

LEARNING AND DEVELOPMENT OF MENTAL MODELS DURING INTERACTIONS WITH DRIVING AUTOMATION: A SIMULATOR STUDY

Yannick Forster^{1,2}, Sebastian Hergeth¹, Frederik Naujoks¹, Matthias Beggiato²,
Josef F. Krems², Andreas Keinath¹

¹ BMW Group, Munich, Germany

² Chemnitz University of Technology, Chemnitz, Germany

Email: yannick.forster@bmw.de

Summary: Higher level cognitive processes such as learning and mental models play a fundamental role in the success of automated driving, as technology can only be as good as our understanding and expectations of it. The present study investigated the development of these processes during interactions with driving automation. In a driving simulator study, $N=52$ participants completed several transitions between manual and Society of Automotive Engineers (SAE) levels 2 and 3 automated driving. Self-reported learning progress and mental model development were assessed via questionnaires. In parallel, eye-tracking data were collected as a behavioral measure of higher level cognitive functions. The results demonstrated that self-reported learning and gaze behavior followed a power-law function; the power-law functions showed task specific parameter manifestations. The evolution of the mental models of the level 2 and level 3 human-machine interface continued up to the fifth contact, indicating a long lasting process. For researchers and practitioners, the present study implies that accurate mental models require up to 5 repeated interactions. Furthermore, learning progress with driving automation can be captured through gaze behavior.

INTRODUCTION

With the feasibility of automated driving functions and their accompanying proliferation of Human-Machine Interfaces (HMI), in recent years research has increasingly focused on the methodological aspects (Naujoks, Hergeth, Wiedemann, Schömig, & Keinath, 2018) of evaluating these HMIs for levels 2 (L2) and 3 (L3) automated driving (SAE, 2018). The present study's objective is to gain insight into the development of higher level cognitive processes when interacting with driving automation. This is crucial in the successful development of automated driving as the evaluation of interfaces frequently only applies initial contact (Forster, Naujoks, & Neukum, 2016; Frison, Wintersberger, Riener, & Schartmüller, 2017/2017). However, if users have not yet fully understood the system, their self-reported evaluation of the system in question might be unreliable and therefore invalid (Beggiato, Pereira, Petzoldt, & Krems, 2015). The present work contributes to the refinement of research methods by (1) examining the behavioral measures describing the learning process and (2) investigating the development of the mental model for L2 and L3 automated driving.

Background

There remains a dearth of research on the development of higher level cognitive processes when interacting with L2 and L3 driving automation. A study by Beggiato et al. (2015) for L1 driving automation illustrated that learning, trust and acceptance can be applied in accordance with the

power law of learning (Newell & Rosenbloom, 1981). The power law is a function with the general form of:

$$P(N|a, b, c) = a + b * N^{-c} \quad (1)$$

The predicted performance P is a function of the number of trials N given the three parameters a, b and c. Parameter a is the asymptote of P with N increasing indefinitely; parameter b describes the amplitude of learning between the initial performance and asymptotical performance; parameter c is the learning rate. Beggiato et al. (2015) investigated higher level cognitive processes by means of the mental model; they found that it takes up to ten repeated encounters with the system for an accurate mental model to evolve.

Besides self-reported measures on learning and the mental model, eye-tracking metrics allow researchers an insight into higher level cognitive processes at the behavioral level (Just & Carpenter, 1984). The present experiment therefore examined gaze behavior to discover whether it could be used to capture the evolution of self-reported learning and mental models in automated driving. Empirical support for this consideration comes from Sarter, Mumaw, and Wickens (2007) who revealed a connection between the mental model and gaze behavior in pilots interacting with flight automation. Similarly, Underwood (2007) observed a lack of both adequacy of mental models and efficiency of visual scanning with inexperienced users. Besides differences in scanning behavior with repeated exposure, there is evidence that gaze behavior is task-dependent (Hayhoe & Ballard, 2005). Due to differences in activation conditions between levels of automation (SAE, 2018), the present study aims to investigate gaze behavior in different transitions (Naujoks et al., 2018).

Bridging the gap between self-reported higher level cognitive processes (Beggiato et al., 2015) and behavioral measures (Sarter et al., 2007) the current research aims at a comprehensive examination in the area of L2 and L3 automated driving.

Research Questions and Hypotheses

Based on the aforementioned considerations and prior research, we derived two research questions: First, how do self-reported measures of higher level cognitive processes evolve over time while interacting with driving automation? To answer this question, we assume that self-reported learning of the interaction follows the power law of learning (Hypothesis 1a). Furthermore, repeated interaction should lead to more accurate mental models (Hypothesis 1b).

Second, how do behavioral measures of higher level cognitive processes evolve after repeated interactions with driving automation? Concerning this question, we expect the number of gaze changes between different areas of interest (AOI) to follow the power law (Hypothesis 2a). The development of gaze behavior should additionally proceed in in a task-dependent manner (Hypothesis 2b).

METHODS

Driving Simulator

The study was conducted in a high-fidelity moving-base driving simulator. The BMW 5 series vehicle mock-up contained a complete instrumentation. Seven 1080p projectors provided a 240° horizontal × 45° vertical frontal field of view. One LCD screen positioned behind the back seats

and two outside projections with the same specifications served as the driver's rear view. The motion system consisted of a hexapod with six degrees of freedom, capable of up to 7 m/s² transitional acceleration and 4.9 m/s² continuous acceleration. The Surrogate Reference Task (ISO, 2012) was displayed on a 12.3" tablet mounted on the center stack console.

Design and Procedure

The present work reports the results of the within-subject factor "expertise" with five levels (Use Case repetition from block 1 – 5).

Prior to the experimental drive, the experimenter outlined the procedure and explained that participants would encounter two automated systems (i.e., L2, L3). They also received information stating that they would not have to constantly monitor the correct functioning of the L3 automated driving system (ADS). No specific additional information about the six items of the mental model questionnaire was given. Prior to each block, participants were instructed that they could decide for themselves whether to engage in a non-driving related task (NDRT) when the automation was active. The subsequent experimental drive included five experimental blocks, each with six interactions between the driver and the driving automation. After each block, participants completed the block inquiry. Subsequent to a 5 minute familiarization drive, the experimental drive on a three-lane highway with moderate traffic density lasted approximately 60 minutes.

Driving Automation and Human-Machine Interface

As soon as the driver activated the L2 or L3 system, it carried out longitudinal and lateral vehicle guidance. The functioning performance of the L2 and L3 automation was identical. During both L2 and L3 automated driving, the visual HMI on the instrument cluster displayed the vehicle and its surroundings. This was based on prior research (Forster et al., 2016; Manca, Winter, & Happee, 2015) and thus represents a prototypical solution. The driving automation systems could be activated with buttons on the left-hand side of the steering wheel. The L2 driving automation could be activated at any time while driving. For the L3 ADS, however, there were several activation conditions (SAE, 2018); these were (1) the availability of the ADS, (2) a velocity below 130 km/h and (3) keeping in line according to a certain lateral margin.

Use Cases

The current experiment included driver initiated transitions between manual, L2 and L3 automated driving based on (Naujoks et al., 2018). Use cases (UC) were automated transitions from L0 to L3 (UC1), and from L0 to L2 (UC2). In addition, we implemented transitions from L2 to L3 (UC3) and vice versa (UC4). Each block included two transitions to manual (i.e., from L2 and L3). The transitions to manual were not analyzed here since the desired system state does not represent a driving automation system. To counteract the experiment's sequential effects, participants were randomly assigned to one of six possible block sequences created using a Latin square. To standardize the instructions, we recorded samples for each UC that were triggered by the experimenter. UCs with a transition to L3 (UC1 and UC3) also included a 10-second delay between the experimenter's instruction and the availability of the L3 ADS. Within this 10-second window, the L3 ADS could not be activated.

Dependent Variables

The mental model questionnaire and items for self-reported learning developed by Beggiato et al. (2015) were adapted for L2 and L3 driving automation. Participants completed the questionnaire after receiving instructions on how to do so, and then again after each experimental block. Self-reported learning was measured with four items on a visual analogue scale ranging from 0-100. The mental model questionnaire included 11 items on a 7-point Likert scale ranging from 1 (“*strongly disagree*”) to 7 (“*strongly agree*”) for both the L2 and L3 automation. Six items covered participants’ understanding of the system, while five items served as distractors. Items were specifically derived to detect differences between L2 and L3. Thus, for each item of interest, the opposite end of the scale was correct for the two levels of automation. If “*strongly disagree*” was the correct answer for L2, “*strongly agree*” was correct for L3 and vice versa.

Gaze data were collected with Dikablis 2.5. Prior to analysis, data quality was ensured by manually recoding all relevant sections. AOIs (i.e., driving scene, instrument cluster, operating element, NDRT) were defined in D-Lab 2.5. Gaze behavior was analyzed from the onset of the standardized experimenter instructions until a UC was solved; a UC was solved successfully if a desired system state was constant for at least five seconds without unintended deactivations or interruptions. For each UC, the number of gaze switches between AOIs was calculated. Two or more glances at the same AOI that were separated by less than 150ms were combined. Glances shorter than 150ms on an AOI were eliminated from the analysis (Irwin, 1992).

Sample

A total of $N=59$ participants took part in the driving simulation experiment. Two participants could not finish the experiment due to simulator sickness and five incomplete datasets were excluded. This left $N=52$ (14 female, 38 male) participants for data analysis. The mean age of the final sample was 31.88 years ($SD=10.09$, $MAX=62$, $MIN=20$). Participants were recruited among BMW Group employees, held a German driver’s license and had normal or corrected to normal vision.

RESULTS

Model fits of the power law for self-reported learning (see Figure 1 and Table 1) demonstrated a steep increase from the prior assessment to the first block and a subsequent slight increase up to the fifth block. Self-reported learning stabilized at a considerably high level, indicating that participants were subjectively familiar with the system.

The change in total gaze switches between AOIs (Figure 2 and Table 2) over time showed that participants’ gaze behavior adhered to the power law of learning. Each UC showed a specific learning pattern: There were lower asymptotes for the transitions to L2 automation (UC2, UC4) than for transitions to L3 (UC1, UC3). The highest learning rate and amplitude was observed in the transition from manual to L2 automation (UC2), while the lowest amplitude was observed in the transition from L2 to L3 automation (UC3). The transition from L3 to L2 automation (UC4) took the longest to stabilize in terms of gaze behavior, which is reflected in its considerably low learning rate of 1.217.

Table 1. Model fit equations and goodness-of-fit indices (R^2 , RMSE) for self-reported learning

	Power Law	R^2	RMSE
Learning 1	$N=86.58-55.36*t^{-1.825}$.9983	1.0924
Learning 2	$N=87.93-45.58*t^{-1.447}$.9965	1.4659
Learning 3	$N=85.18-42.27*t^{-1.278}$.9954	1.2903
Learning 4	$N=84.23-51.94*t^{-1.324}$.9920	2.0737

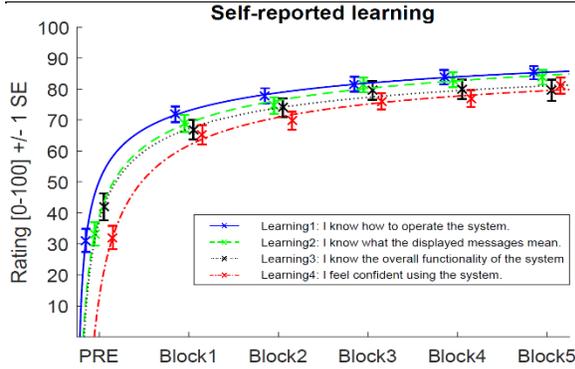


Figure 1. Self-reported learning (M , SE) and according power-law fit

Table 2. Model fit equations and goodness-of-fit indices (R^2 , RMSE) for gaze switches by UC

	Power Law	R^2	RMSE
Use Case 1	$P=24.16+18.53*t^{-2.818}$.9765	1.7311
Use Case 2	$P=10.41+33.71*t^{-4.805}$.9997	0.3519
Use Case 3	$P=21.72+9.54*t^{-3.000}$.9921	0.5146
Use Case 4	$P=6.13+25.42*t^{-1.217}$.9975	0.6307

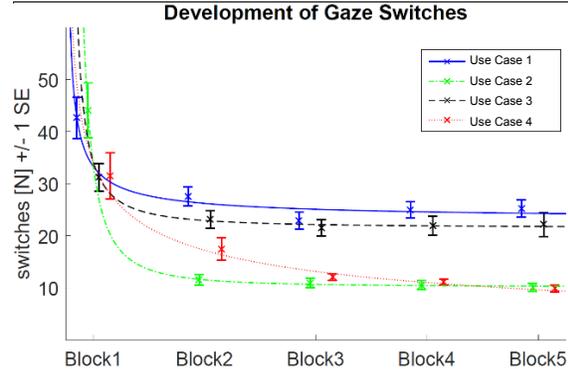


Figure 2. Number of gaze switches (M , SE) and according power-law fit

The results for the mental model questionnaire revealed that development occurs up to the fifth block (see Figure 3). With increasing experience, participants could better discriminate between L2 and L3, as indicated by rising mean differences between the two automation levels. The rising accuracy of the mental models is reflected in the differences between both systems, as well as in smaller mean distances towards their respective ends of the scale. System monitoring related items (i.e., Items 3, 5) already showed large differences between L2 and L3 before the first block. Three variables showed a steadily increasing difference (i.e., Items 1, 4, 6). A difference in the relevance of lane keeping (Item 2) emerged after the fourth block of interactions. After five blocks, there was still room for improving the mental model, as indicated by distances between the mental model mean and the correct answer (1 or 7, respectively).

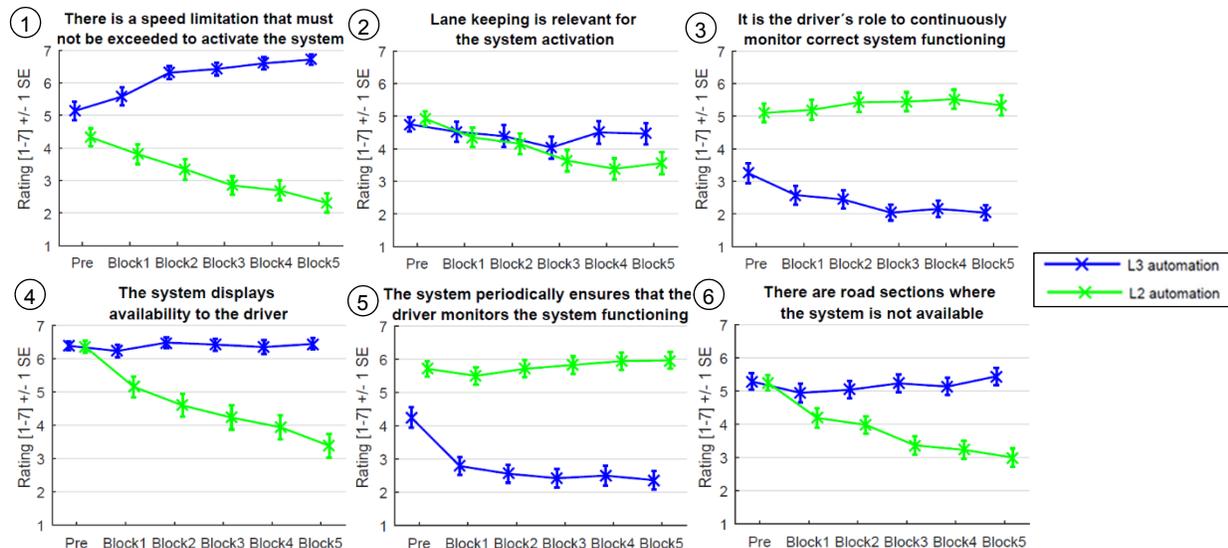


Figure 3. Mental model questionnaire results (i.e., M , SE) by automation level and expertise

DISCUSSION AND CONCLUSIONS

The objective of this study was to examine the effect of increasing practice on higher level cognitive processes at both the self-reported and behavioral levels. The results for the development of both measures over time confirmed Hypotheses 1a and 2a. Both the self-reported learning and the number of gaze switches between AOIs followed the power-law of learning and demonstrated sufficient goodness-of-fit (see Tables 1 and 2). The decrease and subsequent stabilization of the number of transitions between the AOIs across all UCs indicates that scanning efficiency increased. Driving automation users thus seem to improve their visual scanning behavior over time (Underwood, 2007). In addition, the different model fits for the four transitions found in this study support the task-dependency of gaze behavior (Hayhoe & Ballard, 2005). This result supports Hypothesis 2b, as each transition poses different demands to the human operator: Transitions with ongoing manual vehicle guidance (UC1, UC2) exhibited more learning compared to transitions between two levels of automation (UC3, UC4). This result emphasizes that the design process should focus on the task-specific requirements of control transitions. Future research should therefore examine the dispersion of visual attention allocation in transitions between automation levels more closely.

The mental model results support Hypothesis 1b. While participants could not effectively discriminate between L2 and L3 automation before using automated driving, subsequent interactions improved the accuracy of their mental model (see Figure 3). Accordingly, researchers and practitioners in the field of automated driving need to consider that accurate mental models require repeated interactions, even after operational behavior has reached a constant level. Even after five repetitions, there were still participants that did not have an accurate mental model. The difference between L2 and L3 automation was not evident to many participants during their first interactions; for example, the relevance of lane keeping for transitions in L3 took up to four repetitions to become even slightly apparent (see Figure 3, Item 2). Despite the apparent simplicity of interacting with the system (e.g., pressing a button to activate it), these results emphasize the difficulty and long-lasting process of building accurate mental models. As the present experiment only included L2 and L3 automated driving, this effect is proposed to be even stronger with additional automation functions (e.g., L1, L4). The present work supports prior evidence that mental models for driving automation are slow to evolve (Beggiato et al., 2015). In automated driving research, the interpretation of self-reported satisfaction with driving automation must consider that these evaluations were provided during a phase in which the participants' understanding of the system was still evolving (Forster et al., 2016; Frison et al., 2017/2017). Future research efforts are necessary to identify the impact of rising experience with and accuracy of mental models on the development of self-reported satisfaction with HMIs for automated driving.

This study provided the first evidence of the development of attention allocation; however, the total number of switches between AOIs failed to consider a decrease in the time participants spent on task with rising experience. Therefore, future analysis should evaluate behavioral measures in relation to the respective task duration. In addition, the switches between AOIs did not investigate attention allocation to single AOIs. The open question remains how the importance of AOIs changes over time, which is reflected in monitoring frequency (Jacob & Karn, 2003). Besides these concerns, the present work supports the operationalization of learning

through gaze behavior and provides the literature with an insight into the long-lasting process of mental model evolution in automated driving.

REFERENCES

- Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. (2015). Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 75–84.
- Forster, Y., Naujoks, F., & Neukum, A. (2016). Your Turn or My Turn? Design of a Human-Machine Interface for Conditional Automation. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 253–260). <https://doi.org/10.1145/3003715.3005463>
- Frison, A.-K., Wintersberger, P., Riemer, A., & Schartmüller, C. (2017). Driving Hotzenplotz: A Hybrid Interface for Vehicle Control Aiming to Maximize Pleasure in Highway Driving. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 236–244).
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. <https://doi.org/10.1016/j.tics.2005.02.009>
- Irwin, D. E. (1992). Visual memory within and across fixations. In *Eye movements and visual cognition* (pp. 146–165). Springer.
- ISO (2012). *Road vehicles—Ergonomic aspects of transport information and control systems—Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems*. (ISO, 14198). Geneva, Switzerland.
- Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.
- Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. *New Methods in Reading Comprehension Research*, 151–182.
- Manca, L., Winter, J. C. F. de, & Happee, R. (2015). Visual Displays for Automated Driving: A Survey. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Workshop on Adaptive Ambient In-Vehicle Displays and Interactions*.
- Naujoks, F., Hergeth, S., Wiedemann, K., Schömig, N., & Keinath, A. (2018). Use Cases for Assessing, Testing, and Validating the Human Machine Interface of Automated Driving Systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (62nd ed., pp. 1873–1877). Los Angeles, CA: Sage publications.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition*, *1*(1981), 1–55.
- SAE. (2018). *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems* (No. J3016R).
- Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye-tracking data. *Human Factors*, *49*(3), 347–357. <https://doi.org/10.1518/001872007X196685>
- Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics*, *50*(8), 1235–1249.