A Model-driven Tool for getting Insights into Car Drivers' Monitoring Behavior

Sebastian Feuerstack, and Bertram Wortelen

Abstract— **The Human Efficiency Evaluator (HEE) is a model-based tool that predicts car drivers' visual attention based on a variant of the SEEV model. Different to prior research that required individual human factor (HF) expertise to generate valid attention predictions, the HEE enables to collect data from a group of experienced car drivers, to simulate human monitoring behavior, and to end up with valid predictions. We invited two different groups: automotive human factors experts (n=9) and experienced car drivers (n=20) to predict car drivers' monitoring behavior for a highway overtaking scenario with the HEE. Previous research did not detail the amount and experience of the HF experts involved in generating predictions, whereas our study revealed a quite high variance of individual HF experts' predictions about drivers' typical monitoring behavior. We measured car drivers' monitoring behavior using an eye tracking device in a car driving simulator (n=20). The aggregated prediction of the group of car drivers was high (R=0.719) and better than the average prediction of an individual HF experts.**

I. Introduction

A very common approach to the analysis of car drivers' visual attention is to use eye tracking devices to observe them while driving. Such studies offer empirical data, but come with major drawbacks. First, they require a complex study setup in that the machine interface (e.g. an automotive assistance system) and its environment (e.g. the driving situation) need to be realistically simulated to produce valid data. Second, to attain a representative monitoring behavior, a reasonable amount of subjects need to be observed with an eye-tracker in a driving simulation. Effort increases with the number of subjects, because, subjects are tested successively and huge effort needs to be spent on data analysis thereafter (e.g. matching the eye tracking data of each driver to the dynamic traffic situation). Finally, the empirically collected data reflects the observed monitoring behavior but misses' reasons that explain what actually caused the observed behavior.

Model-based attention prediction approaches have been proposed to better understand human behavior but can also complement eye-tracking studies. They are especially helpful in early HMI design phases where only design sketches but no prototypes are available. By simulating human behavior based on psychological and physiological plausible models they can actually be used to predict human behavior such as interface monitoring, which has been already reported in the past for several application domains [1, 2, 3].

Predictive attention models, like e.g. the SEEV (Salience, Effort, Expectancy, Value) model [4] or the application of the analytic hierarchy process by Ha & Seong [3] are promising approaches to actually identify and describe the parameters that influence car drivers' monitoring behavior. Though, the validity of the input parameters that describe a car driver's monitoring behavior is an open issue. To the best of our knowledge visual attention prediction models have been only elaborated by human factors (HF) experts with extensive modelling experience and a strong expertise in the domain. These experts are rare. If such a model is created by just one expert, errors made by the expert can have a huge impact on the predictions. Thus building a model has to be done very thoroughly and is expensive in time and expertise. Our objective is to ease modelling, so that it doesn't require advanced knowledge of cognitive scientists or HF experts.

Therefore, we present the Human Efficiency Evaluator (HEE), a model-based attention prediction tool that combines model-based and empirical methods in a structured process. It eases the modelling process and enables domain experts with no background in human factors to model drivers' monitoring behavior. The HEE structures the modelling process in such a way that models of multiple domain experts can be combined to effectively reduce the impact of individual modelling errors and captures expert knowledge in a reproducible way. The attention predictions with the HEE are generated based on an adapted SEEV model.

We invited HF experts (without prior SEEV model experience) and domain experts (experienced car drivers) to model car drivers' visual attention for an overtaking maneuver. To validate the model predictions, we performed trials in a car driving simulator with eye tracking hardware to measure the actual monitoring behavior for the same highway overtaking scenario that we asked the subjects to model with the HEE. We focus on the following research hypothesis:

H1: *"Aggregating multiple domain expert models ends up in better attention prediction results than those that can be gained by an average individual human factors expert."*

*Research supported by the European Commission (H2020-MG-2014-2015) in the interest of project AutoMate (http://www.automate-project.eu/) – Grant Agreement 690705 and the funding initiative Niedersächsisches Vorab of the Volkswagen Foundation and the Ministry of Science and Culture of Lower Saxony as a part of the Interdisciplinary Research Centre on Critical Systems Engineering for Socio-Technical Systems.

S. Feuerstack is with the OFFIS Institute for Information Technology, 26121 Oldenburg, Germany (phone: +49 441 9722 509; e-mail: feuerstack@offis.de).

B. Wortelen, is with Cognitive Psychology Lab at C.v.O University, 26129 Oldenburg, Germany (e-mail: bertram.wortelen1@uol.de).

In the following, we discuss related work on models for predicting human attention and summarize the experiment and validation setups. We focus on the SEEV model and variants, which seem to be the most popular models based on the amount of publications that we have found. Thereafter, we give an overview about the HEE tool and its underlying knowledge capturing process, before we detail our study setups and report about our findings

II. Attention Prediction Models

The vision of human modeling is to provide methods, techniques and tools to generate predictions of human performance. The SEEV model of attention allocation [4] provides such a promising theory. It describes that "the allocation of attention in dynamic environments is driven by bottom up attention capture of *salient* events, which are inhibited by the *effort* required to move attention, and also driven by the *expectancy* of seeing *valuable* events" [5].

The SEEV model is used to predict the percentage of time, that someone spends looking at an area of interest (AOI). It is typically applied by HF experts that have a deep understanding of human attentional processes. The SEEV model relates the probability P_S of attending a specific AOI s to four factors:

$$P_S = \text{Saliency} - \underline{\text{E}}\text{ffort} + \underline{\text{E}}\text{xpectancy} \cdot \text{Task } \underline{\text{V}}\text{alue} \qquad (1)$$

Saliency and *Effort* are bottom-up factors that describe the saliency of information displayed by an AOI and the effort it takes to obtain the information, e.g., by moving eyes and head or navigating through a menu. *Expectancy* and *Task Value* are top-down factors. They describe how often new information can be expected from an IS and how valuable the information is for accomplishing the tasks of the human operator.

MIDAS [6] for instance is a system developed by NASA since 1985 that integrates the SEEV model. The *saliency* (e.g. contrast), *expectancy* (i.e. how often new information is expected?), and *value* (i.e. how valuable is the information?) coefficients are assigned using HF expert ratings [7]. The *effort* coefficients (e.g. head or eye movements) are calculated automatically based on the distances between the different AOIs. The SEEV model was also integrated in the Attention-Situation Awareness (A-SA) Model, that has been used to predict optimal scanning paths in landing operations of an airplane. System data determines the *expectancy* and regulation data defines the *value* of areas of interest, which have been previously identified using eye-tracking data. "Based on the parameters of *effort*, *expectancy* and *value* [the model] accounted for roughly 30%-80% of the variance in scanning data seen in human data." [8].

Over the last decade, such SEEV model variants have been used to model and predict attention allocations for a wide variety of tasks: To evaluate drivers' monitoring behavior while approaching intersections [9] or to evaluate the influence of secondary tasks [10, 11], for landing an airplane [8], to analyze the influence of specific cockpit instruments [12], or to analyze the allocation of attention of nurses assisting in medical interventions in a hospital [13].

All these studies report moderate up to very high correlations ($0.6 < R < 0.97$) between eye tracking studies and the model predictions. But the number of AOIs and therefore the amount of data points in these studies is quite low, which was also noted by the authors themselves, e.g. in [11], and "may have artificially inflated the model fit" [11]. The AOIs that were distinguished for attention distribution varied between 2 up to 6 per experiment, with the majority modeling and calculating predictions for attention allocations of 3-4 AOIs.

For nearly all of the studies prior experience of the involved experts was not stated (i.e. expert knowledge, e.g. gained by earlier observations of operators in similar situations, or prior eye tracking studies). Most studies rely on explicit methods and techniques like e.g. [14], which uses matrixes and the "least integer ordinal value" heuristic, to end up with parameter values in an ordinal metric use or the "analytic hierarchy process" for quantifying the informational importance [3]. The input data validity often cannot be reproduced easily. Most studies state all the relevant concrete parameter values. This allows to reproduce their predictions, but the prior determination of the parameter values rely considerable on individual expert's expertise. Interestingly, only one study we found [13] reported insights about the amount of experts, their background and prior knowledge, and the method applied to determine and agree on the concrete model input parameter values.

There are individual human factors experts with a very high expertise, such as for instance Wickens et al. that continuously published over a decade high correlating predictions in the aeronautics domain [15]. To which extend Human Factors or SEEV model expertise is required for good predictions and how well experts do concord in their parameter determination has not been researched so far to the best of our knowledge. Research about model-based prediction benefits from empirically reproducible input parameter determination process and from investigating in the expertise required to end up with good and reproducible predictions.

In the subsequent section we describe the Human Efficiency Evaluator (HEE). The HEE is a software tool that we developed to support predictive cognitive modelling in a reproducible way. Some tools already support cognitive model creation. CogTool [10] for instance supports the generation of ACT-R [1] models with deterministic sequences of actions. These models are based on GOMS and KLM and are targeted on evaluating Windows-, Icons, Menus, and Pointer (WIMP) user interfaces. The Distract-R system by Salvucci [16] is also based on ACT-R. It allows to create ACT-R models of in-vehicle, secondary task interactions in a way similar to CogTool. It integrates these models with a detailed driver model to simulate and predict effects of secondary task distraction on driving behavior. COGENT [17] is a graphical modeling editor for psychologists that allows "programming" cognitive models at a higher level of abstraction. It is based on box and arrow diagrams that link to a set of standard types of cognitive modules, which implement theoretical constructs from psychological theory. COGENT, CogTool, Distract-R and HEE share the idea of making cognitive modeling easier by

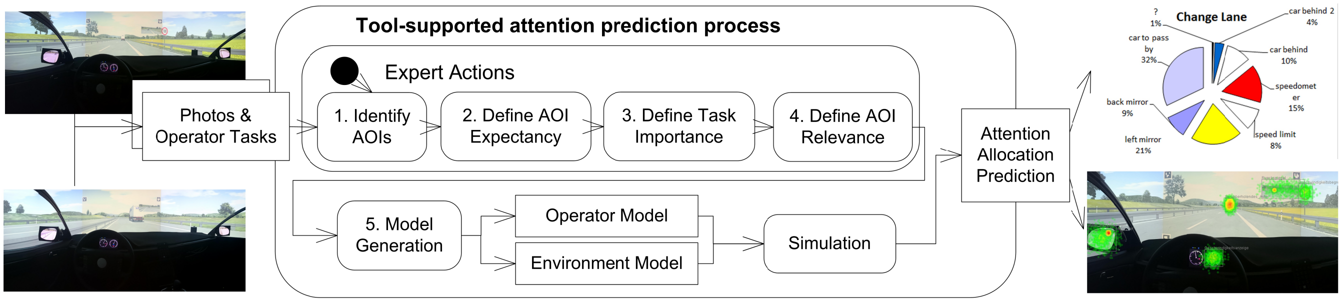

Figure 1. Activities to be performed manually (expert actions) or automatically for predicting attention allocations.

allowing programming on a higher level of abstraction. Whereas COGENT focuses on psychologists and extensive training, the HEE, CogTool and (to a lesser extent) Distract-R do not require any specific expertise to generate cognitive models and can therefore be used by non-experts in cognitive modeling as well.

The following section introduces the HEE and explains the parameter estimation steps incorporated in the tool for generating attention predictions.

III. THE HUMAN EFFICIENCY EVALUATOR (HEE)

For testing our hypotheses, we developed the Human Efficiency Evaluator (HEE), which we use to generate attention models based on the top-down parameters of the SEEV model. For the identification of expectancy and value parameters it implements the lowest ordinal heuristic [8], which is the most common applied process for determining the parameter inputs. The HEE has been carefully designed to be usable even by people without any knowledge in HF or attention prediction modeling. We eliminated any influence of non-reproducible teaching for using the HEE. Therefore we produced a 12 minutes long introduction video that was the only information source that all subjects in our studies had available to learn about the HEE. The video explained the tool based on evaluating monitoring behavior in a soccer game.

Fig. 1 depicts the entire modelling process. It starts by determining the model parameters (steps 1-4), thereafter the tool generates and executes the prediction model simulation, and finally the prediction results are visualized by the tool. The HEE requires photos of HMI monitoring situations that vary either in *design* (e.g. different cruise control assistant systems) or in the reflected point in time (which we call a "*situation*": e.g. an overtaking scenario separated into three consecutive situations: change lane, pass, return to own lane, together with a set of textually defined driver tasks (e.g. "respect speed limits", "overtake slower cars") as input.

The tool supported manual process steps produce input parameter data (expectancy and value for each identified AOI) that are feed into the automated model generation of the tool. Actually, two different types of models are generated: an environment model that defines the physical locations of the AOIs and an operator model that simulates a human's behavior of monitoring an HMI (moving the attention from one to another AOI based on the probability calculated by the SEEV model and based on the input parameter data). The result of a Monte-Carlo simulation of such a non-deterministic model is an attention allocation prediction that the HEE presents either as a heat map or by charts or tables stating the percentage dwell times for each AOI.

For the experiments we were also interested in aggregated predictions of a set of experts. For aggregating individual predictions we were required to extend the tooling because individual experts end up with different AOIs identified, labeled and positioned. We classified each AOI into AOI classes to enable model comparison across individual experts.

Finally, special cases like e.g. overlapping of different AOIs or different levels of details in experts' AOIs (i.e. an AOI marked by one experts is marked by another expert in a more detailed way using several smaller AOIs) needs to be considered. On these aspects we will elaborate later on in greater detail. Thereafter, aggregated predictions can be calculated.

The HEE guides its users through four steps: (1) the identification of AOIs, which are regions on an image of a monitoring situation that communicate a piece of information to the user; (2) the definition of the task importance of each user task, (3) the expectancy of an AOI that describes how often new information can be expected; and (4) the definition of the relevance of each AOI for each task. After having these data collected, the prediction model can be generated (5). The following subsections detail the techniques applied to collect these data from the experts.

A. *Identification of Areas of Interest (AOIs)*

For each design variant, which is in our case represented by a photo, the expert is asked to identify and name all sources of information and their corresponding physical location and dimension as precise as possible. This is done by sketching rectangular areas on each photo representing one design. Fig. 2 depicts a screenshot of the AOI identification step for an overtaking scenario with several AOIs already

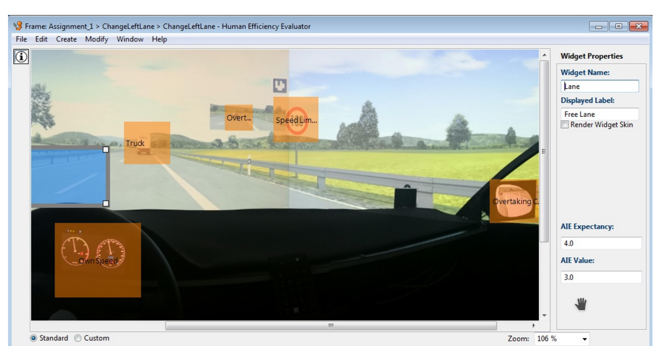

Figure 2. Identification of AOIs within the HEE. The user interface of CogTool [18] was re-used and extended for this step.

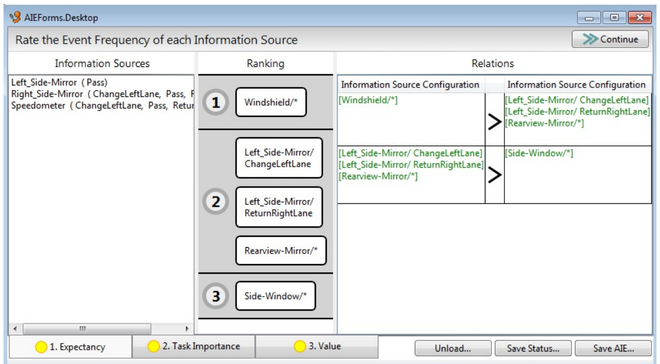

Figure 3. Expectancy definition form of the HEE.

identified, e.g. the right mirror, the left lane and the speedometer.

B. Expectancy Definition

Fig. 3 depicts the screen for defining the expectancy. The subsequent expert action steps are implemented by tabs in the bottom bar. The left side of the window alphabetically lists all identified AOIs together with the corresponding monitoring situation in which the AOI has been identified. The user needs to roughly order these AOIs by dragging them into the list in the middle: AOIs with an expectancy of providing frequently new information are ranked towards the top and those with fewer expected new events towards the bottom of the list. Not all AOIs need to be ranked and no total order needs to be identified. The tool automatically constructs a mathematical relation that reflects the created order by relational statements in the list on the right of Fig. 3. Further relations can manually be added for situations in that a complete order of AOIs cannot be identified (e.g. by defining AOI_a>AOI_b and AOI_a>AOI_c but being unsure about the relation between AOI_b and AOI_c).

C. Task Importance

Monitoring is usually done to collect information relevant for a set of tasks that a car driver has to perform. These tasks are specified at the beginning of the process (e.g. in our experiment: “Keep your car safely on track”, “Respect the speed limits”, and “Overtake slower vehicles”) and differ in their importance and therefore affect the monitoring behavior differently. The rating of the task importance is implemented similarly to the initial expectancy rating step: by asking the experts to rank the list of operator tasks based on their importance.

D. AOI Relevance

The value of each AOI depends on the relevance of an AOI for performing a task and the importance of the task itself [9]. The latter already has been identified by the preceding task importance definition step. The former is identified by filling out a relevance matrix, which is the last step of the guided process. A screenshot of such a relevance matrix form is depicted in Fig. 4. The matrix lists all AOIs as rows and all user tasks as columns. The experts are requested to identify each AOI for every user task either as “necessary” (=1), “helpful” (=0.5) or “not relevant” (=0).

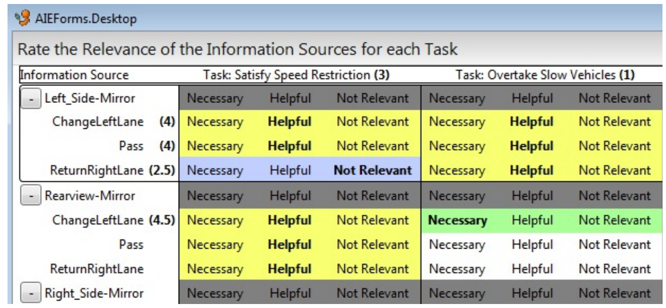

Figure 4. Relevance matrix form of the HEE.

E. Model Generation

From the expectancy relation and the task importance relation two partial ordered graphs are generated automatically. By ranking these graphs using the lowest ordinal algorithm [11] the tool calculates an expectancy coefficient for each AOI and an importance coefficient for each task. Task importance coefficients are multiplied with the relevance matrix as shown in [11] to obtain value coefficients for each AOI. These coefficients are the foundation for generating monitoring behavior models of humans’ attention shifts between different AOIs. It will constantly switch between the AOIs to move visual attention between them, based on the probability calculated with the expectancy and value coefficients [19].

The HEE uses the CASCaS (Cognitive Architecture for Safety Critical Task Simulation) architecture [11] to simulate human behavior. An integral part of CASCaS is an adapted SEEV model, which solely relies on the model’s top-down factors and assigns an expectancy coefficient u_g and a value coefficient v_g to each goal g of a cognitive model executed in CASCaS. A cognitive architecture like CASCaS can be understood as a generic interpreter that executes formalized procedures of a human operator in a psychological plausible way. An overview of cognitive computational models like ACT-R, SOAR, MIDAS and others is provided for instance in [19].

Monitoring involves detecting and reacting to events and is composed of a set of monitoring goals. To execute such a goal the human operator looks to the IS that can signal the event. Upon event detection the operator utilizes the perceived information to react to this event. If no event is detected, the operator’s attention shifts to another monitoring goal probabilistically based on the expectancy and value coefficients. The probability of switching to goal g among a set G of monitoring goals is defined as (cf. [20]):

$$P(g) = \frac{u_g}{\sum_{g_i \in G} u_{g_i}} \cdot \frac{v_g}{\sum_{g_i \in G} v_{g_i}} \qquad (2)$$

For monitoring simulations the HEE generates a probabilistic sequence of actions. Based on CASCaS and the integrated SEEV model, Monte Carlo simulations predict average percentage AOI dwell times and gaze frequencies, which can be visualized e.g. by heat maps.

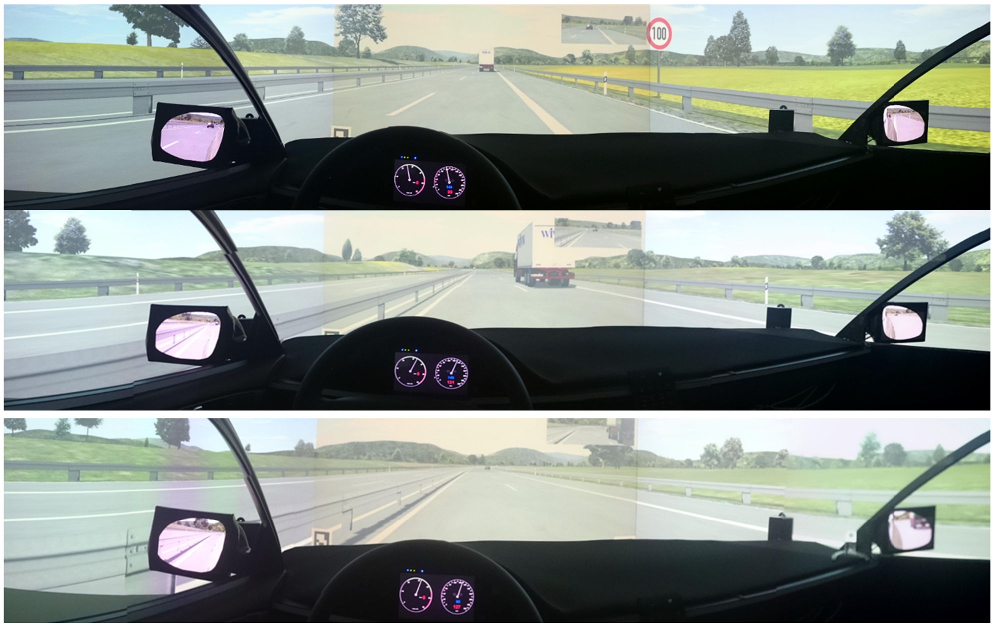

Figure 5. Photos used for modeling the overtaking: Left Merging (LM), Overtaking (OT) and Right Merge (RM)

IV. METHODOLOGY

To test our hypotheses we chose an automotive overtaking scenario on a two lane motorway. We created an HEE project that contains three consecutive situations of an overtaking maneuver. For each phase a representative image was selected (Fig. 5):

1. Moving to the left lane after approaching a slower vehicle (LM)
2. Overtaking the slower vehicle on the left lane (OT)
3. Moving back to the right lane in front of the slower vehicle (RM)

The experiment was performed with two different groups of subjects: On the one hand with experienced car drivers and on the other hand with HF experts in the automotive domain. None of all subjects had ever used the HEE or the SEEV model before.

A public announcement was made in the university to recruit 20 licensed car-drivers who were required to be licensed for at least 3 years, have a minimum driving experience of 2000 km per year and received an expense allowance of 10 EUR/h. The subjects were aged between 21-57 years (median: 23), were licensed between 4-39 years (median: 6), with a driving experience between 2000-40000 km per year (median: 5000). 11 woman and 9 men participated in the study, with the majority of them having a background in social sciences, and the minority (7) in natural sciences, mostly in chemistry and biology. None of them had a background in HF, psychology or computer science - one had a degree in neuroscience.

To compare the modeling quality of the experienced car drivers (ECD) (our non-Human Factors experts) with the modeling quality of Human Factors experts (HFE), we also let 9 HFEs from the automotive domain perform the exactly same experiment part of modeling the overtaking scenario (3 from a large automotive supplier, 3 from a national transportation research institute, 3 from an academic research institute working in human modelling). We required all HFEs to have in depth knowledge of car drivers (they all had experience in performing experiments with car drivers in car simulators for several years).

A. *Procedure*

The ECDs were invited on two different days: At one day they were asked to generate models of their own monitoring behavior for the overtaking scenario and at a second day they were invited for driving the overtaking scenario in a car driving simulator. The order of modeling and driving was randomized and balanced. Between both days there was a break of between 5-10 days. HFEs were invited on just one day to generate models of typical driver behavior. They did not drive the scenario in the driving simulator.

1) *Modelling*

The modeling was performed for the ECDs in a total of 5 and for the HFEs in 3 sessions (around 1 h per session) of 3-5 participants modeling in parallel, each one by its own on a separate computer. One participant of the ECD group failed to create a valid HEE model (just one AOI was identified) and was excluded. All subjects were only instructed by a 15 minute tutorial video[1]. For all subjects in both groups the total modeling time (after watching the video till the prediction model has been finalized) was under one hour.

2) *Car Driving Simulator*

On another day the ECDs drove in a fixed based driving simulator with a 170° field of view, running the SILAB simulation software. First, participants received a short simulator training. Afterwards they were instructed to drive for 20 minutes on a two lane motorway with low to medium traffic and keep a target speed of 130km/h and overtake slower vehicles as necessary. All in all 257 overtaking maneuvers were recorded in the simulator study. Subjects' gaze behavior was recorded using a Dikablis eyetracking system.

[1] http://lnk.multi-access.de/iv17 last checked 30/01/17

A. Data Processing and Analysis

The data processing chain that we implemented considers, among others, the following aspects:

1) Coordinate Transformation

In order to compare the HEE models with actual gaze movements, the coordinates of the AOIs marked using the HEE needs to be mapped to the eye-tracking data. With the Dikablis eye-tracking software AOIs can be defined relative to visual markers placed in the driving simulator. As the driving simulator is a 3D-setup, we placed markers on two layers: (1) within the vehicle cabin and (2) on the forward view projection screen. We calculated the positions of all markers on the 2D HEE background images and created corresponding transformation functions. Using the transformation functions, AOIs defined on the HEE background images can automatically be transformed to AOIs for the analysis software of the eye-tracker. The transformed AOIs, which are mainly marked within the vehicle cabin, were only defined relative to the markers on the cabin layer, while AOIs that are mainly defined on the forward view are only defined relative to the markers of the projection screen layer.

2) Measurement for the Separation of Driving Phases

In order to compare the attention predictions for the three driving phases with eyetracking data, the eyetracking data recorded during the driving simulator sessions needs to be separated into the three driving phases. A naïve approach in separating the driving phases would be to define e.g. a fixed time (e.g. time to collision) or distance to the vehicle driving ahead to separate phases. But such a measure needs to be chosen very carefully, as it might have an impact on the calculated attention allocation for each phase. The ideal measurement would be separating the phases based on the change of the driver's intention to merge left, overtake and merge back right. To not interfere with the driving task, we decided against asking the subjects of performing an additional task to communicate intention changes e.g. pressing a button.

Instead, we selected the driving phases based on the time of lane changes. The left merge phase (LM) starts 4 s before the vehicle center crosses the road marking and ends 0.5 s afterwards. In the same way starts the right merge phase (RM) 4 s before the vehicle returns to the left lane and ends 0.5 seconds after crossing the road marking. In between is the overtaking phase (OT). Thus the LM and RM phase are always 4.5 s long while the duration of the OT phase varies. These times were chosen based on our expertise. We conducted a sensitivity analysis later on to analyze the impact of the driving phase definition. We tested different timing definitions, using start times for the LM and RM phases ranging from 2 s to 6 s before crossing the road marking and end times ranging from 0 s to 1 s after crossing the road marking. None of the definitions had a great impact on the result.

TABLE I. THE 16 AOI CLASSES AGREED BETWEEN THE THREE ANALYSTS

	AOI class		AOI class
1	Left side mirror	10	Slower truck to overtake
4	Dashboard speedometer	11	Traffic ahead
5	Speed limit sign	12	Road directly ahead of ego vehicle
6	Rotation speed indicator	13	Windscreen
7	Distance to traffic ahead	14	Direction indicator
8	Left side window	15	Road condition
9	Right side window	16	Weather

3) AOI Classification

The tooling with the HEE allowed each expert to define an individual set of AOIs and label each of those based on individual preferences. For model comparison of different models a reliable AOI classification process is needed.

We clustered the AOI definitions into AOI classes by applying a two phase classification approach using three analysts. First, we jointly agreed on a set of 16 AOI classes (c.f. Table 1). We implemented an HEE function that generates an excel table with the class names as columns and for each AOI (ECD: 304 / HFE: 200) a row in a randomized order together with one picture for each AOI that showed the location on the photo together with the given label. Second, all three analysts independently classified the AOIs by walking through all of the images and identifying the most appropriate class for each AOI in the excel sheet. We ended up with a high inter-rater concordance (Fleiss kappa: ECD: $\kappa=0.904$ / HFE: $\kappa=0.881$).

4) PDT Calculation

The percentage dwell time (PDT) for an AOI was calculated as the percentage of time the eye tracker identified that a gaze was targeted at an AOI. In some regions AOIs did overlap. If the gaze was located at a region of n overlapping AOIs for some time, the gaze was equally distributed to the AOIs, i.e. the PDT calculation for each of the n AOIs was done as if the subject was looking to each of the AOIs for just 1/n-th of the time duration.

V. RESULTS

R1:"Individual, experienced car drivers are bad in identifying the regions that they are looking at during an overtaking maneuver and also in modelling how they divide attention between these regions."

ECDs are bad in predicting their own monitoring behavior. We took the attention model of each ECD and compared it to the corresponding own eye-tracking data of the same subject. We used the eye-tracking data to calculate the PDT for each AOI defined in the ECD's attention model. Most of the time the ECDs' gazes were not located at the AOIs they defined (75.5% of the time, SD=24.3%). Furthermore the distribution of the time, were the gazes actually were located on any AOI did only weakly to moderately correlate with the PDT predictions of their models. Subjects achieved an average correlation coefficient of R=0.318 with high individual variance (SD=0.292).

R2:"Aggregating the models of multiple experienced car drivers improves attention prediction results compared to an average individual experienced car driver."

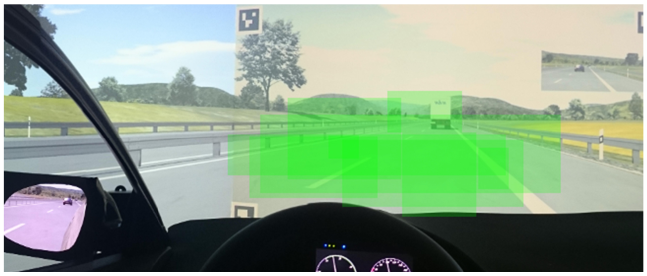

Figure 6. AOIs of class “road directly ahead” of phase “lane change” of all 19 participants

The subjects’ AOI aggregation is done by overlaying all AOIs (thus creating their union respectively their entire outer frame) for those AOIs with clear boundaries in the same AOI class (e.g. all marked left side mirrors) and by combining those without (e.g. the windscreen) also by using the outer frame of all those AOIs. Fig. 6 depicts one such exemplary AOI aggregation for all AOIs identified in the windscreen. The predicted PDT of aggregated AOIs is then calculated by the mean predicted PDT of all AOIs inside the aggregate AOI borders of all subjects.

For such an aggregated model we end up in a high correlation with the observed eye tracking data for which we also calculated the mean observed PDT for each AOI aggregation (R=0.719) while at the same time the amount of time, where the gaze could not be assigned to an AOI was reduced to 18.9%.

R3:”Aggregating multiple experienced car drivers AOIs and averaging their individual corresponding PDTs ends up in better attention prediction results than those that can be gained by an average individual human factors experts.”

The average correlation coefficient of the individual HFE experts was only 0.394 (SD=0.259) which is worse than predictions of the combined ECD group. Surprisingly, their variance was high as well, which seems to indicate that their conformance was lower than we expected based on their expertise.

R4:”Aggregating multiple human factors AOIs and averaging their corresponding PDTs results in a very high correlation compared to the average measured monitoring behavior.”

Combining the models of the HFEs in the same way as it was done for the ECDs results in a very high correlation coefficient of R=0.967. This shows the impact that averaging can have on independently created models with a high variance between the averaged individual models.

VI. DISCUSSION

For the individual predictions of both groups we observed that the majority of gazes (>75% of the time) were not located at the AOIs that the individual experts had defined. In our understanding this is because of two reasons, which are hard to separate based on the data we have recorded:

First, subjects might actually pay attention to information sources that they had not considered in their models. Second, the direct comparison between model prediction and eye tracking data is difficult, because the AOIs for the prediction models can be defined on a much more detailed level than they could be defined to be measured in eye tracking studies with current technology. There, AOIs often are defined much bigger, because of: (1) the noise in the eye tracking measurement, (2) the fact that some information can be perceived in the near periphery, e.g. a brake light flashing up, and (3) not all AOIs are actually static but slightly change in position and size, e.g. a vehicle ahead.

For our study we applied the multiplicative variant of the SEEV model, which connects expectancy and value coefficients multiplicatively. This variant should predict optimal monitoring behavior. However prior research does not agree about, whether humans always monitor in an optimal way. In earlier studies it was shown that often an additive variant with no interaction between expectancy and value yields better predictions [2]. We again conducted a sensitivity analysis with respect to the model formulation. Using the additive variant never changed any of the average correlations by more than 0.097. The combined predictions using the additive SEEV variant is R=0.699 for the ECD group and R=0.944 for the HFE group.

The standard error of averaged values that our prediction approach depends on can be directly controlled by the number of subjects, which is of benefit in safety-critical applications where e.g. missing information can affect a person’s life and therefore predictions can are required to be associated with a certain level of certainty.

With 16 (grouped condition:7) AOI classes covered by the attention prediction based on the best of our knowledge our study is one of the most detailed attention predictions reported so far. Prior research reported about attention predictions of between 3-6 AOIs.

While analyzing the results, we got aware that the good prediction quality of the group of experienced card drivers seems to be in line with the Diversity Prediction Theorem [21], which basically states, that for a given group of predictive models (e.g. the ECD models we collected) the average squared error is equal to the average individual error minus the variance of individual signals. Under the assumption that the models were created independently [22] and ended up in reasonable models, large prediction diversity reduces the collective error.

For the HF expert group the collective prediction error of a group of nine is very small and almost perfectly correlates with the measured attention prediction. Additional the prediction regions for that predictions were made, covered over 81% of the total gazes recorded. On a first glance this high correlation was a huge surprise. But those correlations also seem to be in line with the Diversity Prediction Theorem as we also observed a high diversity in the HF experts data collected via the tool.

Even it might be unrealistic to always have 9 HF experts with a high domain expertise or 20 experts recruited, but we consider this results as very promising for several reasons: (1) the HEE does not need onsite training and therefore can be applied remotely; (2) the effort in time is very low (like around one hour in our studies per expert for the first time usage of the tool); (3) if upcoming studies can confirm these high-correlations, at least HF experts that are following our tool-supported process might be able to completely substitute eye-tracking studies to observe attention allocations. This will bring predictive models of human performance into application, so that they can to guide and constrain user interface designs, which was expensive in time and expertise before and can play their benefits over eye-tracking: they can already be applied in an early design phase where no prototypes are available and also can be inspected to give insights about the underlying causes for the predicted attention allocation.

VII. CONCLUSION

With the help of the Human Efficiency Evaluator, that eases modelling monitoring behavior, for the first time non-experts in human factors can generate valid attention prediction models. Prior research has shown that individual human factors experts are able to predict attention allocations based on the SEEV model for several different application domains. But what kind of expertise was required for generating valid models remained unclear.

We contributed to this field of research with an approach that combines model-based attention prediction and empirical methods in a structured process. A group of 19 experienced car drivers ended up in a prediction model that accounted for 52% of the variance of the average measured attention allocation of three phases of an overtaking maneuver of 19 car drivers. A group of 9 HF automotive was able to generate a prediction model that accounted for 94% of the variance measured. The high correlations are in the line with the Diversity Prediction Theorem since data collected by the tool had a high diversity.

Our approach makes several advances: (1) there is no prerequisite knowledge required (apart from watching a 15 minutes video tutorial); (2) The *complete* prediction generation process including the input parameter generation is reproducible for the first time, since it is an empirical approach eliminating the need of individual HF expertise or knowledge; (3) It is very time efficient, since data collection just requires the software, and can be performed in parallel and remotely, which is especially relevant for niche domains where experts are rare and spread around the world. The results also showed that the variance of predictions made by human factors experts is quite high, which indicates that obtaining valid attention predictions from individual experts depends essentially on their individual expertise.

REFERENCES

[1] C. D. Wickens, J. S. McCarley, A. L. Alexander, L. C. Thomas, M. Ambinder, and S. Zheng, *Attention-Situation Awareness (A-SA) Model of Pilot Error*. CRC Press/Taylor & Francis Group, 2008, ch. 9, pp. 213–239.

[2] B. Wortelen, A. Lüdtke, and M. Baumann, "Simulating attention distribution of a driver model: How to relate expectancy and task value?" in *Proceedings of ICCM 2013 - 12th International Conference on Cognitive Modeling*, R. West and T. Stewart, Eds., Ottawa, Canada, 11-14 July 2013, pp. 269–274.

[3] J. S. Ha and P. H. Seong, "Experimental investigation between attentional-resource effectiveness and perception and diagnosis in nuclear power plants," *Nuclear Engineering and Design*, vol. 278, pp. 758–772, 2014.

[4] C. D. Wickens, J. Helleberg, J. Goh, X. Xu, and W. J. Horrey, "Pilot task management: Testing an attentional expected value model of visual scanning," NASA Ames Research Center Moffett Field, CA, Tech. Rep., 2001.

[5] J. S. McCarley, C. D. Wickens, J. Goh, and W. J. Horrey, "A computational model of attention/situation awareness," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 17. SAGE Publications, 2002, pp. 1669–1673.

[6] K. M. Corker and B. R. Smith, "An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation automation," in *Proceedings of the AIAA Computing in Aerospace 9 Conference*, 1993, pp. 1079–1088.

[7] B. F. Gore, B. L. Hooey, C. D. Wickens, and S. Scott-Nash, "A computational implementation of a human attention guiding mechanism in MIDAS v5," in *Digital Human Modeling*, ser. Lecture Notes in Computer Science, V. G. Duffy, Ed. Springer, Berlin, 2009, vol. 5620/2009, pp. 237–246.

[8] C. Wickens, J. McCarley, and L. Thomas, "Attention-situation awareness (A-SA) model," in *NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*, 2003, p. 189.

[9] A. J. Bos, D. Ruscio, N. D. Cassavaugh, J. Lach, P. Gunaratne, and R. W. Backs, "Comparison of novice and experienced drivers using the SEEV model to predict attention allocation at intersections during simulated driving," in *Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, June 2015.

[10] B. Wortelen, M. Baumann, and A. Lüdtke, "Dynamic simulation and prediction of drivers' attention distribution," *Transportation research part F: traffic psychology and behaviour*, vol. 21, pp. 278–294, 2013.

[11] W. J. Horrey, C. D. Wickens, and K. P. Consalus, "Modeling drivers' visual attention allocation while interacting with in-vehicle technologies." *Journal of Experimental Psychology: Applied*, vol. 12, no. 2, pp. 67–78, 2006.

[12] F. D. H. B. Goodman, A. and J. Wilson, "Characterizing visual performance during approach and landing with and without a synthetic vision display: a part task study." in *Proceedings of the 2003 Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*, A. G. . B. H. D.C. Foyle, Ed., no. NASA/CP-2003-212267, Moffett Field, CA: NASA, 2003, pp. 71–89.

[13] R. Y. I. Koh, T. Park, C. D. Wickens, L. T. Ong, and S. N. Chia, "Differences in attentional strategies by novice and experienced operating theatre scrub nurses," *Journal of Experimental Psychology: Applied*, vol. 17, no. 3, pp. 233–246, 2011.

[14] C. D. Wickens, J. McCarley, L. A. Amy, L. Thomas, M. Ambinder, and S. Zheng, "Attention-situation awareness (A-SA) model of pilot error," University of Illinois at Urbana-Champaign 1 Airport-Road Savoy, Illinois 61874, Technical Report AHFD-04-15/NASA-04-5, January 2005, nASA Ames Research Center, Mosfett Field, CA.

[15] C. D. Wickens, J. Goh, J. Helleberg, W. J. Horrey, and D. A. Talleur, "Attentional models of multitask pilot performance using advanced display technology," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 3, pp. 360–380, 2003.

[16] D. D. Salvucci, "Rapid prototyping and evaluation of in-vehicle interfaces," *ACM Transactions on Computer-Human Interaction*, vol. 16, no. 2, pp. 9:1–9:33, Juni 2009.

[17] R. Cooper and J. Fox, "EnglishCogent: A visual design environment for cognitive modeling," *EnglishBehavior Research Methods, Instruments, & Computers*, vol. 30, no. 4, pp. 553–564, 1998. [Online]. Available: http://dx.doi.org/10.3758/BF03209472

[18] B. N. Harris, B. E. John, and J. Brezin, "Human performance modeling for all: Importing UI prototypes into cogtool," in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '10. New York, NY, USA: ACM, 2010, pp. 3481–3486. [Online]. Available: http://doi.acm.org/10.1145/1753846.1754005

[19] C. Forsythe, M. L. Bernard, and T. E. Goldsmith, Eds., *Cognitive Systems: Human Cognitive Models in Systems Design*. Psychology Press, 2006.

[20] B. Wortelen, "Das Adaptive-Information-Expectancy-Modell zur Aufmerksamkeitssimulation eines kognitiven Fahrermodells," Ph.D. dissertation, Universität Oldenburg - Fakultät II Informatik, Wirtschafts- und Rechtswissenschaften Department für Informatik, 2014.

[21] S. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton, Ed. Princeton University Press, 2007.

[22] J. Surowiecki, *The Wisdom of Crowds*. Doubleday, 2004.