Using Driver Control Models to Understand and Evaluate Behavioral Validity of Driving Simulators

Gustav Markkula , Richard Romano , A. Hamish Jamson, Luigi Pariota , Alex Bean , and Erwin R. Boer

Abstract—For a driving simulator to be a valid tool for research, vehicle development, or driver training, it is crucial that it elicits similar driver behavior as the corresponding real vehicle. To assess such behavioral validity, the use of quantitative driver models has been suggested but not previously reported. Here, a task-general conceptual driver model is proposed, along with a taxonomy defining levels of behavioral validity. Based on these theoretical concepts, it is argued that driver models without explicit representations of sensory or neuromuscular dynamics should be sufficient for a model-based assessment of driving simulators in most contexts. As a task-specific example, two parsimonious driver steering models of this nature are developed and tested on a dataset of real and simulated driving in near-limit, low-friction circumstances, indicating a clear preference of one model over the other. By means of closed-loop simulations, it is demonstrated that the parameters of this preferred model can generally be accurately estimated from unperturbed driver steering data, using a simple, open-loop fitting method, as long as the vehicle positioning data are reliable. Some recurring patterns between the two studied tasks are noted in how the model's parameters, fitted to human steering, are affected by the presence or absence of steering torques and motion cues in the simulator.

Index Terms—Human performance modeling, simulator validation.

I. INTRODUCTION

DRIVING simulators are widely used for various purposes in driver training, traffic research, and automotive development [1]. When using a driving simulator, one does so to target some driving-related objectives (e.g., driver learning out-comes, traffic research questions, and vehicle design decisions) that involve the human driver as a crucial component, but where the use of a real vehicle for attaining the objectives is deemed unsafe, too costly, or otherwise ineffective. The objectives themselves, however, will typically remain focused on real driving, i.e., the use of the simulator is motivated by the assumption that the results will validly transfer to real vehicles and real traffic.

It is, however, clear that driving in a simulator will never be exactly the same as driving in a real vehicle, both in the sense that the exact sensory stimuli will differ—imperfect *physical validity* [2]—and in the sense that drivers are aware that they are not in a real vehicle. Consequently, the validation of driving simulators for various types of applications has become a field of research in its own right, attempting to answer questions such as "Does training in a simulator make drivers better at handling real traffic?" [3], "To what extent do research findings from a simulator say something about real driving?" [4], or "Will the same vehicle design decision be made in the simulator as if using a physical prototype?" [5].

Across all of these contexts, the following question recurs: *Do drivers behave the same way in the simulator as they do in a real vehicle?* If one can devise a convincing, quantitative method for answering this question in the context of a given application, one can evaluate the extent to which a simulator under scrutiny provides *behavioral validity* for that application, supporting decisions on how advanced or expensive a simulator one needs for a given application [2], [5]. Existing methods for assessing behavioral validity conventionally rely on the calculation of various metrics describing the observed driver behavior and comparisons of these between reality and the simulator. However, such analyses do not relate the observed human behavior to the driving situation that triggered it. It has been proposed that a better understanding of behavioral validity can be reached by fitting *driver models* to the observed behavior and by analyzing obtained model parameters [5], [6].

This type of evaluation has been pioneered in the flight simulation domain by Zaal, Pool *et al.* [7], [8]. They proposed the use of quasi-linear models of pilot control behavior, combined with models of the involved perceptual and neuromuscular systems. To allow the fitting of these relatively complex models to human pilot data, they adopted simplified, one-dimensional flying tasks with the visual input constrained to just a roll/pitch-tracking display in the cockpit, and a system identification approach. Pilots were subjected to forcing function inputs, the sum of a low-frequency target signal and high-frequency perturbations, in the same way in both a simulator and in a real aircraft, equipped with

G. Markkula, R. Romano, and A. H. Jamson are with the Institute of Transport Studies, University of Leeds, Leeds LS2 9JT, U.K. (e-mail: g.markkula@leeds.ac.uk; r.romano@leeds.ac.uk; a.h.jamson@leeds.ac.uk).

L. Pariota is with the Institute of Transport Studies, University of Leeds, Leeds LS2 9JT, U.K. and also with the University of Naples "Federico II", 80138 Naples, Italy (e-mail: luigi.pariota@unina.it).

A. Bean is with Jaguar Land Rover Ltd., Coventry CV3 4LF, U.K. (e-mail: abean@jaguarlandrover.com).

E. R. Boer is with the Institute of Transport Studies, University of Leeds, Leeds LS2 9JT, U.K. and also with the Entropy Control Inc., La Jolla, CA 92037, USA (e-mail: erwinboer@entropycontrol.com).

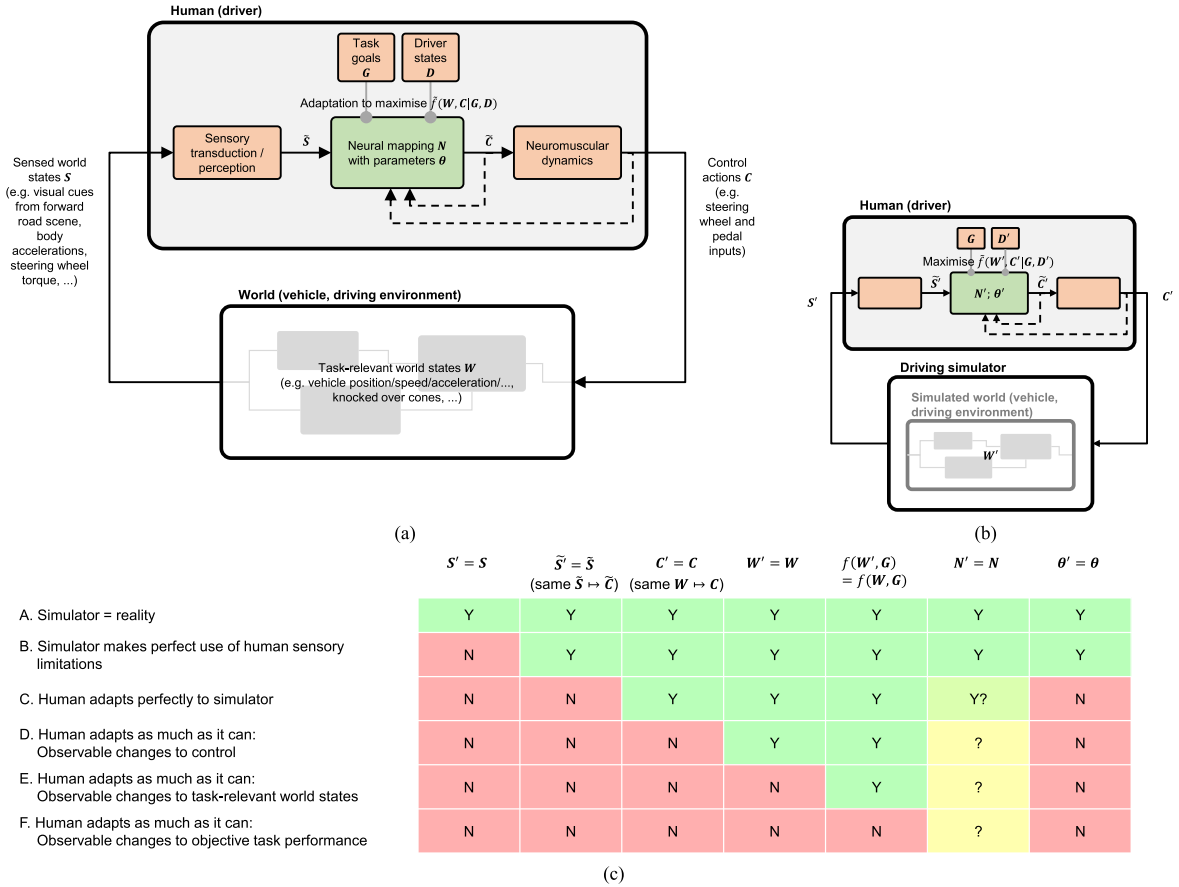|	$S' = S$	$\widetilde{S}' = \bar{S}$ (same $\bar{S} \mapsto \widetilde{C}$)	$C' = C$ (same $W \mapsto C$)	$W' = W$	$f(W', G)$ $= f(W, G)$	$N' = N$	$\theta' = \theta$
A. Simulator = reality	Y	Y	Y	Y	Y	Y	Y
B. Simulator makes perfect use of human sensory limitations	N	Y	Y	Y	Y	Y	Y
C. Human adapts perfectly to simulator	N	N	Y	Y	Y	Y?	N
D. Human adapts as much as it can: Observable changes to control	N	N	N	Y	Y	?	N
E. Human adapts as much as it can: Observable changes to task-relevant world states	N	N	N	N	Y	?	N
F. Human adapts as much as it can: Observable changes to objective task performance	N	N	N	N	N	?	N

(c)

Fig. 1. Conceptual model and taxonomy of behavioral validity in the driving simulator. The first two panels suggest a model of driver–vehicle interaction in (a) a real vehicle and (b) a driving simulator. For a given driving task, the sensory transduction (/perception), task goals, and neuromuscular dynamics are assumed to be constant between the vehicle and simulator, whereas the brain's mapping between internal sensory and motor representations is assumed to be adaptable to keep the internal performance function $\tilde{f}$ maximal (the cost function $-\tilde{f}$ minimal), despite the changes to the task imposed by the driving simulator. Panel (c) provides a taxonomy of the different levels of behavioral validity, based on what parts of the driver–vehicle interaction remain intact in the simulator. Note that the externally defined task performance function f may be related to, but not necessarily identical, to the internal function $\tilde{f}$. "Observable changes" refer to changes that meet some criteria on statistical significance or effect size.

fly-by-wire technology that allowed this type of experimental control.

Damveld *et al.* [6] presented a tentative application of a similar forcing function approach to car steering on curves in a simulator but did not fit any models to the obtained data. They also reported that the perturbations seemed to disturb the driver's normal steering behavior. If this is so, then comparing the behavior in a perturbed version of a driving task between the simulator and reality may not give insights into behavioral validity for the unperturbed task (even disregarding the challenge of achieving perturbations in a real vehicle).

Based on what has been said previously, the two primary aims of the present paper are:

1) to analyze the concept of behavioral validity in some detail, to provide a taxonomy of behavioral validity and a theoretical motivation for why one might want to use driver models to assess driving simulators and what kinds of models are to be used. This is covered in Section II. This conceptual analysis is intended to be general across driving tasks and simulator applications;

2) to explore the feasibility of a model-based simulator assessment by means of an open-loop fitting of steering

models to unperturbed human task performance, i.e., without a forcing function approach. Necessarily, this part of the paper is application-specific; we target the assessment of simulators as an industrial tool for the near-limit, low-friction stability testing of prototype vehicles, a context where a close replication of the human control behavior is particularly relevant but where the necessity of using professional test drivers implies small sample sizes, such that the emphasis here will remain on methodological aspects. The collection of data in a real vehicle and in a number of different simulator configurations is described in Section III, the developed driver models in Section IV, a comparison between them in Section V, and their use in simulator assessment in Section VI.

A general discussion and conclusions are provided in Sections VII and VIII, respectively.

II. Behavioral Validity in the Driving Simulator

A. Conceptual Driver Model

As a basis for the reasoning about behavioral validity, the qualitative model of driver behavior shown in Fig. 1(a) will be

used. This model is intended to be noncontroversial, is compatible with the contemporary literature on driver models (e.g., see [9]–[17]), and summarizes this literature in the form of a few high-level assumptions that most can hopefully accept. The model suggests that for a given state W of the externally observable physical world, the driver's sensory and perceptual systems process S, a (possibly transformed) subset of W, to yield internal representations $\tilde{S}$. For example, while navigating a cone track, the relative positions and velocities of the car and cones (W) might be picked up by the driver as the angular positions and velocities of the upcoming cones in the driver's own field of view (S), which the perceptual system estimates with noise, delays, and possible distortions ($\tilde{S}$). These internal representations are then acted upon in the brain by some mapping $N : \tilde{S} \mapsto \tilde{C}$, with parameters $\boldsymbol{\theta}$, and where $\tilde{C}$ is an internal representation of the control action to be carried out (e.g., a motor command), which is then processed by the motor system and muscles to yield an externally observable control behavior C. Importantly, the mapping N is adaptively selected and tuned by the driver's brain in attempts to maximize some function $\tilde{f}(W, C|G, D)$, rating the success of the behavior, where G denotes the goals of the task at hand and D denotes driver states, such as for example fatigue or stress, which may also affect $\tilde{f}$. This function will clearly be related to externally definable performance measures $f(W|G)$ (e.g., concerning task completion time, knocked-over cones) but might not prioritize the physical outcomes in exactly the same way as an external observer; $\tilde{f}$ is, for example, likely to also consider control effort [10], [18].

Please note that although the conceptual model has here been described in terms of driver control behavior, this can also be interpreted in a rather general sense, depending on the application; the control can, for example, be of a more tactical or strategic nature while monitoring a partially automated vehicle [19] or the control of an in-vehicle human-machine interface (HMI) to carry out some secondary task [4].

B. Taxonomy of Behavioral Validity

Fig. 1(b) shows the same driver model again but now in a driving simulator, with a prime added to all symbols to make the distinction. As mentioned earlier, even if the simulated world states are identical to some real-world situation, i.e., $W' = W$, the sensed world states S' will typically not be exactly the same as S due to the simulator's limited ability to reproduce the sensory (visual, vestibular, haptic, etc.) cues, and the driver states D' may differ, not least in terms of the factual knowledge of being in a simulated vehicle. These differences, small or large, will cascade through the entire control loop in nontrivial ways, becoming amplified or attenuated, leading to more or less changed (or unchanged) perceptual representations $\tilde{S}'$, control action representations $\tilde{C}'$, and overt control C', resulting in updated world states W', which also may or may not depart from what would have occurred in the real vehicle.

Fig. 1(c) depicts some different scenarios of how these deviations might occur along the control loop. Scenario A is a perfect simulator, without any distortion of sensed world states, i.e., achieving $S' = S$. Scenario B is one where $S' \neq S$ but, nevertheless, $\tilde{S}' = \tilde{S}$, i.e., a simulator that does not produce

exactly the same sensed world states (e.g., body accelerations) as in reality but that makes the perfect use of human sensory limitations, such that there would be no way for the driver's brain to tell the difference from S' alone. While Scenarios A and B may be achievable for some constrained tasks, such as the flight tasks referenced earlier [7], [8] or constant speed lane-keeping on a straight road, they are unrealistic for most driving tasks.

In Scenario C, the limitations of the simulator are such that the internal sensory representations have also started diverging, i.e., $\tilde{S}' \neq \tilde{S}$. However, for small divergences, the driver should be able to adapt either just the mapping parameters $\boldsymbol{\theta}$ or possibly also the mapping N itself in such a way that overt behavior stays the same, still maximizing the function $\tilde{f}(W, C|G, D)$ in the same way as in Scenarios A or B (assuming that any differences $D' = D$ at hand do not have a major impact on $\tilde{f}$). In other words, in Scenario C, the mapping $W \mapsto C$ is preserved and, therefore, also all other externally observable aspects of behavior (overt control actions, world states, and objective task performance). Please note that the preservation of $W \mapsto C$ implies the preservation of $S \mapsto C$, since the latter is only a reformulation of the external mapping, operating on more psychologically plausible sensory inputs.

Scenario D occurs if the simulator's limitations make it impossible for the driver to completely succeed in adapting the internal mapping $\tilde{S}' \mapsto \tilde{C}$ (e.g., scaled-down motion cues might fall under perceptual thresholds [20], [21]), such that there are observable changes to the overt control C for a given W, i.e., changes to the external mapping $W \mapsto C$ but without this, yet, causing any observable changes to the external world states or the objective task performance. A typical example might be an increased steering effort that still achieves the same vehicle trajectory [20]. Note that the terms "observable changes" and "the same" are imprecise here. In any practical application, other sources of variability will also be present, meaning that W will never be exactly the same between any two repetitions of a task, and there will be a need for some statistical-level operationalization of what constitutes an acceptably small difference and what does not for the application at hand. We will return to this matter in Section VII. Also, note that in Scenario D, it seems very likely that the neural mapping parameters have changed, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$, but it is not clear whether or not the mapping N itself has changed or if it remains the same.

Scenarios E and F occur as simulator limitations get even more severe, such that the world states begin to observably deviate from reality (Scenario E) and possibly even the objective task performance (Scenario F).

C. Implications for Model-Based Simulator Evaluation

The abovementioned argument suggests that the task of behavioral validity assessment can be regarded as one of distinguishing between the four different feasible driving simulator Scenarios C through F, by answering three questions[1]: Is task performance f (e.g., cone hit frequency) preserved in the simulator?, are the observable world states W (e.g., observed vehicle

---

[1]We have previously referred to these as the *utility triplet* [5] but without the underlying conceptual theory.

trajectories) preserved?, and are the control mappings $W \mapsto C$ preserved?. The first two of these can be addressed with conventional metrics, but for the third question, driver models are needed.

It should be noted that in the flight simulation work referenced earlier [7], [8], the pilot models included an explicit representation of the internal $\tilde{S} \mapsto \tilde{C}$ mapping, as part of the overall $W \mapsto C$ model. This approach makes sense if Scenario B is considered attainable, and it is important to distinguish it from Scenario C, something that may be true in a context where a pilot or driver is to overlearn the exact parameters θ for vehicle control. The extent to which this is important to, for example, driver training is an open question, to which the present framework can hopefully contribute.

As mentioned previously, with the present simulator technology, Scenario B seems unattainable for most driving tasks. If so, the best one can hope for and, therefore, all that needs to be tested is the preservation of the external mapping $W \mapsto C$. This might sound like a severe limitation on the usefulness of driving simulators, but that need not be the case, in part precisely because of the human sensorimotor system's abilities to learn the general aspects of a task's dynamics, with associated quick adaptation to updated dynamics [18]. Furthermore, if $W \mapsto C$ is preserved in a given task, the driver will take the vehicle through the same objective states in reality and in the simulator. For many applications, for example, relating to vehicle testing, it may matter less whether or not drivers have had to adapt their internal mappings somewhat to achieve this consistent performance.

Another point to consider is the actual parameterization of models. As mentioned in Section I, we hope to meaningfully fit models directly to actual, unperturbed driving tasks. Therefore, besides limiting ourselves to models of the $W \mapsto C$ type (since this should be enough to distinguish between Scenarios C and D), we also focus on rather simple models with limited number of parameters.

III. DATA COLLECTION

A. Driving Environments

1) Test Track and Instrumented Vehicle: Real driving data were collected in early 2015 on a test track in the northern parts of Sweden, using an instrumented Jaguar XE prototype. The driving surface was packed, graded snow (and in some cases, polished ice also). By means of deceleration tests, the friction between tyres and snow was estimated as $\mu \approx 0.4$ and that between tyres and ice as $\mu \approx 0.2$. Driver control inputs and vehicle movements were recorded via the vehicle's control area network, an inertial measurement unit, and a differential GPS (DGPS).

2) High-Fidelity Driving Simulator: Simulated driving data were collected during late 2015–early 2016 from the University of Leeds Driving Simulator (UoLDS). The UoLDS features a complete cockpit of a Jaguar S-type vehicle inside a spherical dome with 300° visual projection, mounted on an eight-degrees-of-freedom motion system consisting of a hexapod on an *XY* table providing ±5 m of translation in both longitudinal and lateral directions. The "classical" motion cueing algorithm was used; for further details about this simulator and algorithm, see [22], and see the supplemental material to this paper for the exact motion cueing parameters used. The vehicle dynamics simulation was a Jaguar-developed multibody model of the XE vehicle. This model has been extensively validated to closely capture accurate vehicle behavior on high-friction surfaces. A visual representation of the frozen lake environment was created, and the snow surface was modeled as having normally distributed random variations in height (standard deviation: 1 mm) and friction (standard deviations: 0.02 and 0.005 for snow and ice, respectively) on a linearly interpolated square grid of side 0.5 m.

Besides this standard configuration, two additional simulator configurations were also tested; one with the steering torque feedback to the driver turned off and one with the simulator motion system turned off. The purpose of these rather coarse simulator manipulations was for them to provide clear-cut examples of driving with degraded perceptual cues to support methodological development, including the presently reported model-based methodology, in the preparation for later studies investigating more subtle variations in simulator capability.

B. Tasks

On the test track, data were collected for eight different tasks. After implementing and piloting all of these tasks in the simulator, three tasks were identified as especially relevant for simulator-based testing of vehicle stability and were, therefore, included in the simulator data collection. One of these three tasks, a constant radius circular curve task, will not be considered here due to space limitations. The other two tasks studied in the simulator are described in the following and illustrated in Fig. 3.

1) Lane Change: In the lane change task, drivers were instructed to approach a first cone gate at about 45 km/h and then to make a 12-m wide lane change, of which 6 m in the middle was a lane of polished ice, to pass through another two cone gates at 30-m and 50-m longitudinal distance from the entry gate. If the drivers were not able to successfully complete the maneuver, they were free to reduce the entry speed in subsequent repetitions.

2) Slalom: In the slalom task, the drivers were instructed to maintain a constant speed of 45 km/h through a slalom of eight cones spaced by 25 m.

C. Drivers

Eight drivers took part in this study, all professional test drivers employed by Jaguar Land Rover. Their prior experience in low-friction winter testing, before this visit to Sweden, ranged from one season (two drivers) to two seasons (two drivers) to 15–30 seasons (four drivers).

D. Procedure

The drivers were first briefed, on both the test track and in the simulator, on the tasks that they would be carrying out, and
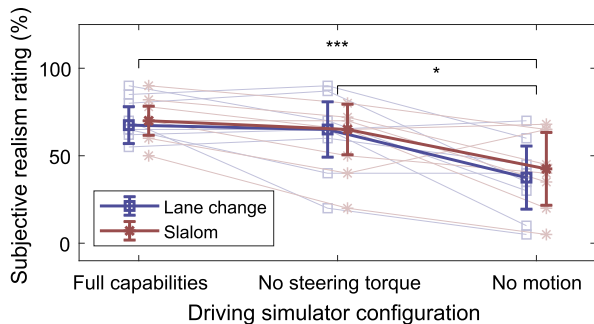
Fig. 2. Driver responses, on a visual analog scale from 0 to 100, to the question: "For the task you just drove, how similar would you say that the experience in the simulator is to reality?". Bold lines show medians with 95% confidence intervals and lighter traces show individual responses. The effect of the simulator configuration was statistically significant (Friedman $\chi^2(2) = 15.79; p < 0.001$), with significance levels for Bonferroni-corrected post hoc tests also indicated in the figure (all three pairwise comparisons tested): * $p < 0.05$; *** $p < 0.001$.

they provided informed consent. On the test track, the drivers familiarized themselves with the instrumented vehicle during a short 1.5-km drive to the first task, whereas in the simulator they were given a more substantial 10-min familiarization drive on a rural road to also get acquainted with the experience of simulated driving as such. On the test track, the per-driver order of the eight driving tasks followed a Latin Square design, whereas in the simulator, per-driver random permutations were drawn for the order of the nine combinations of three tasks and three simulator configurations. On the test track, the lane change task was repeated four times consecutively, whereas due to test track time constraints the slalom was repeated three times, i.e., with eight drivers, $8 \times (3 + 4) = 56$ recordings. In the simulator, each combination of task and simulator configuration was repeated four times consecutively, i.e., $8 \times 2 \times 3 \times 4 = 192$ recordings considered here, and after each task/simulator combination, the drivers also provided subjective feedback on the simulated driving. The drivers' ratings of perceived simulator realism are shown in Fig. 2, depicting a statistically significant decrease in perceived realism when removing motion cues but not when removing steering torque cues.

IV. DRIVER MODELS

Based on the reasoning in Section II-C, two alternative few-parameter models, mapping directly from world state W to control actions C, were investigated.

There is converging support for modeling driver steering not on the level of steering angles but rather on *steering rates* [11]–[15]. Furthermore, there is support for modeling steering rates as delayed, linearly scaled versions of the *yaw rate error* $\omega_{\text{err}} = \omega - \omega^*$, i.e., the deviation between the current vehicle yaw rate ω and currently desired yaw rate ω^* [13], [14], [16]. The two models investigated here were both variations of the simplest possible steering control law of this nature

$$\dot{\delta}(t) = -K \cdot \omega_{\text{err}}(t - T_R) \qquad (1)$$

where δ is the steering wheel angle, K is a gain constant, and T_R is a response delay. Please note that we do not suggest that drivers necessarily perceive or mentally represent actual

and desired yaw rates; the model is equally compatible with the idea that available sensory cues and behavioral heuristics allow drivers to behave *as if* they do [14]. Also, we do not suggest that (1) is a complete account of a closed-loop driver behavior. Instead, what is being investigated here is whether a definition of ω_{err} can be identified such that (1) provides a good approximation of observed human steering rates.

A. Desired Path Yaw Rate Error (DPYRE) Model

There is a long tradition of driver models based on the concept of a *desired path* that the model previews and attempts to follow (see, for example, the review in [10]). Here, one such model was tested, where ω_{err} in (1) was defined as the yaw rate that, starting from the current vehicle position and heading, would make the vehicle's trajectory intersect the desired path after a preview time T_P (see Fig. 3 for an illustration).

In the lane change task, the desired path consisted of straight lines before the first cone gate and after the second one and half a period of a cosine function between the two. The lateral positions of the initial and concluding straight segments were the middles of the first cone gate and the third cone gate, respectively. In the slalom task, the desired path consisted of straight lines before the first cone and after the eighth and last cone, and a sinusoidal path with its extrema at the longitudinal positions of each of the eight cones, and amplitude A, a free model parameter.

B. Modified Gordon & Magnuski (MG&M) Model

While the use of a desired path concept is very common in driver modeling, it is debatable whether drivers are really making use of any mental representations of this nature [14], [18]. Therefore, an attempt was made at replacing the desired path of the desired path yaw rate error (DPYRE) model with something that was purely based on the cone track layouts themselves. This was done by using a model proposed by Gordon and Magnuski [16], who used a control law similar to (1) but calculated the desired yaw rate ω^* as the yaw rate needed to not collide with *boundary points* along the left and right sides of the lane being navigated, here mapping nicely onto cones.

The original authors considered a fixed preview horizon and applied steering rates based on the previewed boundary point with the largest absolute yaw rate error. This does not work well on a cone track, where closer cones have to be passed first even if a later cone suggests a larger yaw rate error. The model was modified accordingly, as illustrated in Fig. 4. At each time step, upcoming cones were considered one at a time, closest first, consecutively narrowing down the range of yaw rates $I = [\omega_{\min}, \omega_{\max}]$ that allowed passing all cones considered so far on the correct side, with an additional safety margin ρ (see [23] for the required mathematics), until a cone requiring a yaw rate outside I was encountered.

If there were no such unfeasible cones, the desired yaw rate was determined exactly as by the original authors, either just maintaining the current vehicle yaw rate ω if $\omega \in I$ (i.e., *satisficing* control [24]) or the closest boundary of I otherwise. If there were unfeasible cones, the desired yaw rate was set to the boundary of I closest to the yaw rate needed to correctly pass
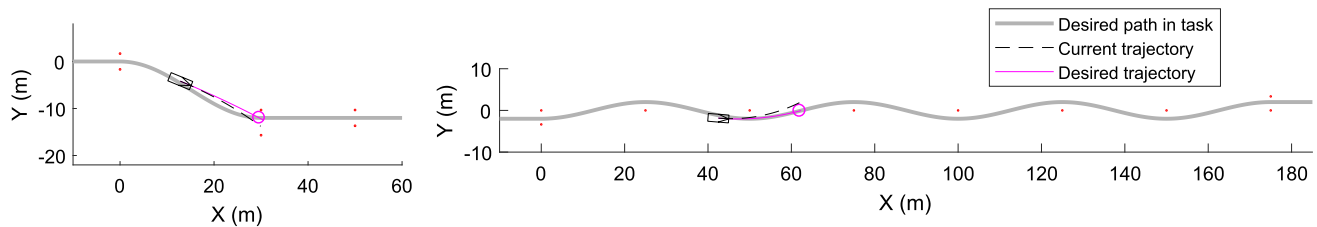
Fig. 3. Illustrations of the two driving tasks, the DPYRE steering model, as well as an example of observed car states. In the DPYRE model, the yaw rate error is the difference between the yaw rates for the desired and current vehicle trajectories, and the desired vehicle trajectory is the trajectory that intersects with a predefined desired path a preview time ahead. The small red rings are cones, shown to scale.
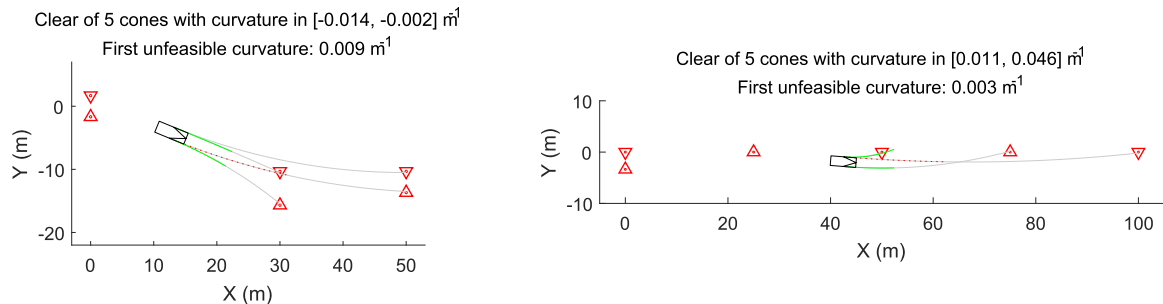

Fig. 4. Illustrations of the MG&M [16] model. The red rings are cones, and the triangles surrounding them indicate the side to which the cone needs to be passed. The gray lines show limit trajectories for correctly passing the cones, the green lines indicate the feasible range of trajectories, and the red dotted line indicates the first cone that cannot be passed with a single circular trajectory that also correctly passes all preceding cones (the lower cone at $X = 50$ m in the left panel; the cone at $X = 100$ m in the right panel).

the unfeasible cone. In the lane change task, for example, this makes the model keep tight toward the rightmost cone at the first cone gate, since at that point the first unfeasible cone is further to the right. Finally, a model parameter T_{pass} was included; if a cone was closer than this time ahead, it was considered as already effectively passed, and it no longer affected steering.

V. MODEL COMPARISON

This section describes the comparison carried out between the two considered models. The described model fitting method will also be used for other purposes in later sections.

A. Model Fitting

1) Target Signal: Both tested models predict steering rates. The human steering rates $\dot{\delta}$ were estimated as follows: For the simulator data, the recorded steering wheel angles were quantized from their original 0.1° resolution to the 1.5° resolution of the instrumented vehicle to make the signals comparable. Then, all steering wheel data were low-pass filtered with a Gaussian kernel of standard deviation 0.1 s, chosen since it was found to reasonably restore the original 0.1° resolution steering wheel angle from the quantized version for the simulator data. These filtered steering wheel angle data were then numerically differentiated to obtain $\dot{\delta}$.

2) Per-Recording, Open-Loop Fitting: The models were fitted to individual task recordings, thus providing a model-based quantification of behavior in each task repetition separately (fitting the models across all repetitions per driving condition was also explored but was found to be overly sensitive to single

TABLE I
MODEL PARAMETERS FITTED USING GRID SEARCH

Parameter	Used for models and tasks	Searched values
A	DPYRE slalom	{0.5, 0.75, 1, ..., 3} m
T_{P}	DPYRE	{0.8, 1.0, ..., 3} s
T_{R}	DPYRE and MG&M	{0, 0.02, 0.04, ..., 0.5} s
T_{pass}	MG&M	{0, 0.25, ..., 1} s
ρ	MG&M	{-0.5, -0.4, -0.3, ..., 0.5} m

outlier recordings with differing behavior or outcome). The fittings were of an open-loop nature; at each evaluated point k in a recording, sampled every 0.05 s, the current situation was fed to the model, including the consideration of any delays, and the model's steering rate $\hat{\dot{\delta}}_k$ in that situation was calculated and compared with the observed steering rate $\dot{\delta}_k$ applied by the driver. Goodness of model fit for a recording was calculated as the coefficient of determination (interpretable as the fraction of observed variance explained by the model) given by

$$R^2 = 1 - \frac{\sum_k (\hat{\dot{\delta}}_k - \dot{\delta}_k)^2}{\sum_k (\dot{\delta}_k - \bar{\dot{\delta}})^2} \qquad (2)$$

with $\bar{\dot{\delta}}$ being the average of $\dot{\delta}_k$ in the recording.

3) Fitting Method: The model parameters were fitted with a combination of an exhaustive grid search and linear least-squares. For each combination of the parameter values listed in Table I, the steering gain K was obtained by least-squares fitting to the human steering, and the final parameterization was selected as the most successful such least-squares fitting
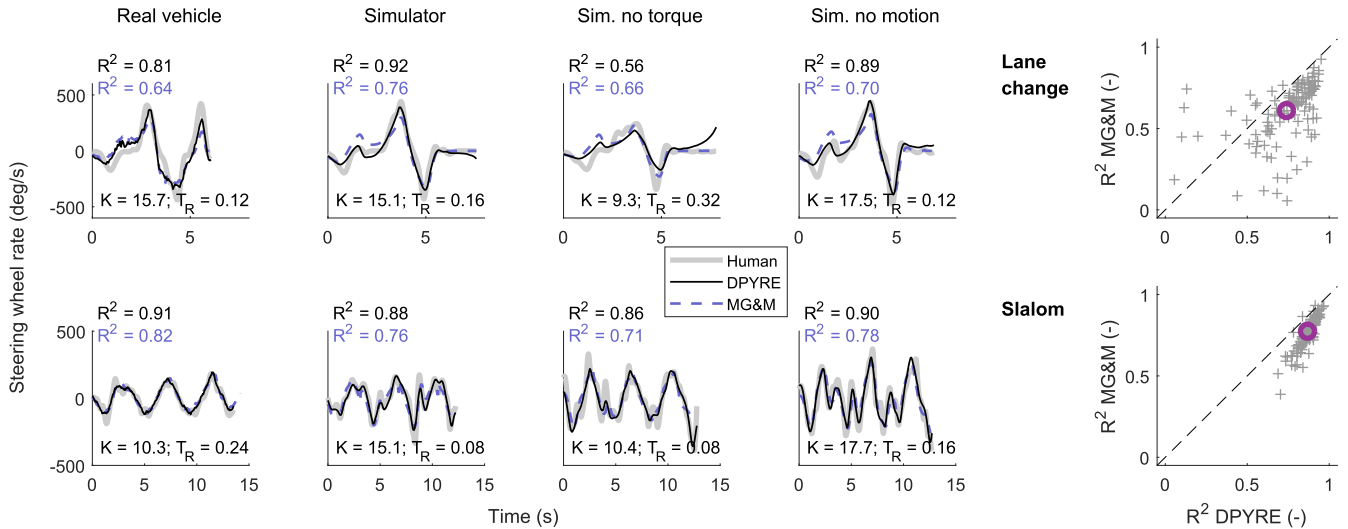
Fig. 5. Comparison of model fit between the DPYRE model and the MG&M model in the lane change (top row of panels) and slalom (bottom row) tasks. The eight smaller panels to the left show examples of typical fits in individual recordings. These are all from the third repetition per driving condition for Driver 3 in the lane change task and Driver 4 in the slalom task. The two larger panels to the right provide overviews of the model fit across all collected data per task. The plus signs show R^2 for individual task recordings, the circles show R^2 averages across recordings, the dashed lines indicate $y = x$, i.e., data points below the dashed line are recordings where the DPYRE model provided a better fit than the MG&M model.

(highest R^2) across the grid search. It can seen from Table I that in all fittings, there were three grid-searched parameters and one gain, except the DPYRE model in the lane change, where only two parameters were grid-searched. Also, note that the ρ parameter of the modified Gordon & Magnuski (MG&M) model was allowed negative values (i.e., a negative safety margin) since this was found to improve the fit in some cases.

4) Data Segments Fitted to: The considered parts of the task recordings were as follows: For the lane change task, from 20 m before the first cone gate to the third cone gate. For the slalom task, from the first to the eighth (last) slalom cone. In practice, these ranges were approximate since for the instrumented vehicle data, the exact relative positions between the car and cones were not recorded as intended due to DGPS limitations. Therefore, the ranges of data for model-fitting were extracted (both for the instrumented vehicle and simulator datasets) using salient features of the vehicle trajectories—the maximum lateral speed as an estimate of the longitudinal midpoint between the first and second cone gates in the lane change, and maximum lateral positions near the second and seventh cones as estimates of these cones' longitudinal positions in the slalom. A similar approach also provided estimated vehicle positions for the models when predicting driver steering in the real vehicle: DGPS positioning was stable between consecutive task repetitions with a single driver (closely overlapping trajectories); thus, the average DGPS location of the abovementioned "salient features" were observed per driver and task, and this average point was then assumed to indicate the same position on the test track as indicated by averaging in the same way over the recordings for the same driver and task in the simulator. This approach is obviously not perfect and adds uncertainty to the real-world model fits.

Out of the total 248 task recordings, 7 (2.8%) were excluded; one because the driver lost control during the slalom task, one because the driver did not carry out the lane change task as instructed, and five lane changes were excluded because data logging in the simulator was incomplete for unknown reasons.

B. Results

Fig. 5 shows typical examples of model fit to individual recordings, as well as overviews of R^2 across the entire dataset. Since the DPYRE model almost always obtained higher R^2 than the MG&M model, with equal number of parameters or fewer, the DPYRE model was adopted as the candidate model for measuring behavior validity (see Section VII-A for further discussion on model performance).

VI. USING THE DPYRE MODEL TO ASSESS BEHAVIOR

A. Accuracy of Parameter Estimates

The system being identified here, i.e., the driver, is being subjected only to the excitation provided by the task itself rather than a more complete set of forcing functions designed for optimal parameter estimation (cf. [6]–[8]). The advantage of this approach is that it lets the driver carry out the task at hand unperturbed, but the drawback is that it becomes less certain that the fitted model parameter values are accurate and meaningful. An additional concern is raised here due to the inexact positioning in the data collected with the real vehicle.

To provide some objective insight into the accuracy of the open-loop fitting method, closed-loop behavior was generated with the DPYRE model itself by simulating it together with a linear vehicle model fitted to the multibody model's observed lateral dynamics on snow (i.e., disregarding, for simplicity, the ice patch in the lane change task). These closed-loop simulations were generated for a full grid of parameter values, with each parameter covering at least the central two quartiles obtained in the fitting to human behavior (four linearly spaced values in each of $T_P \in [1.5, 2.5]$ s; $T_R \in [0.15, 0.3]$ s; $K \in [10, 20]$; $A \in [1.8, 3]$). Three 100-Hz simulations of each task were run per parameterization, with noise added to introduce variability of a magnitude comparable to that observed in the human data: Gaussian noise was added to both steering rates and vehicle yaw
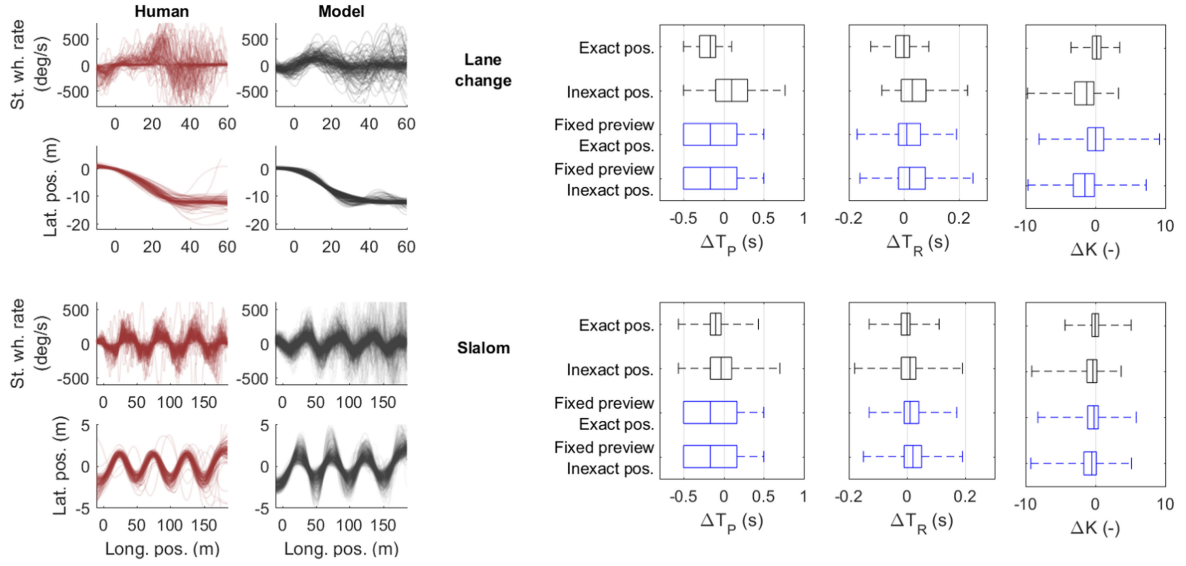
Fig. 6. Closed-loop model simulations to investigate the accuracy of the open-loop model parameter estimation. The two leftmost columns of panels show all recordings of human behavior in the two tasks and the behavior of the driver model, when simulated with noise across a full grid of model parameterizations (i.e., not specifically attempting to reproduce the human behavior). The panels to the right show estimation errors, deviations between the parameter values obtained when applying the open-loop fitting method to the simulated data, and the true parameter values used in the simulations. Each "box plot" shows the median, minimum, maximum, and quartile estimation errors, and each row of a box plot is for a separate variation to the fitting method; see the text for details (the ΔT_P for fits with fixed preview $T_P = 2$ s reflects the deviation from the actual model parameterizations $T_P \in [1.5, 1.83, 2.17, 2.5]$ s; the minimum and lower quartile errors coincide at –0.5 s because the lower preview time values yielded more frequent model control loss and hence, fewer data points).

rates, with standard deviations of 0.2 and 0.05 rad/s, respectively, and low-pass filtered with third-order Butterworth filters with cutoff frequencies of 1.5 and 0.5 Hz, respectively.

The left side of Fig. 6 shows the generated model behavior, alongside the totality of recordings obtained from the human drivers. Despite the DPYRE model not being developed with the aim of closed-loop stability and despite the full-parameter grid covering many types of parameter combinations not adopted by the human drivers, the model was relatively often capable of successfully completing the tasks. If the simulated vehicle's heading relative to the track's forward direction exceeded 70°, then this was judged as a control loss, something that happened in 45 (23%) of the 192 lane change simulations and in 246 (32%) of the 768 slalom simulations; these simulations are not shown in Fig. 6.

The right side of Fig. 6 shows observed parameter estimation errors (ΔT_P, etc.) when applying the open-loop fitting method to the simulated data (here and in the following, results for the A parameter are omitted to save space; this amplitude can easily be investigated with more conventional, nonmodel-based metrics). Results are shown for four variations of the fitting method.

1) Fitting to the exact data generated in the simulations, with exactly the same method as described in Section V.

2) Fitting with the same method, but after repositioning each simulated vehicle trajectory by a separate random vector to emulate uncertainty in vehicle positioning. This repositioning was drawn from a uniform distribution of $\pm(4, 0.7)$ m in longitudinal and lateral directions for the lane change and $\pm(3, 0.25)$ m for the slalom, corresponding to the variability observed in the "salient features" mentioned in Section V-A4.

3–4) Like the first two but fixing $T_P = 2$ s as a possible means of addressing the partial parameter redundancy between

T_P and T_R (increasing or decreasing both at the same time by appropriate amounts can leave the model behavior unchanged to some extent).

Overall, Fig. 6 shows that as long as vehicle position estimates are exact (as in the simulator here), the open-loop fitting method provides unbiased estimates of T_R and K in both tasks (medians close to zero), but with a median downward bias in the preview time T_P of 0.1–0.2 s. Introducing positioning uncertainty, however, introduces additional variability in parameter estimation but, importantly, upward biases also for T_P and T_R and downward biases for K. Fixing the T_P parameter reduces the bias in T_R somewhat, especially for the lane change task, but increases the bias in K.

B. Comparing Model Fits Between Real Vehicle and Simulator

Fig. 7 depicts a comparison of model goodness-of-fit (also including model rms errors, in units of steering rate, as a complement to the dimensionless R^2 values) and obtained parameters across real and simulated driving conditions. As in the previous section, results are provided here both from the fitting as described in Section V, with T_P as a free parameter (black plots), as well as with T_P fixed at the average values obtained in the simulator; 2.2 and 1.7 s for the lane change and slalom, respectively (blue plots). Friedman tests were applied, indicating statistically significant effects of the driving condition in the slalom for the rms errors as well as all three of T_P, T_R, and K, with Bonferroni-corrected post hoc significance levels (all six pairwise comparisons) illustrated in Fig. 7. All tests for the lane change task came out nonsignificant. To avoid trying to interpret parameter values from poorly fit driver models, these analyses excluded recordings with low R^2. Different exclusion thresholds for R^2 in the 0.4–0.7 range were tested, but besides
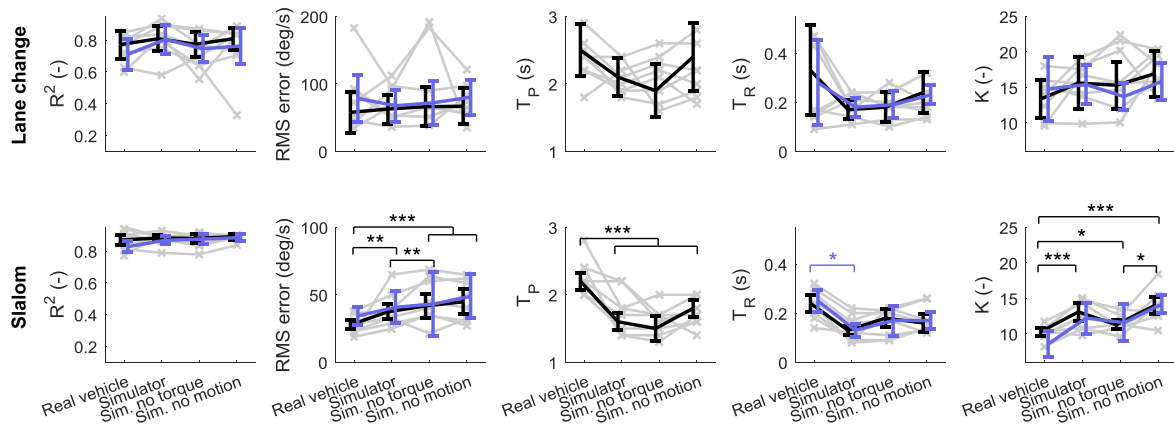
Fig. 7. Comparisons of model fits between driving conditions. The black and light blue thick plots show medians with 95% confidence intervals, from model fits with free and fixed preview time T_P, respectively. The light gray traces show per-driver medians from the fits with free T_P. Significance levels of post hoc tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. It should be noted that the estimated parameter values for the real vehicle driving may have been influenced by inexact vehicle positioning; see the text for further details.

affecting the statistical power, the exact threshold value did not affect the main patterns and conclusions; the results presented here are for $R^2 > 0.6$ (i.e., for example, excluding the no-torque lane change recording shown in Fig. 5). In the analyses of T_R values, the Friedman tests were applied to the parameter values obtained when fitting with a fixed T_P, since the estimation accuracy tests in the previous section indicated that this would provide a fairer comparison of T_R. In all the other tests, the results from model fittings with free T_P were used (and the observed significant effect for T_R actually remains significant also when analyzed in this way).

VII. DISCUSSION

This section will focus, first, on the extent to which the tested models were able to capture the human steering and then on the model-based comparison between driving conditions, including its limitations. Finally, the present contribution will be considered in the wider context of methods for simulator validity assessment.

A. Model Performance

As can be seen from Fig. 5, across both the lane change and slalom tasks, the DPYRE and MG&M models generally predicted very similar time histories of steering rate, but generally also with slightly better fits for the DPYRE model. The bands of fits just under the $y = x$ diagonals (i.e., with slightly better DPYRE fit) in the rightmost panels of Fig. 5 seem to be due to the desired path of the DPYRE model providing a desired yaw rate signal $\omega*$ that is subtly more humanlike than the $\omega*$ obtained with the MG&M model's cone-based approach. This could either be taken as an indication that humans do use a desired path construct, but just as well that the cone-based account needs to be perfected further. Such work could, for example, go in the direction of an optimal control [9] and/or a closer consideration of driver gaze behavior and exact visuomotor heuristics employed by drivers [12].

For the lane change task, at least two other phenomena are at play: 1) notably worse fits for the MG&M model than for

the DPYRE model for some recordings (points far below the $y = x$ diagonal in Fig. 5). Most of these recordings include a cone hit where the MG&M model can sometimes, despite the T_{pass} parameter, provide exaggerated steering rates just before the collision; and 2) poor DPYRE model fits, around or below $R^2 = 0.5$, but with better fits for the MG&M model. All of these are recordings where the driver decelerated to a very low speed or a full stop at or just after the final cone gate. In these situations, the DPYRE model predicts exaggerated steering rates to get to the desired path along a trajectory that is approaching zero length and, therefore, a high curvature (see the example without steering torque shown in Fig. 5).

Overall, the DPYRE model was rather successful at capturing the human steering in the lane change and slalom tasks, with a majority of fits in the R^2 0.7–0.9 range. However, looking closer at the example fits shown in Fig. 5, it is clear that there were also some recurring shortcomings. The initial rightward steering peak in the lane change task (at about 1–2 s in the examples) does not seem well captured in general, and in the slalom there is often a secondary peak of steering rate in each half-period of steering, which tends to be underestimated by the model (e.g., at 4 and 9 s in the third slalom example shown in Fig. 5). In both of these types of situations, it seems that the human drivers are applying an additional burst of steering to "swing wide" (cf. [25]) and "open up" the next cone gate more than suggested by the desired paths used here; this behavior is also to some extent visible in the lane change and slalom snapshots shown in Fig. 3. Again, as mentioned earlier, either optimal control modeling (cf. [25]) or a targeted investigation of exact visuomotor heuristics would seem to provide promising avenues for future work.

As for the closed-loop simulations of the DPYRE model shown in Fig. 6, it should be noted that the dissimilarity between the model and human behavior here can, to some extent, be attributed to the full grid of parameterizations tested, including less humanlike, for example unstable, parameterizations. In any case, as mentioned, the stability or human-likeness in the closed-loop simulation was not a modeling aim here; the aim was instead a model that could reproduce the human steering rate

signal in a way that would allow a meaningful interpretation of fitted model parameters. The visual aspect of how fits, such as those shown in Fig. 5, capture the human steering signal (shape, timing, and amplitude) may be taken as an indicator that this purpose may have been achieved, but a further study of the obtained parameters and of how they vary with experimental manipulations is needed. This leads on to the following section.

B. Effects of Driving Conditions on Model Parameters

1) Comparison Between Simulator Configurations: The parameter estimation accuracy results shown in Fig. 6 provide an additional positive indication. As long as the vehicle positioning is accurate, the open-loop parameter estimation has limited variability, with some limited bias only for the T_P parameter. Also, an estimation bias as such may not be problematic if it is consistent, allowing meaningful relative comparisons. Based on these insights, it is interesting to compare the obtained model fits between the three driving simulator configurations.

The only two statistically significant effects were observed in the slalom; larger model rms errors when turning off steering torques and larger steering gains without motion cues than without steering torques (see further below for possible interpretations). Beyond these findings, there were also trends that aligned in their aspect between the two tasks, with the no-torque configuration possibly standing out as having lower T_P and K than the other two configurations and the no-motion configuration as possibly having the largest T_P. It is interesting that these patterns recur between tasks, but they remain unconfirmed here and require further investigation, preferably with larger samples of drivers if possible. It is not clear why only the slalom and not the lane change would show significant effects of simulator modifications, but previous work has indeed shown the slalom performance to be sensitive to, for example, simulator motion [20].

2) Comparison Between Reality and Simulator: The differences in model parameter values between the real vehicle and simulator were also consistent between tasks and again, statistically significant for the slalom. However, the interpretation of these findings is hampered by the fact that the parameter estimation results (see Fig. 6) showed biases from vehicle positioning inaccuracy, as known to be the case here for the real-world data, in precisely the directions observed here for the real-world driving, a larger T_P and T_R and a smaller K.

The magnitude of differences in parameter values between the simulator and reality is larger than that between the exact and inexact positioning in the estimation tests, which *could* be taken to suggest that the observed effects are not solely due to positioning inaccuracy. Assuming for a moment that this is so and that it can be confirmed in follow-up studies, it is interesting to note that when effects of simulation on delay times and gains were observed in the Delft flight simulation studies [7], [8], these were also typically in the same directions as observed here, shorter delays and larger gains (although often with largest gains for highest-fidelity simulators, in contrast with the large gain for the no-motion configuration here). One possible interpretation of shorter delays and larger gains in the simulator would be that drivers are less able to make use of their normal overlearned

control strategies, making control more inaccurate and, as an adaptive response, more effortful. Such an interpretation would also align with the pattern, here, of increasing slalom model rms errors between reality and simulators, as well as with separate observations of increased cone hits and steering wheel reversal rates in the simulated slalom, when removing steering torques and motion.

It should be noted, however, that any conclusions along these lines are not warranted at this point due to the risk that the presently observed differences between the simulator and reality are artifactual in nature.

C. Methodology for Simulator Assessment

An overall conclusion from the abovementioned results and discussion is that the open-loop fitting of few-parameter models of $W \mapsto C$ type is a promising tool for assessing behavioral validity of simulators, at least in the types of vehicle-control-focused tasks studied here. A suitable next step to further test this tool would be to apply the same models and methods to additional data to see if the rough patterns of effects on parameters observed here can be confirmed. Such work could also extend to a larger variety of driving tasks. Not least, studying tasks that can be carried out by participants from the general public would help address some limitations that arose in the present study from the focus on low-friction vehicle stability testing with professional test drivers, such as small sample size, logistical challenges preventing counterbalancing of real and simulated driving, and a long period of time between the two data collections.

As mentioned in Section II-B, the taxonomy of behavioral validity proposed here, and the usefulness of driver models implied by the taxonomy, should indeed be applicable to a wider range of tasks, such as HMI interactions or the monitoring of a partially automated vehicle, albeit presumably with more advanced models. Interestingly, Ameyoe *et al.* fitted a model, more complex than what has been studied here, to driving with and without secondary tasks and found that episodes of distracted driving could be identified by analyzing the obtained model parameters [17]. Regardless of the driving task being studied, more complex models may be capable of picking up on more subtle behaviors and effects (see the discussion in Section VII-A), but there is of course a tradeoff to be managed, where the parameter estimation may become more difficult or less accurate for more complex models.

With respect to the proposed taxonomy, another important methodological consideration is as to how one should interpret model fitting results, such as those obtained here, to draw conclusions about which behavioral validity scenario (Scenarios A–F) is the case for a given simulator and task. One possibility would be to rely solely on statistical significance tests on model parameters and conclude that $W \mapsto C$ has changed whenever there is a significant difference between parameters fitted to real and simulated driving. However, if doing so, it is to be noted that statistical significance is determined by both effect sizes and sample sizes and that the lack of a statistically significant difference between reality and a simulator configuration does

not permit the conclusion that the behavior was *preserved* in that simulator; large effects with potentially serious applied implications may still be nonsignificant if the sample size is too small. Therefore, if the present experiment had been able to provide reliable real-world model parameter fits, it would have been important to consider also the effect sizes when comparing parameter values between reality and a simulator. Conversely, in a large-sample study, one may find statistical significance for effects that might be small enough to be irrelevant for the application at hand.

Another aspect to note is that the present driver model analyses only addressed one-third of the utility triplet mentioned in Section II-B. More conventional metrics assessing task performance (f) and task-relevant world states (W) remain important, however, not least to clarify whether the model-based metrics provide any insights not provided by the conventional metrics.

We have also not broached the topic of *absolute* behavioral validity, such as studied here, versus *relative* validity [2]. For many applications, it may be enough that the *relative* effect of some manipulations on behavior is preserved in the simulator. For example, if an A–B comparison of behavior between two vehicle design alternatives favors the same alternative in both simulator and reality, it might matter less whether or not the absolute behavior is preserved. The conceptual model and taxonomy proposed here provide some possible handles on this topic that might be useful in future work: Relative behavioral validity would seem more likely if the overall control strategy (the neural mapping N) is preserved, with goodness of model fit as one possible indicator. If so, the lack, here, of any significant effects of driving condition on model R^2 could potentially indicate a good relative validity.

A related matter is the question of how *quickly* drivers adapt to the simulator. We have suggested that for all of the Scenarios C–F, some adaptation will necessarily take place, and if this process and its speed can be observed for a given task (see, e.g., [26]), for example using the model-based methods proposed here (cf. [27]), this could provide another type of measure of behavioral validity. Furthermore, with respect to adaptation, as mentioned earlier in this paper, it remains an open question as to whether the distinction between behavioral validity Scenarios B and C is important for driving simulation, i.e., whether there are driving simulator applications for which it is important that no adaptation whatsoever occurs, not even of the internal $\tilde{S} \mapsto \tilde{C}$ mapping. The taxonomy and models outlined here provide potential stepping stones toward answering such questions. For example, if an imperfect simulator preserves $W \mapsto C$ but is not as efficient a tool as a real vehicle for training in some given task, this suggests that it is important to achieve preserved $\tilde{S} \mapsto \tilde{C}$ mappings, i.e., Scenario B, in this context.

Finally, since the purpose of behavioral validity assessment is always to understand whether a simulator permits some application (driver learning outcomes, traffic research questions, and vehicle design decisions), it is important to combine the behavioral models and metrics with direct comparisons of the application itself between the simulator and reality. This is arguably the only way to determine acceptable thresholds for measures of behavioral validity, model-based or not.

VIII. CONCLUSION

The main aim here has been to advance the state of the art in driver-model-based methods for the assessment of simulators. To this end, we have proposed a task-general conceptual theory and taxonomy of behavioral validity in the driving simulator and have derived from this theory a recommendation of using models mapping directly from observable world states to control actions ($W \mapsto C$ rather than $\tilde{S} \mapsto \tilde{C}$). We have proposed two such models and have shown through open-loop model fittings to unperturbed human steering data from near-limit, low-friction maneuvering that for this purpose one of the models, the "desired path yaw rate error" model, consistently outperformed the other. Furthermore, by means of closed-loop simulations with the DPYRE model, we have shown that its parameters can generally be accurately estimated with the model-fitting method used here, but that vehicle positioning inaccuracies introduce not only an additional variability in parameter estimates but also biases, such that a comparison of fitted-model parameters between driving conditions with differing positioning accuracies is not advisable. Finally, we have demonstrated some seemingly recurring patterns of how the availability of steering torque and motion cues in the simulator affects fitted-model parameters. Based on these findings, we conclude that the models and methods employed here may provide a useful addition to the methods for assessing behavioral validity of driving simulators. An important next step will be to apply these methods in additional empirical studies, to combine them with conventional metrics of behavior and performance, and to correlate them with measures of whether or not the simulator achieves the applied objective in question (e.g., generating the same vehicle design decisions as with a physical prototype).

REFERENCES

[1] D. L. Fisher, M. Rizzo, J. K. Caird, and J. D. Lee, *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*. Boca Raton, FL, USA: CRC Press, 2011.

[2] G. J. Blaauw, "Driving experience and task demands in simulator and instrumented car: A validation study," *Human Factors*, vol. 24, no. 4, pp. 473–486, 1982.

[3] D. L. Fisher, A. P. Pollatsek, and A. Pradhan, "Can novice drivers be trained to scan for information that will reduce their likelihood of a crash?," *Injury Prevention*, vol. 12, pp. i25–i29, 2006.

[4] Y. Wang, B. Mehler, B. Reimer, V. Lammers, L. A. D'Ambrosio, and J. F. Coughlin, "The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing," *Ergonomics*, vol. 53, no. 3, pp. 404–420, 2010.

[5] E. R. Boer, A. H. Jamson, S. K. Advani, and A. J. Horrobin, "Are we there yet? An objective mechanism to support the assessment of driving simulator utility," in *Proc. Driving Simul. Conf.*, Paris, France, Sep. 4–5, 2014, Paper 35.

[6] H. J. Damveld, M. Wentink, P. M. van Leeuwen, and R. Happee, "Effects of motion cueing on curve driving," in *Proc. Driving Simul. Conf.*, Paris, France, Sep. 6–7, 2012, Paper 06.

[7] P. M. T. Zaal, D. M. Pool, M. M. Van Paassen, and M. Mulder, "Comparing multimodal pilot pitch control behavior between simulated and real flight," *J. Guid., Control, Dyn.*, vol. 35, no. 5, pp. 1456–1471, 2012.

[8] D. M. Pool, P. M. T. Zaal, H. J. Damveld, M. M. Van Paassen, and M. Mulder, "Evaluating simulator motion fidelity using in-flight and simulator measurements of roll tracking behavior," in *Proc. AIAA Model. Simul. Technol. Conf.*, Minneapolis, MN, USA, Aug. 13–16, 2012, Paper 2012–4635.

[9] C. MacAdam, "Understanding and modeling the human driver," *Vehicle Syst. Dyn.*, vol. 40, no. 1–3, pp. 101–134, 2003.

[10] M. Plöchl and J. Edelmann, "Driver models in automobile dynamics application," *Vehicle Syst. Dyn.*, vol. 45, no. 7–8, pp. 699–741, 2007.

[11] D. D. Salvucci and R. Gray, "A two-point visual control model of steering," *Perception*, vol. 33, pp. 1233–1248, 2004.

[12] R. M. Wilkie, J. P. Wann, and R. S. Allison, "Active gaze, visual look-ahead, and locomotor control," *J. Exp. Psychol.: Human Perception Perform.*, vol. 34, no. 5, pp. 1150–1164, 2008.

[13] H.-S. Tan and J. Huang, "Experimental development of a new target and control driver steering model based on DLC test data," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 375–384, Mar. 2012.

[14] G. Markkula, O. Benderius, and M. Wahde, "Comparing and validating models of driver steering behaviour in collision avoidance and vehicle stabilization," *Vehicle Syst. Dyn.*, vol. 52, no. 12, pp. 1658–1680, 2014.

[15] G. Markkula, E. R. Boer, R. Romano, and N. Merat, "Sustained sensorimotor control as intermittent decisions about prediction errors: Computational framework and application to ground vehicle steering," *Biol. Cybern.*, to be published. Preprints. [Online]. Available: https://arxiv.org/abs/1703.03030

[16] T. Gordon and N. Magnuski, "Modeling normal driving as a collision avoidance process," in *Proc. 8th Int. Symp. Adv. Vehicle Control*, 2006, pp. 735–740.

[17] A. Ameyoe, P. Chevrel, E. Le-Carpentier, F. Mars, and H. Illy, "Identification of a linear parameter varying driver model for the detection of distraction," *IFAC-PapersOnLine*, vol. 48, no. 26, pp. 37–42, 2015.

[18] D. Franklin and D. Wolpert, "Computational mechanisms of sensorimotor control," *Neuron*, vol. 72, pp. 425–442, 2011.

[19] A. Eriksson, V. Banks, and N. Stanton, "Transition to manual: Comparing simulator with on-road control transitions," *Accident Anal. Prevention*, vol. 102, pp. 227–234, 2017.

[20] A. Berthoz *et al.*, "Motion scaling for high-performance driving simulators," *IEEE Trans. Human–Mach. Syst.*, vol. 43, no. 3, pp. 265–276, May 2013.

[21] C. J. Nash, D. J. Cole, and R. S. Bigler, "A review of human sensory dynamics for application to models of driver steering and speed control," *Biol. Cybern.*, vol. 110, no. 2–3, pp. 91–116, 2016.

[22] A. H. J. Jamson, "Motion cueing in driving simulators for research applications," Ph.D. dissertation, Inst. Transp. Studies, Univ. Leeds, Leeds, U.K., 2010.

[23] S. Chang, "A flexible hierarchical model-based control methodology for vehicle active safety systems," Ph.D. dissertation, Mech. Eng., Univ. Michigan, Ann Arbor, MI, USA, 2007.

[24] M. A. Goodrich, W. C. Stirling, and E. R. Boer, "Satisficing revisited," *Minds Mach.*, vol. 10, no. 1, pp. 79–109, 2000.

[25] E. R. Boer, "What preview elements do drivers need?," *IFAC-PapersOnLine*, vol. 49, no. 19, pp. 102–107, 2016.

[26] D. V. McGehee, J. D. Lee, M. Rizzo, J. Dawson, and K. Bateman, "Quantitative analysis of steering adaptation on a high performance fixed-base driving simulator," *Transp. Res. Part F: Traffic Psychol. Behav.*, vol. 7, no. 3, pp. 181–196, 2004.

[27] D. M. Pool, G. A. Harder, and M. M. Van Paassen, "Effects of simulator motion feedback on training of skill-based control behavior," *J. Guid., Control, Dyn.*, vol. 39, no. 4, pp. 889–902, 2016.

Richard Romano received the M.A.Sc. degree in aerospace studies from the University of Toronto, Toronto, ON, Canada, in 1990 and the Ph.D. degree in industrial engineering from the University of Iowa, Iowa City, IA, USA, 1999.

He has researched and developed driving simulators and simulation technology since 1990. He is currently the Chair of driving simulation at the University of Leeds, Leeds, U.K. His current research interests include motion cueing algorithms, vehicle dynamics, and virtual testing of vehicle technology, including automated systems.

A. Hamish Jamson received the B.Eng. degree in mechanical engineering from the University of Sheffield, Sheffield, U.K., in 1991 the M.Sc. degree in flight dynamics from the College of Aeronautics, Cranfield University, Cranfield, U.K, in 1992 and the Ph.D. degree in motion cueing in driving simulators from Institute for Transport Studies, University of Leeds in 2011.

He is currently a Principal Research Fellow with the University of Leeds, Leeds, U.K. He was responsible for the development of the University of Leeds Driving Simulator, one of the leading such facilities worldwide. His published work involves driver behavior, vehicle design, and driving simulation

Luigi Pariota received the M.Sc. and Ph.D. degrees in transportation engineering from the University of Naples "Federico II," Naples, Italy, in 2009 and 2013, respectively.

He is currently a Postdoctoral Fellow with the University of Naples "Federico II," and a Visiting Research Fellow with the University of Leeds, Leeds, U.K. His current research interests include the modeling of driving behavior and the assessment of the impacts of new technologies on both drivers and traffic flow.

Alex Bean received the B.Eng. degree in automotive engineering from Loughborough University, Loughborough, U.K., in 1996.

He is currently a Technical Specialist in driving dynamics with Jaguar Land Rover, Coventry, U.K., specializing in the application of objective test methods, computer-aided engineering (CAE) simulation, and driving simulators for the development of the vehicle dynamics attribute. He has gained extensive experience in vehicle dynamics development and active chassis systems in the automotive industry.

Gustav Markkula received the M.Sc. degree in engineering physics and complex adaptive systems and the Ph.D. degree in machine and vehicle systems from the Chalmers University of Technology, Gothenburg, Sweden, in 2004 and 2015, respectively.

After having worked in the automotive industry for more than a decade, he is currently an Associate Professor with the University of Leeds, Leeds, U.K. His current research interests include quantitative modeling of road user behavior and virtual testing of vehicle technology.

Erwin R. Boer received the M.Sc. degree from the Twente University of Technology, Enschede, The Netherlands, in 1990, and the Ph.D. degree from the University of Illinois, Chicago, IL, USA, in 1995, all in electrical engineering.

He is currently the President of his own automotive consulting company Entropy Control Inc., La Jolla, CA, USA, and a Senior Research Scientist with NAUTO, Palo Alto, CA, USA. He also holds academic appointments with UCSD, U-Leeds, and TU-Delft. His research interests include human machine cooperation, haptic shared control and human performance, and state quantification across a number of domains.