How Will Drivers Take Back Control in Automated Vehicles? A Driving Simulator Test of an Interleaving Framework

Divyabharathi Nagaraju
University of New Hampshire
Durham, New Hampshire, USA

Alberta A. Ansah
University of New Hampshire
Durham, New Hampshire, USA

Nabil Al Nahin Ch
University of New Hampshire
Durham, New Hampshire, USA

Caitlin Mills
University of New Hampshire
Durham, New Hampshire, USA

Christian P. Janssen
Utrecht University
Utrecht, Netherlands

Orit Shaer
Wellesley College
Wellesley, Massachusetts, USA

Andrew L. Kun
University of New Hampshire
Durham, New Hampshire, USA

ABSTRACT

We explore the transfer of control from an automated vehicle to the driver. Based on data from N=19 participants who participated in a driving simulator experiment, we find evidence that the transfer of control often does not take place in one step. In other words, when the automated system requests the transfer of control back to the driver, the driver often does not simply stop the non-driving task. Rather, the transfer unfolds as a process of interleaving the non-driving and driving tasks. We also find that the process is moderated by the length of time available for the transfer of control: interleaving is more likely when more time is available. Our interface designs for automated vehicles must take these results into account so as to allow drivers to safely take back control from automation.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **Empirical studies in HCI**.

KEYWORDS

automated driving, transfer of control, interleaving framework

ACM Reference Format:
Divyabharathi Nagaraju, Alberta A. Ansah, Nabil Al Nahin Ch, Caitlin Mills, Christian P. Janssen, Orit Shaer, and Andrew L. Kun. 2021. How Will Drivers Take Back Control in Automated Vehicles? A Driving Simulator Test of an Interleaving Framework. In *AutoUI '21: 13th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, Sept 09–14, 2021, Virtual Conference.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3409118.3475128

AutoUI '21, Sept 09–14, 2021, Virtual Conference

ACM ISBN 978-1-4503-8063-8/21/09…$15.00
https://doi.org/10.1145/3409118.3475128

1 INTRODUCTION

The Society of Automotive Engineers (SAE) standardized vehicle automation types into six categories, from level-0 which designates vehicles that have no automation features, all the way to level-5 which designates vehicles where human control is not necessary for any part of the trip [23]. The vast majority of cars on the road today operate at level-0: the driver is fully in charge of all aspects of vehicle control. The vehicles with the most advanced automation features currently operate at level-2. In these vehicles automation can control lateral and longitudinal position of the vehicle, but even while automation is engaged the driver must constantly pay attention to the road and must be ready to intervene immediately if the automation fails.

The next level of automation that we can expect in cars is level-3 automation. Here drivers will not need to drive for extended periods of time, and will be able to engage in non-driving tasks. However, they will also need to be able to return to driving when the system indicates that it can no longer control the vehicle. The SAE standard states that the automation must provide drivers "sufficient time" to take back control - while this language does not prescribe exactly how long or short this time period is, we can expect that the first vehicles with level-3 automation will indeed require drivers to take control of the vehicle quickly, perhaps as quickly as within 10 seconds of the initial signal from the system [9].

We therefore need to understand how best to support drivers as they perform non-driving tasks while automation controls the vehicle such that they can also safely take back control of the vehicle when needed. This is a multi-faceted problem; we need to understand how the characteristics of different non-driving tasks and different user interfaces will influence the ability of drivers to both complete the non-driving tasks and, critically, safely take back control of the vehicle. In this paper we focus on one aspect of this broad problem: how drivers transition from the non-driving task back to driving.

The transition from a non-driving task back to driving has been explored by a number of researchers [21, 22, 24, 26, 27]. We focus on the model proposed by Janssen et al. [11]. This model proposes that the transition is not a one-step process where drivers simply respond to a request to take back control by stopping the non-driving task and taking up the driving task. Rather, Janssen and

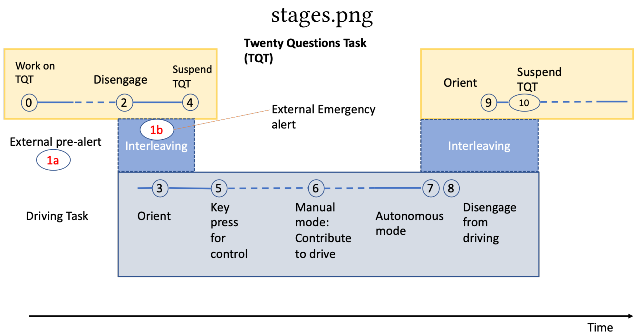

Figure 1: Model showing the stages of transition between driving and non-driving task in a conditionally automated vehicle [11]. In our experiment the non-driving task is playing the twenty questions task (TQT).

colleagues argue that the process has multiple stages, including disengagement from the non-driving task, and ultimately taking control of the vehicle. Furthermore, they argue that for some period of time, drivers will likely interleave the non-driving and driving tasks before completely ending the non-driving task and focusing (hopefully) exclusively on the driving task.

Figure 1 shows the theoretical model [11] that describes the transition stages during the transition of control in an automated vehicle. Figure 1 shows the model in the context of our experiment where participants engage in the twenty questions game during automated driving (see the Tasks section below). The stages of the model are as follows: (0) the driver is working on a non-driving task while automation controls the vehicle, (1) the driver receives external warning that their input is needed for the driving task (a pre-alert and later, if needed, an emergency alert), (2) they disengage for the first time from their original task to start a period of interleaving attention between the original task and the driving task, (3) they orient towards the driving task, (4) they fully suspend their original task, (5) they press a button to indicate that they have taken control of the vehicle, (6) the driver controls the vehicle in manual driving, which is followed by another interleaving period during which (7) the system signals that the driver no longer needs to provide input to the car, (8) the driver disengages from driving, (9) orients to their original non-driving task, and (10) resumes the suspended activities on their original task (the twenty questions game in our experiment).

While the model by Janssen et al. [11] is based on similar models that describe task switching in other domains (e.g., [2, 3]), this particular model has not yet been empirically validated. Thus, we pose the following research questions:

RQ1. Do drivers interleave the non-driving and driving tasks as they take back control of an automated vehicle? If yes, how is the probability of interleaving affected by the time available to take back control?

RQ2. If drivers do interleave, what are they using the time for during interleaving, and how does this depend on the time available to take back control?

2 PRIOR WORK

A number of researchers explored the question of how long it takes for a driver to take back control of a vehicle safely and the various factors that could affect this take-over time (e.g., [5, 8, 25, 27]). The time needed to take over depends on how long the driver needs to gather information from the environment and develop sufficient situational awareness to then act accordingly. Visual-manual non-driving tasks, in particular, seem to lead to poorer driving performance [20].

Gold et al. explored the relationship between the length of time that is available to a driver to take back control, and how quickly they actually take back control [8]. The results show that with a shorter available time, the driver comes to a decision more quickly, reacts faster, but the quality is generally worse. In their experiment, 8 seconds appeared to be sufficient for the driver to understanding the situation and take back control.

Walch et al. recorded driver performance during take-over with different assisting system methods [25]. The experiment tested different warnings and handover procedures, including alerts and take over requests (TOR) with 4 seconds and 6 seconds. Different handover procedures include immediate handover, step-wise handover, and system monitored handover. The results show that participants preferred the combination of auditory and visual handover assistant with 6 seconds of TOR time.

Mok et al. examined the unstructured transition timing for distracted drivers in automated vehicles when the automation is off due to an emergency situation with three different take-over times – 2 seconds, 5 seconds, and 8 seconds [19]. The study recommends that the minimum amount of time needed for the transition of control is between 2 seconds and 5 seconds.

The above efforts provide important information about how take-over time and success are affected by factors such as the type of non-driving task, and type of take-over alert. However, they mostly treat take-over as a single event. Yet from prior work in other domains we know that people often handle interruptions by interleaving the ongoing and interrupting task for a while [2, 3, 6]. In the work presented here we extend our understanding of how drivers take back control from automation by exploring the steps that are involved in switching between the non-driving and driving tasks, arguing that this switch is not a one-step process (cf. [11]).

In our experiment we use a word game, the twenty questions game (described in the next section), as the non-driving task. A version of this game was used in prior driving-related research [15, 18], as were other games (e.g. Taboo [12, 14], and last-letter game [15]). We chose twenty questions as it is an engaging game, it allows scoring participant performance, and it has a clear discourse structure which allows analysis of how the context of the game affected driving behaviors.

3 METHOD

We designed an experiment to study how drivers switch between a driving and a non-driving task in an automated vehicle. Participants were seated in front of a driving set-up where they periodically had to control a simulated vehicle. When not in control of the vehicle they played a word game with an experimenter.

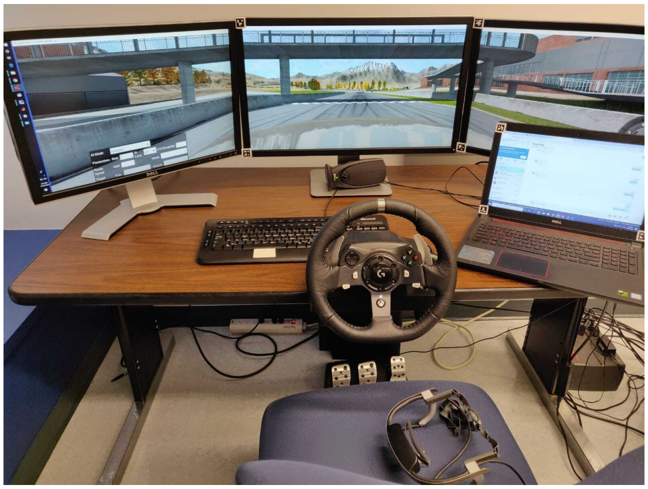

Figure 2: Experiment setup showing the three monitors used as the driving display, the gaming steering wheel and pedals used to control the simulated vehicle, the PC keyboard used to signal that the driver is assuming manual control of the simulated vehicle, the laptop used for the Twenty Questions Task (TQT), and the eye tracker

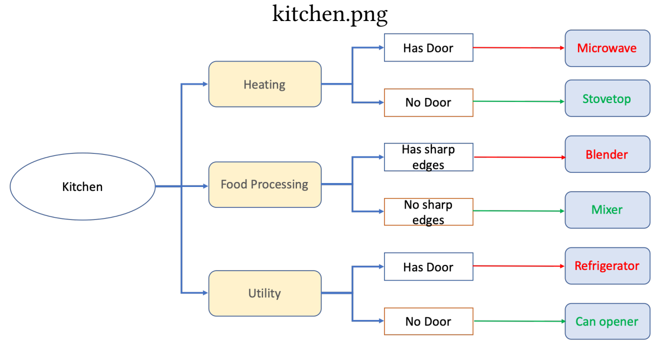

Figure 3: Sample of twenty questions (TQT) task based kitchen items

3.1 Tasks

3.1.1 Driving Task. Participants played the BeamNG.drive game on a PC and had to control a simulated vehicle. This game has also been used by other researchers to simulate driving in an experiment [4, 7]. The simulated road was a single lane rural road with no traffic, in daylight. The setup is shown in Figure 2.

Participants were engaged in two driving modes: manual driving and automated driving. In manual driving mode, participants were tasked to maintain lateral (side-to-side) and longitudinal control of the vehicle by operating the gaming steering wheel and gaming pedals. In automated driving mode, the game controlled all aspects of driving, and participants could completely focus on the non-driving task: the twenty questions game.

The experiment started with a system voice announcement of "start the experiment." This signaled to the participants to initiate manual driving. Participants continued to control the simulated vehicle for 65 seconds. At that point the system issued a beep followed by a voice alert message saying "autonomous mode" and the game's automation took over control of the vehicle.

Once automation was in control the participant was expected to turn to the twenty questions game.

3.1.2 Twenty Questions Task. We used a simplified version of the twenty question task (TQT) as the non-driving task for our experiment. For this version, participants were required to guess a word out of 18 possible words. Each of these words represented a household item that would normally be located in one of three rooms in the house: the kitchen, the bathroom, and the living room. Each item had two additional characteristics to help participants identify them - see for example Figure 3 for a visual representation of kitchen items and their characteristics. Before the start of the experiment, we trained participants to efficiently search for the items by asking questions about the 3 characteristics of each item. Training was done by one experimenter for all the participants in the study.

For the TQT, the participant and experimenter played the game by typing messages using the Skype application. The participant typed on a laptop which was placed next to the gaming PC as shown in Figure 2.

Participants completed multiple games while the simulated vehicle was under the control of automation. The number of games depended on how quickly participants completed each game.

3.1.3 Switching between tasks. After 100 seconds of automated driving, and playing the twenty questions game, the system issued a pre-alert, indicating that soon the driver will be required to take back control of the simulated vehicle. The pre-alert was a beep followed by the voice message saying "there is a narrow road and merging ahead." If the driver did not take back control, the pre-alert was followed by an emergency alert saying "Emergency, take over the control!" This indicated that the driver must take back control in 8 seconds or less.

Participants took back manual control by pressing the space bar on the game PC keyboard (Figure 2). They could take back manual control any time after the pre-alert. If they did not take back control once the maximum total takeover time had passed (15 seconds or 30 seconds, depending on the condition), the system automatically turned on manual control. Participants could continue to play the twenty questions game regardless of alerts and the need for manual driving.

3.2 Experimental design

We conducted a within-subjects experiment that manipulated pre-alert timing with two levels. In one condition ("short") the time between the pre-alert and the emergency alert was relatively short at 7 seconds. In the other condition ("long") this time period was slightly longer at 22 seconds. After the pre-alert there was always an external emergency alert (8 s before the critical event). Thus, for the two conditions the maximum total takeover time from the moment the pre-alert was issued was 7+8=15 seconds (short) and 22+8=30 seconds (long). We counter-balanced the presentation of conditions between participants.

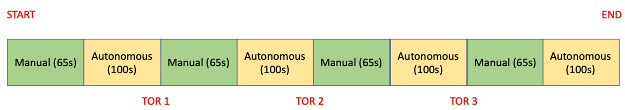

Figure 4: Participants completed two trials in the experiment, one under each alert condition (short and long). This figure shows the sequence of manual and automated driving segments within each trial, and the points of take-over requests where we could observe interleaving between the twenty questions task and the driving task (TOR)

We introduced the two conditions because both of our research questions (RQ1 and RQ2) ask how the available takeover time will affect interleaving behavior. We selected 15 seconds for the short takeover time for two reasons. First, we wanted to make the time realistic, and we know that the first level-3 vehicles will have takeover times in this vicinity (e.g. 10 seconds according to [9]). Second, we took into account that for a 15 second takeover time, the emergency alert will arrive after only 7 seconds, again placing us in the vicinity of 10 seconds. We selected 30 seconds as the longer takeover time because doubling the available takeover time is likely to be at the edge of capabilities for automated vehicles.

Each participant completed two trials, one under each alert condition (short and long), with three take-over request observations per trial. Specifically, each trial started with manual driving for 65 second. Next, game automation took over control and the participant started the twenty questions task. After 100 seconds of automated driving the driver heard the pre-alert and, if needed, the emergency alert. This was followed by the participant taking back manual control and driving for another 65 seconds. After manual driving, automation again took over control of the vehicle and the participant continued the twenty questions game. This pattern repeated for a total of four times as shown in Figure 4, resulting in three take-over requests (TOR) per trial/condition, where we could observe interleaving between the twenty questions task and the driving task.

3.3 Participants

Participation in this experiment was an optional assignment in a course at the University of New Hampshire. Specifically, students had the option of either participating in the experiment or completing a different assignment. Students who participated in the study received course credit for their participation. Twenty-one students opted to participate in the experiment. Due to technical problems that resulted in incomplete data recording, we discarded data from 2 participants. Thus, we present results from N=19 participants. Furthermore, again due to technical problems, for 2 of these 19 participants we discarded data related to the third instance of switching from automated driving to manual driving, both in the short takeover and in the long takeover condition.

Of the 19 participants, 11 participants were women (57.89%) and 8 were men (42.1%). The age of the participants was between 18 - 22 years old (Mean = 19.26 ; SD= 1.02 years of age). Out of the 19, 13 participants (68.42%) had a driver's license and 6 participants didn't have a driver's license.

3.4 Procedure

Participants reviewed an introductory sheet explaining the steps of the experiment. After this, participants read and signed the consent forms, and filled out a demographic survey.

Next, we trained the participants on the driving simulator and twenty questions task. The participants played the TQT game by speaking directly to the experimenter. The participants practiced the tasks before completing the two trials. Eye tracker calibration was done to the driving simulator screen using calibration markers.

After training, the participants completed the two trials, one under each alert condition (short and long).

3.5 Equipment and software

The equipment we used for the experiment is shown in Figure 2. Participants played the BeamNG.drive game on a computer with three 22 inch -displays. They controlled the vehicle using a Logitech G920 Driving Force steering wheel and pedals combination. They typed on a Dell Inspiron laptop with a 15 inch screen to play the twenty questions game. The experimenter typed on another PC (not shown in Figure 2) to play the game.

The BeamNG.Drive game has an option for enabling and disabling automated driving modes. We used python scripts to automate the movement of the cursor on the screen to enable and disable the driving modes (manual or automated).

Participants also wore an Ergoneers Dikablis head-worn eye tracker. The eye tracker was connected to a PC computer by a tether. We placed 2D markers on the three PC displays as well as on the laptop. The eye tracker used these markers to identify areas of interest.

We ran the Ergoneers D-Lab software to log data from the eye-tracker, participants typing data, and the experimenter typing data. D-Lab supports synchronous integration of data from different data sources using IP address connections. Since D-lab logged eye-tracker data, driving related events (e.g. alerts, driving mode switch), and keystrokes from participants and experimenter synchronously, we were able to easily compare timestamps from different sources for our analysis.

3.6 Data collection, processing and analysis

We collected the following data during the experiment:

Driving automation switch and alerts timestamp. At the beginning and end of the experiment, after playing each alert sound, and after each switch between manual and automated driving, a message was sent to the D-lab software. D-lab logged the event with corresponding timestamps.

Gaze position with respect to two areas of interest. We identified two areas of interest (AOI) in the three-dimensional space of our experiment. One was the area related to the driving task. This area consisted of the three PC screens. The second AOI was the area related to playing the twenty questions game. This area consisted of the laptop computer. Using the Ergoneers eye tracker, we collected gaze locations and then processed these using the D-lab software to assess if the participants looked at either of the two areas of interest. We also identified if they looked somewhere else, or if the eye tracker was unable to track their gaze at certain moments. We collected this data at a 60 Hz rate.

Keystrokes. All the key presses on the desktop and laptop by the participants, and on the experimenter's desktop, were logged into D-lab software. The key press on the desktop by the participants represents the event of them taking over the control of the vehicle from automated driving mode. Key presses on the laptop by the participants and on the desktop by the experimenter were used to transcribe each twenty question task. We determined if the participant typed a yes/no question or a guess, and we marked the time they pressed the ENTER key for all questions and guesses.

Using the data we collected we calculated the following measures.

Takeover time. We measured the takeover time as the time it took the participant to take control of the driving after the pre-alert was announced. Participants take control of the driving by pressing the space bar. The takeover time is calculated as the time between the pre-alert and when the space bar is pressed.

Transition stages. We identified which stage of the model [11] the participant was in at any given time as follows:

Stage 0: Performing twenty question task (TQT). The participant plays the twenty questions game while the simulated vehicle is under the control of automation - this mode is on for 100 seconds. The time stamp of the start of automated driving mode is the start of stage 0.

Stage 1: External alert. The participant heard a pre-alert and an emergency alert. We used the timestamp of the pre-alert message as the time of the external alert.

Stage 2: Disengage from TQT. We used the timestamp of the first glance of the driver away from the laptop screen.

Stage 3: Orient to driving. We used the timestamp of the first glance of driver at the driving screen after the pre-alert was issued.

Stage 4: TQT suspension. This is the last moment when the participant played the TQT. We determine this time as the time stamp of either (1) the last glance on the laptop screen, or (2) the last TQT-related keystroke, before the physical take-over of driving. Whichever of (1) and (2) occured later was designated as the timing of Stage 4.

Stage 5: Physical transfer of control. We used either the time stamp of when the participant hit the space bar on the gaming PC keyboard to take back control, or the time stamp when the system was automatically switched to manual control at the end of the take-over period (15 or 30 seconds).

Stage 6: Contribute to driving. This stage starts with Stage 5 and in our experiment lasts 65 seconds.

While our data can be analyzed to assess participant behavior in stages 7-10 in the model proposed by Janssen et al. [11], in this paper we focus only on stages 1-6.

3.6.1 Analytical Approach. We adopted a mixed-effects regression approach (linear or logistic, depending on the DV) based on the repeated measures nature of the data (i.e., more than one data point per person). This approach allowed us to model a random intercept to account for the baseline differences within participants, while also assessing the effect of our independent variable (length of takeover time). For the two key models, we used the lme4 package in R and applied an alpha of .05. We assessed significance using the 'car' package in R, which provides an estimate of significance of the model in the form of a chi-square test (x^2). For each model presented below, we include one fixed effect, which is the length of take-over time. We also present descriptive statistics and effect sizes (d), which we hope will be helpful for informing future work.

4 RESULTS

We validated that participants were engaged in the game by evaluating their performance in the TQT game. Table 1 shows the performance of participants in the Twenty questions task. Participants attempted an average of 6.11 questions in the 15 second scenario and 6.53 questions in the 30 second scenario. Along with attempting numerically more questions overall, participants in the 30 second scenario also appeared to make numerically more correct and less incorrect guesses compared to 15 seconds. Taken together, these results suggest that participants were reasonably engaged in the secondary task during the experiment. This result is comparable to the level of engagement observed in a similar study [15].

4.1 What is the empirical support for the stages from the Janssen model of interleaving? (RQ1)

Overall, we find support for the Janssen model [11]. Out of 110 instances interrupting drivers to take over manual driving, we found that 71 instances (64.5%) followed the stage sequences from the Janssen model. Although participants did engage in the interleaving predicted by Janssen et al. in the majority of transitions, this was not the case for each transition. We identified two patterns of switching from the non-driving task to driving. Specifically, participants either engaged in stages 3-4 in sequential order (interleaving pattern, predicted by Janssen, 64.5% of instances), or in the 4-3 in reverse order (suspension pattern, 35.5% of instances).

In what we call the "interleaving pattern" we observed the canonical order proposed by the Janssen et al. model, where the initial request to take back control (stage 1) is followed by disengagement from the twenty questions task (stage 2), then orienting to the driving task (stage 3), suspension of the twenty questions task (stage 4) and the transfer of control from automation to the driver (stage 5). Note that we observed interleaving of stages between stage 2 and stage 4; this means that after the initial request to take back control the participants looked away from the twenty questions task and toward the driving task, but they also returned to the twenty questions task, interleaving their preparation to take back control with their engagement in the non-driving task. In what we call the "suspension pattern" we observed a different order, one in which participants stopped their engagement in the non-driving task and immediately resumed control of the simulated vehicle. Here, the request to take back control (stage 1) was followed by disengagement (stage 2) and suspension of twenty questions game (stage 4), and then orienting to (stage 3) and immediately resuming (stage 5) the driving task. In this "suspension pattern" participants did not interleave between the non-driving and driving tasks.

4.1.1 Do the patterns depend on allowable transition time? The Janssen model's predictions may also be dependent on how long people are given to regain manual control of the vehicle. Indeed, a logistic (Pattern 1 vs 2, the "interruption" and "suspension" patterns entered as a binary DV) mixed-effects model revealed a significant

Scenario	Mean Attempted items	Mean Correct guesses	Mean incorrect guesses
15 sec	6.11	4.95	0.74
30 sec	6.53	5.64	0.58

Table 1: Mean value of participants' performance on items attempted by end of experiment.

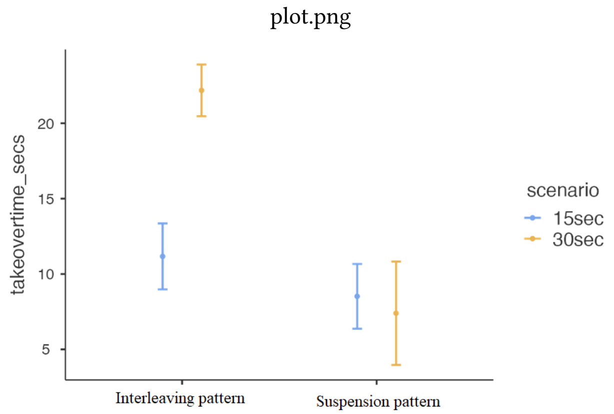

Figure 5: Takeover time for different patterns of switching to manual driving from non-driving task (TQT)

difference in the rates of the patterns across allowable takeover time (15s vs. 30s transition time), x^2 (1) = 11.9, p < .001, B = -1.86 (SE = .539). The differences in interleaving pattern were particularly pronounced in the 15s takeover length, where only 49% engaged in the interleaving pattern. In contrast, 80% of instances were classified as the interleaving pattern in the 30s takeover time.

4.2 How do participants use their interleaving time? (RQ2)

We also tested how allowed transition time impacts the actual manual takeover. We constructed a linear mixed-effects regression where each take-over time was regressed on transition time (15s vs. 30s). Not surprisingly, the 30s transition time resulted in a significantly longer takeover time (M = 19.2; SD = 9.30) than the 15s time (M = 9.82; SD = 3.95), x^2 (1) = 73.2, p< .001, B = 9.4 (SE = 1.10), d = 1.32. This finding is consistent with previous research which suggests urgent situations may lead to shorter takeover time [27].

A more interesting thing to note is that pattern (interleaving vs suspension) seemed to moderate the effect of available transition time on actual takeover time. We tested this by regressing takeover time on the interaction term between pattern and allowable transition time, and observed a significant interaction, x^2 (1) = 20.1, p< .001, B = 10.5 (SE = 2.34). Takeover time was statistically different across the 15s and 30s transition time when drivers followed the interleaving pattern, but were similar in terms of takeover time for the suspension pattern as seen in Figure 5. This implies that drivers do not simply ignore a take over request. If they follow the suspension pattern, they take over the control of the vehicle immediately after the request in both 15s and 30s conditions. If they follow the interleaving pattern, they interleave the driving and non-driving tasks.

Furthermore, we asked: what types of actions do participants take in the twenty questions game after the system issued the pre-warning? As shown in Table 2, we considered four possible options:

- Stopped: the participants stopped participating in the twenty questions game and switched to driving immediately after the takeover request.
- Attempted, not finished: The participant continued the twenty questions game but did not make a guess about the item that the experimenter had in mind.
- Finished and stopped: The participant continued playing the twenty questions game, made a guess (correct or incorrect), and then switched to the driving task.
- New item started: The participant continued the twenty questions game, made a guess, and then started asking questions about the next item, before switching to driving.

Table 2 provides interesting results in terms of participants' choice of action depending on the allowed time before taking over the control of the vehicle, most strikingly in the difference in new items started. We can see that in the 15 second scenario none of the participants started guessing a new item. However, in the longer, 30 second scenario, this happened 11 times, indicating a clear difference between the two conditions. A closer examination of the data shows that eight of the 19 participants started a new TQT task after the takeover request was issued. This indicates that some drivers may continue the non-driving task beyond a natural break point even after the takeover request is issued.

5 DISCUSSION

Our experimental results support the model proposed by Janssen and colleagues [11] in a driving simulator setting. We found empirical support for interleaving between the non-driving and driving tasks as they took back control from automation and commenced manual driving. We also found that the length of time available for takeover has an effect: when only 15 seconds are available to take over control, drivers were not as likely to interleave the tasks as when 30 seconds were at their disposal.

These findings are important as we think about how to design interfaces that (better) support drivers in future automated vehicles. First, we can expect that when the system requests drivers to take back control, drivers will engage in interleaving behavior, at least some of the time. We cannot assume that, once they receive the system request to take back control, drivers will discontinue their non-driving task, and immediately take up the driving task.

From a safety perspective, it is important to note that we observed interleaving even when the maximum takeover time was 15 seconds. In about half of the instances when participants had this short takeover time, they still interleaved the non-driving and

Scenario	Stopped	Attempted Not Finished	Finished and stopped	New Item started
15 sec	19	29	7	0
30 sec	12	19	13	11

Table 2: Number of instances of different actions participants took in TQT task during take-over time.

driving tasks. Clearly, interleaving can have negative safety consequences: instead of shifting their full attention to the driving task, drivers use a portion of their visual, manual, and cognitive resources to work on the non-driving task. When we design interfaces for future automated vehicles, we need to account for the likelihood that drivers will be engaged in non-driving tasks when automation is active, and will need support to quickly shift their attention to driving.

Yet, our data indicate that participants paid attention to maintaining good driving performance. We know from prior work by Janssen and Brumby that people might not wait to reach a sub-task boundary in order to switch from the non-driving task back to driving, if such a delay would results in poor driving performance [10]. The data in Table 2 is consistent with their results: when participants had only 15 seconds to return to driving, there were more instances of stopping or not finishing a game, than when they had 30 seconds to return to driving. Presumably, our participants felt that waiting to stop the twenty questions game at a natural break point would hurt their driving performance in the 15 second scenario.

Nevertheless, it is also interesting to observe that our participants did not always stop the non-driving task once they reached a natural break point in the twenty questions game. In our implementation of the twenty questions game, one such natural break point is when the participant makes a guess about an item. At this point one twenty questions game is over, and the next one has not started yet. A range of prior experiments showed that users can take advantage of such natural break points to switch tasks (e.g. [1, 10, 15]). Thus, our expectation was that if our participants complete a twenty questions game during the transition from automated to manual driving, they will suspend this non-driving task (stage 4 in Janssen model) and transition to the driving task (stage 5). Yet, eight participants in 11 different instances did not do this - rather they started a new twenty questions game. While we only observed this behavior for the longer maximum takeover time (30 second), this is a warning to us as we design in-vehicle user interfaces. Specifically, it is a warning that our users might prioritize the non-driving task over driving, and as Janssen et al. predicted, treat driving as the interruption of the ongoing non-driving task.

6 LIMITATIONS

While our experimental results provide important insights for the design of interfaces for future automated vehicles, there are a number of limitations we need to mention. One limitation is the fact that the experiment was performed in a low-fidelity simulation: our participants played a game on a PC. Their behaviors in an actual vehicle might differ from those we observed in this experiment. For example, they might be more inclined to take back control quickly, given that driving on the road is a safety-critical task. Still, as is broadly the case in driving simulator studies [13], we observed our participants to be engaged in the driving task. In addition to visual observations of this engagement by the experimenter, takeover behavior also supports this conclusion: of the 110 takeovers we recorded, in 91 cases (82.7%) the participants took over before the maximum time expired. Thus, our results can serve as an indicator of what types of behaviors we can expect in real vehicles.

Another limitation is that participants only participated in one experiment. Yet, their behaviors might change over time - research shows that drivers' trust in automated vehicle technology can increase as they use the technology over time [16]. Additional research is needed to assess how interfaces that support interleaving can help users calibrate their trust appropriately [17].

Additional limitations are that we only explored one specific secondary task, and only one implementation of this task. Namely, our participants played a word game by exchanging text messages with an experimenter. Other tasks and implementations might result in different levels of engagement in the non-driving task, as well as different levels of connection to the outside world and specifically the driving environment. Taken together, such differences can have an effect on interleaving behaviors.

Furthermore, we did not account for where in the non-driving task (the twenty questions game) our participants were interrupted. Yet, their progress in the non-driving task might have affected their responses to the pre-alert and emergency alert (see e.g. [15]).

Finally, the experiment was conducted with a homogeneous group of participants - they were all young college students. It will be important to understand how different user groups engage in different non-driving tasks under different contexts in future automated vehicles.

7 CONCLUSION

Our driving simulator experiment provides evidence to support the interleaving model proposed by Janssen and colleagues: namely that in future automated vehicles, when the system requests that drivers take back control from the automation, users will often interleave the non-driving and driving tasks before fully taking control of the vehicle. This result has important implications for the design of in-vehicle user interfaces and driving safety. Specifically, these results indicate that we cannot count on drivers to always shift their full attention to driving immediately after they receive the takeover request, or even immediately after they first disengage from the non-driving task and start orienting themselves to the driving task. Rather, in some cases drivers will be inclined to shift their manual, visual, and cognitive resources between the non-driving and driving tasks. Our interface designs have to take this behavior into account such that drivers can safely take back control from automation.

ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation under grants CMMI-1840085 and CMMI-1840031.

REFERENCES

[1] Brian P Bailey and Shamsi T Iqbal. 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 4 (2008), 1–28. https://doi.org/10.1145/1314683.1314689

[2] Deborah A Boehm-Davis and Roger Remington. 2009. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention* 41, 5 (2009), 1124–1129. https://doi.org/10.1016/j.aap.2009.06.029

[3] Jelmer P. Borst, Niels A. Taatgen, and Hedderik van Rijn. 2015. What Makes Interruptions Disruptive? A Process-Model Account of the Effects of the Problem State Bottleneck on Task Interruption and Resumption. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2971–2980. https://doi.org/10.1145/2702123.2702156

[4] Víctor Corcoba, Zoe Falomir, and Nourelhoda Mohamed. 2018. Using Wearable Sensors to Detect Workload on Driving Simulated Scenarios. In *Artificial Intelligence Research and Development*. IOS Press, Amsterdam, Netherlands, 287–296. https://doi.org/10.3233/978-1-61499-918-8-287

[5] Alexander Eriksson and Neville A Stanton. 2017. Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human factors* 59, 4 (2017), 689–705. https://doi.org/10.1177/0018720816685832

[6] Nadia Fereydooni, Orit Shaer, and Andrew L. Kun. 2019. Switching between Augmented Reality and a Manual-Visual Task: A Preliminary Study. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 99–103. https://doi.org/10.1145/3349263.3351502

[7] Alessio Gambi, Marc Mueller, and Gordon Fraser. 2019. Automatically Testing Self-Driving Cars with Search-Based Procedural Content Generation. In *Proceedings of the 28th ACM SIGSOFT International Symposium on Software Testing and Analysis* (Beijing, China) *(ISSTA 2019)*. Association for Computing Machinery, New York, NY, USA, 318–328. https://doi.org/10.1145/3293882.3330566

[8] Christian Gold, Daniel Damböck, Lutz Lorenz, and Klaus Bengler. 2013. "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1 (2013), 1938–1942. https://doi.org/10.1177/1541931213571433 arXiv:https://doi.org/10.1177/1541931213571433

[9] Bengt Halvorson. 2017. *Traffic Jamming: In the 2019 Audi A8, We Let Automated-Driving Tech Take the Wheel*. Car and Driver. Retrieved 2021-05-13 from https://www.caranddriver.com/news/a15339688/traffic-jamming-in-the-2019-audi-a8-we-let-automated-driving-tech-take-the-wheel/

[10] Christian P Janssen and Duncan P Brumby. 2010. Strategic adaptation to performance objectives in a dual-task setting. *Cognitive science* 34, 8 (2010), 1548–1560. https://doi.org/10.1111/j.1551-6709.2010.01124.x

[11] Christian P Janssen, Shamsi T Iqbal, Andrew L Kun, and Stella F Donker. 2019. Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies* 130 (2019), 221–233. https://doi.org/10.1016/j.ijhcs.2019.07.004

[12] Andrew Kun, Hidde van der Meulen, and Christian Janssen. 2017. Calling While Driving: An Initial Experiment with HoloLens. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. University of Iowa, Manchester Village, Vermont, 200–206. https://doi.org/10.17077/drivingassessment.1636

[13] Andrew L Kun. 2018. Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends® in Human–Computer Interaction* 11, 4 (2018), 201–293. https://doi.org/10.1561/1100000069

[14] Andrew L Kun, Oskar Palinko, Zeljko Medenica, and Peter A Heeman. 2013. On the feasibility of using pupil diameter to estimate cognitive load changes for in-vehicle spoken dialogues.. In *INTERSPEECH*. ISCA, Lyon, France, 3766–3770.

[15] Andrew L Kun, Alexander Shyrokov, and Peter A Heeman. 2013. Interactions between human–human multi-threaded dialogues and driving. *Personal and Ubiquitous Computing* 17, 5 (2013), 825–834. https://doi.org/10.1007/s00779-012-0518-1

[16] David R. Large, Gary Burnett, Davide Salanitri, Anneka Lawson, and Elizabeth Box. 2019. A Longitudinal Simulator Study to Explore Drivers' Behaviour in Level 3 Automated Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 222–232. https://doi.org/10.1145/3342197.3344519

[17] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[18] Natasha Merat, A Hamish Jamson, Frank CH Lai, and Oliver Carsten. 2012. Highly automated driving, secondary task performance, and driver state. *Human factors* 54, 5 (2012), 762–771. https://doi.org/10.1177/0018720812442087

[19] Brian Mok, Mishel Johns, Key Jung Lee, David Miller, David Sirkin, Page Ive, and Wendy Ju. 2015. Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, Piscataway, NJ, 2458–2464. https://doi.org/10.1109/ITSC.2015.396

[20] Oscar Oviedo-Trespalacios. 2018. Getting away with texting: Behavioural adaptation of drivers engaging in visual-manual tasks while driving. *Transportation Research Part A: Policy and Practice* 116 (2018), 112–121. https://doi.org/10.1016/j.tra.2018.05.006

[21] Jonas Radlmayr, Christian Gold, Lutz Lorenz, Mehdi Farid, and Klaus Bengler. 2014. How Traffic Situations and Non-Driving Related Tasks Affect the Take-Over Quality in Highly Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, 1 (2014), 2063–2067. https://doi.org/10.1177/1541931214581434 arXiv:https://doi.org/10.1177/1541931214581434

[22] Sijun Shen and David M Neyens. 2017. Assessing drivers' response during automated driver support system failures with non-driving tasks. *Journal of safety research* 61 (2017), 149–155. https://doi.org/10.1016/j.jsr.2017.02.009

[23] Jennifer Shuttleworth. 2019. *SAE Standards News: J3016 automated-driving graphic update*. SAE international. Retrieved 2021-05-13 from https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic

[24] Remo M.A. van der Heiden, Shamsi T. Iqbal, and Christian P. Janssen. 2017. Priming Drivers before Handover in Semi-Autonomous Cars. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 392–404. https://doi.org/10.1145/3025453.3025507

[25] Marcel Walch, Kristin Lange, Martin Baumann, and Michael Weber. 2015. Autonomous Driving: Investigating the Feasibility of Car-Driver Handover Assistance. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Nottingham, United Kingdom) *(AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/2799250.2799268

[26] Philipp Wintersberger, Andreas Riener, Clemens Schartmüller, Anna-Katharina Frison, and Klemens Weigl. 2018. Let Me Finish before I Take Over: Towards Attention Aware Device Integration in Highly Automated Vehicles. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) *(AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 53–65. https://doi.org/10.1145/3239060.3239085

[27] Bo Zhang, Joost de Winter, Silvia Varotto, Riender Happee, and Marieke Martens. 2019. Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation research part F: traffic psychology and behaviour* 64 (2019), 285–307. https://doi.org/10.1016/j.trf.2019.04.020