Perner's Contacts ***17****(1), 2022*
DOI: 10.46585/pc.2022.1.2130

MEASURING PERFORMANCE DURING A FALLBACK PROCEDURE IN AUTONOMOUS VEHICLES

Adam Gregorovič[1], Miroslava Horáková[2*]

Abstract *The aim of this paper was to assess the functionality of an experiment that was developed for the DRILL project. The experiment used a computer program and a distracting task to measure a driver's ability to take control of an autonomous vehicle (fallback performance). Switching theory and tachistostoscopic traffic tests served as the foundation for this experiment. The experiment measured the subject's reaction time and the correctness of the reaction. The final sample consisted of N = 48 participants, who were self-selected. The results of the statistical analyses suggest that the experiment does successfully induce and register the switching effect. It was confirmed the linear relationship between the results of the experiment and the results of the ATAVT test. Further it was analyzed the reaction time on the item level (e. g. reaction times for an item level depending on its order, effect of correctness of the item´s response on the reaction time). However, the small size of the task-switching effect and the unresolved method of evaluation has not yet allowed us to draw a clear conclusion about the functionality of the experiment. Possibilities for improvements to the experiment design are also presented.*

Keywords *autonomous vehicles, automation, situational awareness, fallback task, fallback performance measurement, experiment*

1 INTRODUCTION

Autonomous mobility is a topical issue. Large automotive companies, such as Tesla, Waymo (i.e., Google) and Honda, are vying to launch the first fully autonomous vehicle (AV) (e.g., Kácovský, 2020; Guthrie, 2020; Kelly, 2020). While these companies deal mainly with the technical side of self-driving vehicles, there is another important factor — humans. People come into contact with autonomous vehicles not only as drivers of other cars, but also as passengers and drivers inside the autonomous cars. The ultimate goal for the carmakers is to achieve complete autonomy, such that vehicles will run entirely without the presence of a human being, and, ideally, no non-autonomous vehicles on the roads. Nevertheless, we are currently, and will for some time to come, be in a transition period (Gessner, 2020). This means that, although vehicles can handle most situations on their own, they do not know how to deal with certain problems on the road and they still require human assistance. This creates a completely new burden for today's drivers because this was not an issue for conventional cars (Maurer et al., 2016). The driver of an autonomous

[1] Centrum dopravního výzkumu, v. v. i. (Transport Research Centre), Líšeňská 33a, Brno 636 00, Czech Republic
[2] Centrum dopravního výzkumu, v. v. i. (Transport Research Centre), Líšeňská 33a, Brno 636 00, Czech Republic
[*] Corresponding author, phone: +420 775 862 791, e-mail: miroslava.horakova@cdv.cz

vehicle must transfer attention from non-driving activities (e.g., reading, typing on a mobile phone) to driving activities, often unexpectedly and as quickly as possible. In order to simulate this change to the driver's attention, we created an experiment to analyze it. This experiment is part of the DRILL project (Driving skill decrease), which deals with the decline in driving ability for a person who has been exposed to a long driving break, which may occur, for example, due to increased automation. In order for this experiment to be used effectively in the project, it was necessary to examine its functionality. In addition, this experimental design could be used in the future for further research into the impact of autonomous driving on drivers' abilities, as this is an emerging area. It could also find application in measuring driving ability within the context of autonomous mobility, the results of which could then be one of the decisive factors in determining a driver's ability to drive an autonomous vehicle.

1.1 Objectives of the study

This work is related to a project that falls under the Technology Agency of the Czech Republic. The researcher of the project is the Transport Research Centre, a public research institution. The project deals with the decline of driving skills, named DRILL. Within this project, an experiment was developed. This work aims to verify its functionality. The experiment is to measure how quickly and with what accuracy a person can move from a non-driving activity to a driving situation that may arise in the event of the failure of the autonomous system in the vehicle. For this purpose, it used a computer program and a distractor.

In order to consider an experiment as functional for the purposes of the DRILL project, the following assumptions needed to be met, based on theory. Respondents should perform worse on the first few photos displayed, due to returning from a non-driving task, particularly for longer reaction times. At the same time, the respondent's performance in the experiment should be positively related to his performance in the ATAVT test, as both methods should, to some extent, measure situational awareness. The respondent's performance in the experiment should also be positively related to his driving experience, as situational awareness is related to driving skills. If the assumptions prove valid, we can say that the experiment is very likely to serve its purpose.

2 THEORETICAL BACKGROUND

2.1 Current state of autonomous vehicles

The National Highway Traffic Safety Administration (NHTSA, 2013) defines an autonomous vehicle as a vehicle in which the control of some aspects of safety-critical control functions (e.g., direction control, throttle control) takes place without the direct intervention of the driver.

Autonomous cars can be of different types, depending on the maturity of their self-steering system. The Society of Automotive Engineers (SAE) (2014) specifically defines **six levels of automation**. The beginning level is zero, where the car is entirely manually controlled by the driver, and only transmission can be automatic. Levels 1 and 2 consist of the features that are available in cars sold today and that assist with the management of the vehicle, like adaptive cruise control to adjust speed and maintain a constant distance according to the vehicle in front and a lane-keeping assistant that can easily correct steering so that there is no deviation from the lane. Level 2 differs only in the ability to have both systems activated simultaneously. From Level 3 onwards the autonomous vehicle, not the human driver, does most of the driving. The driver is an observer who supervises the operation of the autonomous vehicle. At the same time, Level 3 vehicles cannot cope with all situations and occasionally require the driver to take control. The third level is in most of the currently tested autonomous cars and it is expected to enter the mainstream market. This is also the level at which this work will focus. Level 4 includes completely autonomous vehicles that can move without assistance. The last, Level 5, is able to function without restrictions under all conditions.

Autonomous vehicles are not an extremely popular topic today for no good reason. Research shows a generally **positive public inclination** toward this technology (e.g. Dudziak et al., 2021; Schoettle and Sivak, 2014). However, there will be significant regional variation. For example Stoma et al. (2021) discovered that due to many different factors, including costs, legal regulations and conviction, among others, AVs will not appear so soon in common use on Polish roads.There are several reasons to want an autonomous car.

Autonomous vehicles are able to outperform a person in several categories important in critical traffic situations. They can monitor the entire surrounding area using a series of cameras and lidar (i.e., laser radar). They may be able to communicate quickly with other autonomous vehicles and road traffic equipment (e.g., traffic lights). Furthermore, they have faster reaction times than humans and they do not suffer from common human "deficiencies", such as distractions, fatigue, and changes brought on by the ingestion of drugs.

AVs have the potential to make transportation more efficient overall and thus free up congested roads and parking in cities. More efficient transportation would not only be faster but also more environmentally friendly and, most likely, cheaper over time including logistics. In fact, various simulation models have been implemented to plan distribution services carried out by autonomous vans (Caban et al., 2022). At the same time, cars could become available to people who have traditionally not been able to use them due to physical, age, or other limitations. Car drivers could engage in other activities to make their commuting more efficient.

There are countless **benefits** associated with AV, but, of course, these are just predictions that may or may not come true. The main force behind the autonomous-mobility industry are the manufacturers, specifically their vision for profits from the new technology. Manufacturers could not only make money from the sale of the vehicle itself, as has been the case so far, but they could also benefit from software and data, especially those related to advertising. Participating companies stand to gain a lot of money, and this drives the development of autonomous mobility. Not coincidentally, it is those whose financial interest lies in AV that make positive predictions about AV to the media (Litman, 2020). Many articles have, for example, the head of Tesla, the carmaker, predicting a fully autonomous vehicle by the end of 2020 (Marshall, 2019). The truth is, however, that AVs are still far from being able to cope with all of the pitfalls of the road. For example, extreme weather, unpaved roads, and network outages are still potential problems. Tuning this technology must be 100%, because its failure can be life-threatening. Once these issues are resolved, the car manufacturer still needs to complete testing, get approval for normal operation, fulfil cost reductions, and increase customer attraction (Litman, 2020). Therefore, it is very likely that autonomous technology will remain at Level 3 for some time to come, according to the SAE (2014). For this reason, this work is focused on issues related to Level 3.

2.2 Fallback research

To better understand the methodology, it is necessary to **recapitulate the related research**. Fallback studies mainly focus on the speed and quality of the takeovers, as this is important information for the safety of autonomous vehicles. The researchers in the studies mainly used different types of vehicle simulators and different types of distractions. We can also find studies that used real cars with Level 1 or 2 autonomy (i.e., adaptive cruise control, a lane-keeping assistant).

One study that used **adaptive cruise control** was conducted by Stanton and Young (2005). The authors found a negative impact for adaptive cruise control on workload and stress, and also on the driver's situational awareness. Although this technology can be found in cars sold today, the authors chose a fixed simulator for the study, using the Situation Awareness Rating Technique (SART) to measure situational awareness.

Other studies are already dealing directly with **the fallback in a (simulated) autonomous vehicle**. Gold et al. (2013) aimed to find out when the driver's attention needed to be focused back on driving before fallback control. One simulation was driving in the right lane of a multi-lane highway at a speed of 120 kmh. After a while, an accident suddenly appears, which could be avoided by moving into the next lane or braking. The authors used a visual test as a distractor. One group was warned 7 seconds before the accident, the other 5 seconds before, and the control group was devoted to manual steering only. The results showed that faster, but worse, reactions occurred with a shorter reaction time. People checked the mirrors less and braked more. Compared to the control group, which did not use automation, the tested drivers with the longer reaction time also performed worse. They generated more acceleration and more excessive braking. Louw et al. (2015) investigated a similar situation where the tested driver's task was to avoid a collision by changing lanes. Some of the participants controlled the car manually. The other two groups had cars controlled by an autonomous system, with one group with no distractor and the other having the task of reading a tablet. Study participants slowed down less when watching with autopilot compared to both the tablet group and the manual group. The authors did not find differences in the measured variables between manual control and autopilot control without a distractor. But they found that participants responded significantly more slowly when in an autonomous system for one minute (i.e., the groups with and without a distractor). Participants in both autopilot groups also achieved greater lateral acceleration when changing lanes compared to manual steering. Overall, the participants in the groups with autonomous control recognized the situation more slowly, but, after recognition, they reacted faster and more chaotically compared to the group with manual control.

Strand et al. (2014) tested the **differences between a highly autonomous and a semi-autonomous system** for complete, severe, and light braking failures. All of the examined variables indicated that the higher degree of autonomy had a negative effect on the management of a failure and the necessary takeover of the proceedings compared to a lower degree. The analysis showed significant differences in the minimum time to collision and in the number of situations in which the car got into a position such that an accident would occur, and it would not be possible to reverse by the human driver. However, the authors did not find a significant difference for the reaction time. Nor did they find a significant difference for the time it took the vehicle to travel the distance to the position at which the leading vehicle was currently located compared to maintaining the current speed of that first vehicle. Reaction times were also different when comparing the braking failures. This points to the fact that the subjects were guided by their perceived level of urgency rather than by recognizing the failure of the autonomous system, which increased the time needed for them to react.

Other studies included **eye tracking** in the measurements. For example, Dogan et al. (2017) used the simulator to examine taking control when using the assistant for driving in a line of cars. The authors compared unexpected takeovers at 30 kmh and 50 kmh. Participants were divided into two groups, one performing a non-driving activity in the experiment and the other not. The drivers in all of the groups showed relatively short reaction times to take control of the vehicle, but then needed 6-9 seconds to stabilize and gain the necessary lateral control of the vehicle and visual attention. Engaging in non-driving activities increased the initial reaction time, but it did not affect the time to obtain the necessary lateral control of the vehicle and the monitoring of the vehicle's surroundings.

Radlmayr et al. (2014) also addressed **the impact of the traffic situation and non-driving activities** on the process of regaining control of an autonomous vehicle. Participants were divided into a group that performed a cognitive two-back task during autonomous management, a memory-intensive test. Another group focused on the visual Surrogate Reference Task (SuRT) during autonomous driving, where the goal was to select the smallest circle in a picture. The control group managed the vehicle manually, but also had a two-back task. Reaction time correlated with the cognitive load of the parallel tasks. The subjects drove in a simulator on a three-lane road at a speed of 120 kmh and they were to try to avoid an obstacle in their lane. The study included several different variants of this situation with different levels of difficulty. The obstacle always appeared so that the participants had 7 seconds to take control. This time proved to be sufficient. Apart from a significantly higher accident rate, no difference was found between the two groups with different distractions, which means that cognitive activity (i.e., the two-back task) can lead to similar distractions and thus reduced situational awareness as a purely visual activity (i.e., Surrogate Reference Task). The results also show the effect of the difficulty of the situation (i.e., traffic density) on the process of taking control.

The aim of the study by Merat et al. (2014) was to find **differences in the ability of drivers to take control**. The participants drove several kilometers along a highway with various obstacles. In the first version of the experiment, their task was to drive manually. In the next version, autonomous control was switched off at regular intervals. In the last version, autonomous steering was switched off depending on the time the participants were looking away from the road ahead. In addition to eye movements, the time to take control was measured. The results show that, in both versions of the experiment, the examined drivers took about 10 seconds to take manual control when there were not many steering corrections, followed by a few seconds for larger steering corrections and about the 30th to 40th second for these corrections to cease, indicating the achievement of stable lateral control of the vehicle. In the version where the disconnection was regular, and thus more predictable, the driver's attention to the center of the road was higher and more stable. After a 10-second delay, the lateral steering control and steering correction were more stable in the version with the given steering takeover time. The fixation of the driver's eye movements remained variable for some time after the autopilot was switched off and then taken over in a version where the disconnection of the autopilot depended on the time the driver spent observing off-road.

Similar authors have previously conducted studies of driver performance in autonomous cars in reverse (Merat et al., 2012). Participants had two rides, one in fully autonomous mode and the other in manual mode. At the same time, at certain moments, they had the task of completing the Twenty Questions Task, in which one answers Yes or No to questions, and has the task of using the skills of planning, developing solutions, and using working memory. During the ride, an obstacle appeared several times in the lane, about which the participants were warned by an audible signal 1,500 meters in advance. The authors did not find differences in the quality of the evasive maneuver between manual steering and autonomous steering without a distractor. Autonomous drivers, whose attention was diverted by a secondary task, had the worst results.

The study by Dixit et al. (2016) was previously mentioned because it dealt with the process of **disconnecting automatic steering**, which resulted in a forced transition to manual steering. The authors linked the disconnections to reaction time and accidents. The data came from companies operating autonomous vehicles on California's public roads (except for Tesla, which refused to supply data). Only Google and Mercedes measured the exact reaction times for taking control after disconnection at the time. Google and Mercedes reported a very similar and stable reaction time for fallback performance, an average of 0.83 seconds. The number of disconnections averaged 1.1 per 1,000 miles for Google and 980 per 1,000 miles for Mercedes. Manual and automatic disconnections were included in the study. Specifically, Google reported 69 manual disconnections, of which only 13 would have resulted in a collision. The remaining 56 were "safety critical due to the need to properly inspect traffic signs, give preference to pedestrians and cyclists, and violate traffic rules." The authors found differences depending on the type of disconnection, road type, and distance traveled in autonomous mode. The smallest reaction time was achieved on city roads. Confidence also played an important role, influencing how often a person switched to manual mode, and it could also result in a reduction in reaction time.

In a study by Walch et al. (2014), the authors primarily looked at the possibility of **improving the quality of fallback performance of an autonomous vehicle**. The participants of the study watched a video and, after a warning, took control of the autonomous system in the simulator. Fog and three different situations followed: a car accident, a sharp turn, and a control situation without complications. The authors manipulated the method of the warning for taking over control, which was either after 4 seconds, after 6 seconds, or after 6 seconds, with a preliminary warning of fog. On average, participants needed 1.746 milliseconds to physically take control and rated it as positive when they were alerted in advance to the possible danger that a forced takeover could cause. However, the fog warning objectively led to longer takeover times than the other two methods. It also turned out that the participants had more problems with the sharp turn than avoiding the car accident.

By contrast, Erikson et al. (2017) studied **real-life operations** and showed a slightly longer reaction time in fallback performance. The authors focused on comparing the takeover of control on the simulator versus real operation. A total of 26 participants drove in a fixed simulator. Others were driven, in real operation, by the Tesla Model S. This car is equipped with Autopilot technology that enables short autonomous driving. In both cases, participants were artificially asked to take control. Drivers took an average of 3.08 +/- 1.16 seconds to take control in real traffic and 4.56 +/- 1.63 seconds to take over in the simulator. The authors also found a strong correlation between the results of the simulator and real operation.

Fallback studies agree on the **negative impact of the secondary activity on takeover**. At the same time, studies show two examples when control is required to be taken. Dixit et al. (2016) defined the reaction time when the autonomous system was disconnected as the time interval between the moment when the driver of the autonomous vehicle is notified by the system of a system failure and the moment when the driver has taken manual control of the vehicle. According to this definition, it is only a manual takeover, as such. According to the above-mentioned studies, it only takes about 5 seconds on average. Merat et al. (2014) further formulated a "comfortable transition time", which they defined as the time required to achieve adequate stable control of the vehicle when regaining control in manual driving, which is similar for most drivers. This time includes, in addition to manual acceptance, the acquisition of the situational awareness of the vehicle and its surroundings and any maneuvers made on the basis of this awareness. This time turns out to be much longer (about 40 seconds), especially when distracted before driving. The common denominator of the studies is also a relatively small sample of participants, which is usually has a predominance of men. Also the use of simulators was in common.

2.3 Fallback of autonomous vehicles

In addition to the benefits, the introduction of autonomous cars into normal operation brings many potentials problems. Risks are associated with security against cyber-attacks and the (un)reliability of AV sensors (Kockelman et al., 2016). This work will focus on the risks associated with the person and his behavior.

In vehicles with automation Levels 2 and 3, transitions between manual and autonomous steering modes are required. This changes the role of the driver, as such, and the requirements for his cognitive system are different. The driver has fewer responsibilities to make decisions and to react physically (Maurer et al, 2016). The negative consequences of this disconnection may be as follows: excessive or, conversely, insufficient trust in the autonomous system (Lee and See, 2004); a decrease in the manual and cognitive skills associated with traditional driving (Brynjolfsson and McAfee, 2012); and the difficulty of maintaining an optimal level of situation awareness, which will be addressed in this work (Kaber and Endsley, 2004).

When the steering is taken back by the person, the autonomous system that had controlled the car is disconnected. Fallback can be divided into automatic and manual disconnections. Automatic disconnection occurs without human intervention, when the autonomous steering system determines that it is unable to continue driving on its own. This can often be due to a fault in one of the AV systems that is necessary for operation (i.e., sensors, communication with the environment, navigation maps). Manual disconnection is chosen by the driver, often because he does not have confidence in the safety of autonomous driving at the moment or he is dissatisfied with the behavior of the AV. This may be due to poor road and infrastructure conditions or worsened weather (Dixit et al., 2016).

According to an official California DMV report (2019), 9,338 disconnections were reported in 2019 from 676 vehicles, which came from a total of 36 companies and had a total mileage of 4,635,896 km. The number of disconnections per kilometer varied greatly. The best, Baidu, recorded only one disconnection at 29,049 km and the second, Waymo, recorded only one at 21,274 km. On the other hand, Toyota recorded a disconnection every 0.6 km. It is, therefore, necessary to take these numbers with great caution. So far, the data is only about testing. Each carmaker is in a different phase of testing and using a different strategy. Yet, there is evidence that disconnections are common and that drivers of autonomous vehicles are required to take control and drive from time to time. In addition, Favarò et al. (2017), who analyzed AV accidents, noted that, in most cases, autonomous steering was disconnected by the driver just before the collision. Furthermore, two of the four accidents where the AV was identified as the culprit occurred during manual driving after the autonomous system had been disconnected, and the real reason for the accident was identified as the AV driver. This proves that fallback often takes place at important points, which can have a decisive effect on driving safety. Therefore, many authors agree on the need to investigate the fallback of autonomous vehicles and its possible effects on drivers (Cummings and Ryan, 2014; Aria et al., 2016; Anund et al., 2020).

2.4 Situational awareness

In order to better understand the risks associated with AV takeover, it is necessary to clarify the concept of situational awareness. Endsley (1995) defines situational awareness as "the perception of parts of the environment at a given time and place, understanding their meaning and estimating their significance in the near future." **The definition** can be simplified to **"knowing what is happening"** (Salmon et al., 2009). Endsley (1995) includes the perception of the parts of the current situation, its understanding, and the prediction of future states. These components should enable a person to make decisions in a given situation that are aimed at fulfilling his goals. Situational awareness is one of the most important constructs of the human factor that predicts performance and safety (Parasuraman et al., 2008). Research shows that safe driving requires the continuous monitoring of frequent changes on the road and attention to relevant information (e.g., vehicle position, vehicle spacing), and to use this information to anticipate and respond to changes and events in the environment, and to avoid collisions with objects and other users of the road (Gugerty, 2011). And, because situational awareness is a crucial part of human performance in any complex dynamic environment, it is also essential for management (Endsley, 2017). However, the driver's situational awareness level may be compromised at Level 3 of AV. When a driver turns his attention away from the operation of the vehicle while the AV is driving, his situational awareness is reduced because his limited supply of attention is not used to maintain awareness of the vehicle and the overall traffic situation (de Winter et al., 2014). Alerting the driver of an autonomous vehicle to fallback performance may catch the driver unprepared and surprised (Hollnagel and Woods, 2005). The time required to take control depends on how long the driver needs to obtain the necessary information from the environment to create the required situational awareness (Endsley, 1995).

As situational awareness is linked to the perception of relevant objects, it is also linked to driving experience. Kass et al. (2007) showed that novice drivers generally had worse situational awareness than experienced drivers and, as a result, committed many more speeding offenses, had more accidents, and more often failed to stop and stay in their lane.

The driver, while under the control of the autonomous system, may have reduced situational awareness for several reasons. Previous research has shown that the transfer of certain driving tasks from driver to autonomous system makes it tempting for drivers to perform non-driving activities, such as reading or using a mobile phone (Carsten et al., 2012; Llaneras et al., 2013; Jamson et al., 2013). This urge is in line with the vision that automation promises: the time normally spent on driving can be used for more enjoyable activities. If the vehicle's tracking and control activities are automated, the driver will have very little need to be careful, especially if the driver does not have to decide which route to choose (Carsten et al., 2012). Another reason that AV drivers may tend to not monitor driving (and consequently have reduced situational awareness) is the over-reliance on the autonomous system. Even at Level 3, the autonomous system takes complete control of the vehicle, so drivers can learn to overestimate it and rely on it too much. Revaluation occurs when the driver does not question the performance of the autonomous system and insufficiently controls the state of automation (Saffarian et al., 2012). A positive relationship was found between reaction times by the fallback and the number of kilometers traveled in the autonomous mode, which the authors attributed to growing confidence in the autonomous system, which leads to its overestimation (Dixit et al., 2016). Because people do not normally encounter autonomous systems, they have difficulty understanding how they work and where exactly the limits lie. Drivers may think that an autonomous system will alert them in time or even intervene in a crisis situation (Parasuraman and Riley, 1997). If the driver acquires this impression, it further increases the likelihood of engaging in non-driving activities, thus not paying enough attention to driving (Rudin-Brown and Parker, 2004).

In all of these cases, inattention results in an **increased risk for an accident in the event of a sudden reversal**, as distraction from the road is associated with reduced situational awareness (Rogers et al., 2011; Young et al., 2013). The experiment, the functionality of which this work verifies, aims to simulate the situation where the driver's situational awareness is reduced by performing non-driving activities.

2.4.1 Measuring situational awareness

This work is devoted to an experiment that tries to reduce the driver's situational awareness, so it is necessary to know how to measure it. Most **tools for measuring situational awareness** are based on the aforementioned Endsley definition (1995). Thus, it seeks to determine whether the test subject has awareness at a given moment relevant to the activity that is being performed. In other words, a person's ability to perceive the essential elements of the environment, their memorization, or the ability to respond adequately to the current state of the environment are tested. Specifically, while driving, situational awareness can be measured by determining whether a driver currently observes and understands the condition of the vehicle, the road infrastructure, the environmental elements, and the behavior of other road users (De Winter et al., 2014).

Salmon et al. (2006) present the following six types of **tests for situational awareness**: post-trial subjective-rating techniques; observer-rating techniques; process indices; performance tests; performance measures; real-time probe techniques; and freeze-probe recall techniques. Self-assessment techniques are based on the subjective assessment of the testee. Questionnaires are most often completed after the end of the activity and consist of several scales. Observer-assessment techniques are based on predefined observable situational awareness indicators. These are then observed and recorded. Performance tests focus on relevant aspects of the test subject's performance. Process indicators try to capture the process by which the testee achieves the necessary situational awareness. For example, eye-movement sensing or complete verbalization of the activity can be used. When using the real-time-survey technique, the test subject is asked questions during the activity. The correctness of the answers and the time required for the answer are measured. Frozen-remembrance-research techniques work on the principle of stopping the activity of the tested person. Suddenly, everything that the test subject should have perceived, such as display screens, is obscured. Finally, one is asked questions about the current situation during which the test was stopped. This type of technique is also widely used in traffic psychological diagnostics in the form of the Adaptive Tachistoscopic Traffic Perception Test (ATAVT) and the Tachistoscopic Traffic Test (TAVTMB) (Schuhfried, 2011).

Adaptive Tachistoscopic Traffic Perception Test (ATAVT) is a psychological test that uses a series of briefly displayed photographs of road situations intended for passenger cars. The test is mainly focused on the extent of visual perception and the perception of speed. In other words, "it verifies the ability to visually observe and maintain situational awareness and monitors visual orientation and the speed of perception" (Schuhfried, 2011). ATAVT is provided on a computer and contains a total of 84 photos of traffic situations. The test lasts 10 minutes. Individual photos are displayed to the subject for a period of 700 to 1200 milliseconds. After each photo, the person is asked what he saw in the picture, and he can choose from a total of five options. The objects he is asked about are: pedestrians, children, vehicles, cyclists, motorcycles, mopeds, traffic lights, and traffic signs. It is an adaptive test, so the order of items is variable. After the initial phase, only the items that correspond to the performance of the tested person are shown as they advance in the process. The authors declare that the items were created on the basis of a theory that builds on the analysis of the cognitive processes that contribute to performance in the test. This is an extension of the older TAVTMBT test, the functionality of which has been proven (e.g., Karner and Neuwirth, 2000; Sommer et al., 2017). The items in the test were verified using the 1PL Rasch model. The test output is a standard score, a percentile, and a T-score. Reliability, in terms of internal consistency, is based on the validity of the 1PL Rasch model. The measurement accuracy is determined by a critical standard measurement error of 0.49, which corresponds to a reliability of r = 0.80. The specified measurement accuracy is valid for all people at all power levels. The normalization sample was 1,190 people, and the standards were divided by age. Retest validity was tested on a sample of 83 people after three months, r = 0.76 (Schuhfried, 2011).

ATAVT is used as part of a larger battery of traffic psychological tests within the Vienna Test System. In addition to the Czech Republic, the VTS method is widely used in Germany and Austria for traffic psychological diagnostics (Sommer et al., 2008). This ATAVT test battery has been validated many times (e.g., Sommer et al., 2010) and it was used in several studies to assess driving ability (e.g., Bartolacci et al., 2020; Chakrabarty et al., 2015).

Due to its quality and easy availability (for this research), this test was included as a tool to verify the functionality of the experiment.

2.5 Task switching

By fallback one often moves from one activity to another, where the first activity can act as a distractor. In this respect, the situation is similar to task-switching studies (Reichardt, 1998). This collective research usually comes to one main finding: When changing tasks, a person experiences a strong "switch cost", or tax for changing the task, in the form of increased reaction time and sometimes a greater error rate (Kiesel et al., 2010). If a person chooses an activity, according to the theory, he adopts the appropriate mental task set to perform that activity (Allport et al., 1994; Rogers and Monsell, 1995). This mental setting is the arrangement of cognitive processes and the mental representations that allow a person to act according to the requirements of a given activity (Kiesel et al., 2010).

It is the need to recall these relevant mental settings before each task change that is what, according to many authors, is behind the degraded performance of the task change (Rogers and Monsell, 1995). Other authors cite the inability to suppress outdated old mental settings from a previous task as a source for this deterioration. There are differences in the strength of the previous-setting inhibition that depend on the last task performed. Some studies argue that it is more difficult to move from a stronger (i.e., better learned) task to a weaker one, as it is more difficult for a person to suppress the better-fixed task (Allport et al., 1994). However, there are conflicting research results, so it is not clear whether it is more difficult for a person to move from a stronger task to a weaker one, or vice versa (Kiesel et al., 2010). It is likely to be due to degraded performance when changing tasks, both the reconfiguration and inhibition of old mental settings (Monsell, 2003; Ruthruff et al. , 2001). And the tax for the transition between tasks is lower when one has time to prepare. One reason is the reconfiguration of mental settings (Monsell et al., 2003; Hoffmann et al., 2003). But it is also due to the passive fading of old mental settings (Allport et al., 1994; Altmann, 2005). On the other hand, even with early notice of the transition between tasks, and even when one chooses when to make a change, we will never be exempt from tax in the form of longer transition times and degraded performance in the second task (Rogers and Monsell, 1995; Kisel el al., 2010). However, if we repeat the second task several times, the response time decreases (Meiran et al., 2000; Monsell et al., 2003). Thus, theories support what we know from fallback studies, which is the negative impact of secondary activity on the takeover.

3 METHOD

3.1 Data collection procedure

The research group consisted of a total of two groups of people. One group was directly part of the DRILL project. This group was selected through advertising, so it was through self-selection. A financial reward was paid to these participants. The experiment was part of a larger battery of tests. Part of the group was tested again after three months, as required by the DRILL project. The same variant of the fallback experiment was used for the retest. The second group of people was selected by the author of this work on the basis of acquaintance, as it was a relatively long procedure (about one hour). Both collections were negatively affected by pandemic measures related to COVID-19.

The first collection took place in the CDV building on a company laptop. The second collection took place in various quiet places (mostly at the respondents' homes) on the author's personal laptop. Both laptops had similar screen sizes (13") and the same keyboard layout.

3.2 Sample

The final group consisted of 48 respondents, of which 16 were women (33%). The age range of the tested persons was 19-56, for an average age of 27 and a median of 26. In the first sample, there were a total of 25 tested, of which 14 came for the retest. It should be noted that only this part also completed the ATAVT test. Since the retest was after more than three months and the purpose of the work was mainly to verify the functionality of the experiment, the data from the respondents in the retest was treated as data from additional unique respondents. Moreover, these respondents did not drive a car between tests, which could have further change their retest performance. The total number of administered control takeover experiments, including retest, was 62 (i.e., 48 + 14). The sample consisted mainly of university-educated respondents (N = 36). All participants had a driver's license. The average length of their driving experience was 8 years and they had caused an average of 0.5 traffic accidents.

3.3 Measurement tools

3.3.1 Experiment testing the ability of fallback performance

The experiment under investigation consists of a computer program and distraction activities. The computer program was developed by Dr. Michal Šimeček of the Transport Research Centre within the Lazarus development environment. The aim was to create an experiment that would be as similar as possible to the actual traffic situations that can occur in an autonomous vehicle. At the same time, due to the possible future application of the method in diagnostics, care was taken to make the experiment as time-consuming and as easy to perform as possible. The development was based on the theory of task switching (i.e., the alternation of two performed tasks, where one task is considered to be the main one and the other acts as a distractor). Therefore, the main measured variables are the time and accuracy of the answers. Furthermore, the theory for measuring situational insight and tachyoscopic traffic tests, the Adaptive Tachistoscopic Traffic Perception Test (ATAVT) and the Tachistoscopic Traffic Test (TAVTMB), were used. These types of tests use quickly presented photographs to measure the ability to recognize traffic situations. Thus, a database was compiled with a total of 36 different photographs from the perspective of the vehicle driver (plus 12 photographs for training). Half of the 36 photos were from traffic situations in which it was clearly safe for drivers to continue driving. The second half showed situations where the driver should clearly start braking. These photos were available on the internet.

Experimental testing was as follows. The identification of the test subject (e.g., name, code) was entered into the program and one of the 36 offered variants of the experiment was selected. The program could then be activated and the test taker was instructed to place his hand on the directional arrows on the computer keyboard, "up" and "down". It was explained to the respondent that he would see several photos on the computer screen that showed the traffic situation from the driver's point of view. The test subject had the task of responding to the photos by pressing one of the directional arrows. If the situation in the photo meant that the respondent in the hypothetical role of the driver should continue driving, his task was to press the "up" arrow key. On the other hand, if the situation in the photo indicated that he should start braking, his task was to press the "down" arrow. The test taker was instructed to respond to these situations shown in the photographs as quickly as possible, without undue thought, but correctly. The computer program started the training, so the respondent tried the experiment on 12 photographs. Before the training, the test taker was also shown an audible signal (issued by the program itself) to which he is to respond. Then a 15-minute pause followed, during which the person had the task of engaging in non-driving activities. As most of the testing took place within the DRILL project, the activity chosen was to complete paperwork (which provided data then used in the project). In this way, sufficient distraction was achieved in order to imitate the situation for a takeover. At the same time, the length of testing was used effectively. The test taker was instructed to give up any other activity when the beep sounded (in this case, completed the tests) and to return immediately to perform a computer experiment (this time it was counted separately). The experiment had a total of three rounds, between which there was always a 15-minute break filled with another distractor activity (in this case, filling in tests). During each round, the test taker was shown 12 different photographs. The design was set so that the photos were in various orders, so there were 36 versions of the test. This allowed for a subsequent analysis of the difficulty of the items. The program measured the reaction time for each item (i.e., photo) and the correctness of the answers. The first photo was displayed after pressing any key, in order to divide the reaction time and the time to move between the performed tasks. Fig. 1 shows the example of an item "you can continue driving (left)" and an example of an item "where you need to start braking (right)".

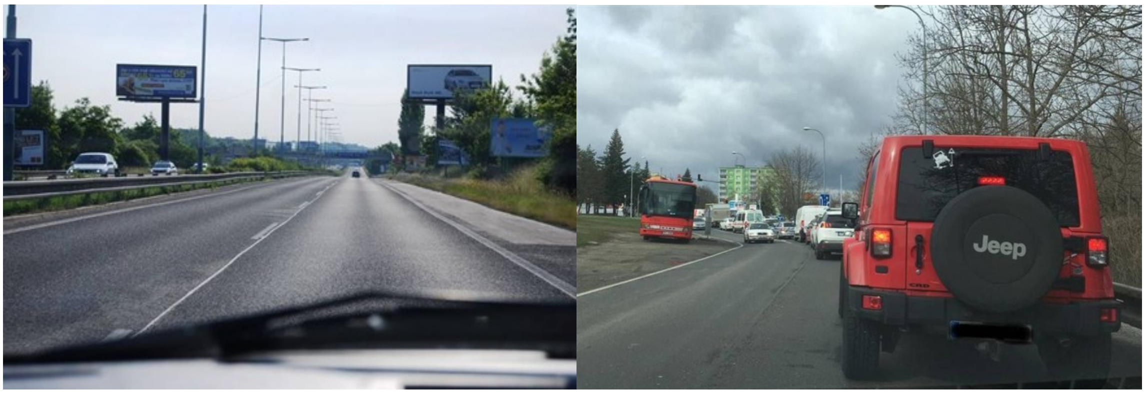

Fig. 1. Example the items

3.3.2 Other tests used in the experiment

Respondents filled in unspecified non-performance tests during the pause that occurred between the individual rounds of the computer experiment. Respondents who were tested within the DRILL project completed these tests mainly on an another computer. The second group of test takers received similar tests in paper form. Information about these tests cannot be published in this work, nor can data from these tests be used.

In order to quantify driving experience, the project allowed the use of driving experience data for analysis, but only from the group of samples that was collected by the author of this study (N = 23).

Furthermore, a test tool from the Vienna Test System, the Adaptive Tachistoscopic Traffic Perception Test, was used to verify the concurrent validity of the experiment. Specifically, it was Version 6 of the Vienna Test System and Version S1 ATAVT (i.e., for right-hand driving). VTS consists of a computer, software, a license on a USB key, and a control device.

MS Excel was used to work with and edit data. Program R 4.0.3 (R Core Team, 2020) was used for data analysis. The readxl package (Wickham & Bryan, 2019) was used to read data from MS Excel. The car package (Fox & Weisberg, 2019) was used for the Levene test and the Breusch-Pagan test; the FSA package (Ogle, Wheeler, & Dinno, 2021) was used to perform Dunn's post hoc test; and the Alboukadel rstatix package was used to calculate η2 (Kassambara, 2020).

4 RESULTS

4.1 Discarded items

The error rate of the items was examined and, on that basis, inappropriate items were excluded. A total of 62 tests were performed, each containing all 36 items. In total, 2,232 responses to items were measured. Of this number, there were 100 erroneous responses to the items and errors were made in 44 tests (71%). On average, 2.27 errors were measured in one test. A total of three items were excluded from further analyses.

Furthermore, it was decided to disable all reactions slower than 4 seconds. Too-long reactions would not necessarily correspond to the theory on which the test was based. Also, a delayed reaction could result in an accident in hypothetical situations, which are indicated in the photographs. A check was also made to see if some reaction times were less than 200 ms, which would indicate that the respondent guessed the answer to the item. Thorough check of the answers to the items of individual respondents did not reveal any suspicious clusters of errors that would indicate intentional inattention or otherwise suspiciously increased error rates. Data from all tested were kept for analysis.

4.2 Influence of error on reaction time (RT)

The error rate was examined as to whether it was related to the reaction time. Due to the small sample and the small number of errors in the data, it was not possible to examine the relationship of the error to RT. Thus, only the reaction time in the error-prone items was compared to the time in the non-error items. A t-test was chosen for this. The group of correct answers consisted of N = 1,982 items (M = 1042.09; SD = 528.63; Md = 876). The group of wrong answers consisted of N = 42 (M = 1497.48; SD = 871.35; Md = 1217). Because the RT item was a time stamp, its distribution skewed to the right. In order to bring the distribution closer to the norm, a data transformation using a logarithm was chosen. After transformation, the Shapiro Wilk test (W = 0.94; $p < 0.05$) was significant, indicating abnormal distribution. Similarly, Levene's test showed a significant result (F = 19.28; $p < 0.05$), indicating violation of the assumption of homogeneity of variances. Thus, pre-logarithmic data were used for analysis and a nonparametric Mann-Whitney U test (i.e., Wilcoxon rank sum test) was used to compare the medians of both groups. The result of this test is significant (Z = -3.12, $p < 0.05$; Cliff delta = -0.28). Thus, the reaction time of items with an error differs from the reaction time of items answered correctly.

4.3 Influence of item order on RT

Due to the results, which indicate the possible effect of errors on RT items, only correctly answered RT items were included in the following analysis. The items were divided into a total of 12 groups, which correspond to their order in the test (see Tab. 1).

Tab 1. RT averages by item order

Order	1	2	3	4	5	6	7	8	9	10	11	12
N	161	165	166	166	167	166	167	164	166	164	165	165
M	1394.6	1118.1	1025.5	1051	983.9	1004	997.1	1006.5	963.5	980.6	990.4	999.8
Md	1164	938	860	907	868	860	860	804.5	817.5	860	829	891
SD	713.7	534.1	527.7	556.7	503.9	492.2	457.9	571.7	437.5	443.4	474.9	446.5

Prior to analysis, the data were transformed with a logarithm to approximate the distribution of data to the norm. After data transformation, the Shapiro Wilk test was used ($W = 0.94$; $p < 0.05$) to indicate an abnormality in the distribution. Furthermore, a significant result in the Levene's test ($F = 2.1$; $p < 0.05$) indicated a violation of the assumption of the homogeneity of variances.

It is possible that this state of data is due to the different levels of difficulty of the items. Therefore, the averages of the individual RTs for each item were calculated and this average was then subtracted from each RT item. This should, to some extent, offset the effect of the difficulty of the items. Even after this transformation, the assumption of data normality, according to the Shapiro Wilk test ($W = 0.94$; $p <0.05$) and the assumption of homoscedasticity according to the Levene's test ($F = 2.42$; $p <0.05$), are violated.

Thus, a nonparametric equivalent of the analysis of variance in the form of the Kruskal-Wallis test was used. This test confirmed the differences in groups $x2\ (11) = 87.96$; $p <0.05$; $\eta2 = 0.04$. A post hoc test was then performed to determine which order differed in the measured RT values. Dunn's post hoc test with the Holm correction was used.

Only the first ranking achieved significant results compared to all other rankings ($p < 0.05$). All other post hoc tests for item combinations were insignificant ($p > 0.05$).

The results of these analyses confirm that there were differences in the reaction times for an item depending on its order. Specifically, the response rate for the first item differed significantly from the response rate for the items in all of the other orders (2-12). For the other items, no significant difference in reaction times was observed.

4.4 Relationship between experimental results and ATAVT

In order to verify the convergent validity of the experiment, the relationship between the results of the investigated experiment and the results of the ATAVT test was investigated. The first variable was the average RT of the person ($N = 39$; $M = 1084.91$; $SD = 295.34$; $Md = 1023.72$) and the second variable was the T score from the same person's ATAVT test ($N = 39$; $M = 57.62$; $SD = 9.46$; $Md = 56$). Only the part of the sample that was collected within the DRILL project was used. Since the data are not normal in nature and a Breusch-Pagan test confirmed the presence of heteroskedasticity ($x2\ (1) = 7.29$; $p <0.05$), a nonparametric method was chosen for analysis, namely the Kendall rank correlation test. This test did not confirm a linear relationship between the mean RT in the experiment and the T score from the ATAVT test, $r\ \tau = -\ 0.4$; $p> 0.05$.

4.5 Relationship between experiment results and driving experience

To examine the external validity of the experiment, the experiment results were compared to driving experience as the value of the years for which the respondent had driven a car. A different part of the sample was used for the analysis than for the correlation of the results with ATAVT. The total number of persons was $N = 23$. The first variable was the number of years of management of the respondent $M = 8.4$; $Md = 7$; $SD = 7.14$. The second variable was the mean RT from the same respondent's experiment ($M = 987.7$; $Md = 950.88$; $SD = 245.11$). The Shapiro Wilk's test revealed that both items violate the assumption of normality of the distribution. (For mean RT $W = 0.9$; $p <0.05$ and for years $W = 0.63$; $p <0.05$.) A nonparametric Kendall's test was then used. This again refuted the hypothesis that there is a linear relationship between the items of average RT and the item expressing the years spent driving, $r\ \tau = 1.12$; $p> 0.05$.

4.6 Comparison of experimental results for retest

Finally, the relationship between the results in the experiment and their retest was examined using a paired t test. The average RT of the respondent was again used as the result of the experiment. The first testing recorded $N = 14$; $M = 1179.67$; $Md = 1124.2$; and $SD = 404.94$. The second test recorded $N = 14$; $M = 1089.33$; $Md = 986.27$; and $SD = 184.01$. The Shapiro Wilk test showed that the data do not meet the assumption of normality for the distribution of transformations ($W = 0.8$; $p <0.05$). A logarithmic transformation of the data was performed in order to bring the data distribution closer to the norm. After data transformation, the Shapiro Wilk test was still significant ($W = 9.12$; $p <0.05$). Therefore, a nonparametric Kendall's test was chosen. It did not find a linear relationship between the results of the experiment and the retest ($r\ \tau = 0.39$; $p> 0.05$).

5 DISCUSSION

Regarding the main results of the experiments, we can discuss five main areas. **First area is the correctness of the answers**. The first analysis performed concerned the effect of the correctness of the item's response on the reaction time. Using the Mann-Whitney U test, it was confirmed that the incorrectly answered items differed in reaction time from the correctly answered items ($Z = -3.12$, $p < 0.05$; Cliff delta $= -0.28$). This result indicates the possible effect of the correctness of the response upon the time required for the response. Unfortunately, due to the violation of the assumptions for parametric tests, a nonparametric method had to be chosen, which negatively affected the strength of the test (Ratcliff, 1993). Data contained only 42 errors, which did not allow for further investigation. Therefore, it was decided to remove these erroneous answers from further analyses. This may have affected the results, as we do not know if there is a relationship between these incorrect answers. It is also possible that, despite the same instructions, different respondents chose different strategies for completing the computer program. For example, some took more risks and made more mistakes at the cost of a faster response, and vice versa.

Second area of the results concerns **the influence of item order on reaction time**. We determined how many items showed the effect of returning from the previous task. Reaction times were compared across all 12 sequences. Due to the non-fulfillment of the preconditions for parametric testing, the Kruskal-Wallis non-parametric test, which negatively affects the strength of the test (Ratcliff, 1993), was chosen. At the same time, the average reaction times of the item were calculated, and this was then subtracted from the individual reaction times of the item so as to limit the effect of the possible different difficulty of the items. The results showed that the individual reaction times differed depending on the order of the item ($x2\ (11) = 87.96$; $p <0.05$; $\eta 2 = 0.04$). A subsequent Dunn's post hoc analysis revealed that only the reaction time when the item is first in order was significantly different from the reaction time of the item in all other orders ($p <0.05$). This indicates the existence of a retuning effect. The result is, therefore, in line with the theory of task switching, which speaks of a delayed response after a task change (Kiesel et al., 2010). However, this effect is only noticeable for the first item. Thus, the other items most likely measure something other than the first item (e.g., situational influence similar to ATAVT.) The respondent's ability of fallback most likely represents only the result of the first item. This is relatively small, as one respondent is exposed to a rotation of tasks only three times. The experiment would have to be extremely accurate in these three measurements for its results to be reliable. Although research reports a fallback performance of about 4 seconds, the effect of retuning in this experiment was only seen on the first item, which lasted an average of about 1.5 seconds (Walch et al., 2014). This difference is probably due to the simplicity of the task, where the respondent needed only to decide between continuing or braking (i.e., pressing the up or down arrow), while most fallback experiments used a driving simulator where a person had many more stimuli and had to do more things at the same time (e.g., Gold et al., 2013). Similarly, thanks to IRT and adaptive input, the ATAVT test allows the respondent to display items exactly according to his performance, thus achieving the required reliability (Schuhfried, 2011).

Third area of the results focuses on **the relationship between experimental results and ATAVT test**. As already mentioned, the experiment is inspired by the ATAVT test. This test measures a situational picture, just as an experiment should measure if we subtract the effect of retuning. From the previous result, it can be seen that the effect of retuning is probably only on the item that is first, and the remaining 11 items (times three sets) should measure the situation overview similarly to the ATAVT test. Thus, the relationship between the results of the experiment and ATAVT was investigated using correlation to determine the convergent validity of these methods. Spearman's correlation ($r\,\tau = -0.4$; $p > 0.05$) was used to violate the assumptions for the parametric method. Results confirmed the linear relationship between the results of the experiment and the results of the ATAVT test. Only a portion of the sample, N = 39, was used in this analysis. A major limit to this finding is the use of the average reaction time as a result of the experiment. This was done because the experiment had not yet been evaluated. Incorrectly answered items should also enter the final score from the experiment, but these were removed before starting these analyses. This is because the relationship between the error and the reaction time is not clear, and the weight for the error in the final score is also unclear. In addition, depending on the version of the experiment, the order changes for how each photo (i.e., item) is displayed. Thus, one respondent could have one photo displayed in his version as the first and another respondent could have the same photo as, for example, the fifth. This fact would not matter if we had real certainty that the difficulty of all items was the same. However, since it is likely that the difficulty is not the same and by removing erroneous items and items to which a response of more than 4 seconds was recorded, different respondents were deprived of results from different items in a different order. This resulted in the average response time for the respondent being further skewed.

Fourth area of results concerns **relationship between experiment results and driving experience.** In this analysis, it was not possible to find a relationship between the driving experience and the result in the experiment in the form of the average RT ($r\,\tau = 1.12$; $p > 0.05$). As in the previous analysis, nonparametric Spearman correlations had to be used because the assumptions for parametric tests were violated. A positive result was expected, as the experiment should measure situational awareness, which is associated with driving skills (Kass et al., 2007). In this analysis, the same limits apply to the average reaction time variable used, as noted above in correlation with ATAVT. In addition, an inappropriate indicator of driving skills in the form of years spent driving was used. Of course, this may not be related to driving skills. The use of data was limited with regard to the data protection of the DRILL project. For the same reason, the sample for this analysis was considerably limited to N = 23. In addition, most of the drivers involved in the research rarely drove. Therefore, the respondent's length of driving experience probably had little value about his driving abilities. Indeed, research shows a link between the quality of fallback performance and driving skills (Young and Stanton, 2007).

Fifth area focuses on the **comparison of experimental results for retest**. This last analysis performed was to examine the relationship between test results and the retest of the experiment. The preconditions for the use of the paramaterial method were violated, so the Kendal test, which negatively affected the strength of the test (Ratcliff, 1993), was used. The analysis did not show a linear relationship between the test results and the retest of the experiment ($r\ \tau = 0.39$; $p > 0.05$). This is contrary to expectations, as the ATAVT test achieved a retest reliability test of $r = 0.76$ (Schuhfried, 2011). Unfortunately, only $N = 14$ respondents completed the experiment twice. The average of the reaction times was again used as a result of the test, so the limitations already mentioned in the correlation of the experiment with ATAVT apply. Due to these limitations, the results of this comparison cannot be taken as an indicator for the unreliability of the experiment. In addition, respondents did not drive cars between tests. It is, therefore, possible that the experiment measured correctly, but the ability to take control had changed. We would recommend a more detailed examination of the reliability of the experiment in the sense of testing the retest for future research.

For completeness it´s necessary to mention that the number of **incorrect answers** to individual items (i.e., photographs) was examined. It was revealed that 3 out of a total of 36 photographs had an error rate that was too high. This may mean that the traffic scene was too ambiguous for the respondent to make the right response. A high error rate could also be caused by the excessive complexity of the item, or a combination of both factors. Whatever the reason, items were removed from all analyses. The experiment assumes the same difficulty of the items, as all items are scored with the same weight, unlike ATAVT, where the difficulty of each item (i.e., photograph) was precisely determined by item analysis (Schuhfried, 2011). The assumption of the same difficulty of the items, of course, brings many potential pitfalls. It is highly unlikely that it would be just as difficult to respond to 36 photos of traffic situations. The effect of the unbalanced difficulty of the items can have a negative effect on the results of the analyses, even though effort was made to eliminate this effect. The influence of the different difficulty of the items is, to some extent, limited by the fact that each item is presented to each respondent in the experiment. However, the items changed in order, so that there was an interaction between the order and the difficulty, about which we have no information yet. This rotation of items was created mainly for a more detailed analysis. However, for use in the DRILL project, each respondent was administered two identical versions of the experiment (i.e., items in the same order). Therefore, the effect of item difficulty did not play such a large role in measuring the differences in the initial test and the retest. However, this effect could play a role in comparing measurements across respondents. Not enough data was collected to further examine the interaction between item difficulty and order, nor was this the primary goal of this work.

In addition to the error rate of responses to the items, **the speed of the response** to individual items was examined. Based on the theory, a maximum time limit for the response was set. Research showed that the rate of fallback usually takes approximately 4 seconds (Walch et al., 2014). It was also necessary to respect the logic of the scenes shown in the photographs, where a longer time delay without a reaction could most likely lead to a hypothetical accident. For this reason, all reaction times over 4 seconds were excluded. This exclusion again brings limitations, as we do not know why these respondents' reactions lasted longer than 4 seconds. It is possible that there is a relationship between these excluded reaction times, which could negatively affect the analysis.

There is also a question about the **overall design of the experiment**, which allowed respondents to respond to a given item (i.e., photograph) for as long as they wanted. For example, if you wait a very long time for a response from the first item after changing tasks, the respondent could experience a complete retuning, so that the retuning effect would no longer be recorded for the second item. On the contrary, a respondent who would respond to the first answer more quickly would not yet have time to retune, and thus the negative effect of retuning would also affect him for the second item. If, due to this retuning effect, the second respondent answered incorrectly to the second item, it could negatively affect the overall result of the experiment. A possible solution would be to determine the maximum response time for the item. After this time, the program would automatically switch to the second item and score the unanswered item as incorrectly answered.

6 CONCLUSIONS

The aim of this work was to verify the functionality of an experiment developed for the DRILL project which aimed to measure the ability of the respondent to take control of a vehicle, which is mainly connected to the issue of autonomous cars (Maurer et al., 2016). The experiment was based on the theory of task switching; in other words, alternation between two performed tasks. In this case, one task was considered major and the other a distractor. During the transition, the effect of retuning is usually observed in humans, which is mainly manifested by a worse reaction to the subsequent task (Kiesel et al., 2010). Furthermore, the experiment was based on the theory of measuring situational insight with the Adaptive Tachistoscopic Traffic Perception Test (ATAVT) and the Tachistoscopic Traffic Test (TAVTMB) (Schuhfried, 2011). These types of tests use quickly presented photographs to measure the ability to recognize traffic situations, and its validity has been demonstrated (e.g., Sommer et al., 2010; Schuhfried et al., 2010). The experiment consists of a computer program that measures the speed and accuracy of the response to the presented traffic situations, and a distractor.

It is necessary to summarize **the main and essential outcomes** of the presented research, which are in detail included in the previous parts:

- The reaction time of items with an error differs from the reaction time of items answered correctly.
- The results of analyses confirm that there were differences in the reaction times for an item depending on its order.
- The results confirmed the linear relationship between the results of the experiment and the results of the ATAVT test.
- It was not possible to find a relationship between the driving experience and the result in the experiment in the form of the average reaction time.
- The analysis did not show a linear relationship between the test results and the retest of the experiment.

Some of project´s outcomes bring **ideas for further research:**

- Based on the results concerning different reaction time of items with an error and items answered corretly, it would be appropriate for future research to investigate this relationship between error and reaction time.
- In particular, we propose to increase the difficulty of the experiment in order to better capture the effect of retuning and to create an evaluation system that would adequately account for errors in the responses to the items. Increased difficulty could be achieved with photographs that showed more complex scenes or by increasing the number of test-subject responses. It would also be good to examine whether the retuning is the effect that the experiment really measures. This could be achieved by comparing the results in the experiment both with the distraction task and without the distractor.
- The relationship between driving skills and the results of the experiment should be explored in the future, as this could be an indicator of the external validity of the experiment.
- Based on the theory, a maximum time limit for the response was set. Research showed that the rate of fallback performance is usually approximately 4 seconds. For this reason, all reaction times over 4 seconds were excluded. This exclusion again brings limitations, as we do not know why these respondents' reactions lasted longer than 4 seconds. For future research, we recommend a closer survey of times over 4 seconds.
- We would recommend a closer analysis of the individual items (i.e., photographs). Knowledge from the development of the ATAVT test, where the authors calculated individual elements in photographs and other qualities, could be used. The experiment should then be presented in a form where all items have approximately the same difficulty, or the difficulty for each item should be known and the formula for the difficulty would be included in the result. If the interaction between

the difficulty and the order of the item were also known, the items could rotate in order to reduce the learning effect of repeating the test for one person.

- The size of the sample is one of the biggest limits of this work. The small number of respondents entailed a large sampling error and limited the possibility for generalizing the results. In addition, the sample was selected (due to the time-consuming nature of the experiment) largely by self-selection. We also recommend replicating some analyses on a larger sample of respondents.
- The participants in the experiment were mostly young and inexperienced drivers, which may have contributed to the bias of the results of the experiment, which aims to examine the driving situation for a wider population.

We can summarize, the results of statistical analyses indicate the ability of the experiment to elicit and capture the effect of retuning. There are questions about the strength of the retuning effect, and the evaluation of this experiment. The results provide great insight into the knowledge of the experiment. From the performed analyses it is evident that the experiment is on a good path to measuring the driver's ability to take control. At the same time, this work offers many suggestions for improving the experiment under study. This experiment, after its future revisions, could provide valuable insights for the DRILL project, as well as find use in further research into the impact of autonomous driving on driving skills. The experiment could also be used to measure driving skills in the context of autonomous mobility. The results of a standardized method based on this experiment could then, for example, be one of the decisive factors in determining a driver's ability to drive an autonomous vehicle.

Acknowledgements

This article is co-financed from the state budget by the Technology agency of the Czech Republic under the ÉTA Progamme.

References

Allport, A., Styles, E. A. and Hsieh, S. **1994**. Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Conscious and nonconscious information processing: Attention and peifonnance XV* (pp. 421-452). Cambridge, MA: MIT Press.

Altmann, E. M. **2005**. Repetition priming in task switching: Do the benefits dissipate? *Psychonomic Bulletin & Review*, 2, pp. 535-540.

Anund, A., Lindström, A., Schagen, I. V. and Linder, A. **2020**. FERSI position paper: Safety through automation? Ensuring that automated and connected driving contribute to a safer transportation system. In *Rethinking transport. 8th Transport Research Arena TRA 2020, April 27-30, 2020.* Helsingfors, Finland (Conference canceled) (p. 5).

Aria, E., Olstam, J., and Schwietering, C. **2016**. Investigation of Automated Vehicle Effects on Driver's Behavior and Traffic Performance, 15, 761–770. <http://doi.org/10.1016/j.trpro.2016.06.063>

Bartolacci, C., Scarpelli, S., D'Atri, A., Gorgoni, M., Annarumma, L., Cloos, C., ... and De Gennaro, L. **2020**. The Influence of Sleep Quality, Vigilance, and Sleepiness on Driving-Related Cognitive Abilities: A Comparison between Young and Older Adults. *Brain Sciences*, 10(6), 327.

Brynjolfsson, E. and McAfee, A. **2012**. Winning the Race with Ever-Smarter Machines. *MIT Sloan Management Review,* 53(2), 53-60.

Caban, J., Nieoczym, A., Dudziak, A., Krajka, T. and Stopková, M. **2022**. The Planning Process of Transport Tasks for Autonomous Vans—Case Study. *Appllied Sciences,* 12(6), pp. 2993.

California DMV. **2019**. Disengagement reports. [Online]. Available at: https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/disengagement-reports/ [Accessed: 20. 1. 2021].

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H. and Merat, N. **2012**. Control Task Substitution in Semiautomated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5), 747–761.

Chakrabarty, N., Gupta, M. K., Velmurugan, S. and Rikku, M. R. **2015**. *Assessment of Driving Behaviour and Skills of Heavy Vehicle Drivers in India.* Paper ID. 451.

Cummings, M. L. and Ryan, J. **2014**. POINT OF VIEW: Who Is in Charge? The Promises and Pitfalls of Driverless Cars. *TR News*, 292, 25-30.

De Winter, J. C., Happee, R., Martens, M. H. and Stanton, N. A. **2014**. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, 27, pp. 196-217.

Dixit, V. V., Chand, S. and Nair, D. J. **2016**. Autonomous vehicles: disengagements, accidents and reaction times. *PLoS one*, 11(12), e0168054.

Dogan, E., Rahal, M. C., Deborne, R., Delhomme, P., Kemeny, A. and Perrin, J. **2017**. Transition of control in a partially automated vehicle: effects of anticipation and non-driving-related task involvement. *Transportation research part F: traffic psychology and behaviour*, 46, 205-215.

Dudziak, A., Stoma, M., Kuranc, A. and Caban, J. **2021**. Assessment of Social Acceptance for Autonomous Vehicles in Southeastern Poland. *Energies,* 14(18), pp. 5778.

Endsley, M. R. **1995**. Toward a theory of situation awareness in dynamic systems. *HumanFactors*, 37, 32–64.

Endsley, M. R. **2017**. From here to autonomy: lessons learned from human–automation research. *Human factors*, 59(1), 5-27.

Eriksson, A., Banks, V.A. and Stanton, N.A. **2017**. Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention*, 102, p. 227-234.

Fagnant, D., and Kockelman, K. **2015**. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy implications. *Transportation Re-search Part* A, 77, pp. 167-81.

Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M. and Varadaraju, N. **2017**. Examining accident reports involving autonomous vehicles in California. *PLoS one*, 12(9), e0184952.

Fox, J. and Weisberg, S. **2019**. *An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gessner, D. **2020, July 22**. Experts say we're decades from fully auto-nomous cars. Here's why. [Online]. Available at: https://www.businessinsider.com/self-driving-cars-fully-autonomous-vehicles-future-prediction-timeline-2019-8 [Accessed: 6. 12. 2020].

Gold, C., Damböck, D., Lorenz, L. and Bengler, K. **2013**. "Take over!" How long does it take to get the driver back into the loop?. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 57, No. 1, pp. 1938-1942). Sage CA: Los Angeles, CA: Sage Publications.

Gugerty, L. **2011**. *Situation awareness in driving*. Handbook for driving simulation in engineering, medicine and psychology, 1.

Guthrie, S (2020, December 4). Elon Musk 'extremely confident' Tesla will launch full-autonomous driving in 2021. [Online]. Available at: https://www.caradvice.com.au/906276/elon-musk-extremely-confident-tesla-will-launch-full-autonomous-driving-in-2021/ [Accessed: 20. 1. 2021].

Hoffmann, J., Kiesel, A. and Sebald, A. **2003**. Task switches under go/no go conditions and the decomposition of switch costs. *European Journal of Cognitive Psychology*, 15, 101-128. <http://doi.org/10.l080/09541440244000003>

Hollnagel, E. & Woods, D. D. **2005**. *Joint cognitive systems: Foundati-ons of cognitive systems engineering.* Boca Raton, FL: Taylor and Francis.

Jamson, A. H., Merat, N., Carsten, O. M. and Lai, F. C. **2013**. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation research part C: emerging technologies*, 30, 116-125.

Kaber, D.B. and Endsley, M.R. **2004**. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomic Science*, 5(2), pp. 113-153.

Kácovský, A. (2020, November 16). Waymo: Naše technologie automa-tického řízení je pokročilejší než Tesly. [Online]. Available at: https://elektrickevozy.cz/clanky/waymo-nase-technologie-automatickeho-rizeni-je-pokrocilejsi-nez-tesly [Accessed: 20. 1. 2021].

Karner, T. and Neuwirth, W. **2000**. Validation of traffic psychology tests by comparing with actual driving. In *International Conference on Traffic and Transport Psychology, 4-7 September*. Berne Switzerland

Kass, S.J., Cole, K.S. and Stanny, C.J. **2007**. Effects of distraction and expe-rience onsituation awareness and simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10, pp. 321–329.

Kassambara, A. **2020**. rstatix: Pipe-Friendly Framework for Basic Sta-tistical Tests. R package version 0.6.0. [Online]. Available at: https://CRAN.R-project.org/package=rstatix [Accessed: 20. 1. 2021].

Kelly, T. **2020, November 11**. Honda says will be first to mass produce level 3 autonomous cars. [Online]. Available at: https://www.reuters.com/article/honda-autonomous-level3-idUSKBN27R0M7 [Accessed: 6. 12. 2020].

Kockelman, K., Avery, P., Bansal, P., Boyles, S. D., Bujanovic, P., Choud-hary, T., ... and Hutchinson, R. **2016**. Implications of connected and automated vehicles on the safety and operations of roadway networks: a final report (No. FHWA/TX-16/0-6849-1).

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M. and Koch, I. **2010**. Control and interference in task switching—A re-view. *Psychological bulletin*, 136(5), 849.

Llaneras, R. E., Salinger, J. and Green, C. A. **2013**. Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway.

Lee, J. and See, K. **2004**. Trust in automation: designing for appropriate reliance. *Human Factors*, 46, pp. 50-80.

Litman, T. **2020**. *Autonomous Vehicle Implementation Predictions Implications for Transport Planning.* Victoria Transport Policy Institute.

Louw, T., Merat, N. and Jamson, H. **2015**. Engaging with Highly Automated Driving: To be or Not to be in the Loop?.

Maurer, M., Christian Gerdes, J., Lenz, B. and Winner, H. **2016**. *Autonomous driving: technical, legal and social aspects.* Springer Nature.

Marshall, A. **2019, February 19**. Elon Musk Promises a Really Truly Self-Driving Tesla in 2020. [Online]. Available at: https://www.wired.com/story/elon-musk-tesla-full-self-driving-2019-2020-promise/ [Accessed: 6. 12. 2020].

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., ... and McGehee, D. **2019**. The "out-of-the-loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work*, 21(1), pp. 87-98.

Merat, N., Jamson, A. H., Lai, F. C., Daly, M. and Carsten, O. M. **2014**. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, 27, pp. 274-282.

Merat, N., Jamson, A. H., Lai, F. C. and Carsten, O. **2012**. Highly automated driving, secondary task performance, and driver state. *Human factors*, 54(5), pp. 762-771.

Meiran, N., Chorev, Z. and Sapir, A. **2000**. Component processes in task switching. *Cognitive Psychology*, 41, pp. 211-253. <http://doi.org/10.10061 cogp.2000.0736>

Monsell, S. **2003**. Task switching. *Trends in Cognitive Sciences*, 7, 134-140. <http://doi.org/10. 10 I/S 1364(03)00028-7>

Monsell, S., Sumner, P. and Waters, H. **2003**. Taskset reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, 31, 327-342.

National Highway Traffic Safety Administration (NHTSA). **2013**. *Preliminary statement of policy concerning automated vehicles.* National Highway Traffic Safety Administration, Washington DC, National Highway Traffic Safety Administration.

Ogle, D.H., P. Wheeler, P. and Dinno, A. **2021**. FSA: Fisheries Stock Analysis. R package version 0.8.32. [Online]. Available at: https://github.com/droglenc/FSA. [Accessed: 6. 2. 2021].

Parasuraman, R. and Riley, V. **1997**. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), pp. 230-253.

Parasuraman, R., Sheridan, T. B. and Wickens, C. D. **2008**. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of cognitive engineering and decision making*, 2(2), pp. 140-160.

Reichardt, D. M. **1998**. Using automated assistance systems-Putting the driver into focus. In IEEE Intelligent Vehicles (IV) Symposium (pp. 413-418).

Radlmayr, J., Gold, C., Lorenz, L., Farid, M. and Bengler, K. **2014**. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2063-2067). Sage CA: Los Angeles, CA: Sage Publications.

Ratcliff, R. **1993**. Methods for dealing with reaction time outliers. *Psychological bulletin*, 114(3), pp. 510.

R Core Team. **2019**. R: A Language and Environmentfor Statistical Com-puting. R Foundation for Statistical Computing. [Online]. Available at: https://www.r-project.org [Accessed: 20. 1. 2021].

Rogers, R. D. and Monsell, S. **1995**. Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, pp. 207-231.

Rogers, M., Zhang, Y., Kaber, D., Liang, Y. and Gangakhedkar, S. **2011**. The effects of visual and cognitive distraction on driver situation awareness. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 186-195). Springer, Berlin, Heidelberg.

Ruthruff, E., Remington, R. W. and Johnston, J. C. **2001**. Switching between simple cognitive tasks: The interaction of top-down and bottom-up factors. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 1404-1419. <http://doi.org/10.1037110096-1523.27.6.1404>

SAE J3016. **2014**. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Society of Automotive Engineers.

Saffarian, M., de Winter, J. and Happee, R. **2012**. Automated Driving: Human-Factors Issues and Design Solutions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), pp. 2296-2300.

Salmon, P., Stanton, N., Walker, G. and Green, D. **2006**. Situation aware-ness measurement: A review of applicability for C4i environments. *Journal of Applied Ergonomics*, 37(2), pp. 225-238.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Raffer-ty, L. and Young, M. **2009**. Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3), pp. 490-500.

Salmon, P. M., Stanton, N. A., Walker, G. H. and Jenkins, D. P. **2009**. *Distributed Situation Awareness: Advances in Theory, Measurement and Application to Teamwork*. Aldershot, UK: Ashgate.

Schoettle, B. and Sivak, M. **2014**. *Public opinion about self-driving vehicles in China, India, Japan, the US, the UK, and Australia*. University of Michigan, Ann Arbor, Transportation Research Institute.

Schuhfried, G. **2011**. *Adaptive Tachstoscopic Traffic Perception Test: Manual*. Mödling: Schuhfried GmbH

Schuhfried, G., Sommer,M., Aniţei, M. and Chraif, M. **2010**. The validation methodology of a psychological assessment battery for drivers' schools in Romania. *Romanian Journal of Applied Experimental Psychology*, 1(1), pp. 5-14.

Sommer, M., Herle, M., Häusler, J., Risser, R., Schützhofer, B. and Chaloupka, C. **2008**. Cognitive and personality determinants of fitness to drive. *Transportation research part F: traffic psychology and behaviour*, 11(5), pp. 362-375.

Sommer, M., Schuhfried, G., Aniţei, M. and Chraif, M. **2010**. *A model of driving test battery validation. Journal of Psychology*, 56, (3–4), pp. 243–250.

Sommer, M., Häusler, J. and Schuhfried, G. **2017.** Non-linear methods for the identification of drivers at risk to cause accidents. *Driver Behaviour and Training*,2 (2), pp. 425.

Stanton, N. A. and Young, M. S. **2005**. Driver behaviour with adaptive cruise control. *Ergonomic*s, 48(10), pp. 1294-1313.

Stoma, M., Dudziak, A., Caban, J. and Droździel, P. **2021**. The Future of Autonomous Vehicles in the Opinion of Automotive Market Users. *Energies,* 14(16), pp. 4777.

Strand, N., Nilsson, J., Karlsson, I. M. and Nilsson, L. **2014**. Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation research part F: traffic psycho-logy and behaviour,* 27, pp. 218-228.

Straka, J. and Pelešková, J. **2020**. *Ročenka nehodovosti na pozemních komunikacích za rok 2019.* Praha: Ředitelství služby dopravní policie Policejního prezidia České republiky.

Vahidi, A. and Eskandarian, A. **2003**. Research advances in intelligent collision avoidance and adaptive cruise control. *IEEE transactions on intelligent transportation systems*, 4(3), pp. 143-153.

Walch, M., Lange, K., Baumann, M. and Weber, M. **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* (pp. 11-18). ACM.

de Waard, D., van der Hulst, M., Hoedemaeker, M. and Brookhuis, K. A. **1999**. Driver behavior in an emergency situation in the Automated Highway System. *Transportation human factors*, 1(1), pp. 67-82.

Wickham, H. and Bryan, J. **2019**. readxl: Read Excel Files. R package version 1.3.1. [Online]. Available at: https://CRAN.R-project.org/package=readxl [Accessed: 5. 12. 2020].

Wojcik, R. (2019, November 21). How Self-Driving Cars Will Change Digital Marketing Forever. [Online]. Available at: https://www.digitalcurrent.com/blog/what-self-driving-cars-mean-for-the-future-of-digital-marketing/ [Accessed: 5. 12. 2020].

Young, K. L., Salmon, P. M. and Cornelissen, M. **2013**. Missing links? The effects of distraction on driver situation awareness. *Safety Science*, 56, 36–43.

Young, M. S. and Stanton, N. A. **2007**. What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics*, 50(8), 1324—1339.