Assessment selection in human-automation interaction studies: The Failure-GAM²E and review of assessment methods for highly automated driving

Camilla Grane

Luleå University of Technology, Division of Human Work Science, 97187 Luleå, Sweden

ARTICLE INFO

Article history:

Received 15 August 2016

Received in revised form

10 August 2017

Accepted 14 August 2017

Available online 31 August 2017

Keywords:

Human-automation interaction

Highly automated driving

Assessment methods

ABSTRACT

Highly automated driving will change driver's behavioural patterns. Traditional methods used for assessing manual driving will only be applicable for the parts of human-automation interaction where the driver intervenes such as in hand-over and take-over situations. Therefore, driver behaviour assessment will need to adapt to the new driving scenarios. This paper aims at simplifying the process of selecting appropriate assessment methods. Thirty-five papers were reviewed to examine potential and relevant methods. The review showed that many studies still relies on traditional driving assessment methods. A new method, the Failure-GAM²E model, with purpose to aid assessment selection when planning a study, is proposed and exemplified in the paper. Failure-GAM²E includes a systematic step-by-step procedure defining the situation, failures (Failure), goals (G), actions (A), subjective methods (M), objective methods (M) and equipment (E). The use of Failure-GAM²E in a study example resulted in a well-reasoned assessment plan, a new way of measuring trust through feet movements and a proposed Optimal Risk Management Model. Failure-GAM²E and the Optimal Risk Management Model are believed to support the planning process for research studies in the field of human-automation interaction.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Technology is constantly evolving, and there have been several occasions throughout history when advances have changed human behaviour dramatically. Over recent years we have seen the start of such a change through the development of highly automated vehicles. The role of the driver is certain to change once the task of driving can be handed over to the vehicle itself. This previously futuristic idea has become a real possibility (Akamatsu et al., 2013; Richards and Stedmon, 2016). Two motives for the development of more advanced automation in vehicles have been improving the driver's well-being and enhancing road safety (Stanton and Marsden, 1996). Automation was believed to significantly reduce human-related errors which are known to be the root cause of many accidents. Hence, one purpose of highly automated driving was, in fact, to change the role of the driver and the driver's behavioural patterns. Although automation is believed to reduce accidents, this effect needs to be verified and possible side-effects need to be identified. As Bainbridge (1983) pointed out early on,

the introduction of automation might introduce additional problems that are difficult to imagine beforehand. The main question is probably not *if* there will be new types of errors but rather *what* types of errors there will be. One challenge lies in making the right error predictions. Another challenge lies in selecting relevant assessment methods that cover the predicted behavioural patterns. Technological development makes it easier and more possible to assess behaviours and reactions that previously were too complicated or too expensive to measure. However, these possibilities do not only aid the planning of studies but also makes it more complex. This paper addresses the process of selecting relevant assessment methods in general and for automated driving in particular. Much can be gained by using a well-designed study with carefully selected and well-motivated assessment methods, especially when exploring new research fields. It is believed that this paper will benefit researchers and vehicle developers exploring new research fields such as highly automated driving.

In this paper, automation at a level above driver assistance is considered. The vehicle is able to drive by itself but the driver is obliged to maintain situation awareness and should be prepared and, if necessary, be able to take over driving at all times. The automation level would be above 7 (executes automatically, then

E-mail address: camilla.grane@ltu.se.

necessarily informs the human) according to the Level of Automation (LoA) proposed by Sheridan et al. (1978), and between 2 and 3 according to the classification proposed by the National Highway Traffic Safety Administration (NHTSA; Richards and Stedmon, 2016). According to the taxonomy proposed by Endsley (1999) the term would be Supervisory Control (SC), one step below Full Automation (FA); the difference between the two is the human's opportunity to intervene. In this paper, the term highly automated driving will be used. The term autonomous will not be used since the driver should be able to take over control (Stensson and Jansson, 2013).

The enhanced safety inherent in fully automated vehicles may, to some extent, depend on how well the driver adapts to the new driver role (Merat and Lee, 2012; Milakis et al., 2017). The introduction of driving assistance functions in vehicles, such as adaptive cruise control, changed the role of the driver slightly in the direction of a more passive and relaxed behaviour, with reduced mental workload as result (Stanton and Young, 1998). At higher levels of automation, the driver-vehicle interaction and control of the vehicle will differ dramatically from traditional driving, while the responsibility of the driver to maintain attention on the road will remain more or less the same (Richards and Stedmon, 2016). Even though automation is introduced in order to replace human manual control, planning, and problem solving, humans will still be needed for supervision and to make adjustments (Brookhuis et al., 2001). The driver will need to detect, understand and correct errors should automation fail (McBridge et al., 2014). Human error includes all planned actions, both mental and physical, that fail to achieve the intended consequences (Reason, 1990; Reason et al., 1990). The transition from manual tasks towards more automation and supervision challenge the concept of human error (Rasmussen, 1990). Rasmussen (1990) found that the chain of actions was better defined, and the cause of errors was easier to identify in manual work tasks than in more complex work tasks involving supervision of an automation process. As Banks et al. (2014) describes the situation, driving will become more of a mind-task than a manual task, and the mental workload might even increase, rather than decrease, due to a more complex monitoring responsibility. A temporarily high workload may also result as an effect of a sudden need to take over driving (de Winter et al., 2016). It is also feared drivers will have problems in maintaining their attention on the road and instead will engage in secondary tasks (Banks and Stanton, 2016). It is anticipated that lack in engagement or situation awareness will affect the ability to assume control if/when needed. Also, at lower levels of automation, when driving with adaptive cruise control, problems in resuming control of the vehicle have been found (Larsson et al., 2014; Stanton and Young, 1998). Also, as could be expected, the ability to regain control in the event of automation failure was found to decrease with increased level of automation (Strand et al., 2014). A lack of situation awareness, or out-of-the-loop performance, was described by Endsley (2015) as one of the most significant human error challenges in the automation domain. Another related issue is trust, which could match automation capabilities but which could also turn into distrust or over-trust (Lee and See, 2004).

At the time of writing this paper, a high level of automation in cars was an uncommon and fairly new concept on actual roads. The number of accidents were naturally also few. Tesla Motors was probably the first company to provide production vehicles with a self-driving mode. According to an ODI Resume (NHTSA, 2017) from the National Highway Traffic Safety Administration in U.S., the population of highly automated Tesla Model S vehicles was estimated to be 43,781. The ODI report considered the first fatal accident during fully automated driving. Automation failed and the Tesla vehicle drove into the side of a truck without braking.

According to the report, the driver was obliged to maintain full attention on the road and be prepared to take over driving at any time. However, "the driver took no braking, steering or other actions to avoid the collision" and appeared to have been distracted for more than 7 s prior to the accident, according to the conclusions made in the ODI Report (NHTSA, 2017). This accident highlights the importance of designing systems with human capabilities in mind. In order to avoid similar accidents, the relationship between the human and the highly automated vehicle and human ability to cope with the new driving role needs to be even better understood and, hence, be studied.

The most common measures in traditional driving safety studies include: vehicle speed, vehicle position in relation to road markings, distance from vehicle in front, angle of the steering wheel position and amount of pressure applied to the brake pedal (Castro, 2009). Young et al. (2009) also add event detection and reaction time as common measures. These measures describe driving performance and have little merit in human-automation studies (Jamson et al., 2013), except for those parts of automated driving that actually include manual driving; as in hand-over and take-over situations. As a consequence, the assessment of driver behaviour will need to adjust to the new driving situation involving automated driving. McBride et al. (2014) specify four categories of human-automation concerns: automation-related (such as reliability), person-related (such as complacency), task-related (such as automaton failure consequences) and so called emergent factors. The emergent factors were described as variables related to the interaction between the human and automation, as in trust, situation awareness and mental workload (McBride et al., 2014). Other factors that should be of special concern in human-automation studies include: behavioural adaptation (as in lowered perceived risk), skill degradation, and inadequate mental model of automation functioning (Saffarian et al., 2012). All of these concerns are not inevitable; they can be mitigated by a well-designed and adapted human-automation interface (Parasuraman, 2000), with a balance between abilities, authority, control and responsibility (Flemisch et al., 2012). A better understanding is required of the driver's relationship with automation and behaviour during automated driving. An important beginning of this understanding was constructed by Heikoop et al. (2016) in their review of causalities between the most commonly studied issues in human-automation research. According to their review the most commonly studied human-automation issues were (presented from most to least frequently studied): Mental workload, Attention, Feedback, Stress, Situation awareness, Task demands, Fatigue, Trust, Mental model, Arousal, Complacency, Vigilance, Locus of control, Acceptance and Satisfaction. These issues are not fully covered by traditional driving performance measures. A similar review of issue-related assessment methods was not found. When planning a study, there is a potential value in obtaining an overview of common measures selected by other researchers in the field. Therefore, one purpose of this paper was to provide a summary of assessment methods used for behavioural studies in the field of vehicle automation.

When planning a study, an overview of possible assessment methods is not enough for the construction of a well-designed study with relevant assessment methods. With such a new field, there may be difficulty in anticipating all issues. It might be difficult to select assessment methods and construction of new assessment methods may also be needed. This challenge was encountered in a Swedish research project called Methods for Designing Future Autonomous Systems (MODAS; Krupenia et al., 2014). In the project, a new information and warning system for highly automated driving was developed, and the aim was that it should be tested in a simulated driving session with a hazardous event. If it had been a

traditional manual driving situation, it would have been interesting to study the drivers' reaction times, steering behaviours etc. In this case, the driving was controlled by the vehicle (simulator) and the driver's role was to supervise and only take over driving if needed. Obviously, traditional assessment methods would only be relevant in case the driver actually decided to intervene. Traditional methods would otherwise not describe the driver's behaviour and reactions. In the project, it was tempting to include multiple assessment techniques, but due to practical and resource-related reasons the assessment methods needed to be narrowed down to a few well-motivated methods. A systematic assessment selection method was required. Stanton et al. (2013) describe 17 methods that systematically aid the identification of human errors and 20 methods that address task analyses. However, no single method was found that covered the assessment selection process completely from errors to assessment methods. As a result one goal with the MODAS project became to create a systematic assessment selection method, the Failure-GAM²E.

The purpose with this paper was threefold. One purpose was to review the selection of assessment methods used in driver behaviour studies during automated driving. A second purpose was to develop and propose an assessment selection method called the Failure-GAM²E that could be useful when a systematic selection is needed or when the research field is new. Finally, a third purpose was to exemplify the usefulness of the assessment selection method by means of a description of the assessment selection process using the Failure-GAM²E in the MODAS project. For clarity, the paper was divided into three Parts: A, B and C. In Part A, the review of assessment methods is summarised. In Part B, the development of the proposed Failure-GAM²E is described. In Part C, use of the Failure-GAM²E is exemplified by means of the assessment selection in the MODAS project.

2. Part A: review of assessment methods for highly automated vehicles

2.1. Introduction

Highly automated vehicles are a relatively new but strongly focused research area. There are several publications available that include investigations of human behaviour in highly automated vehicles. In this section, Part A, a minor review of the investigated issues and selected assessment methods is presented.

2.2. Method

Journal papers were collected via Google Scholar using the search phrase: (assessment OR methods OR measures) AND (automation OR autonomous OR self-driving) AND (driver OR driving OR vehicle* OR car*). From the search hits, 160 papers were downloaded for further reading. Journal papers addressing higher levels of automation in vehicles were prioritised. Finally, 35 journal papers investigating human behaviour in vehicles with automation were selected for the review summary.

2.3. Results

The 35 papers included investigations of the following human factors areas: Mental Workload, Situation Awareness, Trust, Acceptance, Fatigue, Arousal, Mental Model, Information, Feedback and Take-Over Performance. Several different measures were used. A summary of the measures follows.

2.3.1. Mental workload

Parasuraman et al. (2008) describe mental workload as “the

relation between the function relating the mental resources demanded by a task and those resources available to be supplied by the human operator”. According to Hart and Staveland (1988), mental workload is a combination of mental, physical and temporal demands. 13 of the reviewed papers addressed mental workload, stress or arousal, and of these, ten used a subjective rating questionnaire as assessment method. NASA-TLX by Hart and Staveland (1988) was the most commonly used questionnaire (used by Banks and Stanton, 2016; de Winter et al., 2016; Endsley, 1999; Heikoop et al., 2017; Kaber and Endsley, 2004; Sauer et al., 2013; Stanton and Young, 1998). Other questionnaires used for subjective rating of mental workload were: The Rating Scale Mental Effort (RSME; developed by Zijlstra, 1993; used by Brookhuis et al., 2008), the Subjective Workload Assessment Test (SWAT; developed by Reid et al., 1981; used by Baldauf et al., 2009), and the Dundee Stress State Questionnaire (DSSQ, developed by Matthews et al., 2002, and used by Funke et al., 2007; Heikoop et al., 2017). Mental workload was also measured as physiological responses, i.e. heart rate or heart rate variability (Brookhuis et al., 2008; Dehaes et al., 2012; Heikoop et al., 2017; Sauer et al., 2013), electrodermal activity (Baldauf et al., 2009), and eye blink behaviour (Merat et al., 2012). Mental workload or stress was also measured using the PERcentage eye CLOSED measure (PERCLOS; developed by Wierwille et al., 1994; used by Heikoop et al., 2017). Cottrell and Barton (2013) also suggest physiological measurement of cortisol concentration and pupil dilation as indicators of mental workload. Another measure used in several studies was secondary task performance. During high mental load, less mental resource would be left for a secondary task and, hence, affect the result. One secondary task encounter was the Peripheral Detection Task (PDT; developed by Martens and Van Winsum, 2000; used by Brookhuis et al., 2008), another was the Rotated Figures Task (developed by Baber, 1991; used by Stanton and Young, 1998), and a time perception task, i.e. the Current Duration Production (CDP; developed by Zakay and Shub, 1998; used by Baldauf et al., 2009). De Winter et al. (2016) created a distinct secondary task measurement for assessing mental workload.

2.3.2. Situation awareness

Endsley (2006) described situation awareness as the “perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future”. In the reviewed papers, situation awareness was measured in several different ways, including both subjective ratings and objective measures. Ten papers addressed situation awareness, and of them three used the subjective rating scale Situation Awareness Global Assessment Technique (SAGAT; developed by Endsley, 1988; used by Endsley, 1999; Kaber and Endsley, 2004; van den Beukel and van der Voort, 2017). Van den Beukel and van der Voort (2017) also used the subjective Situation Awareness Rating Technique (SART; developed by Charlton, 2002). Three other papers measured the time to resume control, i.e. observed performance in take-over situations (Gold et al., 2013; Merat et al., 2014; Payre et al., 2016). Lu et al. (2017) used an own car placement method as a measure of time perception combined with subjective ratings. Situation awareness was also assessed by means of eye movement patterns and gaze behaviour (Gold et al., 2013; Hergeth et al., 2016; Jamson et al., 2013; Louw and Merat (2017), Lu et al. 2017; Merat et al., 2014). Similar to the measurement of mental workload, two papers also used secondary task performance as a measure of situation awareness (Beller et al., 2013; Kaber and Endsley, 2004). Situation awareness has also been assessed by measuring object detection performance and a voluntary uptake of tasks unrelated to driving (de Winter et al., 2014).

2.3.3. Trust

Trust is defined as “the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability” (Lee and See, 2004). Walker et al. (2016) suggest four different methods of assessing trust: primary task measures, subjective scales, conceptual model building and repertory grids. They raise the problem of using primary task measures in human-automation studies; the measure is not useful when primary tasks is lacking, as in highly automated driving. Seven papers addressing trust in automation were reviewed and all of them included a subjective rating scale. Two papers used the Checklist for Trust between People and Automation (developed by Jian et al., 2000; used by Banks and Stanton, 2016; Beggiato et al., 2015). Other subjective scales were the FAD acceptability scale (developed by Payre et al., 2014; used by Payre et al., 2016), a questionnaire developed by Takayama and Nass (2008; used by Koo et al., 2015), and own-developed questionnaires regarding trust (Beller et al., 2013; Hergeth et al., 2016; Lee and Moray, 1992). In one study, trust was also measured objectively as the rate of unnecessary overtakes (Beller et al., 2013). Trust was also measured by means of gaze behaviour and monitoring frequency (Hergeth et al., 2016).

2.3.4. Acceptance

Acceptance relates to the perceived usefulness and perceived ease of use of a system (Davis, 1993). Nine papers addressed acceptance in technology. All used subjective rating scales or interview questions as measures. Three used a rating scale developed by van der Laan et al. (1997; used by Beggiato et al., 2015; Brookhuis et al., 2008; van den Beukel and van der Voort, 2017; van Driel et al., 2007). Two papers used the Technology Acceptance model, or revised versions (TAM, developed by Davis, 1993; used by Choi and Ji, 2015; Ghazizadeh et al., 2012). In one paper, the Driver Opinion Scale was used (developed by Nilsson, 1995; used by Stanton and Young, 1998). König and Neumayr (2017) used an own set of questions in a web-questionnaire. Payre et al. (2014) also used an own set of questions.

2.3.5. Fatigue and arousal

Fatigue, drowsiness, vigilance, arousal and alertness are definitions that relate to each other or overlap to some extent (Lal and Craig, 2001). The measures are therefore similar and will not be differentiated here. Six papers were reviewed and different measures were used in all of them, with a subjective scale used in one and objective data used in all the others. The subjective scale used was the Stanford Sleepiness Scale (SSS, developed by Hoddes et al., 1973; used by Ting et al., 2008). In one study heart rate measures were used to indicate arousal (Dehais et al., 2012). Alertness was measured by means of eye behaviour (pupil activity and eye closure) and head position in a study by Mbouna et al. (2013). Eye closure was also used to assess drowsiness through the PERcentage eye CLOSeD measure (PERCLOS; developed by Wierwille et al., 1994; used by Jamson et al., 2013). Two papers addressed fatigue or arousal by measuring visual attention; fatigue was assessed using the Psychomotor Vigilance Task (PVT; developed by Dinges and Powell, 1985; used by Baulk et al., 2008), and arousal by measuring eye fixations (Dehais et al., 2012). Heikoop et al. (2017) used a monitoring detection task as a measure of vigilance. In a review of measures of fatigue and alertness, Lal and Craig (2001) also suggest Electroencephalography (EEG) as a reliable method.

2.3.6. Mental model, information and feedback

Two papers addressed Human-Automation Interaction and the user's understanding of the automation system. Beggiato et al. (2015) used an own set of questions based on a mental model

defined by Carroll and Olson (1987). Banks and Stanton (2016) used interviews to investigate how well participants understood the automation and the information provided. Moreover, Davidsson and Alm (2014) have proposed a new method for investigating what type of information drivers need and when.

2.3.7. Take-over performance

Take-Over Performance relates to the situation when the driver needs to take back control from the automated system. It could be due to a change in the driving scenario or weather conditions making manual driving more appropriate than automated driving. It could also be due to automation failure. Take-over situations were studied in five papers. Blommer et al. (2017); Gold et al. (2013); Merat et al. (2014); Payre et al. (2016); and van den Beukel and van der Voort (2017) measured the response type (steering or braking), response time, range to the vehicle in front at take-over and/or collision occurrence.

2.4. Discussion

The review of papers showed a relatively concise selection of addressed human behaviour issues. Most issues were general and not directly related to automation per se. A few assessment methods could be linked to specific automation issues. The most frequently studied issue was mental workload. Mental workload is a common measure in transportation research in general, as is situation awareness. Trust, acceptance and vigilance are not new issues but their relevance may increase with automation. Take-over performance is a new measure directly related to automation and has no relevance during manual driving. The review also showed a large differentiation in the selection of assessment methods. No common method could be identified, except for the mental workload measure NASA-TLX, used in six papers. Both subjective and objective methods were used and many studies included a combination of subjective and objective methods.

3. Part B: a proposed assessment selection method – the Failure-GAM²E

3.1. Introduction

Part A exemplifies that there are several different assessment methods to choose between when designing a study. When the research field is new, the question of selecting relevant assessment methods for a study becomes even more complex since new measurements might be needed. In the MODAS project, the research group consisted of several researchers from different affiliations with previous experience of human behavioural studies in a driving context. Nevertheless, our knowledge was incomplete when it came to anticipating relevant assessment methods for highly automated driving. We had a lot of ideas, but we needed to structuralise the study design process. Several well-known and described methods for structured analysis of human errors and tasks were reviewed. It was noted that many human error analyses are based on Hierarchical Task Analyses (HTA; Annett, 2004). In the MODAS project, there was considered to be some difficulty in starting with an HTA for a situation in which the driver had no manual tasks; the driver's role was to supervise the automated driving. Instead, it felt more natural to investigate possible errors first and then identify relevant actions for the driver to take in order to avoid the errors and thereafter select relevant assessment methods. No such complete method was found. Part B describes the development process of a proposed assessment selection method, The Failure-GAM²E.

3.2. Method

3.2.1. Combination of existing methods

Instead of inventing a method from the beginning, standard methods that met the needs were used and combined to create a new method. A standard Hazard analysis for road vehicles (Road Vehicles – Functional Safety; ISO 26262) was selected for the purpose of identifying the most hazardous human errors, or Failure modes, and convert them into Safety Goals (Fig. 1). The step of identifying failure modes was renamed Failures (Fig. 1). The Goal, Operators, Methods and Selection rules method (GOMS; Card et al., 1983) includes steps for transforming goals into user actions. Goals represent what the user wishes to achieve. Operators stand for the motoric and cognitive actions that the user needs to take. This step was renamed Actions (Fig. 1). In GOMS, Methods stands for the procedures in which the actions could be performed, i.e. not assessment methods. The Selection rules relate to the selection of Methods (i.e. procedures) when there are several options. For the purpose of selecting assessment methods, only the first two steps, Goals and Operators (i.e. Actions), were used. In a refinement process, Safety Goals and Goals were considered closely related and merged into one, named Goals (Fig. 1).

3.2.2. Additional steps

After the specification of actions, steps for identifying suitable methods, both subjective and objective, were added (Fig. 1). Additionally, a step relating to equipment was added (Fig. 1). A first and a last *main* step were also added to the Failure-GAM²E (bottom line in Fig. 1). The first main step related to definition of the situation. The last main step related to final selection of methods and equipment.

3.3. Results – the Failure-GAM²E

3.3.1. The three main steps

The proposed assessment selection method, the Failure-GAM²E, includes three main steps (Fig. 2). In the first step, the problem situation is described and clarified using as many details as

possible, including traffic situation, behaviour of the automatic system, human-automation interface, and the role of the driver. In the second step, the six sub-steps starting with failures and ending with equipment are walked through. In the third step, the selection of methods and equipment is adjusted based on priorities and available resources.

3.3.2. The six sub-steps

The proposed method was called the Failure-GAM²E, based on the six sub-steps of identifying Failures, Goals, Actions, subjective Methods, objective Methods and Equipment (Fig. 2). In the first sub-step (2.1 Failures), the possible failures are defined. At this step several different creative and inspirational methods, such as brainstorming, could be used. The identified failures can be weighted based on specified criteria, such as severity or controllability. The most important failures are selected for further investigation. In the second sub-step (2.2 Goals), each failure is transformed into a driver goal. For example, if the failure is to not detect a hazardous event, the driver goal would be to detect hazardous events. In the third sub-step (2.3 Actions) there is specification of all possible driver actions and responses that are needed or expected in order to achieve the goals. These could be motoric actions, such as moving the eyes towards a hazardous event, cognitive actions, such as thoughts and feelings, and biophysical (psychophysical) responses. In the fourth sub-step (2.4 (Subjective) Methods), there is specification of subjective methods that could be used to assess the actions. These could be interview questions, rating scales, or similar. Both standard methods or own questions could be specified. In the fifth sub-step (2.5 (Objective) Methods), there is specification of objective methods that could be used to assess the actions. These could be measurements of movements, actions or biophysical responses. The measurements could, for example, be of task completion time, eye-fixations or galvanic skin response. The last sub-step (2.6 Equipment) relates to a specification of all materials and equipment needed for both the subjective and objective methods. Many methods could be assessed using different techniques. In this step, alternative solutions can be described. For example, a movement could be assessed by means of

Fig. 1. Development and refinement of the Failure-GAM²E.

Fig. 2. The three main steps and six sub-steps of the proposed Failure-GAM²E assessment selection method.

more advanced motion capture, or by using more simple sensors and an own-written program, or by means of video observations. All these techniques could give the same result with more or less effort and at different costs. In the last main step there is selection of the most suitable methods and equipment for the study, as described in 3.3.1. The three main steps.

3.4. Discussion

The goal was to find a systematic and supporting method for the transition from a problem to a study plan. The underlying standard methods, in this case the hazard analyses ISO26262 and GOMS, were helpful in the process of identifying initial core steps. It is worth noting that other methods for failure identification or task analysis may have led to the same end result. The Hazard Analysis method (ISO 26262) was originally developed for system failures but was found useful also for human-related failures. The method was selected partly because it is a well-used method in the transport sector and partly because it transforms failures into safety goals. The identification of failures is not supported by the hazard analysis method however. Failures need to be identified using other creative methods or based on previous experience or knowledge, i.e. in case there is a specific problem that needs further investigation such as a real accident scenario. In the standard hazard analysis method, failures are ranked according to severity, exposure and controllability. Other criteria-weighting methods might work as well. In Failure-GAM²E, the specific method or criteria for failure selections are therefore not specified. The GOMS was found useful due to the transition from goals to actions. The identification of actions, both minor and more prominent, motoric as well as cognitive, made it easier to identify useful assessment methods, at least in the MODAS project. It seemed important to specifically mention both motoric and cognitive processes since human errors often include both motoric and mental actions (Reason, 1990) and because the tasks are expected to become more mind-related (Banks et al., 2014). In Failure-GAM²E, subjective, objective methods and equipment were allocated to separate steps. This was mainly because human behaviour studies often include both

subjective and objective methods, as the review in Part A showed. They provide different views and a division of methods into subjective and objective may help the researcher to think through values of using both subjective and objective techniques in a study. The review in Part A showed that studies often include both subjective and objective methods, and a single step for all methods was also considered too comprising. In addition, different types of equipment could be used for the same purpose. For example, mental workload could be measured using questionnaires or biophysical sensors such as electroencephalogram (EEG). Equipment was therefore added as a separate step. Failure-GAM²E was developed in parallel with a study plan for a study in the MODAS project. It was noted that an undefined problem perspective made it difficult to identify possible failures. For example, it is necessary to know what the driver is supposed to do before what she or he should not do can be identified. Clearly, some initial definitions were needed before an identification of failures. This was the reason for including a first main step for defining the situation in the assessment selection method. Additionally, the sub-steps in Failure-GAM²E were focused on the identification of methods and equipment that would fit the purpose. The specified possibilities may be far too many for one study. A final decision was needed, in which methods and equipment would be narrowed down to a study plan. This final decision became the last main step.

4. Part C: assessment selection in the MODAS project

4.1. Introduction

The MODAS project involved researchers from a Swedish truck developer (SCANIA), two universities (Uppsala University and Luleå University of Technology) and one research institute (The Interactive Institute) in Sweden. About 15 researchers contributed in the project. The project addressed human-automation interaction during convoy driving with trucks in automated driving mode. A new information and warning system for highly automated driving was developed and the aim was to test it in a simulated driving session with a hazardous event. In the process of selecting relevant

assessment methods, other researchers' selections were reviewed (Part A) and a systematic method for the assessment process was developed, i.e. the proposed assessment selection method described by Failure-GAM²E in Part B. Failure-GAM²E was used to guide the assessment selection in the MODAS project. The procedure and results are described in this part (Part C) as an example of how Failure-GAM²E can be used in practice.

4.2. Method

4.2.1. Process and contribution

In the assessment selection process, all main steps and sub-steps of Failure GAM²E (Part B; Fig. 2) were followed. The first main step, definition of the situation, was a process led by the Swedish truck developer to which all researchers contributed. The first sub-step, identifying failures (2.1 Failures; Fig. 2), was investigated by means of a workshop. The other steps were processed mainly by the author, though the choices were discussed and refined with assistance from project members or colleagues. The final selection of methods and equipment was a joint process in which the research questions, the project budget and other practical limitations were weighed together.

4.2.2. Workshop for identifying failures

A full-day workshop with 6 researchers (three from Luleå University of Technology, one from Uppsala University, one from The Interactive Institute, one from the truck development company, SCANIA) and one aeroplane pilot was arranged. A truck driver was also invited but could not attend. The workshop started with a method description and presentation of the pre-defined specification of situation, level of automation and driver role that was to be in focus during the workshop. The search for possible failures started with private brainstorming. All participants wrote down as many different failures as they could imagine on individual post-it notes. As inspiration, they had a poster containing a lot of images showing different weather and traffic situations, old and young drivers etc. Once all participants had exhausted their ideas, they presented their failures to the group. The post-it notes with failures were put up on a whiteboard. After the presentations, the post-it notes with similar failures were placed together, forming categories. When all post-it notes was presented and grouped into categories, the 24 error mode taxonomies from the Systematic Human Error Reduction and Prediction Approach (SHERPA; Embrey, 1986) were used for additional inspiration. The group members added more failures to the whiteboard inspired by SHERPA and by the other group members' post-it notes. Finally, the failures were rephrased into more strict sentences, as in the example: "fail to trust the system". The most relevant failures for the specific situation to be studied were selected.

4.3. Results

4.3.1. Main step 1: definition of situation

The automation level focused on in the MODAS project was defined as high. According to Level of Automation (LoA) proposed by Sheridan et al. (1978), the automation level was set to 7–10. The LoA differed between conditions in the study. In one condition, the automatic system informed the driver about the driving actions and the causes, for example a hazardous event (LoA 7). In the other condition the automation system did not give any information to the driver (LoA 10). The definition of the situation described a need for the driver to be active, to monitor the situation and to stay in the loop. The driver should be responsible for the driving. The driver could intervene at any time, and should do so if the automation system failed. This definition was based on SCANIA's idea of a future

truck driver at the time. In the defined situation, participants would follow a convoy of trucks. The driving would be fully automatic and without failures, hence the driver would not have to intervene. After 5 min of driving, a hazardous motorbike would drive past the convoy on the wrong side of the truck convoy. The truck moves to the side to make space for the motorbike. The motorbike is passed and the driving scenario ends. The driver should notice the situation as a possible risk and be prepared to take over driving.

4.3.2. Main step 2: failure GAM²E

The failure workshop resulted in three failures of special importance for the project: failure to detect critical obstacle, failure to interpret the situation as critical, and failure to trust the system (Fig. 3). The failures were transformed into goals describing an expectation that the driver should detect the obstacle, prepare for taking over, but not taking over driving (Fig. 3). In order to meet the goals, the driver needed to perform several actions specified in Fig. 3. In short the driver would need to visually focus on the critical event and move a hand to the steering wheel or a foot to the gas or brake pedal without actually turning the steering wheel or pressing down a pedal. The driver was also expected to produce a biophysical response to the hazardous event. Both standard questionnaires and own questions were specified as subjective methods (Fig. 3). The objective methods relate to both observation of movements and biophysical responses (Fig. 3). Both specialised measurement techniques such as eye-tracking and more general techniques such as video observations were identified as possible measurement techniques – hence, the specification of equipment included a variety of tools including eye-trackers and video recording equipment (Fig. 3).

4.3.3. Main step 3: selection of methods and equipment

The equipment and methods were adjusted and selected based on available resources and complications encountered in the study. For example, video data was used instead of eye-tracking due to a too restricted eye-tracking range of the eye-tracking system.

4.3.4. Optimal risk management model

As an additional result, the map of goals and actions defined by means of Failure-GAM²E could be transformed into an optimal risk management model for driving with automation (Fig. 4). The goals from Failure-GAM²E were described as three management tasks: detection, preparation and takeover. The management tasks were further linked to three behavioural issues: situation awareness, risk awareness and trust. Based on the Failure-GAM²E results (Fig. 3), the driver should detect obstacles, i.e. maintain situation awareness. The driver should be prepared to take over, i.e. have risk awareness and not place too much trust in the automation system. Finally, the driver should not take over control, i.e. the driver should trust the automation system, if the automation system is working as it should. However, if the system exhibits a lack of control and the driver notices an automation failure, the driver should intervene and take over control. In other words, the driver should not trust the system when it fails. Fig. 4 shows the relationship between management tasks and different awareness and trust levels. If situation awareness was lacking, the driver would not detect the obstacle, and hence not prepare for takeover or actually take over. If the driver detects the obstacle but does not prepare for takeover, the driver would lack risk awareness due to a misinterpretation of the situation, or too much trust in the system. Finally, if the driver detects the obstacle (has situation awareness) and prepares for take over (has risk awareness), then the decision as to whether or not to take over would be related to the automation system functioning, i.e. if the automation system has control or not.

Failures	Goals	Actions	(Subjective) Methods	(Objective) Methods	Equipment
Fail to detect critical obstacle	Detect critical obstacle	Search for critical events, visually focus critical event.	SA-questionnaires (i.e. SART). Additional question: Describe what happened in the scenario (open). Did you detect the obstacle (direct)?	Registration of eye movements.	Questionnaires, interview recording, eye-tracking or video-recording.
Fail to interpret the situation as critical	Interpret the situation as critical and prepare for take over	Put hands on steering wheel and/or put foot in brake or gas position. Increased attention, concentration and arousal expected.	Question: Did you apprehend the situation as critical? How did you behave in the situation?	Observation of hand and feet movements. Registration of biophysical reaction (HR, muscle tension et c.).	Questionnaires, interview recording, video recording and/or sensors recording movement, and physiological reaction.
Fail to trust the system	Trust the system and not take over driving	Not move the steering wheel, not press down any of the pedals	Trust-questionnaire (i.e. Human-Automation-Trust-questionnaire). Additional question: How did you behave in the situation? Did you want to take over the driving?	Registration of steering wheel or pedal movements	Questionnaires, interview recording, video recording and/or sensors recording steering wheel and pedal movement.

Fig. 3. The result of the Failure-GAM²E process in the MODAS project.

4.4. Discussion

The use of a systematic assessment selection method, in this case the proposed Failure-GAM²E, made the assessment selection in the MODAS project both clearer and easier. By specifying failures, goals, actions, methods and equipment (Fig. 3), the possible assessment methods became well-linked to specific concerns. Hence, the effects of removing a method from the study plan

became clear. If two methods covered the same actions related to a failure, one could be excluded, if a reduction of methods were necessary. But if the method was alone in being linked to a failure, an exclusion would mean that the failure would not be investigated. Also, the specification of failures, goals and actions made it easy to identify unnecessary assessment methods, i.e. methods that had no relation to the specified failures. Hence, the link between failures and methods was found to be very useful when methods

		SA: Situation Awareness RA: Risk Awareness T: Trust; OT: Overtrust		Automation in Control	Automation Failure
		AWARENESS AND TRUST			
		No SA	SA; No RA/OT	SA; RA; T	SA; RA; No T
MANAGEMENT	Detection	No Detection	Detection	Detection	Detection
	Preparation	No Preparation	No Preparation	Preparation	Preparation
	Takeover	No Takeover	No Takeover	No Takeover	Takeover

Fig. 4. Optimal Risk Management Model for driving with automation. While automation is in control, driver management should include hazard detection, preparation for takeover but not overtaking. During automation failure the driver should take over driving.

and equipment were selected for a final study plan.

In the step-by-step Failure-GAM²E process, the identification of failures was experienced as the most difficult step. A workshop with participants from different fields, with different knowledge and perspectives proved useful. In the MODAS workshop the representatives of the academic world were able to add valuable information regarding human errors and cognitive issues in general. The representative from the truck industry added valuable knowledge concerning truck accidents and normal truck driver activities. The pilot contributed valuable insights from the perspective of interacting with automated systems and was also able to relate known human-automation failures in aeroplanes to possible human-automation failures in road vehicles. Where possible failures were identified, the most important failures were to be selected for the study. The ranking of failures depending on severity, controllability and exposure, following the Hazard Analysis method (ISO 26262), was tested during the workshop, though it was found to be too complicated for application to automated driving. Due to the unknown future relationship between human and automation, all failure modes were in one way or another assumed to be able to cause severe hazards with low controllability. Instead the failure selection was made based on what the project members thought was most relevant for a first study. The result of Failure-GAM²E in the MODAS project (Fig. 3) should be read as an example of how Failure-GAM²E can be used. The specifications and selections were subjective and based on what was considered the most important research questions and methods for the MODAS project. The Optimal Risk Management Model is a more general result that reflects driver goals and obligations. Hopefully, it will be found to be of use for other researchers in the process of defining driver responsibilities during automated driving.

5. General discussion

In the MODAS project, researchers with previous knowledge from driver behaviour research experienced what Rasmussen described in 1990, namely that errors and actions can be more complex to define in supervisory tasks than in manual tasks. However, the difficulties encountered in the process of selecting appropriate assessment methods were most likely also an effect of the research field being new and the issues being to some degree unknown. The research group was familiar with traditional assessment methods but lacked knowledge of what may be common assessment methods in vehicle automation studies. An assessment selection strategy was also needed. Surprisingly, it was not possible to identify a method that supported the whole process of selecting assessment methods. Therefore, the purpose of this paper came to be reviewing and summarising assessment methods used for studying driver behaviour during automated driving and to propose an assessment selection method called Failure-GAM²E. Both the review and the assessment selection method were found to be of value in the MODAS project. Using Failure-GAM²E, the large variety of possible assessment methods was narrowed down to a structured and well thought-through study design. The struggles were changed to a systematic step-by-step procedure. It was anticipated that other researchers may face the same struggles. The hope is that the review and method will also be of value for these researchers in their planning of human-automation interaction studies.

Before the development of Failure-GAM²E, the initial idea of which assessment methods to include in the MODAS project was too comprehensive and unstructured. The assessment methods needed to be condensed and more clearly related to specific issues. Time and money also had to be considered. New investments in pre-developed, high standard research equipment such as eye-

trackers, motion capture and advanced biophysical sensors could not be made with the project budget. Building new measurement tools from available sensors and in-house programming was also regarded as demanding of resources. A more simple technique, e.g. entailing video recordings and observations, is relatively easy and inexpensive to set up but is instead time consuming, and hence expensive, to analyse. In the MODAS project, some equipment was pre-installed in the driving simulator. Several actions, such as pedal movements, could be recorded automatically through the existing system. The most resource-effective assessment plan would have been to rely mainly on those data sources. However, when Failure-GAM²E was designed and followed, a different assessment plan developed. The new structured plan had few similarities with the initial sketches and with traditional methods. It became clear that some questions could not be answered using the existing measurement techniques already installed in the simulator. In the MODAS project, data indicating a non-pedal movement was found to be as interesting as a pedal movement. Hence, the foot behaviour and not only the pedal data were of interest. In such a situation, when the decision regarded spending extra resources on developing new measurement techniques, for example an in-house motion capture system for recording foot movements, or adding comprehensive video observation time, the purpose and expected value of the measurement had to be clearly shown. The new assessment method needed to be properly justified. The link between failures, goals, actions and methods in the resulting Failure-GAM²E table (Fig. 3) provided that justification. Interestingly, observation of foot and hand movements, as an indication of risk awareness or trust, was not found in any of the reviewed papers (Part A). Takeover situations was studied using measures such as response time and response type in several papers (Blommer et al., 2017; Gold et al., 2013; Merat et al., 2014; Payre et al., 2016; van den Beukel and van der Voort, 2017), and trust was measured by means of unnecessary takeovers (Beller et al., 2013) but no paper was found that used preparation for takeover as a measure. Clearly, the review of assessment methods (Part A) would not have led to the same assessment selections as Failure-GAM²E (Fig. 3) did. The assessment selection method not only supported the assessment selection process, it also made it easier to think outside the box and select what was appropriate for the research questions rather than what measures were commonly used.

The Optimal Risk Management Model (Fig. 4) was also a result of the Failure-GAM²E process. It was a translation of the identified goals and actions into management tasks and their relation to situation awareness, risk awareness and trust. Optimal management looked different for situations when automation was functioning, as it should (automation in control), and when the automation was malfunctioning (automation failure). This model was based on an idea of a responsible driver (Richards and Stedmon, 2016) who would be needed for supervision and adjustment (Brookhuis et al., 2001) and who would correct errors if automation failed (McBridge et al., 2014). With a different definition of driver responsibilities, the model would change. For example, future legislation might suggest that the system should prevent the driver from intervening. Such legislation would make the allocation of responsibility between vehicle developers and drivers clearer in case of an accident. In such a situation, the driver would not be obliged to maintain situation awareness, have risk awareness or decide whether or not he or she should take over driving. The driver would be transformed into a passenger with no obligations and the Optimal Risk Management Model would not be needed. However, as long as the driver is supposed to keep an eye on the traffic situation and be ready to take over, the Optimal Risk Management Model could be used to define the driver situation.

6. Conclusions

A systematic assessment selection method, such as the proposed Failure-GAM²E, could help researchers to design their study, clearly define the research questions and effectively focus their resources on these questions. The use of a planning tool, such as Failure-GAM²E, can also help the research team to think outside the box and identify new interesting questions and measures instead of using available measures by tradition. And, as a positive side effect, the need of new measurement techniques becomes clear and could push the development of new tools and methods forward. It is believed that Failure-GAM²E and the Optimal Risk Management Model both fill a gap. The hope is that they will become supportive tools for researchers entering the field of human-automation interaction in vehicles. Decisions will still be subjective and limited by the ability to foresee future problems, but with a supporting tool the chances of producing a good study plan are, if not ensured, at least improved.

Acknowledgements

The method and results described in this paper were outcomes of the MODAS project. The researchers that were part of MODAS contributed to these results through discussions and a workshop. Special thanks are due to Stas Krupenia (SCANIA), Håkan Alm (Luleå University of Technology), Jon Friström (Luleå University of Technology), Johan Fagerlönner (The Interactive Institute), and Anders Jansson (Uppsala University) for their valuable comments that contributed to Failure-GAM²E and the Optimal Risk Management Model. The MODAS project was financed by Vinnova-FFI (2012-03678) in Sweden.

References

- Akamatsu, M., Green, P., Bengler, K., 2013. Automotive technology and human factors research: past, present and the future. *Int. J. Veh. Technol.* 2013, 27.
- Annett, J., 2004. Hierarchical task analysis. In: Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H. (Eds.), *Handbook of Human Factors and Ergonomics Methods*. CRS Press, Boca Raton, FL, pp. 329–337.
- Baber, C., 1991. *Speech Technology in Control Room Systems: a Human Factors Perspective*. Ellis Horwood, London.
- Bainbridge, L., 1983. Ironies of automation. *Automatica* 19 (6), 775–779.
- Baldauf, D., Burgard, E., Wittmann, M., 2009. Time perception as a workload measure in simulated car driving. *Appl. Ergon.* 40, 929–935.
- Banks, V.A., Stanton, N.A., 2016. Keep the driver in control- Automating automobiles of the future. *Appl. Ergon.* 53, 389–395.
- Banks, V.A., Stanton, N.A., Harvey, C., 2014. Sub-systems on the road to vehicle automation: hands and feet free but not 'mind' free driving. *Saf. Sci.* 62, 505–514.
- Baulk, S.D., Biggs, S.N., Reid, K.J., van den Heuvel, C.J., Dawson, D., 2008. Chasing the silver bullet: measuring driver fatigue using simple and complex tasks. *Accid. Analysis Prev.* 40, 396–402.
- Beggiano, M., Pereira, M., Petzoldt, T., Krems, J., 2015. Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transp. Res. Part F* 35, 75–84.
- Beller, J., Heesen, M., Vollrath, M., 2013. Improving the driver-automation interaction: an approach using automation uncertainty. *Hum. Factors J. Hum. Factors Ergonomics Soc.* 55, 1130–1141.
- Blommer, M., Curry, R., Swaminathan, R., Tijerina, L., Talamonti, W., Kochhar, D., 2017. Driver brake vs. steer response to sudden forward collision scenario in manual and automated driving modes. *Transp. Res. Part F* 45, 93–101.
- Brookhuis, K.A., van Driel, C.J.G., Hof, T., van Arem, Hoedemaeker, M., 2008. Driving with a congestion assistant; mental workload and acceptance. *Appl. Ergon.* 40, 1019–1025.
- Brookhuis, K.A., de Waard, D., Janssen, W.H., 2001. Behavioural impacts of advanced driver assistance systems- an overview. *Eur. J. Transp. Infrastructure Res.* 1 (3), 245–253.
- Card, S.K., Moran, T.P., Newell, A., 1983. *The Psychology of Human Computer Interaction*. Lawrence Erlbaum, Hillsdale, NJ.
- Carroll, J.M., Olson, J.R., 1987. Mental models in human-computer interaction: research issues about what the user of software knows. In: Committee on Human Factors, Commission on Behavioral and Social Sciences and Education, National Research Council. National Academy Press, Washington, DC.
- Castro, C., 2009. Human Factors of Visual and Cognitive Performance in Driving. CRC Press, Boca Raton, FL.
- Charlton, S.G., 2002. Measurement of Cognitive States in Test and Evaluation. *Handbook of Human Factors Testing and Evaluation*. Lawrence Erlbaum Associates Inc., Mahwah, New Jersey.
- Choi, J.K., Ji, Y.G., 2015. Investigating the importance of trust on adopting an autonomous vehicle. *Int. J. Human-Computer Interact.* 31 (10), 692–702.
- Cottrell, N.D., Barton, B.K., 2013. The role of automation in reducing stress and negative affect while driving. *Theor. Issues Ergonomics Sci.* 14 (1), 53–68.
- Davidsson, S., Alm, H., 2014. Context adaptable driver information – or, what do whom need and want when? *Appl. Ergon.* 45 (4), 994–1002.
- Davis, F.D., 1993. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int. J. Man-Machine Stud.* 38, 475–487.
- Dehais, F., Causse, M., Vachon, F., Tremblay, S., 2012. Cognitive conflict in human-automation interactions: a psychophysiological study. *Appl. Ergon.* 43, 588–595.
- de Winter, J.C.F., Happee, R., Martens, M.H., Stanton, N.A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transp. Res. Part F* 2014, 196–217.
- de Winter, J.C.F., Stanton, N.A., Price, J.S., Mistry, H., 2016. The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *Int. J. Veh. Des.* 70 (4), 297–324.
- Dinges, D.F., Powell, J.W., 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav. Res. Methods Instrum. Comput.* 22 (1), 69–78.
- Embrey, D.E., 1986. SHERPA: a systematic human error reduction and prediction approach. In: *International Meeting on Advances in Nuclear Power Systems*. Knoxville, Tennessee.
- Endsley, M.R., 1988. Design and evaluation for situation awareness enhancement. In: *Proceedings of the Human Factors Society 32nd Annual Meeting*. Human Factors and Ergonomics Society, Santa Monica, CA, pp. 97–101.
- Endsley, M.R., 1999. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics* 42 (3), 462–492.
- Endsley, M., 2006. Situation awareness. In: Salvendy, G. (Ed.), *Handbook of Human Factors and Ergonomics*, third ed. Wiley, New York, pp. 528–542.
- Endsley, M.R., 2015. Situation awareness: operationally necessary and scientifically grounded. *Cogn. Tech. Work* 17, 163–167.
- Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., Beller, J., 2012. Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cogn. Tech. Work* 14, 3–18.
- Funke, G., Matthews, G., Warm, J.S., Emo, A.K., 2007. Vehicle automation: a remedy for driver stress? *Ergonomics* 50 (8), 1302–1323.
- Ghazizadeh, M., Lee, J.D., Boyle, L.N., 2012. Extending the technology acceptance model to assess automation. *Cogn. Tech. Work* 14, 39–49.
- Gold, C., Damböck, D., Lorentz, L., Bengler, K., 2013. "Take over!" How long does it take to get the driver back in the loop? In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2013*, pp. 1938–1942.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), *Human Mental Workload*. North Holland Press, Amsterdam, The Netherlands, pp. 139–183.
- Heikoo, D.D., de Winter, J.C.F., van Arem, B., Stanton, N.A., 2016. Psychological constructs in driving automation: a consensus model and critical comment on construct proliferation. *Theor. Issues Ergonomics Sci.* 17 (3), 284–303.
- Heikoo, D.D., de Winter, J.C.F., van Arem, B., Stanton, N.A., 2017. Effects of platooning on signal-detection performance, workload, and stress: a driving simulator study. *Appl. Ergon.* 60, 116–127.
- Hergeth, S., Lorentz, L., Vilimek, R., Krems, J.F., 2016. Keep your scanners peeled: gaze behavior as a measure of automation trust during highly automated driving. *Hum. Factors* 58 (3), 509–519.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., Dement, W.C., 1973. Quantification of sleepiness: a new approach. *Psychophysiology* 10 (4), 431–436.
- Jamson, A.H., Merat, N., Carsten, O.M.J., Lai, F.C.H., 2013. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transp. Res. Part C* 30, 116–125.
- Jian, J., Bisantz, A.M., Drury, C.G., 2000. Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* 4 (1), 53–71.
- Kaber, D.B., 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. *Theor. Issues Ergonomics Sci.* 5 (2), 113–153.
- Koo, J., Kwak, J., Ju, W., Steinert, M., Leifer, L., Nass, C., 2015. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *Int. J. Interact. Des. Manuf.* 9, 269–275.
- Krupenia, S., Selmarker, A., Fagerlönner, J., Delsing, K., Jansson, A., Sandblad, B., Grane, C., 2014. The 'methods for designing future autonomous systems' (MODAS) project. In: Ahrum, T., Karwowski, W., Marek, T. (Eds.), *Proceeding of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014*. Krakow, Poland.
- König, M., Neumayr, L., 2017. Users' resistance towards radical innovations: the case of the self-driving car. *Transp. Res. Part F* 44, 42–52.
- Lal, S.K.L., Craig, A., 2001. A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* 55, 173–194.
- Larsson, A.F.L., Kircher, K., Andersson Hultgren, J., 2014. Learning from experience:

- familiarity with ACC and responding to a cut-in situation in automated driving. *Transp. Res. Part F* 2014, 229–237.
- Lee, J., Moray, N., 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35 (10), 1243–1270.
- Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. *Hum. Factors* 46 (1), 50–80.
- Louw, T., Merat, N., 2017. Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transp. Res. Part C* 76, 35–50.
- Lu, Z., Coster, X., de Winter, J., 2017. How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Appl. Ergon.* 60, 293–304.
- Martens, M.H., Van Winsum, W., 2000. Measuring distraction: the peripheral distraction task. NHTSA Internet Forum on the Safety Impact of Driver Distraction when using In-Vehicle Technologies. <http://www-nrd.nhtsa.dot.gov/departments/Human%20Factors/driver-distraction/PDF/34.pdf>.
- Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., Warm, J.S., 2002. Fundamental dimensions of subjective state in performance settings: task engagement, distress and worry. *Emotion* 2, 315–340.
- Mbouna, R.O., Kong, S.G., Chun, M.-G., 2013. Visual analysis of eye state and head pose for driver alertness monitoring. *IEEE Trans. Intelligent Transp. Syst.* 14 (3), 1462–1469.
- McBridge, S.E., Rogers, W.A., Fisk, A.D., 2014. Understanding human management of automation errors. *Theor. issues ergonomics Sci.* 15 (6), 545–577.
- Merat, N., Jamson, A.H., Lai, F.C.H., Carsten, O., 2012. Highly automated driving, secondary task performance, and driver state. *Hum. Factors J. Hum. Factors Ergonomics Soc.* 54, 762–771.
- Merat, N., Jamson, A.H., Lai, F.C.H., Daly, M., Carsten, O., 2014. Transition to manual: driver behavior when resuming control from a highly automated vehicle. *Transp. Res. Part F* 27, 274–282.
- Merat, N., Lee, J.D., 2012. Preface to the special section on human factors and automation in vehicles: designing highly automated vehicles with the driver in mind. *Hum. Factors* 54 (5), 681–686.
- Milakis, D., van Arem, B., van Wee, B., 2017. Policy and society related implications of automated driving: a review of literature and directions for future research. *J. Intelligent Transp.* 0 (0), 1–25.
- NHTSA, 2017. ODI Resume. PE 16–1007, 01/19/2017. U.S. Department of Transportation.
- Nilsson, L., 1995. Safety effects of adaptive cruise control in critical traffic situations. In: *Proceedings of the Second World Congress on Intelligent Transport Systems: Steps Forward*, vol. III. VERTIS, Yokohama, pp. 1254–1259.
- Parasuraman, R., 2000. Designing automation for human use: empirical studies and quantitative models. *Ergonomics* 43 (7), 931–951.
- Parasuraman, R., Sheridan, T.B., Wickens, C.D., 2008. Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. *J. Cognitive Eng. Decis. Mak.* 2 (2), 140–160.
- Payre, W., Cestac, J., Delhomme, P., 2014. Intention to use a fully automated car: attitudes and a priori acceptability. *Transp. Res. Part F* 27, 252–263.
- Payre, W., Cestac, J., Delhomme, P., 2016. Fully automated driving: impact of trust and practice on manual control recovery. *Hum. Factors* 58 (2), 229–241.
- Rasmussen, J., 1990. Human error and the problem of causality in analysis of accidents. *Philosophical Trans. R. Soc. B* 327, 449–462.
- Reason, J., 1990. *Human Error*. Cambridge University Press, Cambridge, United Kingdom.
- Reason, J., Manstead, A., Stradling, S., Baxter, J., Campbell, K., 1990. Errors and violations on the roads: a real distinction? *Ergonomics* 33 (10–11), 1315–1332.
- Reid, G.B., Shingledecker, C.L., Nygren, T.E., Eggemeier, F.T., 1981. Development of multidimensional subjective measures of workload. In: *Proceedings of the 1981 IEEE International Conference on Cybernetics and Society*, pp. 403–406.
- Richards, D., Stedmon, A., 2016. To delegate or not delegate: a review of control frameworks for autonomous cars. *Appl. Ergon.* 53, 383–388.
- Saffarian, M., de Winter, J.C.F., Happee, R., 2012. Automated driving: human-factors issues and design solutions. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2012*, pp. 2296–2300.
- Sauer, J., Nickel, P., Wastell, D., 2013. Designing automation for complex work environments under different levels of stress. *Appl. Ergon.* 44, 119–127.
- Sheridan, T.B., Verplank, W.L., Brooks, T.L., 1978. Human/computer control of undersea teleoperators. In: *Proceedings of the IEEE Int. Conf. Cybern. Soc. Tokyo, Japan*.
- Stanton, N.A., Marsden, P., 1996. From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. *Saf. Sci.* 24 (1), 35–49.
- Stanton, N.A., Salmon, P.M., Rafferty, L.A., Walker, G.H., Baber, C., Jenkins, D.P., 2013. *Human Factors Methods. A Practical Guide for Engineering and Design*, second ed. Ashgate, Great Britain.
- Stanton, N.A., Young, M.S., 1998. Vehicle automation and driving performance. *Ergonomics* 41 (7), 1014–1028.
- Stensson, P., Jansson, A., 2013. Autonomous technology – sources of confusion: a model for explanation and prediction of conceptual shifts. *Ergonomics* 57 (3), 455–470.
- Strand, N., Nilsson, J., Karlsson, I.C.M., Nilsson, L., 2014. Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transp. Res. Part F* 27, 218–228.
- Takayama, L., Nass, C., 2008. Assessing the effectiveness of interactive media in improving drowsy driving safety. *Hum. Factors* 50 (5), 772–781.
- Ting, P.-H., Hwang, J.-R., Doong, J.-L., Jeng, M.-C., 2008. Driver fatigue and highway driving: a simulator study. *Physiology Behav.* 94, 448–453.
- Van Driel, C.J.G., Hoedemaeker, M., van Arem, B., 2007. Impacts of a congestion assistant on driving behaviour and acceptance using a driving simulator. *Transp. Res. Part F* 10, 139–152.
- Van den Beukel, A.P., van der Voort, M.C., 2017. How to assess driver's interaction with partially automated driving systems – a framework for early concept assessment. *Appl. Ergon.* 59, 302–312.
- Van der Laan, J.D., Heino, A., de Waard, D., 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transp. Res. Part C* 5 (1), 1–10.
- Walker, G.H., Stanton, N.A., Salmon, P., 2016. Trust in vehicle technology. *Int. J. Veh. Des.* 70 (2), 157–181.
- Wierwille, W.W., Ellsworth, L.A., Wreggit, S.S., Fairbanks, R.J., Kirn, C.L., 1994. *Research on Vehicle-based Driver Status/Performance Monitoring: Development, Validation and Refinement of Algorithms for Detection of Driver Drowsiness*. Final Report, DOT HS 808 247. National Highway Traffic Safety Administration, Washington, DC.
- Young, K.L., Regan, M.A., Lee, J.D., 2009. Measuring the effects of driver distraction: direct driving performance methods and measures. In: Regan, M.A., Lee, J.D., Young, K. (Eds.), *Driver Distraction: Theory, Effects, and Mitigation*. CRC Press, Boca Raton, FL, pp. 85–105.
- Zakay, D., Shub, J., 1998. Concurrent duration production as a workload measure. *Ergonomics* 41, 1115–1128.
- Zijlstra, F.R.H., 1993. *Efficiency in Work Behaviour. A Design Approach for Modern Tools*. PhD thesis, Delft University of Technology. Delft University Press, Delft, The Netherlands.