Estimating the out-of-the-loop phenomenon from visual strategies during highly automated driving

Damien Schnebelen, Camilo Charron, Franck Mars

Estimating the out-of-the-loop phenomenon from visual strategies during highly automated driving

Authors : Damien SCHNEBELEN[1], Camilo CHARRON[1,2], Franck MARS[1]

[1] Université de Nantes, Centrale Nantes, CNRS, LS2N, F-44000 Nantes, France

[2] Université de Rennes 2, F-35000 Rennes, France

Corresponding author: Franck MARS

E-mail address: franck.mars@ls2n.fr,

Postal address:

LS2N - Centrale Nantes

1 rue de la Noë

B.P. 92101

F- 44321 Nantes Cedex 03

FRANCE

Abstract

During highly automated driving, drivers no longer physically control the vehicle but they might need to monitor the driving scene. This is true for SAE level 2, where monitoring the external environment is required; it is also true for level 3, where drivers must react quickly and safely to a take-over request. Without such monitoring, even if only partial, drivers are considered out-of-the-loop (OOTL) and safety may be compromised. The OOTL phenomenon may be particularly important for long automated driving periods during which mind wandering can occur. This study scrutinized drivers' visual behaviour for 18 min of highly automated driving. Intersections between gaze and 13 areas of interest (AOIs) were analysed, considering both static and dynamic indicators. An estimation of self-reported mind wandering based on gaze behaviour was performed using partial least squares (PLS) regression models. The outputs of the PLS regressions allowed defining visual strategies associated with good monitoring of the driving scene. This information may enable online estimation of the OOTL phenomenon based on a driver's spontaneous visual behaviour.

Keywords: mind wandering; gaze behaviour; autonomous vehicles; PLS regression; driver monitoring

1. Introduction

The deployment of highly automated vehicles on the roads is imminent; it could occur anywhere between 2020 and 2030 (Chan, 2017). Among the expected benefits of autonomous vehicles (environmental, societal, etc), road safety is expected to improve (Fitch et al., 2014). The number of crashes caused by human error could be reduced by 90%, according to Stanton and Salmon (2009). However, to meet that target, drivers must be clearly aware of their role in the vehicle. That role partly depends on the level of automation.

There are five levels of automation (SAE International, 2016), corresponding to a different balance of tasks between the automation and the driver. At levels 0 and 1, the driver is in full or partial control of the vehicle commands. Automated driving starts at level 2, when longitudinal and lateral control of the vehicle is performed by the automation, but the driver must continuously monitor the driving scene and intervene when needed, even without a request from the system. At level 3 (conditional automation), the driver may engage in secondary tasks, but must be able to regain vehicle control when required by the system. This implies that monitoring the driving scene is only required starting with the take-over request. At level 4 (high automation), under certain conditions, the automation is able to perform all driving functions and can handle critical situations without requesting a take-over, although driver

override may be possible. In level 5 (full automation), the vehicle is autonomous in all conditions and the driver's action is no longer required.

In all cases, as soon as the operational part of the driving task is automated, drivers become supervisors of the automation system and of the driving scene. The level of expected supervision decreases as the level of automation increases. In manual driving, drivers must gather information from the driving scene and the vehicle (perceptual process), and must interpret that information (cognitive process) and act appropriately (motor process). Their actions in turn generate information. By contrast, starting with level 2 automation, the perceptual-motor loop is neutralized, which has consequences for perception and cognition (Mole et al., 2019). At level 3, driver engagement in secondary tasks may intensify those consequences, with long periods of distraction from the driving scene. These consequences are referred to in the literature as the out-of-the-loop (OOTL) phenomenon.

The OOTL phenomenon was first observed in the aviation field (Endsley and Kiris, 1995), where automated piloting has long existed. Parasuraman and Riley (1997) showed that human pilots may be poor supervisors of the system. They may enter into a passive state when interacting with highly automated systems, causing a lack of situation awareness (Endsley, 1995). Recently, Merat et al. (2019) proposed an operational definition of OOTL in the context of automated driving. To be OOTL, drivers must lack physical control of the vehicle (motor process) and must not be monitoring the driving scene (perceptual or cognitive process). By contrast, when the driver is in manual control, they are considered to be in-the-loop. An intermediate state, namely on-the-loop (OTL), was introduced to designate cases where the driver correctly monitors the situation during autonomous driving.

In cars, the OOTL phenomenon has mainly been investigated through comparisons of driver behaviour between automated and manual driving. As most of the information processed during driving is visual (Sivak, 1996), the analysis of gaze behaviour has received much interest. Compared to manual driving, simulated automated driving involves more horizontal dispersion of gaze (Mackenzie and Harris, 2015; Louw and Merat, 2017) and a lower percentage of glances towards the road centre (Louw et al., 2015a; Mackenzie and Harris, 2015). Similarly, in curve driving, automated driving was shown to enhance long-term anticipation through look-ahead fixations, to the detriment of short-term anticipation used to guide the vehicle (Mars and Navarro, 2012; Schnebelen et al., 2019). When a secondary task was performed by drivers, automated driving was associated with relatively frequent fixations on those tasks (Merat et al., 2012).

When drivers are required to regain control of the vehicle, their behaviour after the take-over request is also considered to be an indicator of the OOTL phenomenon. Differences between automated and manual driving, as observed in critical scenarios, indicate that automated driving leads to impaired

visuomotor coordination during take-over (Mole et al., 2019). Navarro et al. (2016) showed that gaze distribution was widely dispersed, resulting in difficulties in steering around unexpected obstacles. Furthermore, drivers had longer reaction times to critical events, and vehicular control was impaired (Neubauer et al., 2012; Gold et al., 2013; Saxby et al., 2013; Louw et al., 2015b; Zeeb et al., 2015, 2017; Eriksson and Stanton, 2017). Such changes in driver behaviour during take-over were attributed to drivers being OOTL during automated driving.

In addition, the OOTL phenomenon seems to increase with a prolonged period of automation. In the aviation field, Molloy and Parasuraman (1996) showed that monitoring performance decreased with time. In the automated driving context, some studies have indicated that prolonged periods of automated driving rendered drivers further OOTL (Körber et al., 2015; Feldhütter et al., 2017; Bourrelly et al., 2019). For example, Bourelly et al. (2019) observed longer reaction times (+0.5 s) to a critical event after 1 h of automated driving, compared to reactions to the same event after 10 min.

According to the definition of OOTL in automated driving (Merat et al. 2019), OOTL drivers do not correctly monitor the driving situation. The OOTL state may be experimentally induced by modifying perceptions of the driving environment or through instructions given to the drivers (i.e. using a secondary task). Louw et al. (2015b, 2016, 2017) reduced the visual information available for drivers using simulated fog to examine whether the driver was further OOTL when the fog was dense. In that sense, the OOTL state mostly occurred through impairment of perception. In other studies (Carsten et al., 2012; Merat et al., 2012), monitoring of the driving environment was altered by a secondary task. In these cases, the degradation of both perceptual (eyes off-road) and cognitive processes (mind off-road) yielded the OOTL phenomenon.

However, OOTL may also spontaneously and progressively occur without any modification of the driving environment or the presence of a secondary task. Due to lack of activity, drivers can experience mind wandering (MW), progressively disengaging from the supervision task even in the absence of an external source of distraction (Körber et al., 2015; Feldhütter et al., 2017; Burdett et al., 2019). Gouraud et al. (2017) proposed a detailed review of the links between the OOTL phenomenon and MW. Both phenomena are characterized by a decoupling of the immediate task and leads to similar safety issues. The decoupling can be the result of a reduction in the perception of the environment relevant to the task (sensory attenuation), which can have a negative impact on the construction of an accurate situation model. Both MW and the OOTL phenomenon have been associated with a slower response or detection failure when a critical event occurs. Gouraud et al. (2017) conclude that MW markers could help study OOTL situations. MW can be assessed by various means, including physiological and behavioural indicators, although self-report measures remain widely used because of their robustness. Specific to the case of autonomous driving, MW may occur at level 2, when the driver is supposed to monitor the

driving scene. It may also happen at level 3 if the driver chooses not to engage in secondary tasks. In both cases, MW can be more difficult to assess than visual distraction, since gaze is not diverted from the driving scene.

The OOTL state was defined in previous studies in a relative way: either drivers are more OOTL when automation is on, compared to manual driving; or they are more OOTL during prolonged automation, compared to a shorter duration. In all cases, the OOTL phenomenon impacts driver safety during take-over, especially during long automated drives. Hence, determining whether a driver is OOTL or OTL must be performed before any situation requiring a take-over.

This study estimated the driver's state based on the observation of spontaneous gaze behaviour. The participants experienced an 18-min drive of automated driving (similar to Feldhuetter et al. 2017), with the OOTL phenomenon occurring spontaneously. Quantitative assessment of the OOTL state was based on the self-reported time of MW. The drivers' gaze behaviour was analysed by considering 13 areas of interest, using static and dynamic indicators. Static indicators refer to the percentage of time the gaze is directed to one AOI and dynamic indicators refer to transitions from one AOI to another. An original method was used that involved multiple partial least squares (PLS) regression analyses. The goals were 1) to identify and select the most important gaze indicators, and 2) to generate models for the relationship between gaze behaviour and MW score. The principle of this methodological choice was to identify the essential indicators from the overall spontaneous gaze behaviour, without relying on theoretical preconceptions.

The research was guided by two research questions:

1. Is it possible to identify gaze behaviour that is characteristic of OOTL drivers? (i.e. what constitutes inadequate monitoring of a driving situation?)
2. Is it possible to estimate the driver's OOTL state from the observation of spontaneous gaze strategies?

2. Materials and Methods

2.1. Participants

The study involved 12 participants (N = 12; 3 female, 9 male), with a mean age of 21.4 years (SD = 5.34 y). To facilitate the recording of accurate gaze data, volunteers were required to have either normal

vision or vision corrected with contact lenses. They all held a valid driver's license, with average driving experience of 9950 km/year (SD = 5500). They signed written informed consent to participate.

2.2. Experimental setup

Figure 1 presents the driving simulator setup. This fixed-base simulator consisted of an adjustable seat, a steering wheel with force feedback, a gear lever, clutch, accelerator and brake pedals. The driving scene was generated with SCANeR Studio (v1.6) and displayed on three large screens in front of the driver (field of view ~= 120°). A dashboard screen indicated the speed of the vehicle. An HMI screen was added to the right-hand side of the driver, approximately where a vehicle's centre console is located. The description of the HMI can be found in the Procedure section.

Figure 1: Driving simulator setup

Gaze data were recorded using a Smart Eye Pro (V5.9) eye-tracker with four cameras; two were below the central screen and one below each peripheral screen. The calibration was performed in two steps. First, a 3D model of the driver's head was computed using an 11-point head calibration procedure with the head and gaze oriented toward the points. Then, the gaze was calibrated using 15 points: nine on the central screen, two on each peripheral screen, one on the dashboard and one on the HMI screen; with the head oriented to the central screen and gaze directed to the points. Gaze data were synchronized and recorded with vehicle data at 20Hz by the driving simulator software.

Most of the road was a 40-km two-lane dual carriageway, with a speed limit of 130 km/h in accordance with French regulations. Occasional changes in road geometry and speed limits were included to make the driving less monotonous. This included temporary 3-lane traffic flow, highway exits, slope variation

and variation of the speed limit (130 km/h to 110 km/h). In both directions on the highway, traffic was fluid, with eight overtaking situations.

2.3. Procedure

The participants first adjusted the seat position, and the gaze calibration procedure was performed. Then, they drove manually along a training track to become accustomed to the driving environment and the vehicle's reactions. Once this training session was completed, instructions for automated driving were given orally.

Drivers were told that the automated function would only be available for a portion of the road. The distance and time remaining in the autonomous mode were displayed on the left of the HMI. When activated by pressing a button, the automation controlled the lateral position and speed of the vehicle appropriately, accounting for traffic, speed limits and other conditions, and overtaking other cars if necessary. Participants were instructed to take control of the vehicle when requested by the system.

Two possible use cases were presented to the participants. In the first case, the vehicle was approaching the end of the automated road section. The drivers would receive mild auditory and visual warning signals and would have 45 s to regain control. The second use case was an unexpected event, such as the loss of sensors. In this case, an intense auditory alarm would sound and a new pictogram would be displayed, and drivers would have only 8 s to resume control. All the pictograms and sounds used by the HMI were presented to participants before they began a second training session. The pictograms are shown in Figure 2.

Figure 2: Pictograms displayed on the HMI. A, autonomous driving available; B, autonomous driving activated; C, critical take-over request (8 s); D, planned take-over request (45 s).

The second training session allowed participants to experience semi-automated driving (SAE level 1): cruise control with the driver in charge of the steering wheel; and level 3: conditional automation. At level 3, there were four transitions to manual control, two in each use case presented in the instructions. All take-overs were properly performed during the training session.

The experiment itself then commenced. The study followed a within-participant design, with all participants driving under both automated and semi-automated conditions. These conditions were similar, but no critical case happened in the semi-automated condition. However, because this paper concerns the analysis and modelling of gaze behaviour during automated driving, only the results obtained during the automated condition are presented here.

In the automated driving condition, participants activated the automated driving mode just before entering a highway. Gaze data were recorded as soon as the vehicle was correctly inserted in the lane and had reached 130 km/h. No major driving events or take-over request occurred in the first 15 min on the road, to allow enough time for the driver to become OOTL. The driver did not perform any secondary task during that time.

Between minutes 16 and 17, the two vehicles that would be involved in the critical case appeared in the driving scene. One overtook the participant's vehicle on the left and positioned itself in the right lane, 300 m ahead. The other vehicle remained in the left lane, slowly approaching the participant's vehicle. During minute 18 of automated driving, participants experienced a critical take-over request occurring in response to unexpected braking from the lead vehicle. The warning signals were delivered as soon as the lead vehicle started to brake, with a time-to-collision of 8 s. At that moment, the lead vehicle was in the adjacent lane, in the blind spot of the participant. Changing lane would lead to a collision. To successfully handle the critical situation, drivers had to brake, remain in the right lane until the overtaking vehicle had passed, and then change lanes to avoid the lead vehicle. The scenario ended 30 s after the critical case.

Participants were then asked to report on a continuous Likert scale the proportion of time they had spent thinking about something other than the driving task, throughout the trial. This simple method of MW self-assessment has been shown to be sensitive to driver disengagement during prolonged driving sessions (Mars et al., 2014).

2.4. Data structure and annotations

2.4.1 Definition of the MW score Y

In the absence of a secondary task, the evaluation of the percentage of time spent thinking about something other than the driving task was regarded as a self-assessment of the MW. The higher the percentage, the more the driver had estimated being OOTL. Percentages for all participants were stored

in a vector with 12 elements. After standardization of the data (conversion to a z score), the vector was denoted as Y and named MW score.

2.4.2 Definition of the matrix of gaze behaviour X

The driving scene was divided into 13 areas of interest (AOI), as shown in Figure 3. These are described below.

a) The central screen contained six AOIs:

- central mirror (CM)
- road centre (RC), defined as a circular area of 8° radius in front of the driver
- four additional areas, defined relative to the road centre (Up, Left, Down, Right).

The percentage road centre (PRC) is defined as the proportion of time spent in RC, as introduced by Victor (2005). A decrease in PRC was found to be a reliable indicator of distraction during driving; drivers reduced their PRC when visually or auditorily distracted (Victor et al., 2005).

b) Each peripheral screen contained two areas, with two items in each:

- lateral mirror (LM, RM)
- the remaining peripheral scene (LS, RS)

c) The dashboard (D)

d) The HMI (HMI)

e) All data for gazes directed outside all the above areas were grouped as “other areas” (Others).

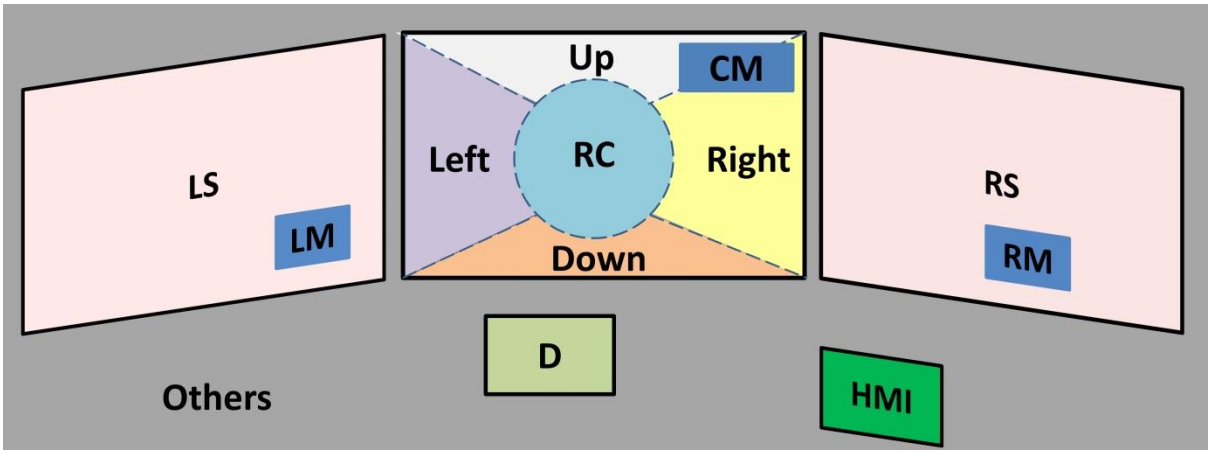


Figure 3: Division of the driving environment into 13 areas of interest

The percentage of time spent gazing at each AOI was computed, as was a matrix of transitions between AOIs. Using Markov logic, the matrix of transitions between AOIs corresponded to the probability of shifting from one AOI to another. The probability that gaze remained in the same AOI was also computed. This constituted the diagonal of the transition matrix. Probabilities were estimated by the observations of the participants (see Gonçalves et al., 2019, for an application of Markov chains in case of a lane change manoeuvre). Before a given AOI was intersected for the first time, a prior probability was associated with it. A uniform law was used for prior probabilities.

The 13 AOIs defined the entire world. Thus, the transition matrix was a 13*13 matrix. If rows contained the current AOI and columns the probabilities, the sum of each row was equal to 1. In all cases, next gaze intersection appears in one of the 13 areas of interest.

In this study, the driver's gaze behaviour for each participant was considered to be the combination of static and dynamic indicators of gaze behaviour. Static refers to percentage of time in one AOI; dynamic refers to the transition matrix. Thus, the gaze strategy of a participant was represented by a vector of 182 numerical indicators (= 13*13 transitions + 13 percentage of time on each AOI). When considering all participants, the matrix of gaze strategies was named X, and its size was 12 (participants) * 182 (visual indicators).

Supplementary indicators that were sensitive to drivers' drowsiness were also computed. These included percentage of eye closure (PERCLOS; see Wierwille et al., 1994) and the blink-rate (Stern et al., 1994).

2.4.3 Computation of training and validation datasets

The objective was to predict the MW score as a function of gaze behaviour. Therefore, a training dataset (i.e., a gaze-behaviour matrix to create the model) and a validation dataset (a gaze-behaviour matrix to evaluate the model) were required.

The gaze-behaviour matrix obtained during the last two minutes (16 and 17) of automated driving was chosen as the validation dataset. The rationale was that because OOTL increases with automation duration, the final two minutes might reflect the visual consequences of the OOTL phenomenon most accurately. The validation dataset did not include the gaze data recorded once the take-over request was initiated.

Nevertheless, the question of the speed of appearance of OOTL and the observation time required to model it was of interest. To answer this question, 15 training datasets (i.e. 15 gaze-behaviour matrixes), labelled X_t, were calculated. The difference between the sets was the integration time – that is, the duration of automated driving while the matrix of gaze behaviour was computed. The integration time varied from 1 min to 15 min. The reference for the time window was the 15th minute of simulation (i.e. immediately before the validation dataset). Thus, X_1 considered gaze behaviour during one minute (the 15th minute), whereas X_7 was computed from seven minutes of automated driving (between the 9th and 15th minutes). The rationale was to evaluate whether using a short time window was enough to capture the consequences of OOTL to create a satisfactory model, or whether aggregating more data by enlarging the time window would make the model more robust.

In summary, 15 matrixes of gaze behaviours (labelled X_t for t between 1 and 15 min) were calculated, and these constituted the training datasets for the model. Once the best model had been selected, we validated it by comparing the MW score to the model's prediction based on the two final minutes of driving (16th and 17th minutes).

2.4.4 Choice of model for prediction: PLS regression

Regarding the data structuring described above, the aim was to predict Y (MW score) from X_t (gaze-behaviour matrix), given the following conditions:

- Visual indicators (X_t) are correlated. The driving environment was divided into 13 AOIs, hence, the percentage of time spent on 12 AOIs enabled calculating the percentage of time spent on the 13th AOI. In mathematical terms, X_t might not be full rank.
- The number of visual indicators (182) that could explain Y is higher than the number of observations made on the participants (12).

Considering these constraints, the PLS regression model was selected. PLS regression yields the best estimation of Y available with a linear model given the matrix X_t (Abdi, 2010). All PLS regressions performed in this study used the PLS regression package (Wehrens and Mevik, 2007) from R (Core Team et al., 2013).

2.5. Data analysis

Four sequential stages composed the analysis of the training datasets (X_t) (see Figure 4):

1. Step A entailed finding the optimal structure (i.e. the optimal number of components) of the PLS regression model for predicting the MW score.

2. Step B entailed selecting the visual indicators that significantly contributed to MW score prediction.
3. In Step C, we considered the optimal parameters (components and visual indicators) of the prediction model, and evaluated the model's accuracy using the mean square error of prediction (MSEP) for both the training and validation datasets.
4. In Step D, we selected the model with the least validation error.

The step-by-step procedure of the data analysis is presented in the appendix. Only the final results, which lead to model selection, are presented in the next section (Results).

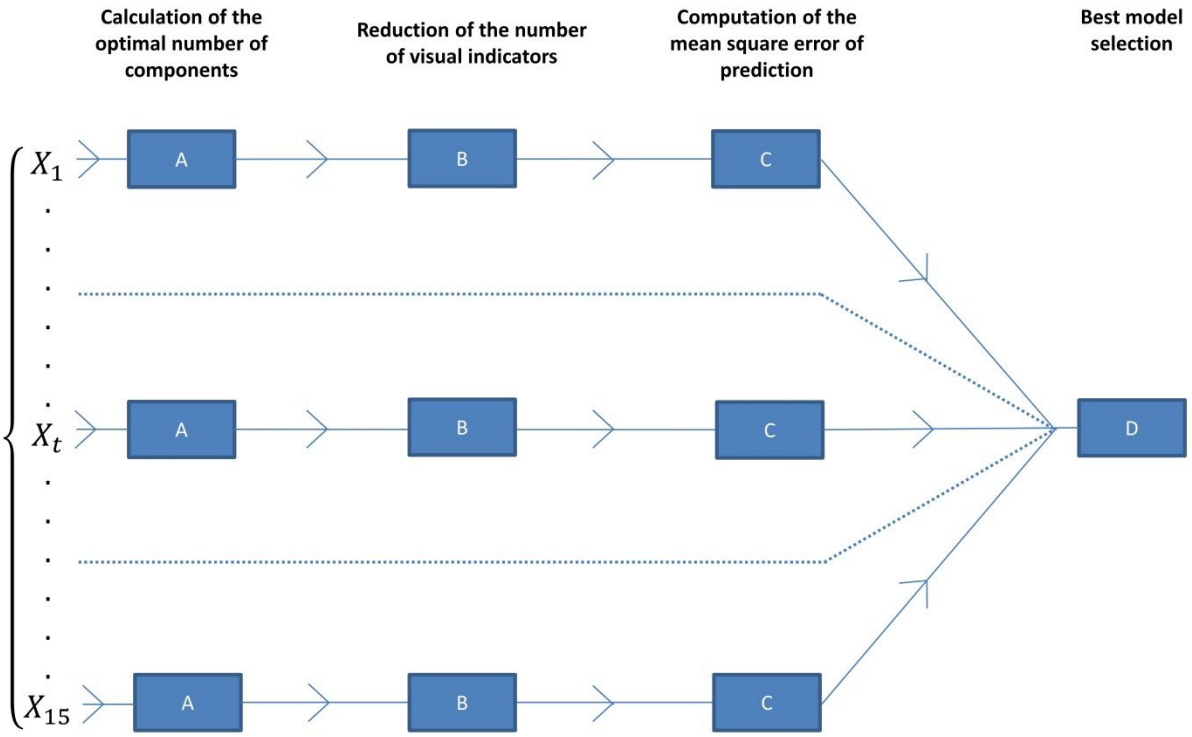


Figure 4: Data analysis was performed in four sequential steps. First, the best parameters of the PLS regression models were identified. This entailed finding the optimal number of components and reducing the number of visual indicators (Steps A and B). Then, in this optimal configuration, the accuracy of the models was considered by computing the mean square error of prediction, for both training and validation datasets (Step C). In the last step (Step D), the model with the lowest validation error was selected and interpreted.

3. Results

3.1. MW scores and drowsiness indicators

Figure 5 shows that the self-reported MW scores varied widely among participants. The median score was 43%, but four participants (participants 9 to 12) estimated that they had spent more than 70% of the time thinking about something other than the driving task. Those four participants failed in the critical case. Two other participants (3 and 4) also failed, with moderate MW scores of around 30%. All other participants managed to avoid a collision.

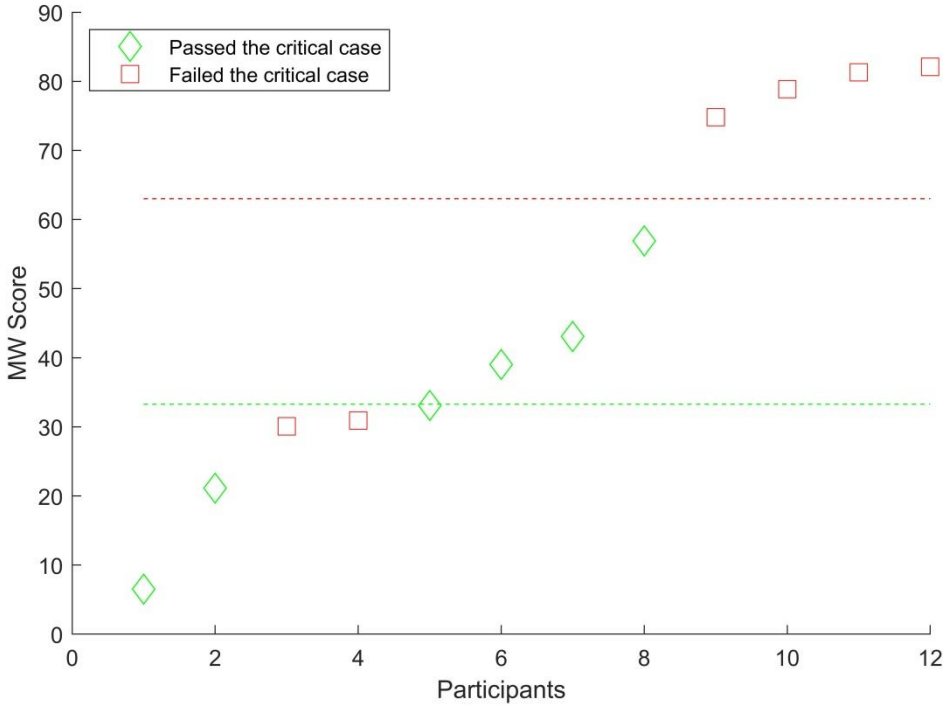


Figure 5: MW scores reported by the participants according to the outcome in the critical case. The colour of the marker indicates whether participants passed (green) or failed (red) the critical case. The dashed lines represent the average of the successful or failed groups. Participants are sorted according to their MW score, not their order of passing the experiment.

Mean PERCLOS and blink-rate per participant showed no significant correlation with MW score ($r = 0.25$ and $r = -0.49$ respectively). The highest PERCLOS score was 4.73%, obtained by participant 2. Furthermore, a paired t-test revealed no significant differences for either PERCLOS or blink-rate between the first and last five minutes of automated driving ($p = 0.28$ and $p = 0.11$ respectively).

3.2. Selection of best model for MW score prediction

Results obtained for all different training datasets at the end of the PLS regression procedure (see appendixes for details) are presented in Table 1.

Table 1: Optimal number of components, number of selected visual indicators and mean square error of prediction (for both training and validation datasets) as a function of the time window. A minimum of validation error was found for gaze data computed over 10 minutes of automated driving (0.419).

Time Window (min)	Optimal number of components (step A)	Number of selected visual indicators (step B)	Mean square error of prediction with training dataset (step C)	Mean square error of prediction with validation dataset (step C)
1	3	56	<0.001	0.478
2	1	26	0.200	0.625
3	1	26	0.122	0.697
4	1	16	0.054	0.915
5	3	52	0.002	0.721
6	1	19	0.233	0.542
7	4	88	<0.001	0.563
8	5	80	< 0.001	0.522
9	1	11	0.224	0.443
10	**1**	**12**	**0.211**	**0.419**
11	1	8	0.130	0.446
12	1	7	0.131	0.496
13	1	57	0.248	0.533
14	1	3	0.141	0.598
15	1	3	0.147	0.595

Table 1 shows that the training error was small (<0.25) in all cases and it depended on the structure of the model (i.e. number of components). The models having the most components (time windows of 1,

5, 7 and 8 minutes) predicted the learning dataset almost perfectly (MSEP <0.001). However, MSEP for the validation dataset was comparatively high. As the time window increased, the validation error decreased, attaining the minimal value (0.419) for 10 min of data. Thereafter, adding more data by expanding the time window increased the MSEP. Thus, ultimately the best model for predicting MW score from the last two minutes of driving was obtained by aggregating the gaze data for the 10 min that preceded those final minutes.

3.3. Final prediction of MW score

After identifying the best model, it remains to be determined what visual indicators were most important for predicting the MW score and how well the predicted scores were correlated to the actual scores.

3.3.1 Visual indicators

The best PLS model retained only 12 visual indicators to predict the MW score. The PLS regression coefficients and the correlation coefficients between MW score and the 12 selected visual indicators are presented in Table 2. Figure 6 illustrates the visual indicators retained by the model.

Table 2: Coefficients of the PLS regression and correlation coefficients between the MW score and the visual indicators (* p<0.05; ** p< 0.01). The 12 first indicators correspond to those shown in Figure 6, which were used for the final prediction of MW score. The last four static indicators were not selected by the PLS regression, but are reported for discussion.

Impact on the prediction	Visual indicators	PLS regression coefficients	Correlation coefficients
Decreases the MW score	Transition from the Central Mirror to the Left Screen	-0.098	-0.65*
	Transition from Others to the Left Screen	-0.096	-0.63*
	Transition from the Others to the Road Centre	-0.093	-0.61*
	Transition from the Road Centre to the Left Mirror	-0.091	-0.63*
	Percentage of time spent in the Left Mirror	-0.089	-0.58*
Increases the MW score	Percentage of time spent in the Others Area	0.087	0.58*
	Percentage of time spent in the Dashboard	0.094	0.62*
	Transition from the Central Mirror to the Others Area	0.099	0.66*
	Transition from the Road Centre to the Others Area	0.101	0.67*
	Transition from the Road Centre to the Down area	0.102	0.66*
	Multiple gazes in the Down area	0.108	0.71**

	Percentage of time spent in the Down area	0.115	0.76**
Non-selected static indicators	Percentage of time spent in the Road Centre (PRC)		-0.16
	Percentage of time spent in the Central Mirror		-0.14
	Percentage of time spent in the Right Mirror		-0.09
	Percentage of time spent in the HMI		0.47

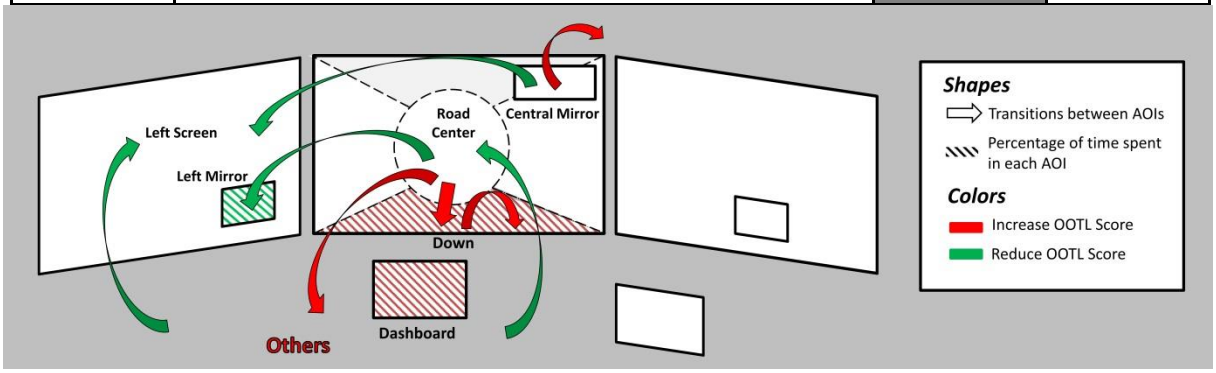


Figure 6: Visual indicators relevant for MW score prediction.

The influence of a visual indicator on MW score is displayed in colour: red shows an increase in MW score, and green shows reduced MW score. Arrows represent transitions between AOIs. Filled areas, or name written in red (in the case of Others) means that percentage of time spent gazing in the AOI was selected by PLS regression.

Of the 12 indicators, eight were dynamic (transitions between AOI) and four were static (percentage of time spent in the area). The signs of the coefficients (see Table 2) indicate that seven of them contributed to an increase of the MW score estimation. These are shown in red in Figure 6, and can be summarized as follows: 1) taking the gaze off the central mirror to look away from the driving scene, 2) taking the gaze off the road centre area, to look down or away from the driving scene, and 3) spending too much time in the down area or on the dashboard.

By contrast, five indicators contributed to a reduction in MW score estimation, shown in green in Figure 6. These consisted of 1) redirecting the gaze to the road from any area outside the driving scene (Others), 2) regularly checking the surroundings by looking at the left view mirror or the left side screen.

3.3.2 Prediction of MW score

Figure 7 presents the results of estimating the MW scores using the best model (time window = 10 min; 12 visual indicators; 1 component) for the training dataset (A) and the validation dataset (B).

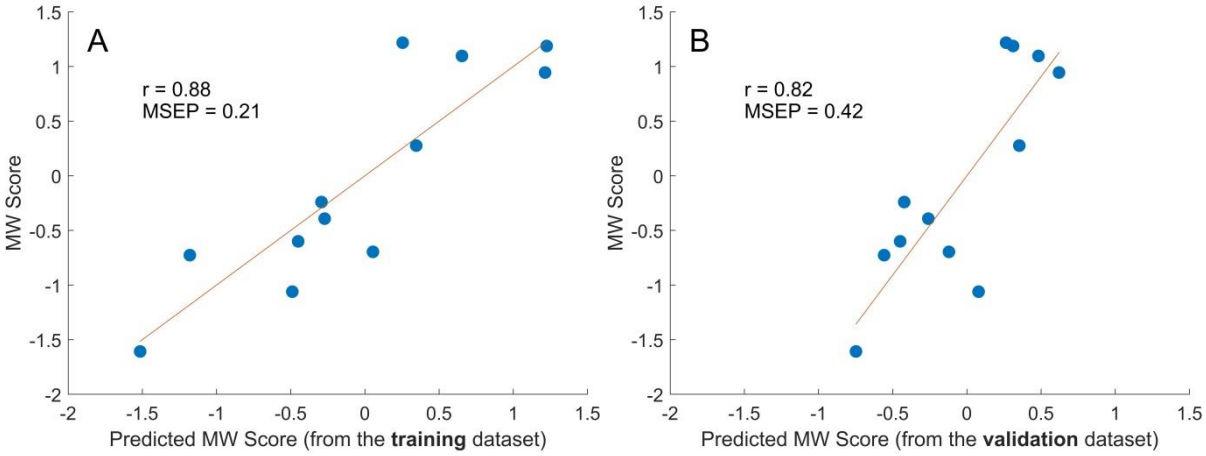


Figure 7: Correlation plots between the (real) MW score and the predictions of the MW score by PLS regressions on the training dataset (A) and the validation dataset (B). All the values presented here are standardized.

As discussed in the previous section, the PLS model with the training data provided an accurate estimation of the MW score. The mean square error of prediction was low (MSEP=0.21) and a significant positive correlation was noted between the predicted and real values ($r=0.88$, $p<0.01$). The prediction was not as accurate with the validation dataset (MSEP=0.42), but a significant positive correlation was still obtained ($r=0.82$, $p<0.05$). The goodness-of-fit of this model was computed using R-squared values for training and validation, which were 0.76 and 0.67, respectively. The assessment of goodness-of-fit can also be performed using the model residuals. These results are provided in the Appendix.

4. Discussion

The OOTL phenomenon results from the combination of a driver not having physical control of their vehicle and incorrectly monitoring the driving situation (Merat et al, 2019). The alternative state, OTL, indicates a driver who satisfactorily monitors the driving environment during automated driving. Monitoring of the driving situation may be impaired when actively engaged in a secondary task. It can also take the form of MW, i.e cognitive disengagement from the driving task due to lack of activity. This study specifically investigated the second form of the OOTL phenomenon. Whatever the condition under which the OOTL phenomenon occurs, the question of how to model and quantify what constitutes proper monitoring of the driving scene remains debatable. The study addressed this question by distinguishing OTL and OOTL drivers in a highway driving context.

The method consisted of using multiple PLS regressions to identify characteristic elements of the gaze behaviour of OTL and OOTL drivers. The multi-step approach began with 182 indicators as an input matrix, accounting for both static elements (cumulative time spent in 13 AOIs) and dynamic elements (transitions between AOIs). These visual indicators were computed for different time windows to assess the evolution of the OOTL phenomenon over time. Once the optimal parameters of each model were calculated, the MW scores were predicted. Finally, a linear combination of the most important indicators enabled estimating the driver's MW score accurately.

4.1. What constitutes good monitoring of a driving situation?

The results revealed that drivers with relatively low MW scores made many transitions from the road centre to the left rearview mirror. In total, they spent more time looking at the mirror. They also looked at the left screen, where they could monitor traffic, immediately after gathering information from the central mirror. After spending time looking at areas unrelated to driving ("Others"), they frequently returned their gaze to the road (road centre area) or to the left screen. By contrast, drivers with relatively high MW scores made many transitions from the road centre to areas irrelevant to driving, where they gazed for a considerable time. They also gazed often at the lower part of the front screen and the dashboard.

These findings can be interpreted in terms of the adequacy of the driver's gaze strategy to maintain sufficient situation awareness (Endsley, 1995) in autonomous mode. According to Merat (2019), situation awareness during automated driving involves three dimensions: perception, comprehension and projection. In the current study, OTL drivers remained dynamically aware of their surroundings by regularly checking the road ahead, the left lane and the left rearview mirror. Doing so helped them to anticipate future hazards and avoid difficulties when reacting in the final critical case. These gaze strategies allowed them to perceive, comprehend and project the future state of the driving situation appropriately. In other words, they had adequate situation awareness.

By contrast, the OOTL drivers' gaze was more strongly attracted by irrelevant information outside the simulator. Even when looking at the driving environment, the driver favoured gazing at the road immediately in front (Down Area) and monitored their vehicle speed on the dashboard. Although OOTL drivers continued to correctly perceive the state of their vehicle and its position in the lane, they showed relatively few signs of visual anticipation by looking far ahead and they paid relatively little attention to other vehicles on the road.

Previous studies (Carsten et al., 2012; Louw et al., 2016) showed that manual drivers displayed a higher percentage of looking at the road centre (PRC) than drivers with automation. In these studies, secondary tasks were allowed or required by the experimental procedure. By contrast, in the current study, PRC did not emerge as a critical indicator (see Table 2). This was the case for most static indicators. This finding indicates that that a good monitoring of the situation is not so much a matter of spending considerable time looking at the road. It looks more important to regularly check relevant objects in the scene. Thus, considering the transitions between areas is more important than the amount of time spent on a particular area. For instance, moving the gaze away from the road centre might indicate a relatively high level of disengagement if the driver repeatedly looks at irrelevant areas. However, it could also contribute to situation awareness if the gaze moves to the left rearview mirror.

The multiple PLS regression method allowed us to identify markers of the OOTL phenomenon from a rich set of data on drivers' gazes. This identification might have been achieved with simpler approaches, for example by examining the mean difference in visual indicators between two groups of participants, namely those having high or low MW scores. It might also have been possible to examine only the correlations between MW score and individual indicators. In this case, however, an a-priori selection should have been made, given the number of possibilities (182 indicators, with 13 AOI). Furthermore, individual correlations would not account for relationships between indicators. The advantage of PLS regression analysis is that all visual indicators are considered together to optimize the prediction of MW scores.

4.2. Is it possible to estimate drivers' OOTL state from the observation of spontaneous gaze strategies?

The results of the modelling work show that the best estimation of the MW score was obtained by considering 10 minutes of gaze data. The least prediction error occurred when the training dataset was used; however, performance was also good for the validation dataset (i.e. only two minutes of driving data, which were not used for the model determination).

It can be concluded that the influence of the driver's state on their gaze strategy was qualitatively similar during the last two minutes of automated driving as it had been during the previous 10 minutes. This point suggests that detecting the OOTL state could perhaps be performed using a shorter time window than 10 minutes. This is interesting from the perspective of defining an algorithm for real automated vehicles to monitor the driver's state. However, further tests are necessary, notably with other participants and in other driving contexts.

The results also suggest that the OOTL phenomenon took some time to appear and that it increased with the duration of automated driving (Körber et al., 2015; Feldhütter et al., 2017; Bourrelly et al., 2019). The prediction error decreased when more gaze data were considered, but the minimal value occurred for 10 minutes. Beyond this point, the error gradually increased. This finding can probably be explained by the fact that considering more than 10 minutes of gaze data meant aggregating data from the beginning of the scenario. During the initial minutes, drivers had not had enough time to become OOTL. In order to evaluate the actual evolution of MW over time, an online measurement would have been preferable. Probe techniques that interrupt the participant could hardly have been considered under our driving conditions, as they could have prevented the OOTL phenomenon from developing. Physiological indicators such as pupil diameter, skin conductance or cardiac measurements could be considered in the future, although their robustness remains to be demonstrated.

The influence of the OOTL phenomenon has often been assessed in terms of its consequences in a take-over situation. Drivers who are OOTL usually react relatively late and inefficiently when a critical event occurs, especially after a long time spent in automation (Neubauer et al., 2012; Gold et al., 2013; Saxby et al., 2013; Louw et al., 2015b; Zeeb et al., 2015, 2017; Eriksson and Stanton, 2017, Bourrelly et al., 2019). More direct assessment of the OOTL phenomenon may also be performed using post-trial questionnaires (see for instance Lu et al., 2017, for an interesting approach in evaluating situation awareness). In the present study, the drivers' state was assessed by asking the participants to report their level of MW. The results showed that drivers with the highest level of MW all failed to take over adequately, which implies that they were indeed OOTL. Nevertheless, failure to correctly manage the critical situation also occurred for two participants, who reported only a moderate level of MW. This finding suggests that the quality of a take-over is not entirely determined by the degree of OOTL before the take-over request. The skills of the driver to quickly recover situation awareness and to handle the vehicle in the time allocated may also be essential aspects.

A last point worth mentioning is that the MW scores were not correlated to PERCLOS and blink-rate indicators, which have been shown to be highly predictive of drowsiness (Jacobé de Naurois et al., 2019) . The two indicators remained lower than values typically associated with drowsy states during driving. (PERCLOS of >12.5% was used to categorize drowsy drivers in Hanowski et al. 2008). Thus, we concluded that the MW score did not reflect driver fatigue in our experiments. This result might be different for an extended period of automated driving, during which participants might fall asleep (Vogelpohl et al., 2019). This was observed by Bourrelly et al. (2018) in a 1-h test run of automated driving. It remains to be determined whether the multiple PLS approach we propose can discriminate between the two phenomena. The value of the PLS approach to realize an integrated diagnostic must also be determined.

5. Conclusion

We investigated whether drivers' gaze behaviour could be used to detect the OOTL phenomenon during automated driving. The results indicate that the gaze dynamics appear to be a crucial point: being OTL required frequent gaze shifts to the road while also obtaining tactical information about the oncoming situation.

It remains to be determined whether this conclusion can be generalized to other conditions under which the OOTL phenomenon may occur. It was induced in this study by a relatively monotonous driving task in the absence of external distraction. However, being out of the loop may also be due to the engagement of attention in a secondary task, an expected consequence of Level 3 automation. In this case, changes in driver behaviour will most likely be characterised by a massive redirection of attention to external displays. It would be interesting to assess whether dynamic indicators will be essential to improve the estimation of the driver's state or whether static indicators, such as the time spent without looking at the road, are sufficient. The method used in this study could help to answer this question.

To more accurately detect the OOTL phenomenon during automated driving, the analysis of gaze behaviour could perhaps be coupled with other approaches. For example, physiological measurements or the analysis of the driver's posture could be incorporated in the diagnosis.

6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. Acknowledgments

The authors would like to thank Denis Creusot for his invaluable help on the driving simulator, as well as the participants involved in the experiment. This study was supported by the French National Research Agency (Agence Nationale de la Recherche, AUTOCONDUCT project, grant n ◦ANR-16-CE22-0007-05).

8. References

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression

(PLS Regression): PLS REGRESSION. *WIREs Comp Stat* 2, 97–106. doi:10.1002/wics.51.
Bourrelly, A., Jacobé de Naurois, C., Zran, A., and others (2018). Gaze behavior during take-over after a long period of autonomous driving: A pilot study'. in *Proc. Int. Conf. Driving Simulation Conference Europe VR, Antibes, France*.
Bourrelly, de Naurois, C. J., Zran, A., Rampillon, F., Vercher, J.-L., and Bourdin, C. (2019). Long automated driving phase affects take-over performance. *IET Intelligent Transport Systems*. doi:10.1049/iet-its.2019.0018.
Burdett, B. R. D., Charlton, S. G., and Starkey, N. J. (2019). Mind wandering during everyday driving: An on-road study. *Accident Analysis & Prevention* 122, 76–84. doi:10.1016/j.aap.2018.10.001.
Carsten, O., Lai, F. C., Barnard, Y., Jamson, A. H., and Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? *Human Factors* 54, 747–761. doi:10.1177/0018720812460246.
Chan, C.-Y. (2017). Advancements, prospects, and impacts of automated driving systems. *International Journal of Transportation Science and Technology* 6, 208–216. doi:10.1016/j.ijtst.2017.07.008.
Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors* 37, 32–64.
Endsley, M. R., and Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors* 37, 381–394. doi:10.1518/001872095779064555.
Eriksson, A., and Stanton, N. A. (2017). Driving performance after self-regulated control transitions in highly automated vehicles. *Human Factors* 59, 1233–1248. doi:10.1177/0018720817728774.
Feldhütter, A., Gold, C., Schneider, S., and Bengler, K. (2017). How the duration of automated driving influences take-over performance and gaze behavior. *Advances in Ergonomic Design of Systems, Products and Processes*, 309–318. doi:10.1007/978-3-662-53305-5_22.
Fitch, G. M., Bowman, D. S., and Llaneras, R. E. (2014). Distracted driver performance to multiple alerts in a multiple-conflict scenario. *Human Factors* 56, 1497–1505. doi:10.1177/0018720814531785.
Gold, C., Damböck, D., Bengler, K., and Lorenz, L. (2013). Partially automated driving as a fallback level of high automation. in *6. Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren.* (TÜV SÜD Akademie GmbH).
Gonçalves, R., Louw, T., Madigan, R., and Merat, N. (2019). Using markov chains to understand the sequence of drivers' gaze transitions during lane-changes in automated driving. in *Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (Leeds).
Gouraud, J., Delorme, A., and Berberian, B. (2017). Autopilot, mind wandering, and the out of the loop performance problem. *Frontiers in Neuroscience* 11, 541. doi:10.3389/fnins.2017.00541.
Hanowski, R. J., Bowman, D., Alden, A., Wierwille, W. W., and Carroll, R. (2008). PERCLOS+: Development of a robust field measure of driver drowsiness. in *15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting, New*

York, NY.
Jacobé de Naurois, C., Bourdin, C., Stratulat, A., Diaz, E., and Vercher, J.-L. (2019). Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis & Prevention* 126, 95–104. doi:10.1016/j.aap.2017.11.038.
Körber, M., Cingel, A., Zimmermann, M., and Bengler, K. (2015). Vigilance decrement and passive fatigue caused by monotony in automated driving. in *6th International Conference on Applied Human Factors and Ergonomics*, 2403–2409.
Louw, Kountouriotis, G., Carsten, O., and Merat, N. (2015a). Driver Inattention During Vehicle Automation: How Does Driver Engagement Affect Resumption Of Control? in *4th International Conference on Driver Distraction and Inattention*, 1–16.
Louw, Madigan, R., Carsten, O., and Merat, N. (2016). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury Prevention* 23, 281–286. doi:10.1136/injuryprev-2016-042155.
Louw, and Merat, N. (2016). A Methodology for Inducing the Out of the Loop Phenomenon in Highly Automated Driving. in *International Conference on Traffic and Transport Psychology, Brisbane, Australia*.
Louw, Merat, N., and Jamson, H. (2015b). Engaging with Highly Automated Driving: To be or Not to be in the Loop? in *Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 190–196. doi:https://doi.org/ 10.17077/drivingassessment.1570.
Louw, T., and Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies* 76, 35–50.
Lu, Z., Coster, X., and de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics* 60, 293–304. doi:10.1016/j.apergo.2016.12.003.
Mackenzie, A. K., and Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual cognition* 23, 736–757. doi:10.1080/13506285.2015.1079583.
Mars, F., Deroo, M., and Charron, C. (2014). Driver adaptation to haptic shared control of the steering wheel. in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE), 1505–1509. doi:10.1109/SMC.2014.6974129.
Mars, F., and Navarro, J. (2012). Where we look when we drive with or without active steering wheel control. *PLoS One* 7, e43858. doi:10.1371/journal.pone.0043858.
Merat, N., Jamson, A. H., Lai, F. C. H., and Carsten, O. (2012). Highly Automated Driving, Secondary Task Performance, and Driver State. *Human Factors* 54, 762–771. doi:10.1177/0018720812442087.
Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., et al. (2019). The "out-of-the-loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work* 21, 87–98. doi:10.1007/s10111-018-0525-8.
Mole, C. D., Lappi, O., Giles, O., Markkula, G., Mars, F., and Wilkie, R. M. (2019). Getting Back Into the Loop: The Perceptual-Motor Determinants of Successful Transitions out of Automated Driving. *Human Factors* 61, 1037–1065. doi:10.1177/0018720819829594.
Molloy, R., and Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors* 38, 311–322.

682 doi:10.1177/001872089606380211.
683 Navarro, J., François, M., and Mars, F. (2016). Obstacle avoidance under automated steering:
684 Impact on driving and gaze behaviours. *Transportation Research Part F: Traffic Psychology*
685 *and Behaviour* 43, 315–324. doi:10.1016/j.trf.2016.09.007.
686 Neubauer, C., Matthews, G., Langheim, L., and Saxby, D. (2012). Fatigue and voluntary
687 utilization of automation in simulated driving. *Human Factors* 54, 734–746.
688 doi:10.1177/0018720811423261.
689 Parasuraman, R., and Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse,
690 Abuse. *Human Factors* 39, 230–253. doi:10.1518/001872097778543886.
691 SAE International (2016). Taxonomy and Definitions for Terms Related to On-Road Motor
692 Vehicle Automated Driving Systems. Washington, DC: SAE International.
693 Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., and Neubauer, C. (2013). Active
694 and Passive Fatigue in Simulated Driving: Discriminating Styles of Workload Regulation and
695 Their Safety Impacts. *J Exp Psychol Appl* 19, 287–300. doi:10.1037/a0034386.
696 Schnebelen, D., Lappi, O., Mole, C., Pekkanen, J., and Mars, F. (2019). Looking at the Road
697 When Driving Around Bends: Influence of Vehicle Automation and Speed. *Front. Psychol.*
698 10. doi:10.3389/fpsyg.2019.01699.
699 Sivak (1996). The information that drivers use: is it indeed 90% visual? *Perception* 25, 1081–
700 1089. doi:10.1068/p251081.
701 Stanton, N. A., and Salmon, P. M. (2009). Human error taxonomies applied to driving: A
702 generic driver error taxonomy and its implications for intelligent transport systems. *Safety*
703 *Science* 47, 227–237. doi:10.1016/j.ssci.2008.03.006.
704 Stern, J. A., Boyer, D., and Schroeder, D. (1994). Blink rate: a possible measure of fatigue.
705 *Human Factors* 36, 285–297. doi:10.1177/001872089403600209.
706 Team, R. C., and others (2013). R: A language and environment for statistical computing.
707 Victor (2005). Keeping eye and mind on the road. *Digital Comprehensive Summaries of*
708 *Uppsala Dissertations from the Faculty of Social Sciences*, 83.
709 Victor, Harbluk, J. L., and Engström, J. A. (2005). Sensitivity of eye-movement measures to
710 in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*
711 8, 167–190. doi:10.1016/j.trf.2005.04.014.
712 Vogelpohl, T., Kühn, M., Hummel, T., and Vollrath, M. (2019). Asleep at the automated
713 wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis &*
714 *Prevention* 126, 70–84. doi:10.1016/j.aap.2018.03.013.
715 Wehrens, R., and Mevik, B.-H. (2007). The pls Package: Principal Component and Partial
716 Least Squares Regression in R. *Journal of Statistical Software* 18, 1–24.
717 Wierwille, W. W., Wreggit, S., Kirn, C., Ellsworth, L., and Fairbanks, R. (1994). Research on
718 vehicle-based driver status/performance monitoring; development, validation, and refinement
719 of algorithms for detection of driver drowsiness. *National Highway Traffic Safety*
720 *Administration Final Report*.
721 Zeeb, K., Buchner, A., and Schrauf, M. (2015). What determines the take-over time? An
722 integrated model approach of driver take-over after automated driving. *Accident Analysis &*
723 *Prevention* 78, 212–221. doi:10.1016/j.aap.2015.02.023.
724 Zeeb, K., Härtel, M., Buchner, A., and Schrauf, M. (2017). Why is steering not the same as
725 braking? The impact of non-driving related tasks on lateral and longitudinal driver

interventions during conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 50, 65–79. doi:10.1016/j.trf.2017.07.008.

9. Appendix 1: Step-by-step procedure of PLS regression

This appendix develops step-by-step the procedure to predict the MW score (Y) using PLS regressions. We used a training dataset (matrix of gaze behaviour computed over a given time window t, X_t) and a validation dataset (matrix of gaze behaviour computed over the two final minutes of automated driving, X_{val}).

Steps A and B were performed with the training datasets and enabled computing the optimal parameters (number of components and relevant visual indicators) of the prediction models. With that configuration, the accuracy of the model was tested for both the training and validation datasets (Step C).

Step A: Calculating the optimal number of components

Principle

PLS models are based on several orthogonal components, which constitute the underlying structure of the prediction model. With many components, the model will be complex and highly accurate but also very specific of the data. By contrast, few components will mean a simpler model structure. The model may lose accuracy but may be more generalizable to other datasets. Thus, an optimal compromise in the number of components can prevent data overfitting while maintaining high accuracy.

Application

This compromise was sought by testing several numbers of components (from one to 10 components). The optimal number of components would minimize the mean square error of prediction (MSEP), with a leave-one-out procedure.

Step B: Reducing the number of visual indicators

In the previous step, an optimal structure of the prediction model was found, considering all possible visual indicators (182) to predict the MW score. The aim of this next step was to increase the predictive power (i.e. the percentage of variance in Y that was explained) by selecting fewer indicators.

Principle

Because PLS regression is a linear statistical model, the relationship between the training dataset (X_t) and the dependent variable to estimate $\widehat{Y}_{t,\ train}$ was linear:

$$\widehat{Y_{t,\ train}} = X_t * C_t$$

where C_t is the matrix of the regression coefficients.

Coefficients can be interpreted as follows:

- Coefficient signs indicate the direction in which a visual indicator (from X_t) influenced the estimation of the MW score ($\widehat{Y}_{t,\ train}$). If positive, the MW score increased. By contrast, a negative coefficient meant that the MW score decreased.
- A coefficient's magnitude (absolute value) indicates the importance of each indicator relative. If the magnitude of a coefficient was close to zero, the contribution of this visual indicator to the prediction would be negligible. By contrast, a large magnitude indicated a crucial indicator in the prediction.

Application

To reduce the number of visual indicators of X_t, coefficient magnitudes were compared with an increasing threshold value, which ranged from 0.01 to 0.2. A new PLS regression was computed for each partial matrix (i.e. a matrix comprising only the indicators whose coefficient magnitude exceeded the threshold value). The threshold was increased by steps of 0.005 until the percentage of variance in Y that was explained by the partial model no longer increased.

At the end of this step, a partial matrix of gaze behaviour, which included only the selected visual indicators, was computed for the training and validation datasets. These partial matrixes are denoted X_t^p and X_{val}^p for the training and validation datasets, respectively.

Step C: Computing the mean square error of prediction

The previous steps found the most appropriate parameters for PLS regression models (number of components in Step A and relevant visual indicators in Step B) for predicting the MW score. A new model that considered those parameters was set up, with its coefficients denoted C_t^p. Then, the estimations from the training and validation datasets for a given time window t were calculated as follows:

$$\begin{cases} \widehat{Y_{t,\ train}^{p}} = X_t^p * C_t^p \\ \widehat{Y_{t,\ val}^{p}} = X_{val}^p * C_t^p \end{cases}$$

From those estimations, the mean square error of prediction was:

$$\begin{cases} MSEP_t^{train} = \overline{\left(Y - \widehat{Y_{t,\ train}^{p}}\right)^2} \\ MSEP_t^{val} = \overline{\left(Y - \widehat{Y_{t,\ val}^{p}}\right)^2} \end{cases}$$

where C_t^p are the coefficients of the PLS regression, computed from the partial matrixes with the optimal number of components for time window t; and the terms X_t^p and X_{val}^p refer to the partial matrixes of gaze behaviour for the training and validation datasets, respectively.

10. Appendix 2: goodness-of-fit based on the model residuals

Due to the standardization of the data, MW score may be interpreted as Z-scores. Consequently, the residuals are significantly different from 0 with a 95% confidence level if their value is outside the range [-1.96, 1.96]. Figure 8 shows that residuals for each different participant are within this range for both the validation and training datasets. This demonstrates that the model is well-fitted for all participants in all cases.

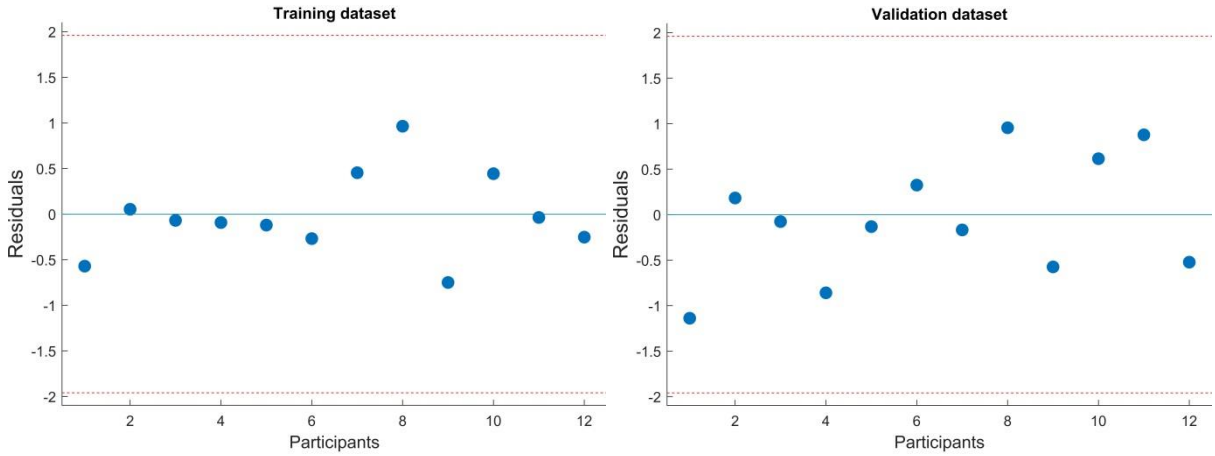


Figure 8: Residuals from the PLS regression models for the training (left) or validation (right) datasets. All residuals are within the confidence interval [-1.96, 1.96] (red dotted lines).

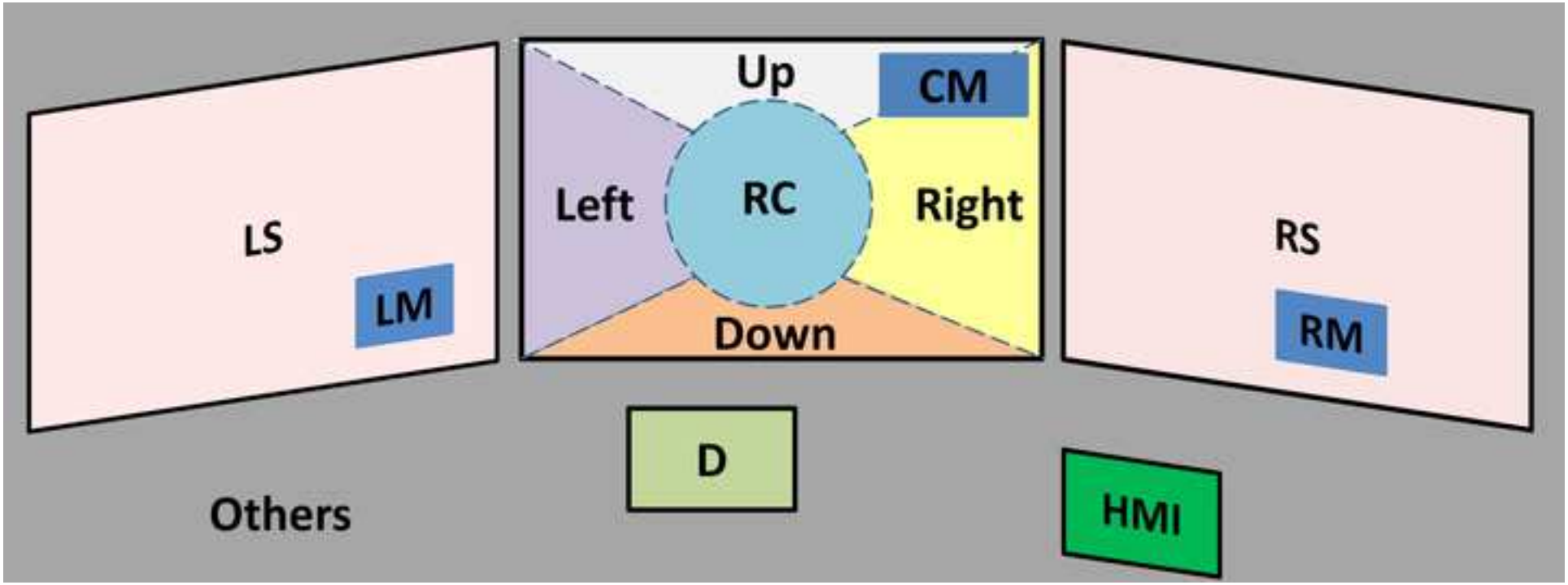
Figure

Click here to access/download;Figure;Figure 1.jpg

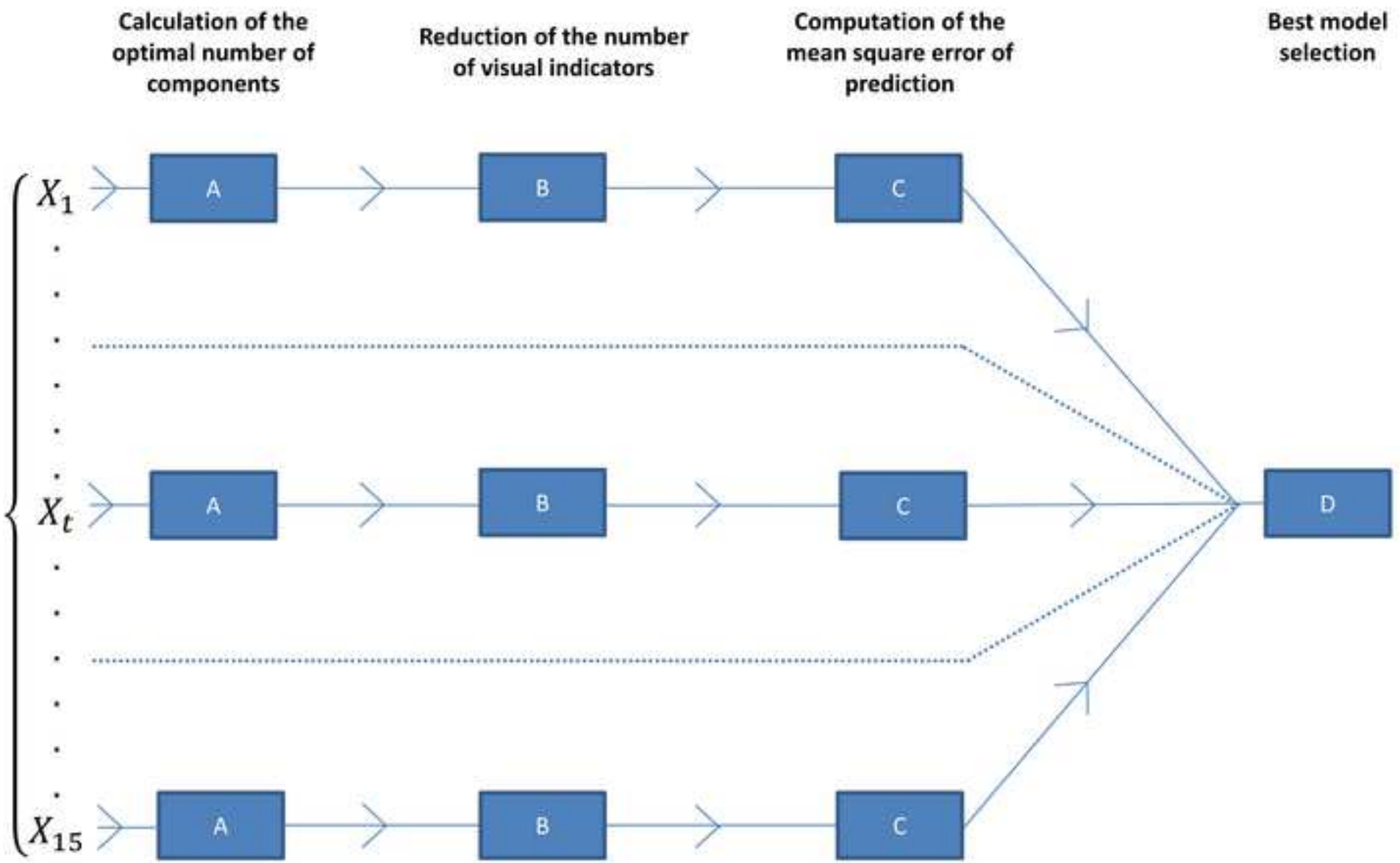
Figure

Click here to access/download;Figure;Figure 2.jpg

Figure

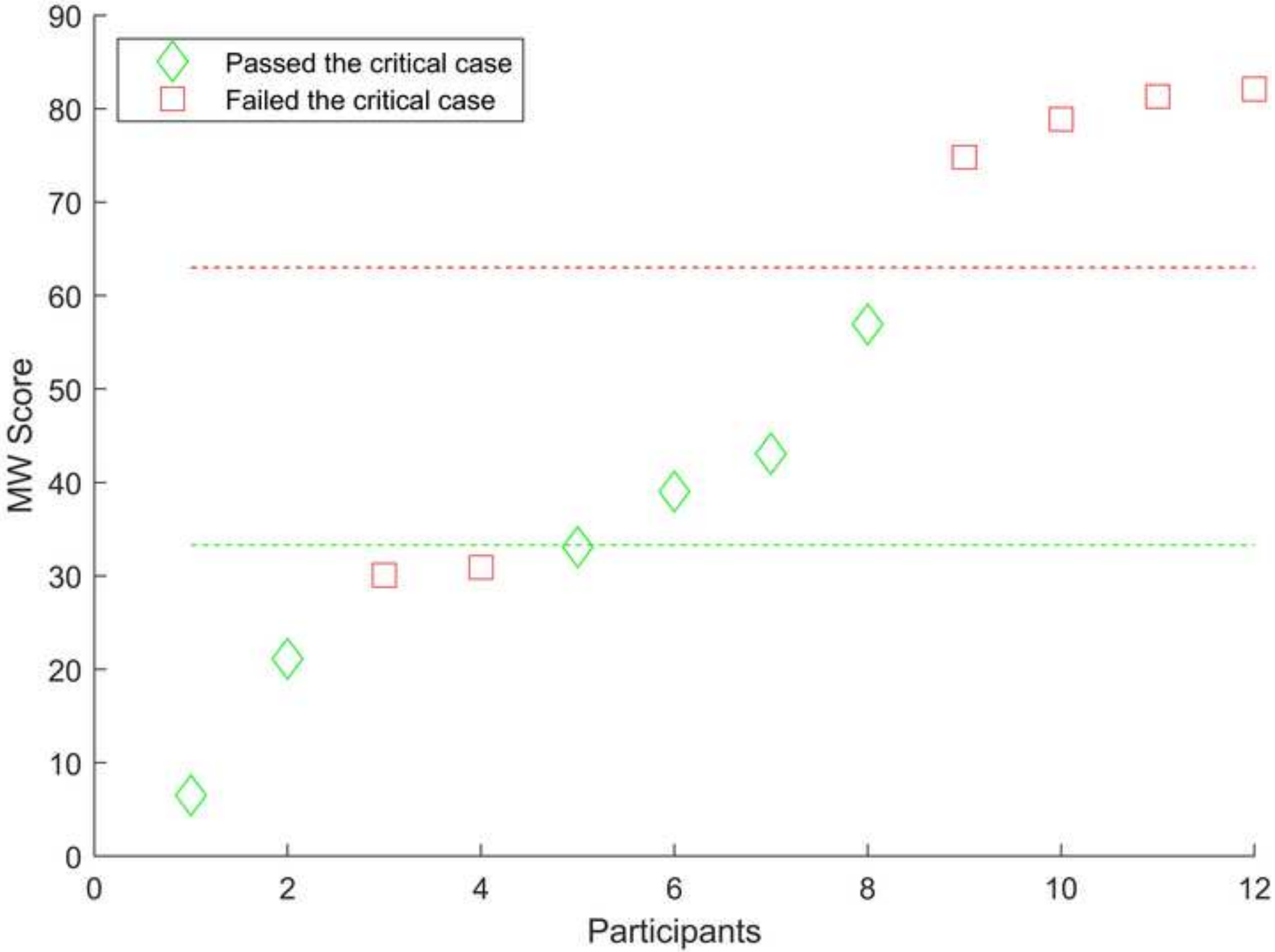


Figure

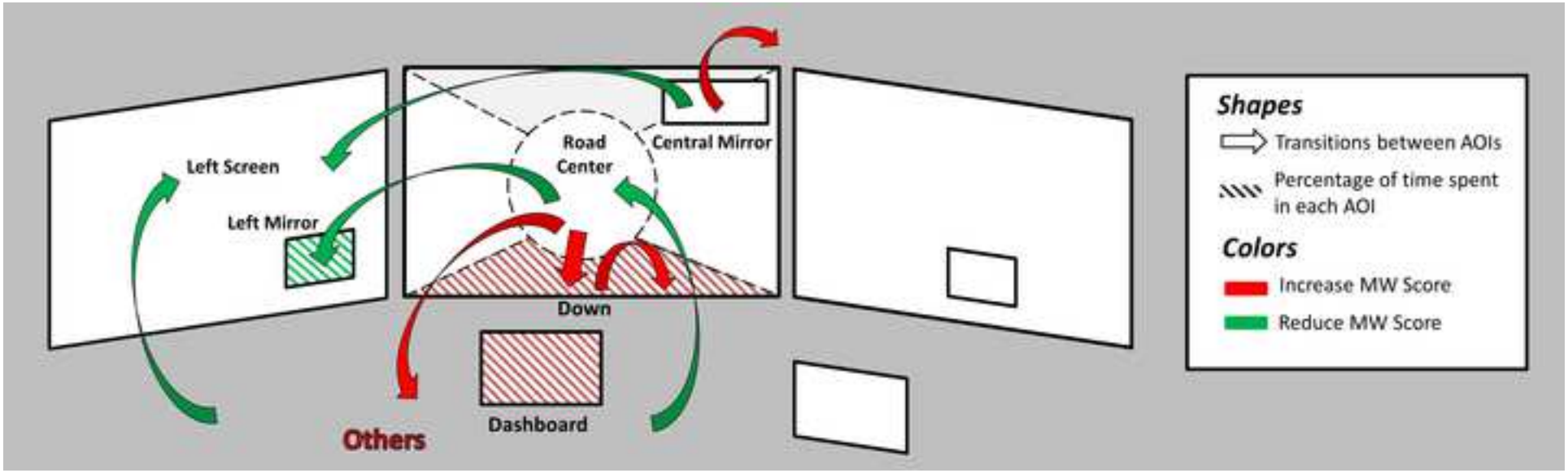


Figure

Click here to access/download;Figure;Figure 5 R1.jpg

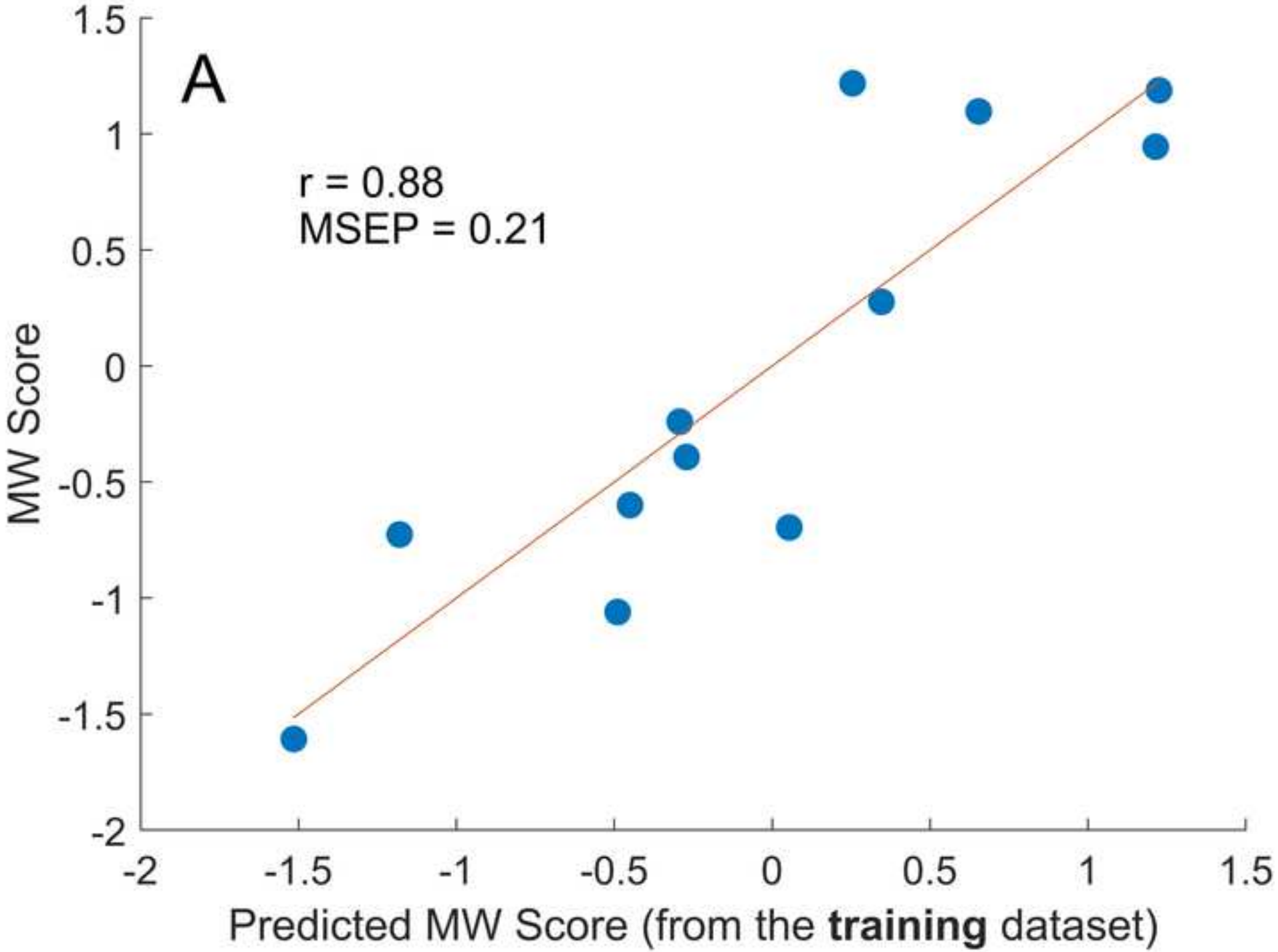


Figure



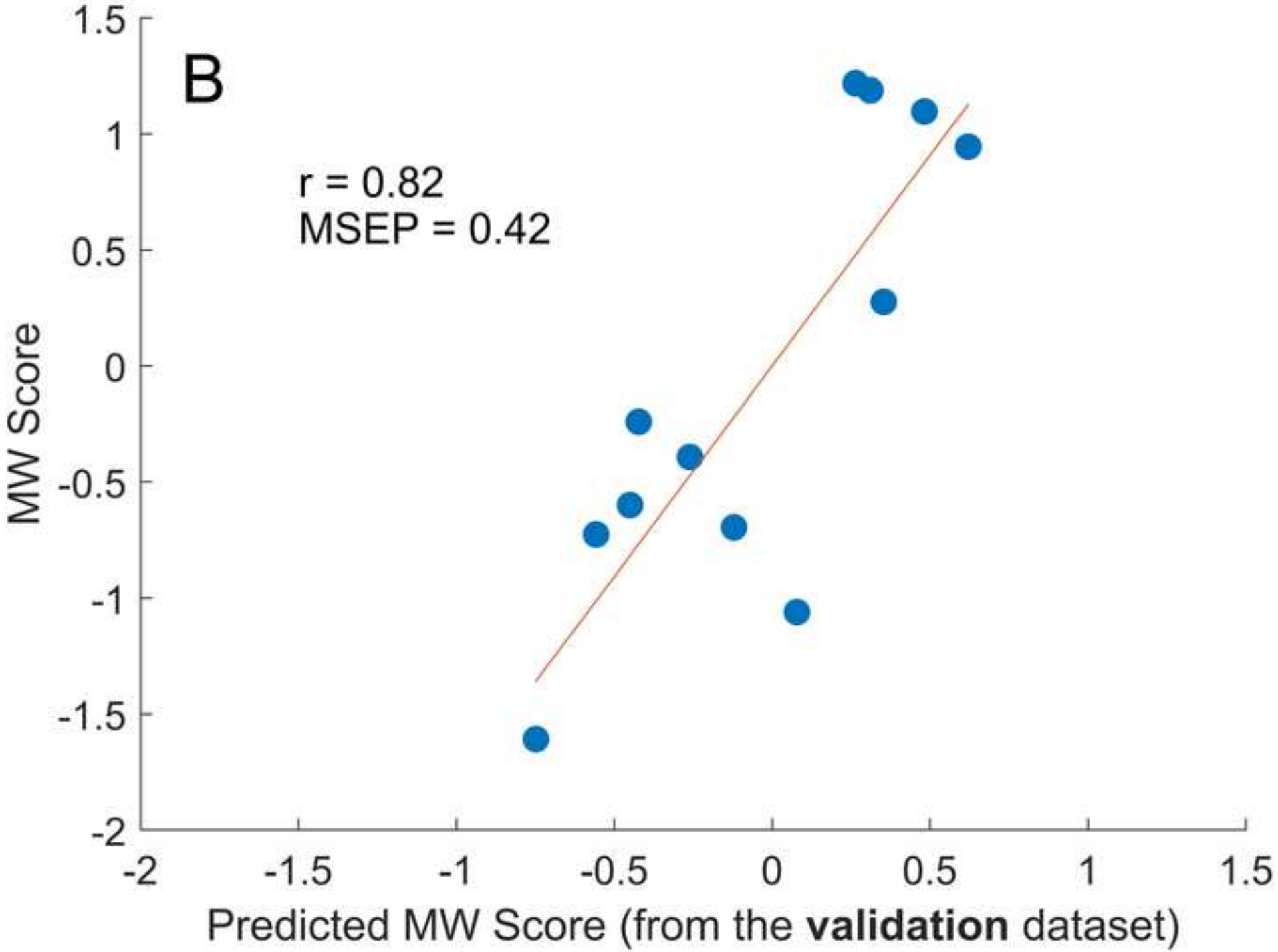
Figure

Click here to access/download;Figure;Figure 7a R1.jpg



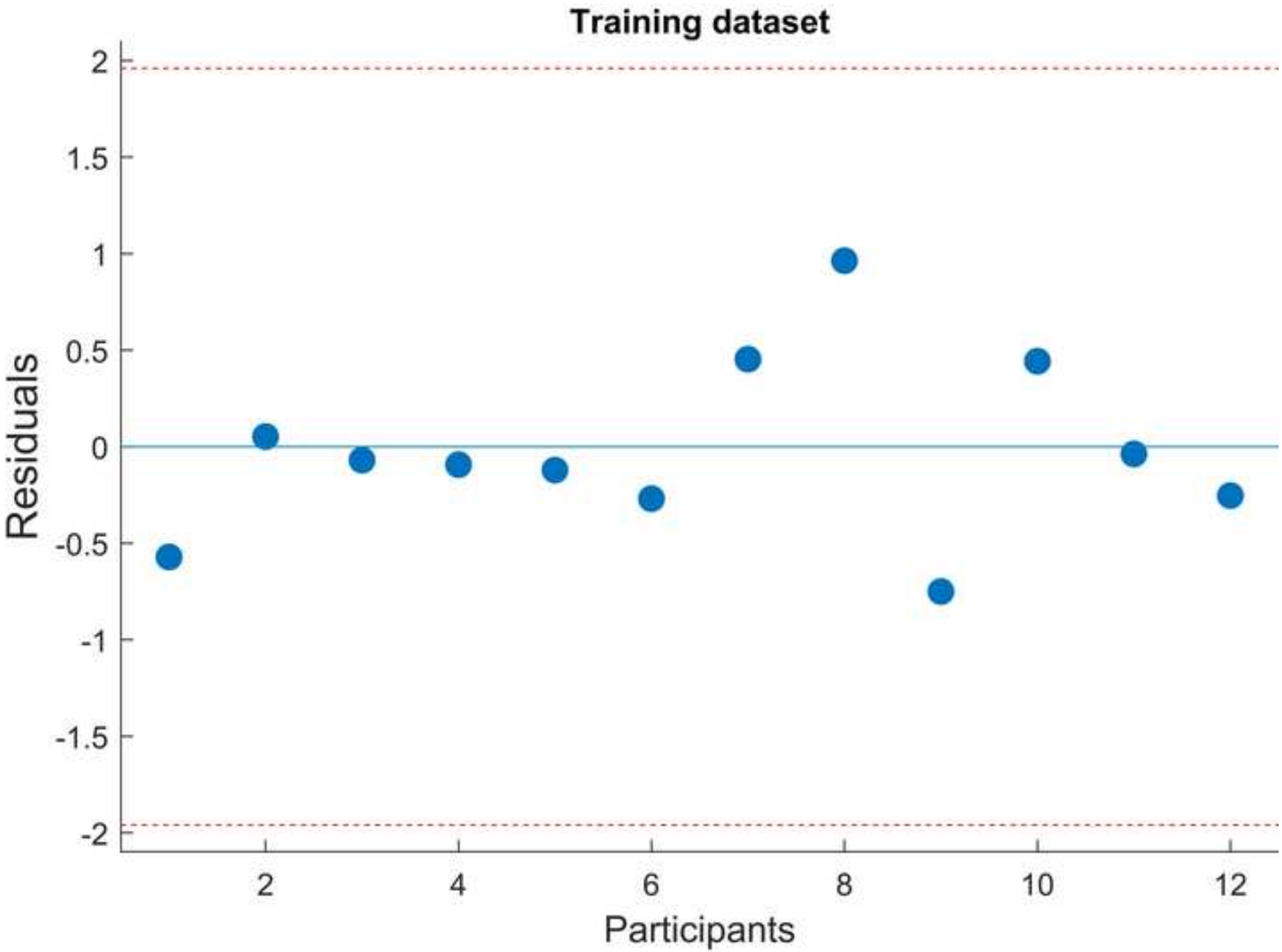
Figure

Click here to access/download;Figure;Figure 7b R1.jpg



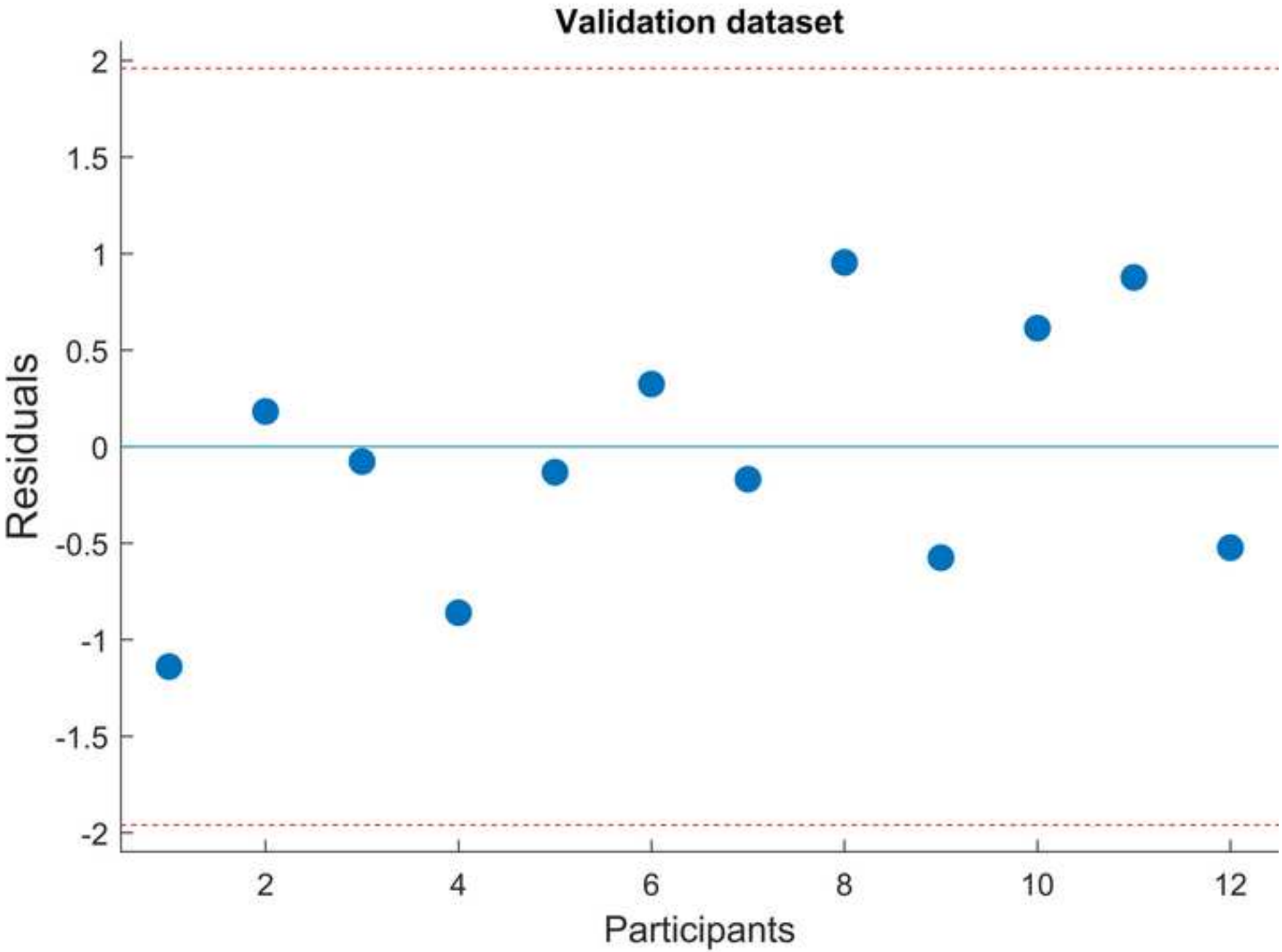
Figure

Click here to access/download;Figure;Figure 8a.jpeg



Figure

Click here to access/download;Figure;Figure 8b.jpeg



The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Damien Schnebelen: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - Original Draft, Visualization **Camilo Charron:** Conceptualization, Methodology, Formal analysis, Writing – review and editing **Franck Mars:** Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Supervision, Project administration, Funding acquisition