to alerts, leading to fewer escalated sequences in the latter half of the study. However, drivers also spent more time engaging in non-driving-related activities and glancing off-road, which corresponded with more initial alerts of the attention reminder sequence as time went on. Prolonged disengagement culminated in 16 lockouts across the sample, although in general, drivers responded faster and had fewer lockouts over time.

Conclusion: Our findings demonstrate the human ability to learn system constraints and thus illustrate that it is possible to shape safer driving behavior with robust safeguards. User-centric design considerations for driver support strategies are presented in this paper.

Keywords: Tesla; Autopilot; attention reminders; slowdown; lockout

Introduction

A quick glance at the vehicle market shows just how rapidly technologies are evolving (Highway Loss Data Institute [HLDI], 2023). Some of them improve safety by preventing certain types of crashes, such as automatic emergency braking and lane departure warning (Cicchino, 2017, 2018). Others enhance driver convenience, but one should not assume they have the same benefits as crash avoidance features. Convenience features may have indirect advantages, such as how hands-free connectivity features can reduce visual-manual activity in smartphone and vehicle infotainment use; however, in the end, it depends on the implementation (Reimer, Mehler, Muñoz, Dobres, Kidd, & Reagan, 2021). Designated as Level 2 driving automation according to SAE International's taxonomy (2021), partially automated systems are highly advanced driver convenience features and, while widely available, there is a need to assess their net current safety benefits and to ensure they are implemented as safely as possible.

Partial driving automation offers vehicle control assistance for extended periods. It manages the vehicle's speed and automatically reduces it to maintain a set headway timing when there is a slower vehicle in front. At the same time, such systems provide persistent steering support to keep the vehicle centered within the lane. With this capability, naming conventions (Abraham et al., 2017; Teoh, 2020), and related advertisements, some drivers may expect the vehicle to drive itself; however, no automaker currently offers vehicles that truly drive themselves for private ownership. The driver is expected to stay engaged in the driving task while using these systems, but that does not always happen. Drivers tend to do things they are not supposed to do more often and for longer periods with the system's support, which they believe are safer to do than while driving under manual control (Mueller, Cicchino, & Calvanelli, 2024). This can have undesirable and sometimes serious consequences (National Transportation Safety Board, 2017, 2019, 2020).

Most vehicles equipped with partial automation have safeguards intended to minimize driver disengagement—disengagement in this case is an umbrella term that includes distraction, drowsiness, and inattention. However, just as no two systems are the same because of brand-specific idiosyncrasies, safeguards also differ based on what they are designed to address and how they address it; for a review, see Mueller, Reagan, and Cicchino (2021). The safeguard mechanisms of interest for this study are the ones that activate based on what the driver is doing, which is detected through various in-cabin monitoring features. No technology on the market is capable of determining what a driver is actually thinking. The best a vehicle can do is infer driver engagement through behavior, such as whether hands are on the wheel or gaze is to the road; some do this through steering torque or capacitive touch sensors, and/or others through eye- or head-tracking cameras, respectively. The purpose of these sensing approaches is to inform reactive mechanisms that aim to shape behavior in real-time so that the driver meets the system's user operation requirements. To date, there is limited information about how these safeguards affect driver behavior in the long run.

Attention reminders

When the vehicle detects that the driver is no longer fully engaged in the driving task, as defined by the monitoring and driver support system, it will begin communicating through alerts to get the driver reengaged. These alerts are known as attention reminders. Depending on the monitoring strategy, the vehicle will initiate attention reminders with instructions to put hands back on the wheel, steer the vehicle, or look back to the road. If the driver does not respond with the required behavior within a defined interval, attention reminders often escalate by becoming increasingly more salient (Mueller et al., 2021). While by no means an exhaustive list, some examples of salience escalation for visual alerts include changes to text, color, size, contrast, location within the display, and motion. Audible and haptic/tactile alerts can likewise escalate by, for instance, increasing in frequency and amplitude. Escalating from mild to intense communication is a design strategy often intended to minimize initial annoyance and the possibility that the driver could be startled and consequently swerve or brake harshly.

Attention-reminder escalation usually incorporates more modalities as the sequence progresses because the likelihood of drivers responding increases when more modalities are used simultaneously (Politis, Brewster, & Pollick, 2013). Visual alerts are useful at the initial phase, particularly if the driver is already paying attention and can respond quickly. However, visual communication on its own is ineffective if the driver is looking away from where it is displayed. Escalation to other modalities, such as auditory and haptic alerts, is often employed to increase the probability that the vehicle can capture the driver's attention.

Emergency countermeasures

Alerts without physical consequences may be ineffective for some people, especially those who deliberately misuse the technology. Some automakers use last-resort countermeasures towards the end of the escalation sequence to address this possibility, which is sometimes referred to as the emergency escalation phase (Mueller et al., 2021). One strategy is for the vehicle to slow down to either a complete standstill or a crawl to reduce potential collision severity. The slowdown can also help to physically motivate the driver to take over if they are just ignoring the attention reminders. Regardless of the reason for noncompliance, as the driver has not been maintaining the system-defined steering control, most systems do maintain lane-centering support during the slowdown so that the vehicle does not leave its lane. With the slowdown occurring in an active roadway, some vehicles will make an SOS call to help limit their crash exposure, which can also help motivate drivers to take over if they are just disregarding the alerts.

Drivers can intervene at any point by resuming control of the vehicle, which will cancel the escalation for most systems. However, given that the slowdown is only initiated when the driver has been nonresponsive for a relatively long time, some vehicles lock out the driver's access to the partially automated system once they resume control. Lockout refers to the driver being unable to reactivate the system for a certain period of time or until the next ignition/power cycle. While these last-resort countermeasures may seem severe for most people, research on closed-course tracks has shown that some drivers disengage from driving long enough to necessitate their activation (e.g., Llaneras, Cannon, & Green, 2017). Such observations support the idea that a lockout could have protective value for those who are prone to misusing the technology, although it remains to be seen whether and how driver behavior changes after experiencing a lockout event in the real world.

Study objectives

There is limited research on how different mechanisms within an attention reminder system correspond with changes in driver behavior over time, and so it is unclear whether current designs work effectively to support driver engagement and readiness. It has been observed that people can develop internal timers for how long it takes for attention reminders to start after fairly brief exposures (Atwood, Guo, & Blanco, 2019), and therefore drivers may learn the onset and duration timing of each alert phase as they become familiar with the partially automated system. This would mean that they can learn to respond to attention reminders more quickly over time. However, there may be unintended consequences to this internalization. Many people who regularly use these systems feel that attention reminders are only mildly annoying, whereas emergency escalation countermeasures (e.g., lockout) are far more aversive (Mueller et al., 2024). Hypothetically, with enough system exposure, people may learn how long they can do things that they are not supposed to before the system escalates to the more aversive phases. This may lead drivers to become disengaged more often, but for shorter periods, which should correspond with more initial attention reminder alerts but fewer escalated alerts over time.

Longitudinal naturalistic observational data is the key to answering these questions. As such, data from the Massachusetts Institute of Technology (MIT) Advanced Vehicle Technology (AVT) Consortium's field operational study were used to investigate how user safeguard activation corresponds with changes in driver behavior while using partial driving automation over time. The present study considered Tesla's partially automated system, called Autopilot, which has a multimodal attention-reminder-escalation process with vehicle slowdown and lockout as last-resort countermeasures.

Method

Data source

The MIT-AVT data collection effort involves a fleet of production vehicles that were given to volunteers to use as their personal vehicles for a month (Gershon et al., 2021). The present study used data from 14 licensed drivers who had no previous partial driving automation experience. Twelve of them were male and the average age of the sample was 39 years (SD = 12, min = 25, max = 58). Upon vehicle delivery, participants received training on the equipped technology. The training session started with a 30-minute in-vehicle instruction period with the vehicle at a standstill, followed by an hour of on-road training, during which participants interacted with the different partially automated systems. Drivers received a monetary incentive of $80 to complete a post-drive interview.

Study vehicle

The vehicles used in this study were model year 2020 Tesla Model 3s with Autopilot, which is a partially automated system. The vehicle's driver-monitoring system tracks steering torque through sensors within the steering wheel to infer that the driver's hands are on the wheel. This means that the driver must physically move the wheel for the sensors to register any hand-on-wheel input. Although Model 3 vehicles are equipped with driver-facing cameras, at the time of study those cameras were not part of Autopilot's driver-monitoring strategy. Tesla regularly pushes over-the-air updates to their entire vehicle fleet, and those updates sometimes concern Autopilot. While it is beyond the scope of the study to examine how these updates correlate with changes in Autopilot use and behavior, the following operating system versions were installed in the AVT Model 3 vehicles during the study:

- 2021.4.12
- 2021.4.15
- 2021.4.18.2
- 2021.12.25.7
- 2021.32.22
- 2021.44.25.2
- 2021.44.30
- 2022.4.5
- 2022.4.5.3
- 2022.8.2
- 2022.12.3.2
- 2022.16.3

Attention reminders and emergency escalations. Table 1 and Figures 1a and 1b describe the individual phases of the attention reminder and the emergency escalation sequences.

Table 1. Individual phases in Autopilot's escalation sequence in response to driver noncompliance.

Phase	Visual alert	Audible alert	Emergency countermeasures
A1	A small icon showing two grey hands holding a steering wheel at the bottom of the display. Text notification in white font within a black bubble textbox: "Apply slight turning force to steering wheel." Flashing blue light at the top of the display.	None	None
A2	A small icon showing two red hands holding a steering wheel at the bottom of the display. Text notification in white font within a black bubble textbox: “Apply slight turning force to steering wheel." Blue light at the top of the display flashing with increased frequency.	One short tone burst	None
A3	A small icon showing two red hands holding the steering wheel at the bottom of the display. Text notification in white font within a black bubble textbox: "Apply slight turning force to steering wheel." Blue light at the top of the display flashing with increasing frequency.	Two short tone bursts	None
A4	A large icon showing two red hands holding the steering wheel in the center of the display. Text notification in white within a red bubble textbox: "⚠ Autosteer unavailable for the rest of this drive. Hold steering wheel to drive manually."	Continuous tone burst	Slowdown and lockout
Lockout alert	Alert occurs when the driver attempts to reactivate Autopilot after the A4 lockout event. Text notification in white font within a black bubble textbox: "⚠ Autosteer unavailable for the rest of this drive. Hold steering wheel to drive manually."		

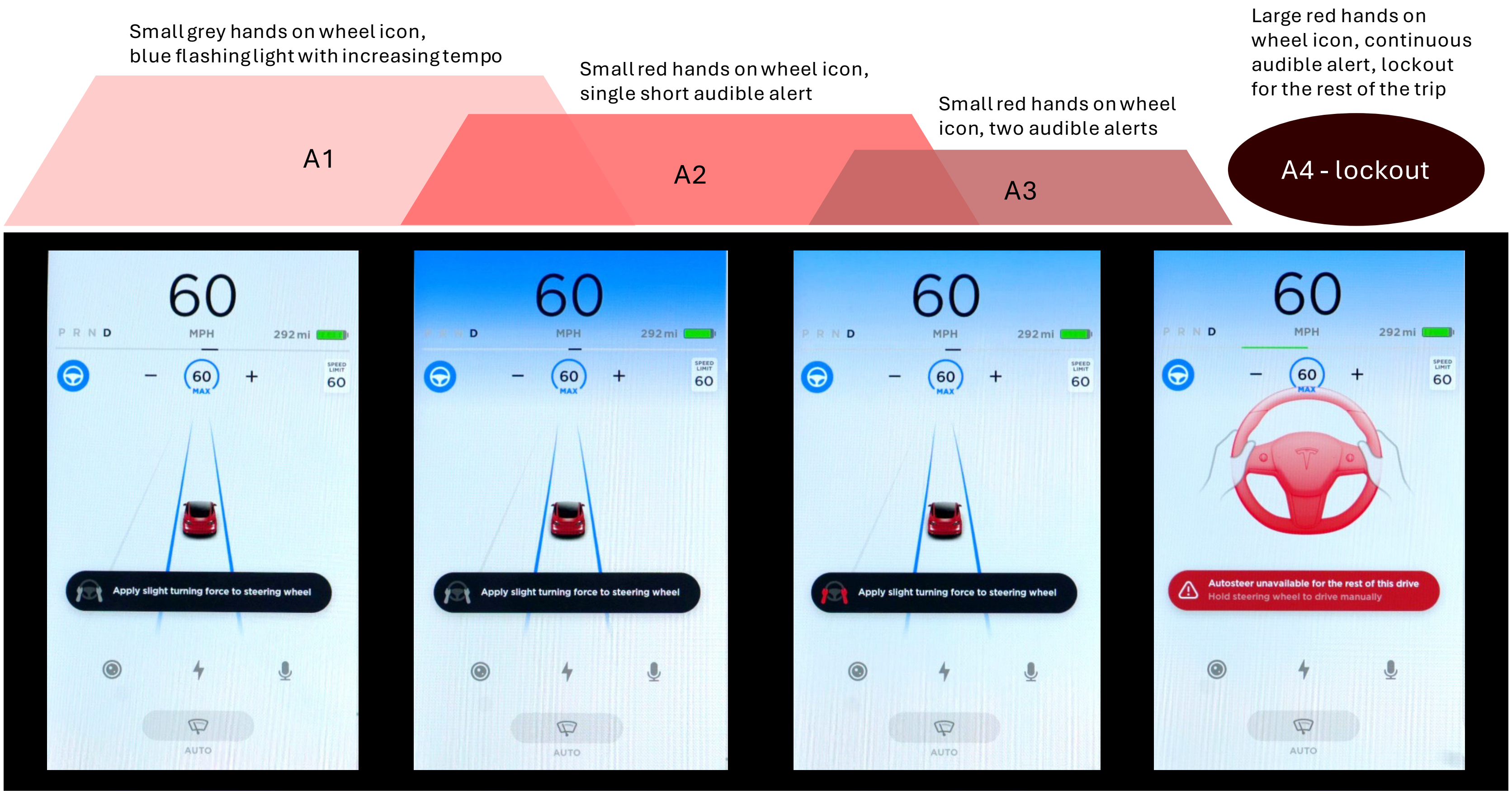

Figure 1a. Video stills depicting each phase of the attention reminder sequence.

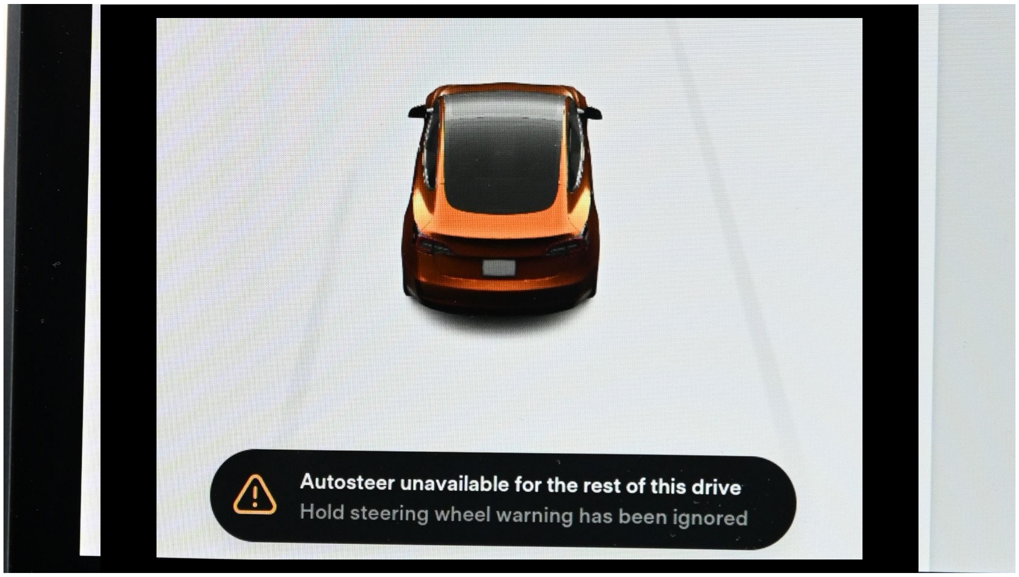

Figure 1b. Video still depicting the lockout notification that occurs when the driver attempts to reactivate Autopilot after experiencing an A4 lockout event.

The instrument panel display is a large tablet-like touchscreen display in the center stack area as depicted in Figure 2. It is the only visual interface within the cabin, serving to provide information that traditionally would be distributed between the instrument panel and center stack display. It shows all vehicle operation information, vehicle control and menu options, infotainment content, and navigation content, as well as Autopilot visual communication, including attention reminders. Information specific to advanced driver assistance features (www.tesla.com/support/autopilot) is presented on the left side of the display. The default Autopilot information shown includes travel speed, adaptive cruise control (ACC) activity, lane-centering support activity, and automated lane-change activity, which is available while using Enhanced Autopilot functionality. Tesla's brand name for ACC is Traffic-Aware Cruise Control and its lane centering is called Autosteer.

Figure 2. Photo of Tesla Model 3 interior showing the combined instrument panel and center stack display next to the steering wheel.

Vehicle instrumentation

The Model 3 vehicles were instrumented with the following recording equipment: cameras with audio recording, GPS, and CAN bus recorders. As shown in Figure3, the camera angles recorded included the driver's face, the driver's seat area, the center stack/instrument panel display, and the forward field of view of the road. The video, GPS, and CAN bus data were synced and analyzed at 30 Hz.

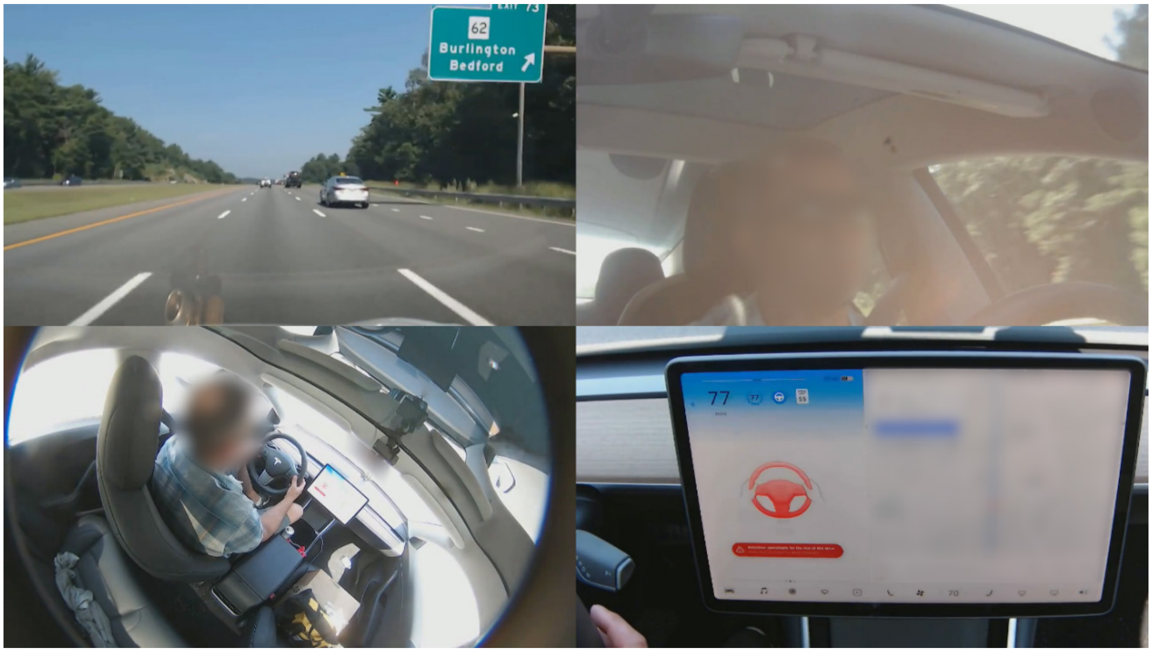

Figure 3. Still frames from video footage demonstrating the camera angles.

Data processing

Attention reminder events. Segments of interest contained attention reminder events and continued until the attention reminder sequence was terminated either through driver intervention or system lockout. These video segments of interest are hereafter referred to as epochs. Each epoch included 10 seconds (sec) prior to the start of the attention reminder sequence and 10 sec after the sequence terminated. Some epochs had multiple independent alert sequences that occurred within 20 sec of each other; for example, an A1 alert could be triggered and shortly afterwards terminated by the driver nudging the wheel, and then 10 sec later another A1 alert could be initiated again followed by an A2 after which the driver nudged the wheel to terminate the escalation—these two independent sequences (A1 and A1 followed by A2) would be combined into a single epoch. The analysis occurred at the alert sequence level, meaning that in this example the two sequences would be analyzed separately. Epochs were nevertheless tracked to ensure there was no overlap with other alerts when examining behavior before and after the alerts.

Trained annotators labeled driver eye-glance behavior, hand-on-wheel activity, and engagement in secondary activities at a frame-by-frame level. A quality control process was subsequently implemented whereby 30% of the annotated epochs were randomly selected and reviewed by a second senior off-site annotator, and an additional 10% of the epochs were randomly selected for review by an annotation lead at MIT. A review of the annotated data was then conducted to identify inconsistencies or missing annotations. In total, the dataset included 3,858 individual attention reminder events from 1,233 independent trips in which Autopilot had been in use.

Automation state. Autopilot state was identified from vehicle CAN data and included: (1.) Autopilot off (manual or Traffic-Aware Cruise Control support only) and (2.) Autopilot on. The Autopilot on state could contain additional features such as Navigate on Autopilot, which incorporates automated wayfinding and lane-changing functionality that is not available in the conventional Autopilot mode.

Alert sequences. Autopilot's timing and manner of escalation vary depending on what the driver is doing and what is happening on the road; however, this is by no means unique to Tesla as many automakers have similar adaptability built into their systems. Although the exact nature of the escalation sequence's adaptability is proprietary to each automaker, an initial investigation of the data revealed that Autopilot's attention reminder sequence can bypass its initial or middle phases to accelerate its alert intensity. Using the definitions in Table 1, the following attention-reminder-escalation sequences were observed in this study:

- A1 only
- A1 followed by A2
- A1 followed by A2 followed by A4
- A1 followed by A4
- A1 followed by A2 followed by A3 followed by A4

For the current analysis, we separated the sequences into two groups: A1-only alerts and Escalated alerts, which are made up of all the other alert phases and sequence combinations. We also examined alert sequences that ended in lockouts (A4 phase), which we refer to as lockout events. No attention reminder sequences ending in the A3 phase were observed, most likely because of the timing proximity between phases A3 and A4. Phase A3 represents the auditory tone burst change that rapidly transitions to A4's continuous burst.

Hand activity. Hand activity was coded in relation to the steering wheel. The main outcome was the number of hands the driver had on the wheel (no hands, one hand, two hands). A driver can rest their hand(s) on the wheel without actually holding it. Although this offers little steering control, such contact may allow drivers to move the wheel enough for Autopilot to stop its escalation. Therefore, any hand touching, regardless of level of control, was coded as being hands-on. We also coded how drivers stopped the escalation process. Possible codes were as follows: toggling the scroll wheels, pressing the turn stalk/gear-shift stalk, moving the steering wheel (with hands or other body part, such as the knee), and unknown (i.e., hand activity not codable).

Glance behavior. Glances to the following on-road areas of interest (AOI) were coded: left window and mirror, right window and mirror, rear mirror, and forward roadway. Off-road AOI included down and over the shoulder because some over-the-shoulder glances occurred while the driver conversed with passengers or reached for objects in the rear seats. Display AOI were comprised solely of glances to the center stack. A fourth category of "other" was defined as regions unrelated to driving and not covered by the other labels.

Secondary activity. Secondary activity refers to any activity not immediately pertinent to the physical control of the vehicle (i.e., steering and speed). However, some driving-related secondary activities, such as glances at the center stack display or using steering wheel controls cannot be easily distinguished from non-driving-related activities involving the same interface. Therefore, activities involving the center stack and steering wheel controls were categorized as driving-related. This is because the Tesla Model 3 provides scroll wheel buttons on the steering wheel for drivers to perform various functions. For ACC and Autopilot, the scroll wheels are used to change the vehicle's set speed and following distance and are also used as an indicator to Autopilot that the driver is engaged in the driving task—it is treated the same way as steering torque in how the system infers that the driver's hands are on the wheel. Meanwhile, those buttons can be used in a completely different capacity to control the center stack display. Due to limitations of camera angle and the Model 3's display complexity, it was not possible to reliably determine when those buttons were used for driving-related versus non-driving-related purposes. For similar reasons, it was not feasible to reliably distinguish when the driver was looking at the display for driving-related versus non-driving-related reasons as both types of information can be presented simultaneously on the screen. Nevertheless, the following behaviors were coded as driving-related: touching the scroll wheels on the steering wheel and looking and touching the center stack display. A description of each category is provided in Table 2.

Non-driving-related activities were more easily identified. They included holding, reaching for, putting down, or manipulating a personal electronic device, such as a smartphone, or having a phone call while holding that device. Similar non-driving-related activities were also coded when the driver was reaching for, putting down, looking at, writing on, or otherwise interacting with an object that is neither integral to the vehicle nor a personal electronic device. Hands-free activities such as phone calls and using the voice command features of the vehicle or their smartphone were coded as non-driving-related, as were talking with passengers, grooming/hygiene tasks, eating, and drinking. Non-driving-related activities were further subdivided into voice-based, visual, manual, and visual-manual activities. Voice-based activities included speaking on a personal electronic device hands-free, using voice commands, and speaking with a passenger. Visual activities predominantly concerned reading in areas other than the center stack display. Manual activities consisted of holding personal electronic devices and talking on those devices while holding them. Visual-manual activities involved reaching for those devices, manipulating them, talking on them while holding them, interacting with other objects using hands, grooming/hygiene tasks, and eating/drinking.

Table 2. Description of the driving-related and non-driving-related categories of secondary activities.

Secondary activity category	Description
Driving-related tasks	Interacting with the center stack; interacting with the steering wheel controls; and reading/watching the center stack
Non-driving-related tasks	
Voice-based	Talking on a PED hands-free; using voice commands; interacting with a passenger
Visual only	Reading, non-center stack
Manual only	Holding a PED; talking on a handheld PED
Visual-manual	Reaching for a PED; manipulating a PED; talking on a handheld PED; grooming/hygiene tasks; interacting with a non-PED object; and eating/drinking

Note: PED = personal electronic device (i.e., cellphone, tablet, etc.).

Statistical analysis

Given that attention reminder activation and escalation are indicators of driver disengagement from the driving task and its severity (Atwood et al., 2019), we investigated how the frequency and duration of each type of alert sequence, including slowdown and lockout events (which include an A4 phase), changed across and between the study weeks. Poisson mixed-effects models were used to evaluate the incidence and linear mixed-effect models were used to evaluate the duration of A1-only alerts and Escalated alert sequences over time in the study (study weeks). Lockout events were relatively rare (a total of 16 events), and as such we report central tendency measures and frequency counts for the prevalence and duration of those events.

Hand activity for epochs with A1-only and Escalated alert sequences was modeled using linear mixed-effect models. For each sequence type (A1-only and Escalated), we calculated the percentage of time when the driver had no hands, one hand, or two hands on the steering wheel in the 10 secs before the first alert phase in the sequence, during the sequence, and 10 secs after the last alert phase of the sequence. If there were multiple sequences in an epoch, then the percentage of time was calculated over the time between the sequences, thereby ensuring that other sequences were not included in the outcome. In other words, if the time between two sequences was less than 10 sec, the hands-on-wheel percentage used that less-than-10-second period as the denominator.

To assess driver response to the first alert in a sequence we used generalized linear mixed-effects models assuming a gamma distribution. We calculated "Time to hands on wheel after alert onset" as the time from when the first alert in the sequence began to the time when a hand on wheel was first annotated. For epochs containing multiple sequences, the reaction time analysis only considered the first alert sequence in the epoch, thereby limiting the potential priming effects of multiple alert sequences in close succession. Epochs in which the driver did not put their hand on the wheel before the alert sequence ended were excluded from that analysis. In addition, it was common for drivers to have at least one hand touching the wheel at the start of the alert sequence, and those epochs were also excluded from the "Time to hands on wheel after alert onset" analysis. Similarly, "Time to hands off wheel after alert sequence end" was calculated as the time from the end of the alert sequence (in the case of a multi-sequence epoch, the end of the alert sequence was the end of the first alert sequence in that epoch) to the time when both driver's hands left the steering wheel was first annotated. Alerts were excluded if a driver did not take their hand off the wheel in that epoch or after the sequence had ended, if their hands were already off the wheel when the alert sequence ended, or if that alert sequence followed another sequence in a multi-sequence epoch. If hand activity could not be determined, the epoch was excluded from the "Time to hands on wheel after alert onset" and "Time to hands off wheel after alert sequence end" analysis.

Glance behavior for A1-only and Escalated alert sequence epochs was modeled using linear mixed-effect models. For each alert sequence type, we calculated the percentage of time in an epoch in which the driver's glances were directed to on-road (i.e., forward roadway, windows, and mirrors), the display, and off-road (i.e., down and over the shoulder) AOI in the 10 secs before the first alert phase in the sequence, during the sequence, and 10 secs after the last alert phase in the sequence. For epochs containing multiple sequences, the percentage of time in the post-alert and subsequent pre-alert phase were calculated over the time between the sequences.

Secondary activity for A1-only and Escalated alert sequence epochs was modeled in a similar fashion using linear mixed-effect models. For each alert sequence type, we calculated the proportion of secondary activity engagement for driving-related and non-driving-related tasks in an epoch (see Table 2) in the 10 secs before the first alert phase of the sequence, during the sequence, and 10 secs after the last alert phase of the sequence. For epochs that contained multiple sequences, the percentage of time was calculated over the time between the sequences.

Finally, we examined how driver behaviors were associated with the probability of alert escalation with a logistic mixed-effect model. Driver behavior was defined as the proportion of time engaged in secondary activities, the proportion of time spent looking at the on-road and off-road AOI, as well as the proportion of time spent with no hands, one hand, and two hands on the wheel. Glances to the display were not included in this analysis due to their high correlation with driving-related secondary activities—82% of driving-related secondary activities were comprised of reading/watching the display. Proportions were calculated for the 10 sec before the first alert phase of a sequence and during the alert sequence. For epochs containing multiple sequences, the proportion of time was calculated over the time between the sequences. All models assumed a driver-specific random intercept to account for between-driver variability and defined statistical significance at a 0.05 confidence level with 95% confidence intervals (CIs).

Missing data. As can happen with naturalistic observation research in the real world, there were times when the videos were not codable for certain behaviors. This was sometimes due to environmental conditions, such as nighttime, which degrades the video quality and limits the codability of all behaviors. At other times it was the driver's actions that were responsible; for example, wearing dark sunglasses prevented reliable coding of eye gaze. For the 3,786 A1-only alert epochs, 1,229 had missing glance behavior, 704 had missing hand activity, and 586 had missing secondary activity data. For the 72 epochs that contained Escalated alert sequences, 42 had missing glance behavior, 26 had missing hand activity, and 23 had missing secondary activity data. Lastly, for the 16 epochs that contained lockouts, 11 had missing glance behavior, 7 had missing hand activity, and 7 had missing secondary activity data. The analysis of hand activity, glance behavior, and secondary activities included only complete cases; in other words, no epochs with missing data were included. Note that including epochs with missing data in the analysis did not yield substantially different patterns.

Results

Prevalence of attention reminders

Overall, we observed 3,858 attention-related alerts during 12,161 miles of driving with Autopilot engaged. The most common alert type belonged to the A1-only phase, which accounted for 98% of the attention reminders. The average duration of A1-only alerts, characterized solely by visual icon presentation in the center stack, was 3.07 sec (*min* = 0.33 sec, *max* = 15 sec). The average duration of Escalated alerts was 18.02 sec (*min* = 1.33 sec, *max*= 37.9 sec) (see Table 3). The number of A1-only alerts increased by 39% from week 1 to week 4, while the number of Escalated alerts decreased by 76% over the same time interval (Table 3). While drivers had quite a bit of variability in both cumulative Autopilot miles (*m* = 653 miles, *SD* = 517 miles) and number of alerts (*m* = 276, *SD* = 381), we found that drivers who experienced Escalated alerts typically had more overall Autopilot miles than those who only experienced A1-only alerts.

Table 3. Count, mean duration (SD), and duration range (minimum to maximum) of A1-only and Escalated alerts across the 4 study weeks.

	A1-only alerts (sec)			Escalated alerts (sec)		
Week	**Count**	***M* (*SD*)**	**Range**	**Count**	***M* (*SD*)**	**Range**
1	793	3.78 (2.9)	0.33, 14.93	38	19.16 (5.7)	3.47, 34.90
2	866	2.89 (2.0)	0.33, 13.50	17	17.86 (9.9)	1.5, 37.90
3	1,023	2.60 (1.8)	0.37, 14.63	8	15.96 (4.9)	5.67, 23.03
4	1,104	3.13 (2.0)	0.33, 15.00	9	15.36 (9.1)	1.33, 16.63
Total	3,786	3.07 (2.2)	0.33, 15.00	72	18.02 (7.3)	1.33, 37.90

Note. Poisson mixed-effects models with a driver-specific random intercept were used to estimate the A1-only and Escalated alert rates.

Table 4 summarizes the frequencies of attention reminder alerts per 1,000 miles with Autopilot over the 4-week study period. Incidence rate ratios (IRRs) revealed that the prevalence of A1-only alerts increased significantly over time relative to week 1. By the second week drivers were 8% (IRR= 1.08, p = 0.39) more likely to experience A1-only alerts than in the first week, and that increase in likelihood jumped to 25% (IRR = 1.25, $p < 0.001$), and 26% (IRR = 1.26, $p < 0.001$) by weeks 3 and 4, respectively. On the other hand, Escalated alerts followed an opposite trend, decreasing in likelihood by 41% by week 2 (IRR= 0.59, $p = 0.28$), 49% by week 3 (IRR= 0.51, $p = 0.38$.), and 64% by week 4 (IRR = 0.36, p = 0.04) relative to week 1.

Table 4. Estimated attention-reminder alert rates per 1,000 miles with Autopilot as a function of week in the study.

	A1-only alerts		**Escalated alerts**	
Week	**Rate per 1,000 miles**	**IRR (ref. week 1)**	**Rate per 1,000 miles**	**IRR (ref. week 1)**
1	262 (158, 435)	-	4.2 (1.3, 13.3)	-
2	284 (171, 470)	1.08 (0.98, 1.19)	2.5 (0.7, 8.1)	0.59 (0.33, 1.06)
3	328 (198, 543)	1.25^{***} (1.13, 1.38)	2.1 (0.6, 7.4)	0.51 (0.22, 1.17)
4	331 (200, 547)	1.26^{***} (1.15, 1.38)	1.5 (0.4, 5.2)	0.36^{*} (0.17, 0.77)

* p value < 0.05, ** p value < 0.01, *** p value < 0.001

Note. Incidence rate ratios (IRR) relative to week 1 are provided and 95% confidence intervals (CI) for each value are included in parentheses.

Estimated alert duration over time was calculated with 95% CIs using linear mixed-effects models with a driver-specific random intercept. The duration of A1-only alerts decreased significantly after the first week in the study, where the mean difference was 0.45 sec (95% CI [0.66, 0.24], p < 0.001), 0.46 sec (95% CI [0.67, 0.24], $p < 0.001$), and 0.37 sec (95% CI [0.58, 0.17] $p < 0.001$) shorter by weeks 2, 3, and 4, respectively. There was no significant change in the duration of Escalated alerts over time in the study (see Table 5).

Table 5. Estimated duration of A1-only and Escalated alerts.

	A1-only alerts		**Escalated alerts**	
	Estimated mean (sec)	**CI**	**Estimated mean (sec)**	**CI**
1	3.45	3.01, 3.90	15.64	9.87, 21.41
2	3.00^{***}	2.56, 3.44	16.30	10.26, 22.34
3	3.00^{***}	2.56, 3.44	15.02	8.21, 21.82
4	3.08^{***}	2.64, 3.52	16.28	9.50, 23.06

* p value < 0.05, ** p value < 0.01, *** p value < 0.001

Note. Denoted significance indicates statistically significant differences relative to week 1.

Escalated alerts

In total, 72 A1 attention alerts escalated to other phases. In those sequences, on average, the A1 phase of the sequence was 13.17 sec (*min* = 0.07 sec, *max* = 15.27 sec) and the A2 phase lasted on average 3 sec (*min* = 0.03 sec, *max* = 10.17 sec). The A3 to A4 (ending in a lockout) phases lasted on average 8.87 sec (*min* = 2.47 sec, *max* = 19.8 sec) (see Table 6). Although the A1 and A2 alert phases seem to have a maximum duration of about 15 sec and 10 sec, respectively, we did observe instances where those phases escalated before 15 sec had elapsed. In three Escalated alert events, the sequence bypassed the A2 and A3 phases altogether, escalating directly from the A1 phase to the A4 phase. Summaries of the number of alert phases, mean, minimum, and maximum duration of each phase in the sequence by study week, are provided in Table 6.

Table 6. Count, mean duration, standard deviation, and duration range of individual alert phases across escalated alert sequences.

	A1-> A2 -> A3/A4 **(sec)**			A1 -> **A2** -> A3/A4 **(sec)**			A1 -> A2 -> **A3/A4 (sec)**		
Week	**Count**	***M* (*SD*)**	**Range**	**Count**	***M* (*SD*)**	**Range**	**Count**	***M* (*SD*)**	**Range**
1	38	13.92 (2.9)	2.73, 15.27	36	3.02 (2.5)	0.03, 10.00	9	10.03 (4.7)	5.37, 19.80
2	17	12.49 (5.1)	0.20, 15.27	17	3.27 (3.5)	0.03, 10.17	4	5.22 (4.1)	3.33, 12.60
3	8	13.25 (4.4)	2.57, 15.27	7	1.61 (1.1)	0.27, 3.10	2	5.20 (3.9)	2.47, 7.93
4	9	11.24 (6.1)	0.07, 15.20	9	3.49 (3.3)	1.13, 10.00	1	5.70	5.70, 5.70
Total	72	13.17 (4.1)	0.07, 15.27	69	3.00 (2.8)	0.03, 10.17	16	8.87 (4.4)	2.47, 19.80

Lockout escalation as a last-resort countermeasure. A lockout event could only happen once per trip and four drivers experienced 16 events in total. One participant had 12 lockouts and they attempted to reengage Autopilot afterward in three of those 12 trips (2 sec and 142 sec after lockouts in week 1, and 265 sec after lockout in week 2). Two other participants had one lockout each and both tried to reengage the system within 5 minutes of being locked out; one participant attempted 2 sec and again 150 sec after their lockout and the other participant attempted 312 sec after their lockout. The fourth participant experienced two lockouts but did not try to reengage Autopilot in either trip. As shown in Table 6, the number of lockout events dropped dramatically over the weeks in the study. There was also a general decreasing trend in the duration of the lockout events over time ($R^2_{\text{adj}} = 0.15$, $p = 0.08$) (see Table 6). The duration of Escalated alerts that ended with a lockout decreased from an average of 10.03 sec (*SD* = 4.7) in week 1 to 5.70 sec in week 4.

Response to attention reminders

Among the 3,858 attention reminder alerts, drivers used various system-accepted behaviors to stop the alerts. Those behaviors included nudging the steering wheel (n = 2,550, 66.1%), toggling the scroll wheels (n = 554, 14.4%), and pressing the turn stalk/gear-shift stalk (n = 21, 0.5%). With respect to nudging the wheel, drivers did not always use their hands to do that (n = 2,424, 63%); sometimes they used other body parts, such as a knee (n = 126, 3%), to activate the torque sensor. In 19% (n = 733) of the alerts, the disengagement method could not be determined.

Hand activity. Alerts were often initiated while drivers were behaving in accordance with Tesla's hands-on-wheel requirement while using Autopilot. In 52% (n = 1,600) of the A1-only alert epochs and 67% (n = 31) of the Escalated alert epochs, drivers had at least one hand on the steering wheel prior to the first alert onset. For all other epochs (among those that did not have missing data), drivers had both hands off the steering wheel prior to the alert onset (A1-only alerts: n = 1482; Escalated alerts: n = 15). The driver's hands were already on the wheel when the first alert in the sequence started in eight out of the nine epochs that ended in a lockout. Epochs in which hand activity was missing were excluded (Al-only alerts: n = 704, Escalated alerts: n = 26, Escalated alerts ending in lockout: n = 7).

Linear mixed-effects models with a driver-specific random intercept were used to estimate the differences between the week 4 and week 1 percentage of time in an epoch where drivers had one, two, or no hands on the steering wheel. For A1-only alert epochs, no-hands-on-wheel time increased by 33 percentage points (95% CI [30, 36%], $p < 0.001$) in week 4 compared with week 1, one-hand-on-wheel time decreased by 26 percentage points (95% CI [23, 29%], $p < 0.001$), and two-hands-on-wheel time decreased by 7-precentage points (95% CI [4, 10%], $p < 0.001$) over the same study period (see Figure 4, left). For Escalated alert epochs, there was an estimated 57 percentage point increase (95% CI [24, 90%], $p < 0.001$) in no-hands-on-wheel time in week 4 compared with week 1, an 18-precentage point decrease (95% CI [−15, 50%], $p = 0.49$) in one-hand-on-wheel time, and a 39 percentage point decrease (95% CI [7, 72%], $p = 0.012$) in two-hands-on-wheel time over the same time period (see Figure 4, right). The percentage of epochs in which the drivers' hands were on the wheel at the beginning of the alert decreased over time in the study from 80% in week 1 to 43% in week 4.

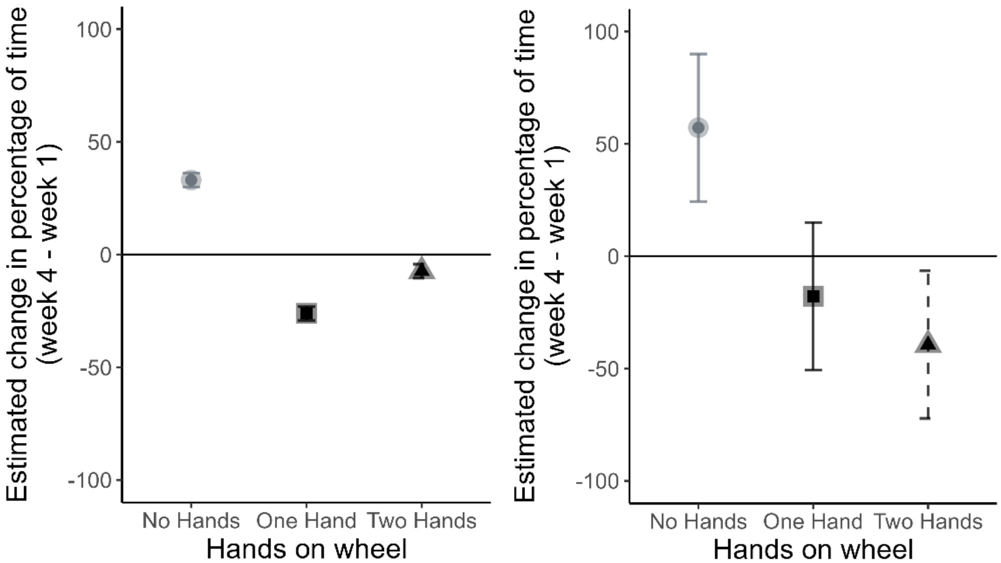

Figure 4. Estimated difference between week 4 and week 1 percentage of driving with hands on the wheel (none, one, two) for A1-only epochs (left) and Escalated alert epochs (right).

Note. Error bars represent 95% lower control and upper control limits (i.e., confidence intervals). Error bars are very small for the A1-only alerts and may not be visible in the figure.

Linear mixed-effects models with a driver-specific random intercept were used to examine how hand activity changed within the alert epoch. For A1-only epochs, the estimated percentage of time drivers had no hands on the wheel decreased over the course of the alert ($p_{before\text{-}during} < 0.001$, $p_{during\text{-}after} < 0.001$, $p_{before\text{-}after} < 0.001$), from 48% before the alert (95% CI [47, 50%], $p < 0.001$) to 40% during it (95% CI [39, 42%], $p < 0.001$), and 35% after (95% CI [34, 37%], $p < 0.001$) (see Figure 5, left). For Escalated alert epochs (see Figure 5, right), the percentage of time drivers had no hands on the wheel was estimated at 33% before the alert (95% CI [20, 45%], $p < 0.001$), and 33% during the alert (95% CI [20, 45%], $p < 0.001$), then decreased substantially after the alert ended to 17% (95% CI [4, 30%], $p = 0.008$), but this difference was not statistically significant ($p = 0.14$).

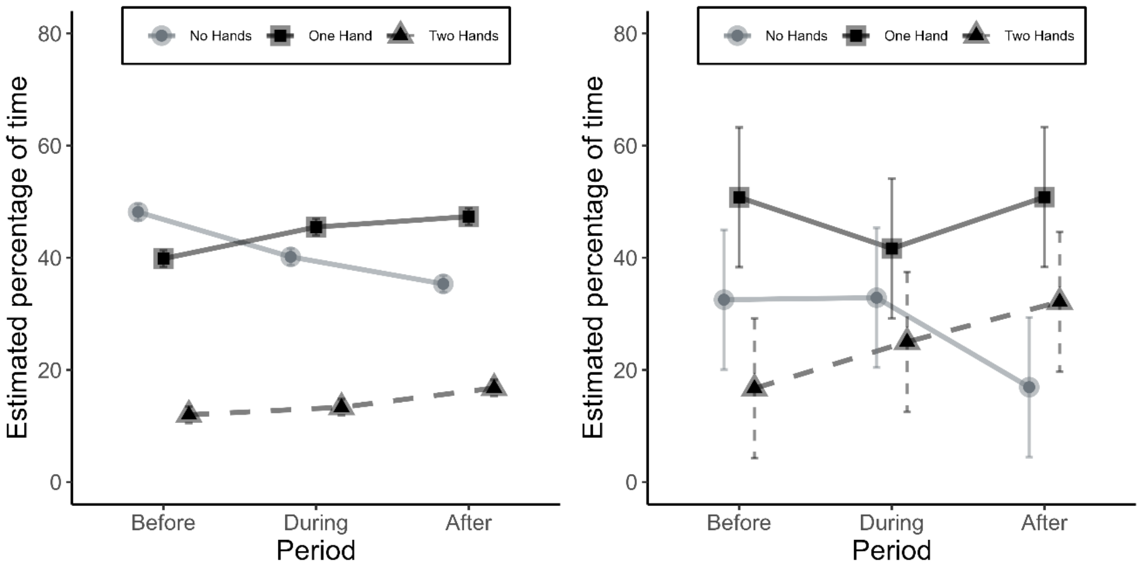

Figure 5. Estimated percentage of time with no hands, one hand, and two hands on the wheel before, during, and after A1-only alert (left) and Escalated alert sequences (right).

Note. Error bars represent 95% confidence intervals. Error bars are very small for the A1-only epochs and may not be visible in the figure.

Among all the epochs in which drivers' hands were off the wheel at the first alert's onset, drivers took an average of 2.6 sec (*SD* = 2.2 sec) to place at least one hand on the wheel. Linear mixed-effects models with a driver-specific random intercept were used to estimate the time to hands on wheel across study weeks. We found that response time from the first alert onset to when drivers placed their hands on the wheel remained relatively consistent over the study weeks, from 2.77 sec in week 1 (95% CI [2.21, 3.33 sec], $p < 0.001$) to 2.81 sec in week 4 (95% CI [2.21, 3.33 sec], $p < 0.001$). Week 3 was associated with the shortest response time, 2.6 sec (95% CI [1.99, 3.20 sec], $p < 0.001$), although there was no significant difference between weeks (see Figure 6, left).

In 37% of the epochs (n = 1,155) drivers removed both of their hands from the steering wheel during the period between after the alert sequence ended and before the epoch concluded. In those epochs, the average time it took drivers to take their hands off the wheel following the alert sequence was 1.60 sec (*SD* = 1.92 sec). Analysis of time to hands-off-wheel following the end of an alert sequence across the 4 weeks indicated that hands-off-wheel time following an alert decreased over time, with a significant decrease between week 1 and week 4 (week 4, week 1 = −1.09 sec, 95% CI [−1.63, −0.53 sec], $p < 0.001$) (see Figure 6, right). This meant that, while drivers responded to the alert in the same amount of time, as time went on drivers were quicker to remove their hands from the wheel once the alert had ended.

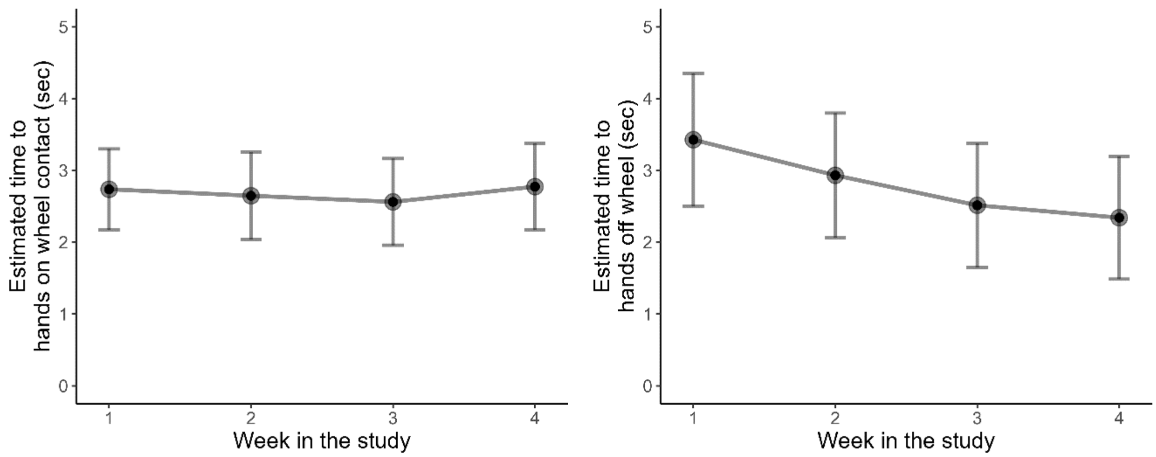

Figure 6. Estimated response time for drivers to put their hands on the wheel after alert onset (left) and to take both off the wheel at the end of the alert sequence (right) as a function of week in the study.

Note. Error bars represent 95% confidence intervals.

Glance behavior. Drivers spent the majority of their time from before the alert sequence started to after it ended (m = 74%, SD = 21%) with glances directed to the on-road AOI (at the forward roadway and mirrors/windows). Glances to the off-road (m = 11%, SD = 21%) and display AOIs (m = 16%, SD = 15%) made up most of the other glance locations. A1-only alerts were associated with a higher proportion of time glancing at the road than Escalated alerts (A1-only alerts, Escalated alerts = 18% difference, 95% CI [13, 23%], $p < 0.001$).

Linear mixed-effects models with a driver-specific random intercept were used to estimate the percentage of time drivers glanced at the on-road, display, and off-road AOIs. For A1-only epochs, off-road glances increased by 9 percentage points (95% CI [7, 11%], $p < 0.001$) in week 4 compared with week 1, and glances to the display decreased by 8 percentage points (95% CI [6, 10%], $p < 0.001$) over the same period (see Figure 7, left). For Escalated alert epochs, there was a 26 percentage point increase (95% CI [−6, 59%], $p = 0.12$) in off-road glance behavior in week 4 of the study compared with week 1 and a 23 percentage point decrease (95% CI [−55, 10%], $p = 0.19$) in glances on-road over the same period, although these results were not statistically significant (see Figure 7, right).

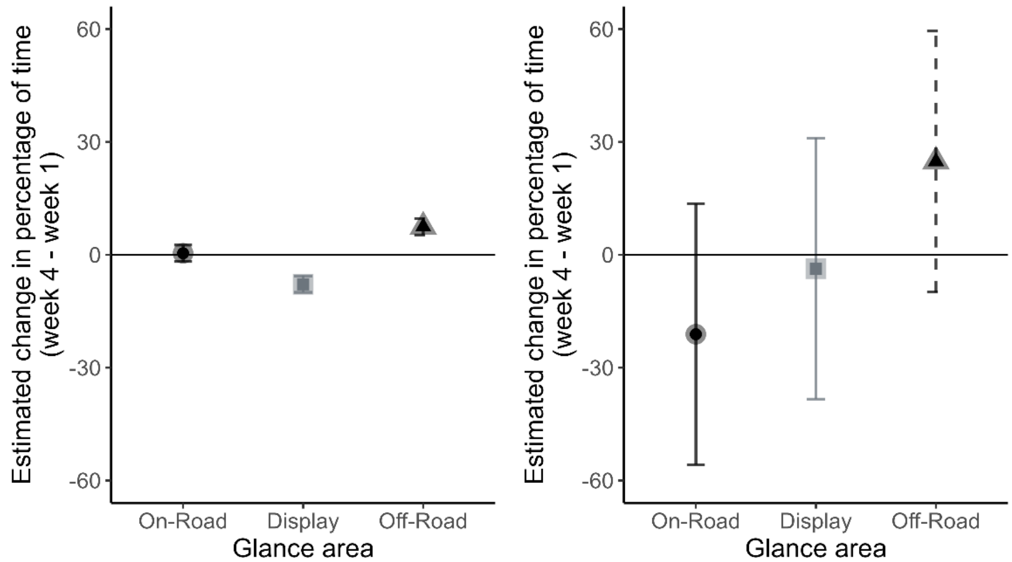

Figure 7. Estimated difference between week 4 and week 1 in the percentage of time looking to the on-road, display, and other off-road AOI for A1-only epochs (left) and Escalated alert epochs (right).

Note. Error bars represent 95% confidence intervals. Error bars are very small for the A1-only epochs and may not be visible in the figure.

Linear mixed-effects models with a driver-specific random intercept were used to estimate the percentage of time looking at the on-road, at the display, and off-road AOIs before, during, and after the alert sequence. In A1-only epochs (Figure 8, left), drivers spent an estimated 77% (95% CI [76, 78%], $p < 0.001$) of their time looking towards the road before the alert began. During the A1-only alert period, on-road glances significantly decreased ($p < 0.001$) to an estimated 56% (95% CI [56, 57%], $p < 0.001$), but glances to that AOI returned to near before-alert-onset values once the alert ended ($m = 74\%$, 95% CI [73, 75%]). The decrease in glances to the road coincided with a significant increase in glances to the display ($p < 0.001$), increasing from an estimated 12% of glances before the alert (95% CI [11, 13%], $p < 0.001$) to 32% of glances during it (95% CI [31, 33%], $p < 0.001$). Glances to other off-road locations did not change substantially ($p_{before\text{-}during} = 0.11$, $p_{during\text{-}after} < 0.001$, $p_{before\text{-}after} = 0.10$) over the alert periods ($m_{\text{before}} = 11\%$, 95% CI [10, 12%], $p < 0.001$; $m_{\text{during}} = 12\%$, 95% CI [11, 13%], $p < 0.001$; $m_{\text{after}} = 9\%$, 95% CI [8, 10%], $p < 0.001$). This means that drivers tended to look away from the road towards the display during the A1-only alerts, but quickly returned to their pre-alert glance behavior once it ended.

For Escalated alert epochs, glances to the on-road AOI decreased ($p = 0.12$) with alert onset ($m_{\text{before}} = 62\%$, 95% CI [50, 74%], $p < 0.001$; $m_{\text{during}} = 45\%$, 95% CI [33, 57%], $p < 0.001$) and glances to the display increased compared with their pre-alert levels ($p = 0.30$) ($m_{\text{before}} = 19\%$, 95% CI [7.5, 30%], p

= 0.002; m_{during} = 31%, 95% CI [20, 42%], $p < 0.001$) (Figure 8, right). Drivers spent 19% of their time before the alert sequence (95% CI [8, 31%], $p = 0.002$) and 24% during it (95% CI [12, 35%], $p < 0.001$) looking towards off-road locations. Once the alert sequence ended, off-road glances returned to pre-alert phase levels at an estimated 20% of glances directed off-road (95% CI [8, 32%], $p = 0.001$), and on-road glances increased slightly to an estimated 50% (95% CI [39, 62%], $p < 0.001$). Although these differences were not significant, the trends indicate that there may be changes to glance behaviors, especially off-road glances, associated with Escalated alerts.

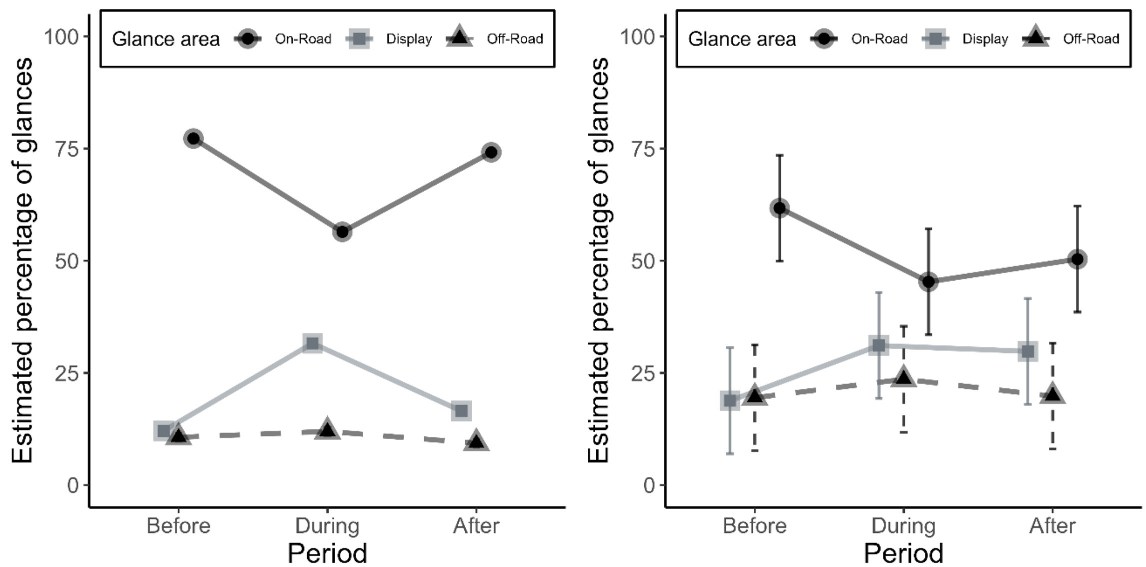

Figure 8. Estimated percentage of time glancing at on-road, display, and off-road AOI before, during, and after A1-only alert (left) and Escalated alert sequences (right).

Note. Error bars represent 95% confidence intervals. Error bars are very small for the A1-only epochs and may not be visible in the figure.

Secondary activity. Driving-related secondary activities were the most prevalent activity type in the dataset, with 72% of the epochs (n = 2,346) containing at least one driving-related activity. Non-driving-related activities were less prevalent, with visual-manual activities present in 46% (n = 1,505), voice-based activities in 16% (n = 526), manual activities in 9% (n = 307), and visual-only activities in less than 1% of the epochs (n = 2). Due to the limited data for some non-driving-related activity categories, for the remaining analysis, we grouped the non-driving-related activities into one category. In total, non-driving-related activities were present in 55% of the epochs (n = 1,783). On average, drivers spent 13% (SD = 16%) and 35% of their time (SD = 42%) engaged in either driving-related or non-driving-related secondary activities, respectively. Epochs in which secondary activity information was missing were excluded from analysis ($n = 609$).

Over time there was a shift in driver secondary activity engagement associated with A1-only alerts with a 9 percentage point increase (95% CI [6, 12%], $p < 0.001$) in non-driving-related activity engagement and a 4 percentage point decrease (95% CI [1, 7%], $p < 0.001$) in driving-related activity engagement in week 4 compared with week 1 (Figure 9, left). We did not observe a significant change in secondary activity behavior within Escalated alert epoch over time in the study (Figure 9, right).

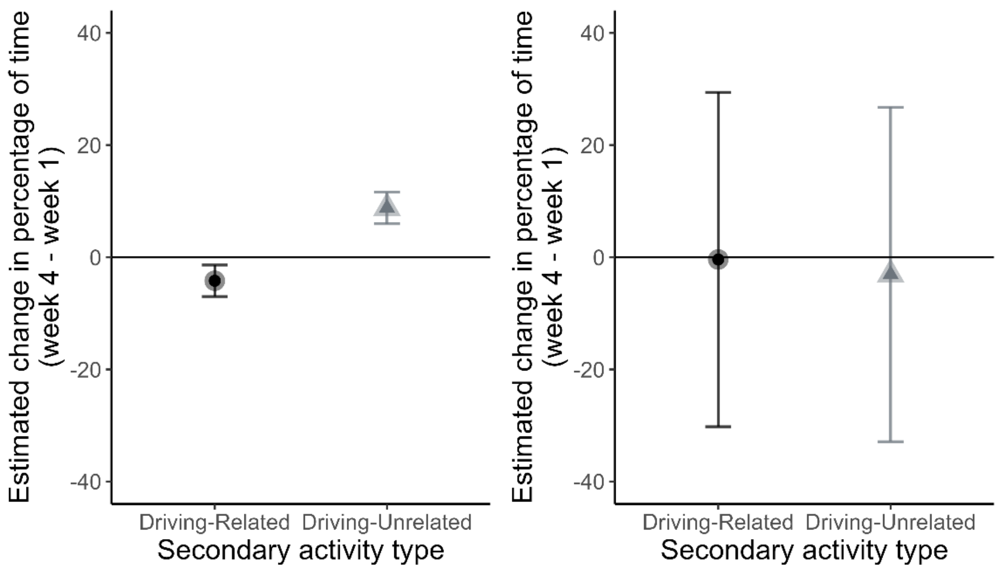

Figure 9. Estimated percentage of time engaged in driving-related and non-driving-related tasks before, during, and after A1-only alert epochs (left) and escalated alert epochs (right).

Note. Error bars represent 95% confidence intervals.

Linear mixed-effect models with a driver-specific random intercept were used to estimate the proportion of driving-related and non-driving-related secondary activities before, during, and after A1-only and Escalated alerts. In A1-only alert epochs (Figure 10, left), drivers were least likely to engage in driving-related secondary activities prior to the alert ($m = 4\%$, 95% CI [-0.7%, 8%], $p = 0.11$), but significantly more likely ($p < 0.001$) to engage in driving-related secondary activities during the alert ($m = 16\%$, 95% CI [11, 20%], $p < 0.001$). This finding aligns with the increase in the proportion of glances to the center stack. Engagement in non-driving-related activities decreased significantly ($p < 0.001$) over the alert period from before the alert ($m = 26\%$, 95% CI [21, 31%], $p < 0.001$) to after the alert ($m = 23\%$, 95% CI [19, 28%], $p < 0.001$).

In Escalated alert epochs shown in Figure 10 (on right), engagement in driving-related activity increased slightly over the alert period with an estimated engagement of 17% before the alert (95% CI [-0.2, 34%], *p* =0.05) to 27% after it (95% CI [10, 44%], *p* = 0.003). Engagement in non-driving-related secondary activities remained constant with no significant change in engagement before the alert (*m* = 49%, 95% CI [32, 66%], *p* < 0.001) to after the alert (*m* = 47%, 95% CI [30, 64%], *p* < 0.001).

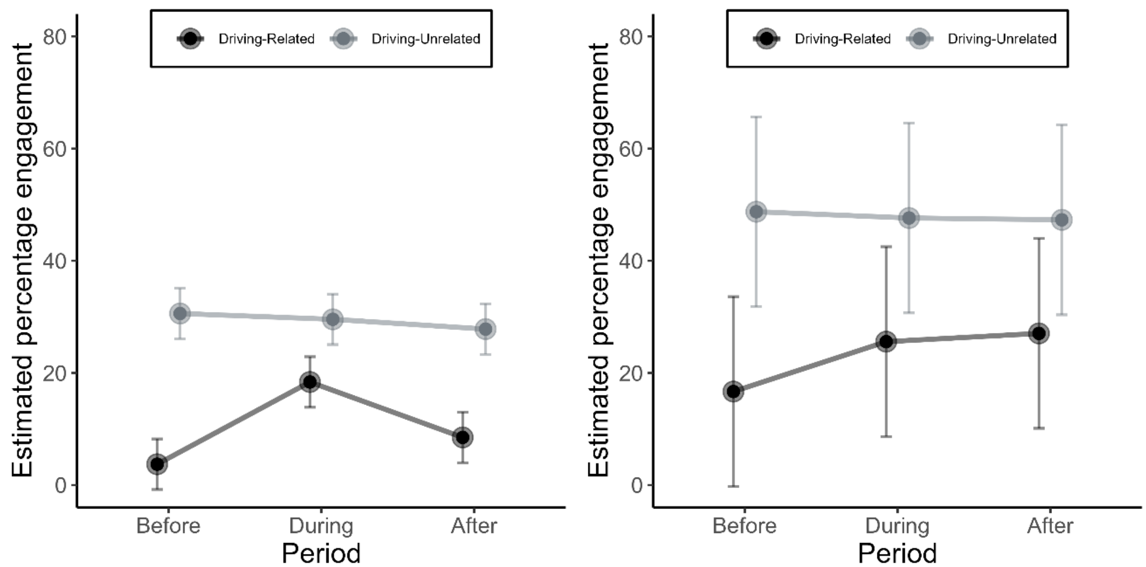

Figure 10. Estimated percentage of time engaged in driving-related and non-driving-related secondary activities before, during, and after A1-only alert sequences (left) and Escalated alert sequences (right).

Note. Error bars represent 95% confidence intervals.

To determine whether the probability of an alert escalating varied with overall driver behavior, we estimated escalation probability using a logistic mixed-effects model with a driver-specific random intercept (Table 7). Glances to the display were not modeled in this analysis due to their high correlation with driving-related secondary activity: 82% of driving-related secondary activities were associated with reading/watching the display. Only epochs in which hand activity, glance behavior, and secondary activities were annotatable were included in this analysis and, therefore, the model was fit on 2,485 alerts with 30 escalated alert sequences. Odds ratios larger than 1 indicate that the predictor was associated with a higher likelihood of escalation, while odds ratios less than 1 indicate the predictor was associated with a lower likelihood of escalation. Hand activity, glance behavior, and secondary activities were modeled as a proportion of time, meaning that the odds ratios can be interpreted as the change in the probability of alert escalation given that the predictor of interest changes in 1 unit—in this case, 1 unit is 10 percentage points of the proportion of time performing a behavior. For example, if the driver increases the percentage of their two-hands-on-wheel time from 10% to 20%, they will decrease the odds ratio for alert escalation by 0.15. Epochs containing escalated alert sequences were characterized by a higher proportion of non-driving-related secondary activity engagement. Conversely, we saw that higher proportions of time with two hands on the wheel corresponded with a lower likelihood of escalation. Escalated alert sequences were also less likely to occur over time.

Table 7. Odds of alert escalation based on driver behavior.

Predictors	**Odds ratios**	**CIs**	***p* value**
Week in the study	0.64	0.47, 0.88	0.006
10 percentage-point increase in proportion of hands activity			
One hand	0.94	0.86, 1.03	0.163
Two hands	0.85	0.76, 0.97	0.015
10 percentage-point increase in proportion of glance behavior			
Off-road	1.11	0.86, 1.44	0.436
On-road	0.88	0.67, 1.15	0.339
10 percentage-point increase in proportion of secondary activity			
Driving-related	1.02	0.81, 1.29	0.855
Non-driving-related	1.09	1.02, 1.17	0.010

Note. CIs = confidence intervals.

Discussion

There are many reasons why people multitask while driving, such as addictive behavior around smartphones (Sapacz, Rockman, & Clark, 2016) and the need to maintain cognitive arousal when conditions are under stimulating (Cunningham & Regan, 2018; Gershon et al., 2009; Yin, Shao, & Zhang, 2024). Although many partially automated systems have safeguards intended to limit multitasking, many drivers continue to engage in these behaviors (Insurance Institute for Highway Safety [IIHS], 2024). In addition to having the lowest fatal crash rates of any driving environment, limited-access highways are considered to be the easiest for both humans and machines to drive in because of the wide, fairly straight roadways, with usually well-marked lane boundaries and consistent traffic patterns (Chen et al., 2019; National Center for Statistics and Analysis, 2022). As such, it makes sense that most partially automated systems have been developed to operate primarily in these conditions. The downside to this, though, is that because automation is meant to make driving even easier—and therefore can be under stimulating—its use tends to correspond with increased non-driving-related multitasking (Dunn et al., 2021; Noble et al., 2021; Reagan et al., 2021). This has been observed with different drivers and systems, indicating that it is a widespread phenomenon not unique to a single automaker.

The current study has shown that driver interactions with partial automation are dynamic. Some of the changes we observed indicate that system safeguards can beneficially shape behavior both immediately and in the longer term, whereas other patterns revealed potentially unintended consequences. It is important to note that these findings are likely not unique to Tesla's Autopilot, as many systems on the market have overtly similar safeguard designs. As such, some observations from this study maybe relevant to other driver assistance technology that still requires the driver to be engaged in the driving task.

Attention reminders

Certain behaviors correspond with changes in alert sequences over time. Having two hands on the wheel decreased the likelihood of attention reminder escalation, and fewer escalated attention reminders were observed as time went on. As there was a notable change in driver behavior after the first week, we conclude it was within a relatively brief period that participants learned how Autopilot responds to hands-off-wheel behavior. This internalization of the attention reminder sequence and timing revealed a trade-off in behavior over time. Drivers exhibited more instances of sustained disengagement and escalated attention reminders at the beginning of the study than at the end of it. As the study progressed, drivers changed their behavior to more frequent, but briefer periods of disengagement that were still sufficiently long to lead to an alert. This corresponded with an increase in A1-only alerts towards the end of the study compared with the beginning of it. We also saw a tendency for some drivers to become increasingly quicker to again remove their hands once the alerts stopped.

In general, the longer people are disengaged from the driving task, such as when looking away from the road, the more likely they are to be involved in a crash (e.g., Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). Given that partial driving automation is not a substitute for the driver's role in monitoring and maintaining situation awareness, the same principle likely applies to its use. In fact, there might be added risk when combining disengagement from driving with overreliance on and complacency with automation (Lin et al., 2018; Schneider et al., 2022; Victor et al., 2018) because these systems may abruptly require rapid driver intervention (American Automobile Association, Inc, 2020; IIHS, 2018). Nevertheless, while the increase in driver disengagement is not desired, our results show that escalating multimodal attention reminders do modify behavior in the moment and over time.

Last-resort countermeasures

When we began this study, we did not expect to see many, if any, emergency-slowdown-to-lockout events in the data; yet we observed 16 lockout events among four of the 14 participants. The driver with 12 lockouts demonstrates just how persistent the tendency for sustained disengagement can be in some individuals, as the litany of videos on social media can further attest. Three drivers also attempted to reactivate the system after receiving the lockout notification. The reactivation attempts show that the emergency slowdown on its own may not be as effective for some people as when it is paired with a lockout. Moreover, the reactivation attempts also suggest that communication around the lockout may either not be seen or it is not intuitive enough for some drivers to comprehend, which would not be surprising because it is a challenging concept to convey in a visual display. Many vehicle owners, including ones that have the lockout safeguard in their personal vehicles, misunderstand the countermeasure in general (Mueller et al., 2024).

Nevertheless, the learning curve we observed with attention reminders also exists for the last-resort countermeasures, as lockouts occurred predominantly within the first half of the study, with only one lockout event happening in the fourth week. The driver who had the most lockout events also grew quicker to respond to them over time. As the emergency slowdown automatically removes access to the system for the rest of the trip, these findings suggest that those last-resort safeguards are effective deterrents of excessive system misuse in both the heat of the moment and the longer term, especially among individuals who are prone to repeated instances of prolonged disengagement from driving. While speculative, it may be worth investigating whether, within a given trip, these safeguards might further benefit from factoring in alert activation frequency, how far those instances escalated, monitored behavior post-alert sequence termination, and time since the last attention reminder activation. These factors could potentially help to refine the subsequent onset and escalation of not only the alerts but also of the last-resort countermeasures in a way that adapts to the individual driver as opposed to a one-size-fits-all escalation sequence approach. Moreover, it would seem that overall earlier activation and more rapid escalation of the alerts might help to minimize windows of opportunity where drivers become disengaged from the driving, although further testing is required to confirm this hypothesis.

The nuances of system communication may depend on the vehicle's design; however, putting brand-level idiosyncrasies aside, the automotive industry as an ecosystem may benefit from federal or other guidance on the best design practices to promote safe and effective return-to-the-loop behavior. The National Highway Traffic Safety Administration's (NHTSA's) official guidance on human-machine interfaces for driver assistance technology is almost a decade out of date (see Campbell et al., 2016), but they are presently leading a multiyear data collection effort to understand the real-world effects of commercially available safeguards (NHTSA, 2022). In the meantime, IIHS (2024), which is an independent non-profit scientific and educational organization devoted to traffic safety, has released a vehicle ratings program on partial automation safeguards to push for empirically supported design solutions that address driver disengagement and system misuse (Mueller et al., 2021). The intent of this effort is to proactively encourage constructive dialog and action on this topic. Leveraging existing data from the MIT AVT Consortium, this paper details the real-world effects of the countermeasures associated with the most frequently used partial driving automation system with an aim to further advance constructive international dialog and action.

Limitations

A limitation of this study, and any other similar investigation, is that we could not disentangle the influence of familiarity with Autopilot and the vehicle itself. Our participants were inexperienced with both the Tesla Model 3 vehicle and Autopilot, and this simultaneous exposure to technology and vehicle is consistent with what any new vehicle owner experiences. Although the system's behavior and communication affect how people drive, the driving interaction itself is constrained by the vehicle's configuration and controls. We can therefore assume that the changes observed in this study reflect the influence of both types of familiarity and are more likely reflective of new owners than experienced ones. Without longer observation periods to compare our data with, we do not know whether the behavior we observed by week 4 reflects stabilized system and vehicle usage patterns; however, other research from the MIT AVT Consortium is currently underway to address this question through naturalistic observation of Tesla vehicle owners in their personal vehicles over extended periods of time.

Although we showed that there was an increase in A1 alerts and the duration of secondary tasks within the epochs examined over time, we did not examine how driver disengagement manifests over the course of the whole trip. Another limitation was that the analysis does not lend itself to assessing the overall degree and nature of the consequences of the driver disengagement over time. In addition, our analysis of behavior surrounding emergency-slowdown-to-lockout events was limited because the majority of those events came from one driver. Nevertheless, we have shown that last-resort countermeasures might activate more frequently than previously believed among select vehicle users. Our findings suggest that while drivers who use the technology appropriately will be unaffected by the implementation of those safeguards, other drivers—possibly high-risk drivers— evidently may need them.

Missing data also restricted the analysis, particularly for eye glances. Unfortunately, this is often the case with naturalistic observation research. Sunglasses and nighttime driving conditions were among the many circumstances that limit the ability to accurately code eye glances, and we erred on the side of caution when it came to making determinations about where the driver was looking in those epochs. Fortunately, camera technology continues to evolve, but precise gaze analysis still requires eye-tracking equipment that is not currently practical to implement in a field study like this. More research is needed to better understand the nuances around the relationship between hand activity and gaze patterns, especially when it comes to vehicle-monitoring strategies to detect driver disengagement.

Conclusions

Learning is not an inherently good or bad thing, but it does happen with exposure to partial automation over time. Drivers in this study demonstrated the ability to identify windows of opportunity for non-driving-related activities while learning to avoid the escalation process of the attention reminders. Part of the safeguard solution may include adaptive activation and escalation upon the detection of persistent driver disengagement. Continuous monitoring and adaptation of countermeasure onset and escalation may be instrumental in addressing the idiosyncratic disengagement tendencies of a diverse driving population.

Acknowledgments

Data collection was supported by the MIT AVT Consortium. This analytic study was funded by the Insurance Institute for Highway Safety. We are grateful to John Liebengood, Dan Purdy, Skye King, and Jeff Babcox for their help with producing video stills displayed in several figures of this paper.

References

Abraham, H., Seppelt, B., Mehler, B. & Reimer, B. (2017). What's in a name: Vehicle technology branding & consumer expectations for automation. *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicle Applications (AutomotiveUI '17)*, Oldenburg, Germany. pp. 226–234.

American Automobile Association, Inc. (2020). *Evaluation of active driving assistance systems.* https://newsroom.aaa.com/asset/report-fact-sheet-evaluating-active-driver-assistance-systems/

Atwood, J. R., Guo, F., & Blanco, M. (2019). Evaluate driver response to active warning system in level-2 automated vehicles. *Accident Analysis & Prevention*, *128*, 132–138. https://doi.org/10.1016/j.aap.2019.03.010

Campbell, J. L., Brown. J. L., Graving, J. S., Richard, C. M., Lichty, M. G., Sanquist, T., Bacon, P., Woods, R., Li, H., Williams, D. N., & Morgan, J. F. (2016). *Human factors design guidance for driver-vehicle interfaces* (Report No. DOT HS 812 360). National Highway Traffic Safety Administration.

Chen, S., Usman Saeed, T., Alinizzi, M., Lavrenz, S., & Labi, S. (2019). Safety sensitivity to roadway characteristics: A comparison across highway classes. *Accident Analysis & Prevention*, *123*, 39–50. https://doi.org/10.1016/j.aap.2018.10.020

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

Cicchino, J. B. (2018). Effects of lane departure warning on police-reported crash rates. *Journal of Safety Research*, *66*, 61–79. https://doi.org/10.1016/j.jsr.2018.05.006

Cunningham, M. L., & Regan, M. A. (2018). Driver distraction and inattention in the realm of automated driving. *IET Intelligent Transport Systems*, *12*(6), 407–413. https://doi.org/10.1049/iet-its.2017.0232

Dunn, N., Dingus, T., Soccolich, S., & Horrey, W. (2021). Investigating the impact of driving automation systems on distracted driving behaviors. *Accident Analysis &Prevention*, *156*, 106152. https://doi.org/10.1016/j.aap.2021.106152

Gershon, P., Ronen, A., Oron-Gilad, T., & Shinar, D. (2009). The effects of an interactive cognitive task (ICT) in suppressing fatigue symptoms in driving. *Transportation research part F: traffic psychology and behaviour*, *12*(1), 21–28. https://doi.org/10.1016/j.trf.2008.06.004

Gershon, P., Seaman, S., Mehler, B., Reimer, B., & Coughlin, J. (2021). Driver behavior and the use of automation in real-world driving. *Accident Analysis & Prevention*, *158*, 106217. https://doi.org/10.1016/j.aap.2021.106217

Highway Loss Data Institute [HLDI]. (2023). Compendium of HLDI collision avoidance research. *Bulletin*, *40*(4), 1–23.

Insurance Institute for Highway Safety [IIHS]. (2018). IIHS examines driver assistance features in road, track tests. *Status Report*, *53*(4), 3–6. https://www.iihs.org/news/detail/iihs-examines-driver-assistance-features-in-road-track-tests

Insurance Institute for Highway Safety [IIHS]. (2024). *First partial driving automation safeguard ratings show industry has work to do.* https://www.iihs.org/news/detail/first-partial-driving-automation-safeguard-ratings-show-industry-has-work-to-do

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Report No. DOT HS 810 594). Virginia Tech Transportation Institute.

Lin, R., Ma, L., & Zhang, W. (2018). An interview study exploring Tesla drivers' behavioral adaptation. *Applied Ergonomics*, *72*, 37–47. https://doi.org/10.1016/j.apergo.2018.04.006

Llaneras, R., Cannon, B., & Green, C. (2017). Strategies to assist drivers in remaining attentive while under partially automated driving: Verification of human-machine interface concepts. *Transportation Research Record*, *2663*, 20–26. https://doi.org/10.3141/2663-03

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*, *15*, 3–27. https://doi.org/10.1177/1555343420983126

Mueller, A. S., Cicchino, J. B., & Calvanelli, J. V. Jr. (2024). Habits, attitudes, and expectations of regular users of partial driving automation systems. *Journal of Safety Research*, *88*, 125–134. https://doi.org/10.1016/j.jsr.2023.10.015

National Center for Statistics and Analysis. (2022, May). *Early estimates of motor vehicle traffic fatalities and fatality rate by sub-categories in 2021* (Crash•Stats Brief Statistical Summary. Report No. DOT HS 813 298). National Highway Traffic Safety Administration.

National Highway Traffic Safety Administration. (2022, September 28). *Contract with Virginia Polytechnic Institute and State University: Field data collection with consumer available driver support systems* (Award ID 693JJ922F00192N). Department of Transportation. https://www.fpds.gov/fpdsng_cms/index.php/en/

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).

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).

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* (Report No. Highway Accident Report NTSB/HAR-20/01).

Noble, A., Miles, M., Perez, M., Guo, F., & Flauer, S. (2021). Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention*, *151*, 105959. https://doi.org/10.1016/j.aap.2020.105959

Politis, I., Brewster, S., & Pollick, F. (2013). Evaluating multimodal driver displays of varying urgency. *AutomotiveUI '13: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 92–99. https://doi.org/10.1145/2516540.2516543

Reagan, I. J., Teoh, E. R., Cicchino, J. B., Gershon, P., Reimer, B., Mehler, B., & Seppelt, B. D. (2021). Disengagement from driving when using automation during a 4-week field trial. *Transportation Research Part F*, *82*, 400–411. https://doi.org/10.1016/j.trf.2021.09.010

Reimer, B., Mehler, B., Muñoz, M., Dobres, J., Kidd, D., & Reagan, I. (2021). Patterns in transitions of visual attention during baseline driving and during interaction with visual-manual and voice-based interfaces. *Ergonomics*, *64*, 1429–1451. https://doi.org/10.1080/00140139.2021.1930197

SAE International. (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles* (SAE Standard J3016, Report No. J3016_202104). https://doi.org/10.4271/J3016_202104

Sapacz, M., Rockman, G., & Clark, J. (2016). Are we addicted to our cell phones? *Computers in Human Behavior*, *57*, 153–159. https://doi.org/10.1016/j.chb.2015.12.004

Schneider, N., Ahrens, L., & Pruksch, A. (2022). Controllability of lateral drift failures while driving with SAE Level 2 advanced driver assistance systems. *14th Uni-DAS e.V. Workshop on Driver Assistance and Automated Driving*, 183–193. https://www.uni-das.de/fas-workshop/2022.html

Teoh, E. (2020). What's in a name? Drivers' perceptions of the use of five SAE Level 2 driving automation systems. *Journal of Safety Research*, *72*, 145–151. https://doi.org/10.1016/j.jsr.2019.11.005

Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Ljung Aust, M. (2018). Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel. *Human Factors*, *60*, 1095–1116. https://doi.org/10.1177/0018720818788164

Yin, J., Shao, H., & Zhang, X. (2024). The monitoring requests on young driver's fatigue and take-over performance in prolonged conditional automated driving. *Journal of Safety Research*, *88*, 285–292. https://doi.org/10.1016/j.jsr.2023.11.015