Prediction of take-over time in highly automated driving by two psychometric tests

Moritz Körber a, Thomas Weiβgerber a, Luis Kalb a, Christoph Blaschke b & Mehdi Farid b

a Institute of Ergonomics, Technische Universität München, Garching, Germany. {koerber, weissgerber, kalb}@lfe.mw.tum.de
b BMW Group {christoph.blaschke, mehdi.farid}@bmw.de

Received: February 18th, 2015. Received in revised form: March 16th, 2015. Accepted: September 29th, 2015

Abstract
In this study, we investigated if the driver’s ability to take over vehicle control when being engaged in a secondary task (Surrogate Reference Task) can be predicted by a subject’s multitasking ability and reaction time. 23 participants performed a multitasking test and a simple response task and then drove for about 38 min highly automated on a highway and encountered five take-over situations. Data analysis revealed significant correlations between the multitasking performance and take-over time as well as gaze distributions for Situations 1 and 2, even when reaction time was controlled. This correlation diminished beginning with Situation 3, but a stable difference between the worst multitaskers and the best multitaskers persisted. Reaction time was not a significant predictor in any situation. The results can be seen as evidence for stable individual differences in dual task situations regarding automated driving, but they also highlight effects associated with the experience of a take-over situation.

Keywords: Automated driving; out of the loop; dual task; multitasking; reaction time; take over time.

1. Introduction

1.1. Automation Effects in Vehicle Control

Technological progress in advanced driver assistance systems (ADAS; [9]) is currently initiating a shift in vehicle control from manual driving to automated driving since current sensory technology and data processing now provide the ability to allow longitudinal control as well as lateral control be carried out by an automation [13]. In this case, the driver is completely removed from the task of driving in such a way that, in contrast to manual driving, a vehicle automation system fully operates the vehicle. This change in vehicle control is accompanied by a change in the driver’s tasks and the resulting task demands. Firstly, in case of partial automation ([11]; level 2 in [27]), the driver has to constantly monitor the automation, i.e. the active role of driving is replaced with a passive role as a monitor. Secondly, if a
system limit or failure occurs, the driver has to switch from passive automated control to manually steering the vehicle. Thirdly, the automation provides the driver with the ability to engage in non-driving-related activities since vehicle control is carried out by the automation. As a result, the driver now has to switch between two tasks if he needs to regain control of the vehicle. The goals of introducing vehicle automation are to reduce the driver’s workload [34] and to increase traffic safety [22,29] and comfort [35]. Since possible problems in the interaction between human and automation have already been found in other fields, such as aviation [28,39], it is necessary to review not only the technological but also the human aspect of safety. Issues associated with human-automation interaction are subsumed under the term automation effect. According our definition, this is:

an effect that is caused by the difference in demands and tasks of the operator between automation and manual operation and is detrimental to the operator’s capability to perform.

1.2. The Out-of-the-loop state as a Consequence of Automated Vehicle Control

Vehicle control can be seen as an interaction between human, machine and environment [21] and, therefore, the paradigm of a feedback loop of a human-machine system is applicable [3,33]. The driving task represents the input parameter and the set point is represented by the target speed and route. Vehicle movement is the resultant output parameter. The driver acts as a controller of the loop in order to minimize the discrepancy between actual and target speed or route. To successfully undertake this controller task, the driver has to continuously observe the environment, traffic and his own vehicle movement. This loop is shown in Fig. 1.

In case of an automated drive, the vehicle automation takes over the task of the controller so that the driver is taken out of the (feedback) loop. Negative consequences of this state are subsumed under the term out-of-the-loop [8] state. We define out-of-the-loop as:

a driver state of readiness in which the driver is not able to immediately intervene in the feedback loop comprised of controller and vehicle. In this state, the driver does not have up-to-date knowledge of the parameters that are relevant for the controlling task, e.g. his own speed, position, or a headway vehicle. He is also not able to predict the situation insofar as to create a time window for himself that is long enough to react to events in a manner that is safe for road traffic.

Figure 1. Model of driver-vehicle feedback loop
Source: created by the authors; adapted from Bubb [4].

The state is not binary but a continuous dimension with in-the-loop and out-of-the-loop as poles. Therefore, drivers can be out-of-the-loop to varied extents. The consequences of this state are longer reaction times [24,36], omission of a reaction [7], or errors in information collection [2]. A high out-of-the-loop state can be reached as a result of different causes, but increased engagement in a secondary activity as a behavioral adaption to automation is the automation effect with the strongest link to it. Contrary to manual driving, where the driver’s attention on the road is constantly required, automated driving allows the driver to engage in other activities, such as reading the newspaper or playing a video game. Marras [23] considers boredom that arises during a drive to be a consequence of not undertaking the driving task, which could lead to increased engagement in secondary activities. Accordingly, Carsten and Colleagues [6] found that the engagement in a secondary activity increases as the level of automation rises. As a consequence, the driver allocates at least a part of his attention to a non-driving-related task and no longer completes the aforementioned task of updating the relevant situation parameters and therefore reaches, to a certain degree, the out-of-the-loop state.

1.3. Evidence for Individual Differences

In their literature review, Körber and Bengler [19] point out that potential inter-individual differences could exist in automation effects and imply that they should be taken into account in sampling and should be investigated in empirical studies. In order to keep the out-of-the-loop-state low, the driver has to have the ability to constantly update the relevant parameters for driving safety (e.g. road, other traffic, his own movement), while being engaged in a secondary task at the same time. This ability can be seen as an application of the construct multitasking ability: the ability to work on two tasks at the same time. Previous research has revealed evidence for stable individual differences in multitasking: Bühner and colleagues [5] found that working memory performance was the best predictor of multitasking, ahead of reasoning and attention. Accordingly, Morgan and colleagues [25] also found working memory and scholastic aptitude to be significant predictors of multitasking in a flight simulation. Working memory load also induces an attentional blindness [10] and could therefore be detrimental in terms of detecting hazards in traffic while driving. Kahneman, Ben-Ishai, and Lotan [17] were able to link the ability to relocate attention with a driver’s accident history as lower performance was associated with higher accident frequency. Alzahabi and
Becker [1] split their participants into light and heavy multitaskers based on their frequency of engaging in two media activities at the same time. Although no difference was found with regard to working on two tasks simultaneously, heavy multitaskers were more capable of switching between two tasks. Beyond this, Watson and Strayer [38] found no performance decrement in a difficult dual task setting for 2.5% of their participants, who they named *supertaskers*. This evidence suggests that drivers vary in their multitasking abilities and thus differ in their potential to reach a critical out-of-the-loop state by engaging in a secondary task. Since the out-of-the-loop state lead to longer reaction times, we therefore expect that the ability to multitask is directly related to the time needed to take over an automated vehicle.

**H1:** The performance in a multitasking test is negatively correlated to take-over time.

In their work, Körber and Bengler [19] list the *individual reaction time* as another factor influencing take-over time. This seems intuitive as, since even if the driving task is carried out by the vehicle automation, the driver is required to quickly take back control as a response to a take-over request (TOR; e.g. an earcon) by the vehicle if the automation reaches a system limit or fails. We therefore predict the following relationship:

**H2:** The individual reaction time is positively correlated to take-over time. Since we expect two different mechanisms in relationship to take over time, we assume that both multitasking and individual reaction times have independent unique influences on it. Therefore, we predict the following relationship:

**H3:** The performance in a multitasking test and the individual reaction time are in an independent relationship with take-over time.

2. **Method**

2.1. **Sample**

Originally, 30 participants were recruited through a written announcement. Due to data logging problems, 7 participants had to be excluded, leaving the sample size for data analysis at n = 23, comprised of 13 (56.5 %) males and 10 (43.5 %) females. The mean age was $M = 34.7$ ($SD = 13.27$), with a range of 21–59 years. Of all participants, 11 (47.83 %) were students, 3 (13.04 %) were research assistants and 9 (39.13 %) were employed. All participants had held a driving license for a minimum of 4 years, with a mean of $M = 16.70$ years ($SD = 12.51$). The subjects reported to have driven $M = 20304.34$ km ($SD = 21274.04$) over the last year. They rated their experience with driver assistance systems on a 5-point Likert-scale with a mean of $M = 3.35$ ($SD = 0.93$). Participation was rewarded with 20 euros.

2.2. **Measures**

2.2.1 Take over time (TOT)

The dependent variable of the experimental design is the take over time (TOT): the point in time a subject consciously took over control, i.e. either braked or started to steer. This time point is indicated relative to the TOR and, as such, represents the time span before or after the TOR signal. A subject with a TOT of 0 ms has, therefore, taken over exactly at the same moment as the TOR was emitted.

2.2.1. **Multitasking Test (MT)**

The multitasking test was conducted on two separate monitors, on the left a Fujitsu Siemens P17-1 (screen size: 17") and on the right a Samsung Syncmaster 245B (screen size: 24"). The two monitors were placed on a table at a distance of 60 cm (frame to frame). The subjects sat in seats about 40 cm in front of the monitors, which were set at an angle of 45° to the subject who was facing straight ahead. On the left monitor, the subjects had to perform a reaction time task that is a modified version of the PEBL Perceptual Vigilance Task [18, 26]: A white fixation cross appeared for 400 ms in the center of the screen on a black background. Next, a red dot appeared at random intervals from the set [4, 5, …, 8] s. Subjects had to respond by pressing the space bar as quickly as possible. Upon pressing the space bar, the dot disappeared and a new trial started.

On the right screen, a version of the PEBL Visual Search Task [37] was used. Subjects had to search for the letter X, which was presented next to a random selection of 10, 20 or 30 distractor letters (“U”, “D”, “G”, “C”, “Q”). All letters were written in white color, the background was black. If the subject found the letter, he had to respond with a left click on the mouse. All of the letters shown turned into white circles and participants had to left click on the spot where the target was previously displayed. The resultant measure was the sum of both the reaction time of the left and the right test. Participants had to work at both tasks at the same time for 3 min. The dependent variable was the combined reaction time of both tasks.

2.2.2. **Reaction Time Test (SRT)**

We used a modified version of the PEBL Simple Response Time (SRT) [30] to measure the individual reaction time. For this task, the Fujitsu Siemens P17-1 monitor was used again. Subjects were presented with a black letter “X” on a grey background at random inter-stimulus intervals from the set [500, 750, 1000, …, 2500] ms. The required response was to press the “X” key on the keyboard as quickly as possible. Each key stroke started a new trial. Reaction times shorter than 150 ms and longer than 3000 ms were excluded from the analysis. The test was run for 75 trials.

2.2.3. The Secondary Task: Surrogate Reference Task (SuRT)

Driver distraction is often caused by texting with a cell phone, using a navigation system or a media system, all of which can all be subsumed under the “visual-manual” category [15]. To simulate visual-manual distraction we used the *Surrogate Reference Task* (SuRT; a detailed description can be found in [16]). Subjects had to solve this task in the *hard* mode. The task was implemented on a Lenovo ThinkVision LT1421 (screen size: 14") placed atop of the
central information display of the mockup. The subjects input their responses by a special keyboard that only contained the cursor keys. The task is easily interruptible and requires the subject to switch their visual focus between road/environment and the task. In order to motivate the participants to engage in the SuRT, they were told that each solved task would be rewarded with 10 cents, whereas their rewards would be halved if they collide with an obstacle.

2.2.4. Eye Tracking

We recorded each participant’s eye movements using the eye tracking system Dikablis from Ergoneers GmbH. We set up two areas of interest: one area of interest was the screen of the secondary task, to be referred to as SuRT. The other area of interest was the road and the surrounding environment, to be referred to as road/environment. The analyzed parameter was the glance location probability.

2.2.5. Driving Simulation Scenario

The study was conducted in a static driving simulator that provided a front field view of approximately 180° and three additional screens for the rear mirrors. The participants drove highly automated, i.e. longitudinal as well as lateral control was carried out by the vehicle automation for about 38 min at a speed of 80 km/h in the middle lane of a six lane highway. The automation could be switched on by pressing a button and off, by steering, or by using the gas pedal or brake. The subject’s vehicle was surrounded by 12 other road users driving at varying distances between 40–125 m. Five situations were set up in the simulator track where the participants were requested to take back vehicle control when hearing an acoustic signal. The reasons for take-over situations were obstacles in the middle lane, e.g. three rear-end collision accidents and two cars that had broken down. The view of the obstacles was obstructed by two headway vehicles until the time to collision (TTC) was 10 s. At a TTC of 3 s (66.67 m), the automation signaled that a system limit is reached and that the subject has to take over control. The appropriate reaction in this situation is to slow the vehicle down until two vehicles driving on the left and right lane respectively have passed the subject’s vehicle and then to change lanes in order to pass the obstacle. Driving time between the situations was approximately 5 min. The subjects were requested to turn on the automation again and to return to the middle lane after each situation.

2.3. Procedure

After being welcomed by the experimenter, the participants filled out a demographic questionnaire. Next, the subjects were introduced to the SRT and could complete trials until they felt comfortable starting the task. Then, they performed the SRT. Following this, they were introduced to the multitasking test. They could try out both tasks separately performed the SRT. Following this, they were introduced to the multitasking test. They could try out both tasks separately and then take a seat in the driving simulator and performed an introductory drive that consisted of manual driving, driving

with high automation and a take-over situation. After that, subjects could practice the SuRT until they felt comfortable performing it during a drive. The experimental drive then started. The introduction to the experimental drive stated that vehicle control is carried out by the automation if activated, but the responsibility for safe driving still lies with the driver. After the experimental drive was completed, the subjects received their monetary compensation.

3. Results

For all statistical tests, a significance level of $\alpha = .05$ was set. Table 1 shows the results of the multitasking test (MT) and the SRT. To determine the performance in the tests, the mean reaction time was calculated. Table 2 shows the results for the secondary task SuRT. We analyzed the mean number of solved trials, the mean number of errors made and the mean processing time for each trial. Fig. 2 shows the mean TOT in Situations 1–5. There were significant differences between the means of the situations ($F(4, 88) = 17.15, p< .001, \eta^2 = .44$), and the means decreased in a linear manner ($F(1, 22) = 36.83, p<.001, \eta^2 = .63$). We further investigated the significant differences in means by post-hoc tests with adjustment of the significance level following the Bonferroni method. Significant post-hoc tests are also marked in

<table>
<thead>
<tr>
<th>Situation 1</th>
<th>Situation 2</th>
<th>Situation 3</th>
<th>Situation 4</th>
<th>Situation 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean TOT [ms]</td>
<td>Mean number of solved trials</td>
<td>Mean number of errors</td>
<td>Mean RT [ms]</td>
<td></td>
</tr>
<tr>
<td>Situation 1</td>
<td>1950</td>
<td>1980</td>
<td>2000</td>
<td>2020</td>
</tr>
<tr>
<td>Situation 2</td>
<td>1920</td>
<td>1950</td>
<td>1980</td>
<td>2000</td>
</tr>
<tr>
<td>Situation 3</td>
<td>1900</td>
<td>1930</td>
<td>1960</td>
<td>1980</td>
</tr>
<tr>
<td>Situation 4</td>
<td>1880</td>
<td>1910</td>
<td>1940</td>
<td>1960</td>
</tr>
<tr>
<td>Situation 5</td>
<td>1860</td>
<td>1890</td>
<td>1920</td>
<td>1940</td>
</tr>
</tbody>
</table>

Figure 2.
Mean TOT in every situation in ms; negative TOT indicate a take-over before the TOR signal; bars mark significant results of post-hoc tests with Bonferroni adjustment; ** $p < .01$, *** $p < .001$; error bars mark standard deviation; Source: The authors.
In order to investigate why the correlation decreases, we divided the subjects into four quartiles based on their multitasking performance and plotted their TOT in the course of the five situations (see Fig.3). The addition of this between factors increased the explained variance of the within factor situation ($F(4, 76) = 17.47, p < .001$, $\eta^2_p = .48$), but only a trend for quartile group ($F(3, 19) = 3.25, p = .09, \eta^2_p = .15$) and no significant interaction effect ($F(12, 76) = 1.47, p = .16, \eta^2_p = .19$) was found. However, it can be seen that the means of quartile 1–3 converge until Situation 3 and the difference disappears. Nevertheless, although the worst multitaskers also decrease their TOT in the course of the experiment, their means do not converge to the other quartiles and a gap remains. Furthermore, we investigated the relationship between MT results and glance distribution and calculated Pearson’s correlation coefficient (results listed in Table 4). For Situation 1 and 2 we observed medium to large positive correlations with the SuRT and medium to large negative correlations to the road/environment. That means that the worse the performance in the MT was, the more the subjects looked at the secondary task and the less they scanned the road and environment.

![Figure 3](image)

For the othersituations the same trend was found, but the correlations were not significant. The correlations decreased linearly from Situation 2 to Situation 5 for both AOIs.

4. Discussion

The aim of this study was to investigate if take over time can be uniquely predicted by performance in a multitasking test and in a reaction time test. The take-over time decreased linearly with every take over situation (except for Situation 4). Hence, the participants learned how to cope with the dual task situation and became quicker at switching tasks.

We found a significant relationship between the multitasking test and the take over time, even when we controlled for individual reaction time. The worse the participants performed in this test, the longer was the needed time to take-over Situations 1 and 2. This finding is supported by the eye tracking data: subjects with low multitasking concentrated their gaze more on the secondary task and less on the road and environment. It is conceivable that subjects who have difficulties performing two tasks at the same time alleviated task induced stress by prioritizing one task. However, both relationships diminish in the course of the other three situations. To investigate the reason for this, we divided the subjects into four quartiles regarding their multitasking performance. For quartile 1–3, the 75% best multitaskers, the mean take-over time converges until Situation 3 and is then on an equal level. It could be possible that the participants in the 2nd and 3rd quartile either changed their strategy, increased their effort or improved their multitasking skill by learning. Improvement of their multitasking skill seems unreasonable since the 4th quartile (the lowest 25% of performers) also experiences a decrement in take over time, but there remains a gap between the other groups. Quartiles 2 and 3 would, therefore, also have to differ in a covariate that allows them to learn faster than quartile 4. Moreover, the quartiles’ performance is equivalent for the first time in Situation 3, thus it can be seen that it took more than one situation to adapt to the task. The experience of one take-over seems that it should have been enough to adapt the glance strategy. Nevertheless, stable differences in

<table>
<thead>
<tr>
<th>Situation</th>
<th>$\beta$</th>
<th>SE</th>
<th>Beta</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>0.60</td>
<td>0.25</td>
<td>.47*</td>
<td>.24°</td>
</tr>
<tr>
<td>SRT</td>
<td>3.23</td>
<td>9.19</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>0.76</td>
<td>0.22</td>
<td>.61**</td>
<td>.40*</td>
</tr>
<tr>
<td>SRT</td>
<td>4.78</td>
<td>8.00</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>0.54</td>
<td>0.36</td>
<td>.31</td>
<td>.15</td>
</tr>
<tr>
<td>SRT</td>
<td>12.86</td>
<td>13.39</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>0.38</td>
<td>0.33</td>
<td>.24</td>
<td>.13</td>
</tr>
<tr>
<td>SRT</td>
<td>14.14</td>
<td>12.10</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>.45</td>
<td>0.36</td>
<td>.27</td>
<td>.08</td>
</tr>
<tr>
<td>SRT</td>
<td>-7.46</td>
<td>13.34</td>
<td>-.12</td>
<td></td>
</tr>
</tbody>
</table>

## Table 4

Correlation between MT and glance location probability on secondary task or road/environment. Source: the authors.

** $p < .01$, * $p < .05$, ° $p < .10$; Source: authors.
multitasking ability appear to exist, since quartile 4 also became accommodated to the dual task situation (and therefore the take over time decreases), but never achieve the same performance level that the good multitaskers demonstrate. A subsequent study should clarify the reasons for the differences in performance development.

Individual reaction was in not a significant predictor of take-over time in any situation. The reason for this could lie in the difference in the actions required for the tasks. While the SRT required only a simple button press upon the appearance of a letter, a take-over is more complex, since it requires the driver to relocate their attention to the road, to process and interpret the situation, to choose a reaction, to locate the steering wheel or brake and then to execute a maneuver [12]. Another point of consideration is the low standard deviation of the SRT reaction times. Without an existing variance, no correlation can be found. A more difficult task or a larger sample could have precluded this limitation.

Certainly, the study’s conclusions have limitations. The SuRT is a very artificial secondary task that is not very interesting or distracting. Since the task was new to the subjects, they had no practice of it previously, not had they developed strategies as one would expect when using a navigation device. Moreover, the subjects engaged in the task because of the reward and compliance, but in everyday real-life automated driving, the motivation for engagement in other activities is rather intrinsic (e.g. enjoyment of a game). Beyond this, the sample’s mean age was quite low. As Körber and Bengler [19] pointed out, the effects of age could be very relevant to a take-over situation. Other researchers have already found difficulties for elderly people in working memory tasks [31], task switching [20], multitasking [14] and prolonged reaction times [32]. These skills are required for a safe engagement in a secondary task. Therefore, elderly people should be investigated as subjects in future studies on take-over times. In addition, secondary tasks with greater external validity such as games or difficult topic conference calls could be used in studies. Inter-individual differences with respect to trust in automation, attention allocation (e.g. complacency) or proneness to boredom in topics that have not yet been well researched in an automotive context should therefore also be considered in further studies.

In conclusion, a multitasking test can predict initial take over time and initial attention allocation when a driver is engaged with a secondary task. This relationship diminishes in the course of the experiment for the majority of participants, possibly due to training or a change in strategy. For the worst multitaskers, a stable difference in take-over time remains throughout every situation and can be seen as evidence for stable individual differences in dual task performance. A reaction time task, however, cannot predict take over time.

References


[22] Makishita, H. and Matsunaga, K., Differences of drivers’ reaction times according to age and mental workload. Accident Analysis & Prevention 40 (2), pp. 567-575, 2008. DOI: 10.1016/j.aap.2007.08.012


M. Körber, is a graduate research assistant under Professor Dr. Klaus Bengler at the Institute of Ergonomics at the Technische Universität München. In 2012, he earned his diploma (German equivalent to a Master’s degree) in psychology and business at the University of Regensburg. His Thesis topic was “Ethical leadership and its influence on employees’ challenging citizenship behavior”. After working on several User Experience projects, his primary research interests are vigilance, fatigue, and automation effects on driver attention.

ORCID: 0000-0002-5839-8643

T. Weißgerber, graduated with a diploma in Mechanical Eng. at the Technische Universität München. Since 2011 he has been a graduate research assistant at the Institute of Ergonomics (Professor Dr. Klaus Bengler). His main research interest is automated driving with augmented reality.

L. Kalb, is currently studying mechanical engineering at the Technische Universität München. He wrote his BSc. Thesis on an examination of eye tracking and automated driving.

C. Blaschke, graduated with a diploma in Psychology at the Philipps Universität Marburg and received a PhD in Eng. at the Universität der Bundeswehr München. Since 2013 he has been working on “Safety in use” at the BMW group.

M. Farid, earned his diploma in Electrical Eng. at the Technische Universität München. He has been working on “Safety in use of Highly Automated Driving” at the BMW Group since 2009.