Benchmarking Behavior Prediction Models in Gap Acceptance Scenarios

Supplementary Materials - Implementation Details

Julian F. Schumann, Arkady Zgonnikov, Aravinda R. Srinivasan, Jens Kober, Gustav Markkula

textwidth in cm: 18.13275cm
linewidth in cm: 8.85553cm

I. Notation

The notations inside this supplementary material have been defined inside the paper presenting the original framework for for benchmarking gap acceptance models [1].

II. Adaptions to the framework for benchmarking gap acceptance problems.

A. Varying the number of input time steps

In the previous version of the framework for benchmarking gap acceptance models [1], when dealing with the problem that the available data (provided only at $\boldsymbol{T}$) is insufficient to provide enough data for all n_I required input time steps of size δt for a prediction to be made at t_0 (i.e., $\min\{\boldsymbol{T}\} > t_0 - (n_I - 1)\delta t$), the same adjustment was always made:

$$\widehat{t}_0(n_I) = \max\{t_0, \min\{\boldsymbol{T}\} + (n_I - 1)\delta t \tag{1}$$

It is easy to see that this allows for cases where prediction might have to be made at a different time if a different n_I is chosen. Additionally, as the condition

$$\widehat{t}_0(n_I) < \min\{t_A, t_{\text{crit}}\} \tag{2}$$

still has to be fulfilled for a sample to be included in a final dataset, this approach would also allow for differently sized dataset. Consequently, a valid evaluation of the influence of the number of input time steps on the predictive performance of a model is not possible. Furhtermore, if a stricter selection of t_0 would be pursued (i.e., $\widehat{t}_0(n_I) = t_0$), than the different number in input time steps would result in a number of samples being reject for a lack of input data for higher n_I.

To overcome these issues, we provided the framework with the additional input information $n_{I,\max}$. Using this, a sample than has to fulfill the adjusted requirement of

$$\max\{t_S, \min\{\boldsymbol{T}\} + (n_{I,\max} - 1)\} \leq t_0 < \min\{t_A, t_{\text{crit}}\} \tag{3}$$

instead, which should guarantee otherwise identical datasets for all $n_I \leq n_{I,\max}$. The only exception here is the case of $t_0 = t_S$ and $n_I = 2$ for the *L-GAP* dataset (where $\min\{\boldsymbol{T}\} = t_S$), with the relaxation of $\widehat{t}_0 = t_S + \delta t$.

B. Extracting 1D data from 2D data

To allow for a transformation from two-dimensional (2D) trajectories to quasi-one-dimensional (1D) ones, it is important to give the respective agents (besides the ego vehicle V_E that offers the gap and the target vehicle V_T thas has to deceide whether to accept it) certain roles.

- V_1: The vehicle driving along V_E's path P_E in front of V_E. Inclusion is justified because of the need to define t_S.
- V_2: The vehicle following V_E along P_E. This is justified by studies showing that the size of the next gap can influence the decision making of the target vehicle V_T.
- V_3: The agent preceding V_T along its path P_T. It can for example block V_T from accepting a gap, or force a decision as an upcoming roadblock (this can also for instance be the end of the merging lane on highways), but also goad V_T into accepting the gap by following a specific example. Therefore, its inclusion is justified.

Based on those roles, the following one dimensional distances can be defined:

- D_C: The distance of V_E to the contested space. It is positive if V_E is still offering the gap. It is measured from V_E's front bumper.
- D_A: The distance of V_T to the contested space. It is positive as long the gap is not accepted. It is again measured from V_T's front bumper.
- D_1: The distance between V_1 and V_E along P_E from bumper to bumper. It is positive if V_1 is in front of V_E
- D_2: The distance between V_2 and V_E along P_E from bumper to bumper. It is positive if V_E is in front of V_2.
- D_3: The distance between V_3 and V_T along P_T from bumper to bumper. It is positive if V_3 is in front of V_T.
- L_E: The size of the contested space along P_E.
- L_T: The size of the contested space along P_T. While this concept is fairly easily to understand for clearly crossing trajectories, determining L_T is more difficult if P_E and P_T start to continuously overlap after entering the contested space. Examples are highway lane changes or cases where V_T enters a roundabout. In such situations, the size of L_T might depend on current kinematic information, i.e. $L_T = L_T(t)$, and might be defined as the distance for which V_T would have to accelerate to to prevent being rear-ended by V_E.

Those seven distances than from the input provided to the prediction models unable to deal with two dimensional inputs, such as the *Commotions* model [2]. Their definition in regards to the two dimensional trajectory data has to be defined for each dataset individually. However, those transformations from 2D to 1D data is not invertable, which limits such model to the prediction of the binary choice (accept/reject a gap) and the time of accepting the gap.

III. The used modules

The simulations and comparisons performed are facilitated by the previously developed framework for benchmarking gap acceptance models [1]. In the following section, the detailed implementation of the required modules inside this framework is discussed.

A. Datasets

Three datasets were used in the experiments discussed in the main part. Their respective size can be seen in Tab. I. While their are differences in the implementation of those datasets, some aspects of their implementation, which follow below, are general.

- The following parameters are identical accross all datasets: $a_{\text{brake}} = 4\,\mathrm{ms}^{-2}$, $\delta t = 0.2\,\mathrm{s}$, $t_\epsilon = 0.01\,\mathrm{s}$, and $n_p = 100$.
- The Δt needed to define constant sized gaps is always created by maximizing the term
$$\min\left\{N_A,\ N_{\neg A}\right\} , \tag{4}$$
where
$$N_A = \sum_i a_i\ ,\ N_{\neg A} = \sum_i (1 - a_i) \tag{5}$$
are the number of samples that fulfill all requirements to be included in the final dataset based on the current choice of Δt, which influences the chosen t_0.
- A common definition of C_S, C_C, and C_A (which are used to extract t_S, t_C, and t_A respectively) is used, based on the extracted 1D data :
$$\begin{aligned} &C_S(t) : D_1(t) - D_C(t) = L_E \ \wedge\ \dot{D}_1(t) > \dot{D}_C(t) \\ &C_C(t) : D_C(t) = 0 \ \wedge\ \dot{D}_C(t) < 0 \\ &C_A(t) : D_A(t) = 0 \ \wedge\ \dot{D}_A(t) < 0 \end{aligned} \tag{6}$$
- For all datasets, the following definitions are used for the variables $t_{\underline{C}}$ and t_{brake} required for extracting t_C and t_{crit}:
$$\begin{aligned} t_{\underline{C}}(t) &= \frac{D_C(t)}{\max\left\{-\dot{D}_C(t), 0\right\}} + t \\ t_{\text{brake}}(t) &= \frac{\max\left\{-\dot{D}_C(t), 0\right\}}{2a_{\text{brake}}} \end{aligned} \tag{7}$$
- A lane width of $w = 3.5\,\mathrm{m}$ and a vehicle length of $l = 5\,\mathrm{m}$ are always assumed.
- When additional agents V_i are unavailable but required as inputs, vehicles in this role are created. These are placed far away from the ego and target vehicles, so their influence on the prediction model should be minimal. For example, a missing V_1 would be assumed to drive in front of the ego vehicle V_E, but at a very large distance of $500\,\mathrm{m}$.

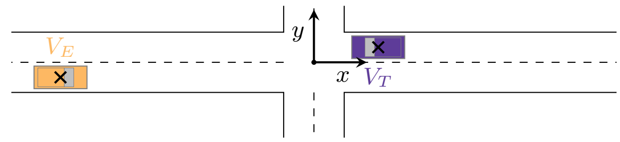

Fig. 1. The left turn. The black crosses symbolize the position vector $\boldsymbol{x}_i$ belonging to agent V_i.

1) Left turns: The *L-GAP* dataset covers left turns at intersections, where the target vehicle V_T turns left across the planned trajectory of the ego vehicle V_E, which wants to drive straight ahead across the intersection (Fig. 1). Here, V_E appears when V_T has slowed down sufficiently during its approach to the intersection (t_S is set at this point). As there is no position data for V_E before t_S, special adjustments are needed ($t_0 = t_S + (n_I - 1)\delta t$) to permit prediction at the opening of the gap. However, this adjustment is only allowed for $n_I = 2$, so no onset of movement should be observable in V_T due to limited human reaction times.

The following transformation is used to convert the two dimensional position data into the one dimensional form:
$$\begin{aligned} D_C(t) &= -\left(x_E(t) + w + \frac{1}{2}l\right) \\ D_A(t) &= \operatorname{sgn}\left(y_T(t)\right)\sqrt{y_T(t)^2 + \Delta x_T(t)^2} \\ &\qquad \Delta x_T(t) = \min\left\{\frac{y_T(t)\dot{x}_T(t)}{\dot{y}_T(t)},\ x_T(t) + \frac{1}{2}w\right\} \\ D_1(t) &= D_2(t) = D_3(t) = 500\,\mathrm{m} \\ L_E(t) &= 2w \\ L_T(t) &= w \end{aligned} \tag{8}$$

For this dataset, the optimization of equation (4) results in $\Delta t = 3.36\,\mathrm{s}$.

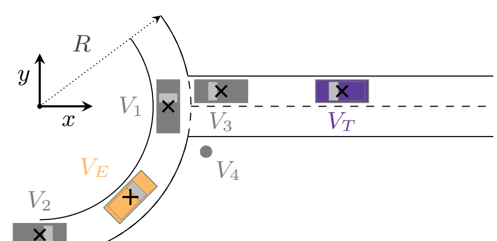

Fig. 2. Entering the roundabout. The black crosses symbolize the position vector $\boldsymbol{x}_i$ belonging to agent V_i.

2) Roundabouts: The *rounD* dataset is another naturalistic dataset covering the scenario of roundabouts. There, V_T wants to enter the roundabout, either in front of or behind the ego vehicle V_E already inside the roundabout with an outer radius R. Further agents are V_1 and V_2 driving respectively in front of or behind V_E inside the roundabout. V_3 might enter the roundabout directly in front of V_T, while V_4 is the pedestrian most likely to interact with V_4 (Fig. 2).

TABLE I
THE NUMBER OF ACCEPTED GAPS N_A AND REJECTED GAPS $N_{\neg A}$ IN THE IMPLEMENTED DATASETS, IN THE FORM: $N_A - N_{\neg A}$. THE NUMBERS DEPEND ON THE METHOD FOR CHOOSING THE TIME OF PREDICTION t_0

Dataset	Initial gaps $t_0 = t_S$	Constant sized gaps $t_0 = \min\{t \,\vert\, t_{\underline{C}}(t) - t = \Delta t\}$	Critical gaps $t_0 = t_{\text{crit}}$
L-GAP (Left turns)	$700 - 716$ ($n_I = 2$)	$375 - 205$	$227 - 397$
rounD (Roundabout)	$789 - 2164$	$381 - 556$	$147 - 2810$
Leeds (Straight intersection)	$99 - 942$	$41 - 108$	$0 - 942$

The following transformation is used to convert the two dimensional position data into the one dimensional form:

$$\begin{aligned} D_C(t) &= d_{I,E}(t) - \frac{1}{2}l \\ D_A(t) &= d_{O,T}(t) \\ D_1(t) &= d_{I,E}(t) - d_{I,1}(t) - l \\ D_2(t) &= d_{I,2}(t) - d_{I,E}(t) - l \\ D_3(t) &= d_{O,T}(t) - d_{O,3}(t) - l \\ L_E(t) &= w \\ L_T(t) &= \max\left\{w, \frac{1}{2}a_{\text{esc}}t_{\text{esc}}(t)^2 - \dot{D}_A(t)t_{\text{esc}}(t)\right\} \end{aligned} \tag{9}$$

Here, a transformation from a Cartesian to a cylindrical coordinate system is used (when driving in a counter clockwise circle, the angle $\alpha_i \in [-\infty, \infty]$ is not reset but keeps increasing):

$$r_i(t)\exp\left(i\alpha_i(t)\right) = x_i(t) + iy_i(t)$$

Meanwhile, one also uses the 1D distance of a vehicle inside (d_I) and outside (d_O) the roundabout

$$\begin{aligned} d_{I,i}(t) &= \ln\left(\frac{1}{1+\exp\left(r_i(t) - R_L\right)}\right)\tanh\left(5\alpha_i(t)\right) \\ &\quad - R_L\alpha_i(t) \\ d_{O,i}(t) &= R_L\left(\frac{w}{R} - \alpha_i\right)\max\left\{0, 1 - \frac{1}{32}\exp\left(5h_i(t)\right)\right\}^2 \\ &\quad + h_i(t) - \ln(2) \\ h_i(t) &= \ln\left(1+\exp\left(r_i(t) - R\right)\right)\tanh\left(5\left(\frac{\pi}{4} - \alpha_i(t)\right)\right) \\ R_L &= R - \frac{1}{2}w\,, \end{aligned} \tag{10}$$

as well as the escape acceleration $a_{\text{esc}} = 2\,\text{ms}^{-2}$ and escape time t_{esc}

$$t_{\text{esc}}(t) = \frac{\max\left\{\dot{D}_A(t) - \dot{D}_C(t), 0\right\}}{a_{\text{esc}}}\,. \tag{11}$$

For this dataset, the optimization of equation (4) results in $\Delta t = 2.53\,\text{s}$.

3) Intersection: This dataset created in the Leeds driving simulator [3] cover the behavior of vehicles at an intersection, where both V_E and V_T want to go straight through the intersection. Like in the dataset above, further agents are included as well. This are V_1 and V_2, respectively driving in front of or behind V_E inside the roundabout (Fig. 3). This dataset was created under assumption of vehicles driving on the left side of the road, in opposition to the previous two datasets.

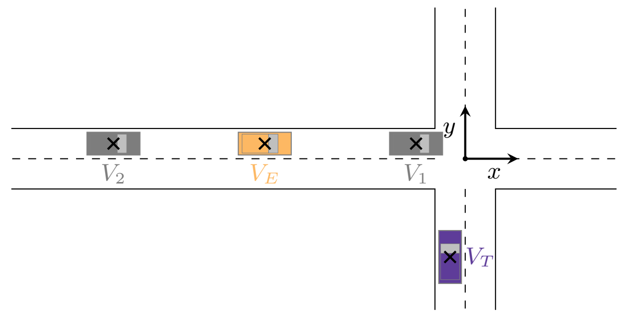

Fig. 3. Entering the intersection. The black crosses symbolize the position vector $\boldsymbol{x}_i$ belonging to agent V_i.

The following transformation is used to convert the two dimensional position data into the one dimensional form:

$$\begin{aligned} D_C(t) &= -\left(x_E(t) + w + \frac{1}{2}l + 1\,\text{m}\right) \\ D_A(t) &= -\left(x_E(t) + w + \frac{1}{2}l + 1\,\text{m}\right) \\ D_1(t) &= x_1(t) - x_E(t) - l \\ D_2(t) &= x_E(t) - x_2(t) - l \\ D_3(t) &= 1000\,\text{m} \\ L_E(t) &= w + 1\,\text{m} \\ L_T(t) &= w \end{aligned} \tag{12}$$

For this dataset, the optimization of equation (4) results in $\Delta t = 6.22\,\text{s}$.

B. Splitting methods

1) Random splitting: The splitting is done separately for samples with accepted and rejected gaps so that the bias toward one of these groups should be roughly equal in both the training and testing set. It has to be noted that ten separate random splittings are used in this work, to avoid the case that a single model seems superior because it is especially well suited to that specific splitting. By repeating the trials multiple times, it is also possible to estimate if differences between models performances are indeed statistically significant.

2) Extreme splitting: Here, the goal is to select those cases as testing samples, where the decision by the human actor V_T is most unintuitive. As for the previous method, the splitting is done separately for accepted and rejected gaps.

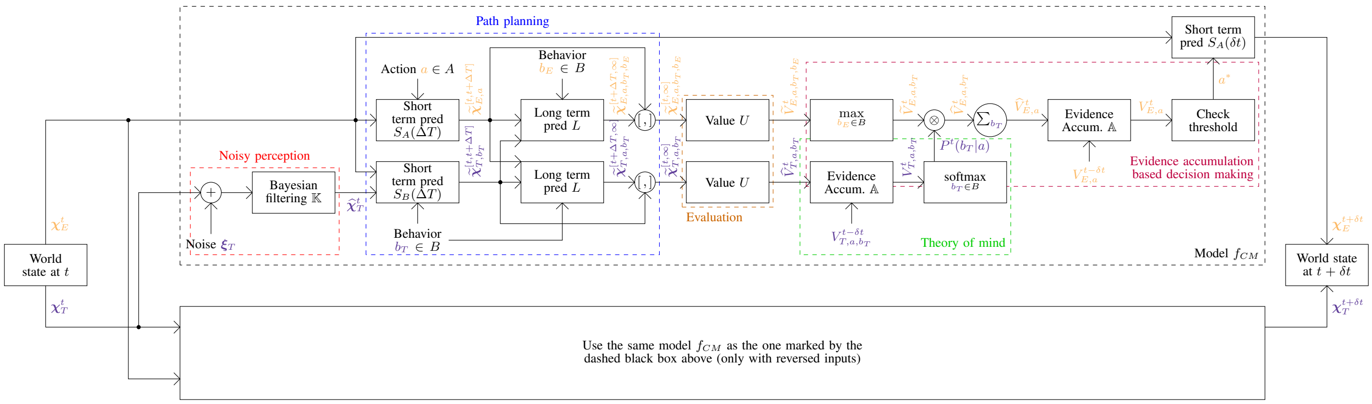

Fig. 4. The detailed version of the *Commotions* model. The theoretic parts are highlighted. For a full prediction, this process has to be repeated multiple times. The parameters $\boldsymbol{\theta}$ of the function f_{CM} are described in Tab. II.

A rejected gap is considered more unintuitive if the available gap is larger. Accordingly, the samples with the highest values of $t_C - t_0$ become part of the test set. Meanwhile, accepting a gap is deemed more unintuitive for smaller gaps. Thus, those accepted gaps where $t_{\underline{C}}(t_A) - t_A$ is lowest are used for testing.

C. Models

1) Trajectron++ (T+): The *Trajectron++* model uses many different neural network, such as long-short-term memories (LSTM) and convolutional neural networks (CNN). It requires velocity and acceleration data as inputs, necessitating some preprocessing of the given position data $\boldsymbol{X}_{\boldsymbol{T}_I}$ and $\boldsymbol{T}_I$. For the hyperparameters, the values used by the authors when applying the model to the *nuScenes* dataset [4] are selected [5], as this dataset also mainly includes motor vehicles. Furthermore, an attention radius between vehicles of $150\,\mathrm{m}$ was chosen. No other changes have been made to the model.

2) Logistic Regression (LR): While normally specifically selected properties are used as inputs of a logistic regression model, the positional data in $\boldsymbol{X}_I$ is given here. This results in an input with a dimensionality of $2n_I|\boldsymbol{V}|$. Although this is not done normally in the literature, where velocity data is common as well, this is no contradiction (as long as $n_I \geq 2$) since, during training, the model can extract the velocity data from the difference in positions at different time-steps:

$$\begin{aligned} a_x x_1 + b_x x_2 &= a x_1 + b v_1 \\ v_1 &= \frac{x_2 - x_1}{\delta t}\ ,\ a = a_x + b_x\ ,\ b = b_x \delta t\,. \end{aligned} \tag{13}$$

The same can be said for both relative distances and velocities and even accelerations (for the latter as long as $n_I > 2$). Due to the convexity of the optimized cost function during the model training, it will ultimately arrive at the same result, despite the linear transformation of the inputs.

3) Commotions model: The *Commotions* model is a model consisting out of multiple parts (Fig. 4), trying to fully model the human decision making process, applying cognitive theory. While the logic behind this can be found in more detail in the original work [2], to understand the specific modifications made, a short introduction is nonetheless needed.

Each of the two agents in the *Commotions* model are assigned a state variable $\boldsymbol{\chi}$

$$\boldsymbol{\chi}_i^t = \{\chi_i(t), \dot{\chi}_i(t), \ddot{\chi}_i(t)\}\ , \tag{14}$$

where

$$\begin{aligned} \chi_E(t) &= D_C(t) + \frac{1}{2}\left(L_E(t) + l\right) \\ \chi_T(t) &= D_A(t) + \frac{1}{2}\left(L_T(t) + l\right)\,. \end{aligned} \tag{15}$$

To get a simulated prediction, one has to apply the model transfer function f_{CM}, which is parameterized with $\boldsymbol{\theta}$:

$$\begin{aligned} \boldsymbol{\chi}_E^{t+\delta t} &= f_{CM}\left(\boldsymbol{\chi}_E^t, \boldsymbol{\chi}_T^t | \boldsymbol{\theta}\right) \\ \boldsymbol{\chi}_T^{t+\delta t} &= f_{CM}\left(\boldsymbol{\chi}_T^t, \boldsymbol{\chi}_E^t | \boldsymbol{\theta}\right)\,. \end{aligned} \tag{16}$$

By using this model (Fig. 4 shows exemplary how to calculate $\boldsymbol{\chi}_E^{t+\delta t}$), one can simulate the future behavior over multiple time-steps. A binary prediction a as well as a prediction of t_A for a single sample can then be extracted (using linear interpolation between simulated time steps if necessary):

$$a_{\text{pred}} = \begin{cases} 1 & \chi_T(t_C) < \frac{1}{2}\left(L_T(t_C) + l\right) \\ 0 & \text{else} \end{cases} \tag{17}$$

$$t_{A,\text{pred}} = \min\left\{ t \,|\, \chi_T(t) < \frac{1}{2}\left(L_T(t) + l\right) \right\} \tag{18}$$

The parameters $\boldsymbol{\theta}$ and their purpose and range can be found in Tab. II. If their range is given as a power of 10, $\ln \theta_i$ will be optimized instead of θ_i, which results in small values of the parameter being searched with higher likelihood during optimization.

The following two loss functions were then used to find the optimal values of those parameters $\boldsymbol{\theta}$:

$$\begin{aligned} \mathcal{L}_1 &= \sum_i \frac{1}{n_p} \sum_{p=1}^{n_p} \begin{cases} \left(t_A - t_{A,\text{pred},i,p}\right)^2 & a_i = 1 \\ \left(t_C - \min\left\{t_C, t_{A,\text{pred},i,p}\right\}\right)^2 & a_i = 0 \end{cases} \\ &\qquad + 4\left|a_i - a_{\text{pred},i,p}\right| \\ \mathcal{L}_2 &= \mathcal{L}_1 + \sum_i 100\mathbb{V}\left(t_{A,\text{pred},i}\right) - 20\sqrt{\mathbb{V}\left(t_{A,\text{pred},i}\right)} + 1\,. \end{aligned} \tag{19}$$

TABLE II
THE PARAMETERS OF THE *Commotions* MODEL (INSIDE f_{CM}).

θ_i	Range	Description
β_V	$10^{[0,2.3]}$	This is the multiplication in the soft-max function that calculates $P(b_T\|a)$ (Eq. (22)[1]).
$\Delta V_{\text{th, rel}}$	$10^{[-3,-1]}$	The relative factor used to calculate ΔV_{th} in Eq. (3), the threshold checked to decide on a^*.
T	$[0.1, 0.5]$	In Eq. (4), this time constant is used to calculate the evidence accumulation damping factor in $\mathbb{A}$.
T_δ	$10^{[1,2]}$	This is the scale factor for the time discount inside L (Eq. (16)).
σ_v	$10^{[-2.3,0.196]}$	This is the angular noise at the retina of a human, used to scale the added noise $\boldsymbol{\xi}_T$.
$\sigma_{\dot{\chi}}$	$10^{[-2,0]}$	The standard deviation of the process noise of the velocity needed in the Bayesian filtering $\mathbb{K}$.
ΔT	$[1.01\delta t, 5\delta t]$	The time period over which a given action a^* is applied in S_A (Eq. (2)).
T_s	$[0, 2]$	The safety time margin – used in short and long term planning (S_B, L) – for a certain behavior b to be considered safe.
D_s	$[0, 5]$	The safety distance margin – used in short and long term planning (S_B, L) – for a certain behavior b to be considered safe.
$u_{0,\text{rel}}$	$10^{[0,2]}$	The relative factor used to calculate the scale factor u_0 applied to the unrestricted value in U (Eq. (8)).
k_{da}	$10^{[-1,1]}$	The factor in U that is used to punish high acceleration in a motion plan (Eq. (10)).
$\sigma_{R_0\chi,\text{rel}}$	$10^{[-1,1]}$	The relative factor used to scale the initial observation noise $\sigma_{R_0\chi} = \|\chi(t_0)\|\sigma_{R_0\chi,\text{rel}}$ of the position in the Bayesian filtering $\mathbb{K}$.
$\sigma_{R_0\dot{\chi},\text{rel}}$	$10^{[-1,1]}$	The relative factor used to scale the initial observation noise $\sigma_{R_0\dot{\chi}} = \|\dot{\chi}(t_0)\|\sigma_{R_0\dot{\chi},\text{rel}}$ of the velocity in the Bayesian filtering $\mathbb{K}$.
$v_{\text{free,rel}}$	$10^{[-0.3,0.3]}$	The relative factor used to scale the desired free speed.
ΔT_{free}	10	The time that is assumed (in S_B and L) a vehicle would need to accelerate to free speed, if there is no chance of a dangerous interaction.
h_E	1.5	The eye height of an observer that is used to scale the added noise $\boldsymbol{\xi}_T$.

[1] The equation references given in the descriptions reference equations in the original source [2].

Here, the second loss function additionally forces the model to produce more varied predictions.

As f_{CM} and thus those loss functions are non-differentiable in regard to $\boldsymbol{\theta}$, zeroth-order optimization methods have to be used to find the best model parametrization. We used Bayesian optimization [6] to this end, as this algorithm is capable to also deal with random loss functions, which is here the case due to the influence of $\boldsymbol{\xi}_T$ in f_{CM}. For this Bayesian optimization, we generated 210 initial samples using based on sobol sampling [7]. For the one step optimization (*1O*), then 150 additional parameter settings where samples (using expected improvement as an acquisition function). Meanwhile, for the two stage method (*2O*), only 100 such samples where created in each stage. In this method, the search space was restricted to one fifth of each original along each dimension, centered around the optimized parameter setting from the first stage.

As mentioned in the main article, there are two methods to calculate v_{free}, the original one using identical speeds (*IS*) with

$$v_{\text{free}} = v_{\text{free,rel}} 10\frac{m}{s}\,, \tag{20}$$

and the adjusted one (*AS*) with

$$v_{\text{free, i}} = v_{\text{free,rel}} \max\left\{10\frac{m}{s}, \dot{\chi}_i(t_0)\right\}\,. \tag{21}$$

D. Metrics

1) AUC:

$$F = \frac{1}{N_A N_{\neg A}}\left(\left(\sum_i a_i r_i\right) - \frac{N_A\left(N_A+1\right)}{2}\right) \tag{22}$$

Here, r_i is the position of sample i according to its value $a_{\text{pred},i}$ ($r_i = 1$ for the lowest value of a_{pred} and $r_i = N_A + N_{\neg A}$ for the highest value). Here, one always gets $F_r = 0.5$ for a random predictor.

2) Average displacement error (ADE):

$$\begin{aligned} F = &\frac{1}{\left(N_A + N_{\neg A}\right)\lceil n_p \beta\rceil}\sum_i \sum_{p=1}^{\lceil n_p\beta\rceil} D_{i,p} \\ D_{i,p} &= \frac{1}{n_{O,i}}\sum_{t\in \boldsymbol{T}_{O,i}} \left\|\boldsymbol{x}_{T,i,p}(t) - \boldsymbol{x}_{T,i}(t)\right\| \end{aligned} \tag{23}$$

with

$$D_{i,p} \leq D_{i,p+1}\; \forall\, p \in \{1, \dots, n_p - 1\}\,.$$

3) True negative rate for perfect recall (TNR-PR):

$$F = \frac{\left|\left\{i \,|\, a_i = 0 \wedge a_{\text{pred},i} < \min\left\{a_{\text{pred},i} \,|\, a_i = 1\right\}\right\}\right|}{N_{\neg A}} \tag{24}$$

For a uniformly random predictor, one would get

$$F_r = \min\left\{a_{\text{pred},i} \,|\, a_i = 1\right\} = \frac{1}{N_A + 1}\,. \tag{25}$$

REFERENCES

[1] J. F. Schumann, J. Kober, and A. Zgonnikov, "Benchmarking Behavior Prediction Models in Gap Acceptance Scenarios," *IEEE Trans. on Intell. Veh.*, pp. 1–12, 2023. Conference Name: IEEE Transactions on Intelligent Vehicles.

[2] G. Markkula, Y.-S. Lin, A. R. Srinivasan, J. Billington, M. Leonetti, A. H. Kalantari, Y. Yang, Y. M. Lee, R. Madigan, and N. Merat, "Explaining human interactions on the road requires large-scale integration of psychological theory," June 2022. PsyArXiv.

[3] "Car Simulator (UoLDS)."

[4] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A Multimodal Dataset for Autonomous Driving," in *IEEE/CVF Conf. on Comput. Vis. Pattern Recognit. (CVPR)*, pp. 11621–11631, 2020.

[5] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data," Aug. 2022.

[6] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions," *J. Global Optim.*, vol. 13, pp. 455–492, Dec. 1998.

[7] I. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," *USSR Comput. Math. Math. Phys.*, vol. 7, pp. 86–112, Jan. 1967.

[8] A. Szczepanek, "Critical Value Calculator."

IV. Results

A. Statistic significance

In both cases, it might be possible to test if the difference in models is statistically significant, especially if a variance can be extracted from the 10 different random splitting. There should be to sources of variance in the result, namely the variance in the training test split and variance in the model training, where for example either the model itself or the optimization algorithm might include randomness. While the latter is likely uncorrelated between different models, the former most likely introduces some form of correlation between models. Consequently, a paired Student t-test is used to assess statistic significance. There, one then gets for ten random splittings the following condition to reject that null hypothesis that F_1 is not larger than F_2 with a significance level of $\alpha = 0.05$ (see [8] for the critical value):

$$t = \sqrt{\frac{10}{\mathbb{V}\left[F_{R,1} - F_{R,2}\right]}}\mathbb{E}\left[F_{R,1} - F_{R,2}\right] > 1.833 \tag{26}$$

This is equivalent to the condition

$$\frac{\mathbb{E}\left[F_{R,1} - F_{R,2}\right]}{\sqrt{\mathbb{V}\left[\right]F_{R,1} - F_{R,2}]}} > 0.5796 \tag{27}$$

When comparing the values for critical splits, this is not as easy, as there is only one value to consider. To still allow for a significance estimation, we assume that the variance for critical values is similar to the one for random ones, resulting in:

$$\frac{F_{C,1} - F_{C,2}}{\sqrt{\mathbb{V}\left[F_{R,1} - F_{R,2}\right]}} > 2.92 \tag{28}$$

B. *Comparing the* Commotions *model*

In Tab. III, IV, V, and VI, only $n_I = 2$ is used. The upper value in each of the fields is the average result as well its standard deviation on the ten different random test sets. Meanwhile, the lower value is the result on the test set that includes the most unintuitive samples.

TABLE III
AUC FOR PREDICTIONS AT INITIAL GAPS ($t_0 = t_S$)

Dataset	**Models**										
	$T++$	LR_{2D}	LR_{1D}	$CM_{IIJ1,\mathcal{L}_1}$	$CM_{INJ1,\mathcal{L}_1}$	$CM_{IIJ2,\mathcal{L}_1}$	$CM_{INJ2,\mathcal{L}_1}$	$CM_{AIJ1,\mathcal{L}_1}$	$CM_{ANJ1,\mathcal{L}_1}$	$CM_{AIJ2,\mathcal{L}_1}$	$CM_{ANJ2,\mathcal{L}_1}$
L-GAP	$0.737^{\pm 0.040}$	$\underline{0.830}^{\pm 0.013}$	$0.829^{\pm 0.013}$	$0.599^{\pm 0.042}$	$0.615^{\pm 0.045}$	$0.579^{\pm 0.090}$	$0.541^{\pm 0.082}$	$0.684^{\pm 0.041}$	$0.668^{\pm 0.041}$	$0.678^{\pm 0.067}$	$0.662^{\pm 0.059}$
	0.396	0.092	0.095	0.385	$\underline{0.500}$	0.374	0.489	0.327	0.322	0.381	0.493
rounD	$0.990^{\pm 0.004}$	$\underline{0.991}^{\pm 0.002}$	$0.988^{\pm 0.004}$	$0.958^{\pm 0.009}$	$0.958^{\pm 0.009}$	$0.955^{\pm 0.013}$	$0.958^{\pm 0.011}$	$0.957^{\pm 0.010}$	$0.958^{\pm 0.009}$	$0.958^{\pm 0.010}$	$0.947^{\pm 0.040}$
(V_E and V_T)	0.855	0.570	0.770	0.786	0.862	0.771	0.859	0.862	$\underline{0.865}$	0.862	0.784
Leeds	$0.987^{\pm 0.017}$	$0.986^{\pm 0.017}$	$0.986^{\pm 0.017}$	$0.969^{\pm 0.020}$	$0.962^{\pm 0.016}$	$0.967^{\pm 0.024}$	$0.973^{\pm 0.019}$	$0.961^{\pm 0.018}$	$0.969^{\pm 0.023}$	$0.972^{\pm 0.020}$	$0.972^{\pm 0.022}$
(V_E and V_T)	$\underline{0.990}$	0.777	0.777	0.956	0.937	0.954	0.978	0.961	0.970	0.955	0.937

	$CM_{IIA1,\mathcal{L}_1}$	$CM_{INA1,\mathcal{L}_1}$	$CM_{IIA2,\mathcal{L}_1}$	$CM_{INA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_1}$	$CM_{ANA1,\mathcal{L}_1}$	$CM_{AIA2,\mathcal{L}_1}$	$CM_{ANA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_2}$	$CM_{ANA1,\mathcal{L}_2}$	$CM_{AIJ2,\mathcal{L}_2}$
L-GAP	$0.770^{\pm 0.042}$	$0.780^{\pm 0.031}$	$0.787^{\pm 0.020}$	$0.791^{\pm 0.021}$	$0.679^{\pm 0.112}$	$0.768^{\pm 0.027}$	$0.724^{\pm 0.030}$	$0.769^{\pm 0.030}$	$0.650^{\pm 0.111}$	$0.765^{\pm 0.026}$	$0.707^{\pm 0.101}$
	0.318	0.361	0.355	0.415	0.066	0.252	0.086	0.123	0.098	0.084	0.381
rounD	$0.946^{\pm 0.009}$	$0.946^{\pm 0.009}$	$0.948^{\pm 0.010}$	$0.951^{\pm 0.010}$	$0.948^{\pm 0.007}$	$0.949^{\pm 0.008}$	$0.948^{\pm 0.012}$	$0.947^{\pm 0.009}$	$0.980^{\pm 0.009}$	$0.986^{\pm 0.005}$	$0.963^{\pm 0.066}$
(V_E and V_T)	0.829	0.827	0.811	0.828	0.824	0.834	0.834	0.825	0.846	0.827	0.778
Leeds	$0.981^{\pm 0.025}$	$0.985^{\pm 0.013}$	$0.940^{\pm 0.123}$	$0.979^{\pm 0.022}$	$0.987^{\pm 0.013}$	$0.988^{\pm 0.012}$	$0.986^{\pm 0.018}$	$0.981^{\pm 0.016}$	$0.994^{\pm 0.008}$	$0.993^{\pm 0.005}$	$\underline{0.996}^{\pm 0.004}$
(V_E and V_T)	0.909	0.911	0.935	0.917	0.890	0.937	0.904	0.865	0.936	0.922	0.952

TABLE IV
AUC FOR PREDICTIONS AT CONSTANT-SIZED GAPS ($t_0 = \min\{t \,|\, t_{\underline{C}}(t) - t = \Delta t\}$)

Dataset	**Models**										
	$T++$	LR_{2D}	LR_{1D}	$CM_{IIJ1,\mathcal{L}_1}$	$CM_{INJ1,\mathcal{L}_1}$	$CM_{IIJ2,\mathcal{L}_1}$	$CM_{INJ2,\mathcal{L}_1}$	$CM_{AIJ1,\mathcal{L}_1}$	$CM_{ANJ1,\mathcal{L}_1}$	$CM_{AIJ2,\mathcal{L}_1}$	$CM_{ANJ2,\mathcal{L}_1}$
L-GAP	$0.978^{\pm 0.014}$	$\underline{0.978}^{\pm 0.013}$	$0.971^{\pm 0.015}$	$0.940^{\pm 0.015}$	$0.936^{\pm 0.016}$	$0.945^{\pm 0.018}$	$0.937^{\pm 0.016}$	$0.953^{\pm 0.026}$	$0.941^{\pm 0.025}$	$0.948^{\pm 0.023}$	$0.942^{\pm 0.032}$
	$\underline{0.885}$	0.838	0.811	0.674	0.687	0.674	0.727	0.681	0.681	0.681	0.681
rounD	$0.986^{\pm 0.007}$	$\underline{0.992}^{\pm 0.005}$	$0.990^{\pm 0.004}$	$0.955^{\pm 0.010}$	$0.957^{\pm 0.007}$	$0.955^{\pm 0.006}$	$0.955^{\pm 0.013}$	$0.954^{\pm 0.012}$	$0.955^{\pm 0.011}$	$0.957^{\pm 0.010}$	$0.956^{\pm 0.011}$
(V_E and V_T)	0.917	0.933	0.835	0.697	0.690	0.706	0.716	0.699	0.703	0.733	0.716
Leeds	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$0.999^{\pm 0.002}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$
(V_E and V_T)	$\underline{1.000}$	0.938	0.938	$\underline{1.000}$	$\underline{1.000}$	0.938	0.938	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$

	$CM_{IIA1,\mathcal{L}_1}$	$CM_{INA1,\mathcal{L}_1}$	$CM_{IIA2,\mathcal{L}_1}$	$CM_{INA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_1}$	$CM_{ANA1,\mathcal{L}_1}$	$CM_{AIA2,\mathcal{L}_1}$	$CM_{ANA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_2}$	$CM_{ANA1,\mathcal{L}_2}$	$CM_{AIJ2,\mathcal{L}_2}$
L-GAP	$0.947^{\pm 0.021}$	$0.949^{\pm 0.018}$	$0.943^{\pm 0.022}$	$0.948^{\pm 0.024}$	$0.965^{\pm 0.012}$	$0.965^{\pm 0.017}$	$0.943^{\pm 0.048}$	$0.951^{\pm 0.046}$	$0.966^{\pm 0.013}$	$0.965^{\pm 0.013}$	$0.967^{\pm 0.014}$
	0.800	0.800	0.799	0.793	0.806	0.791	0.812	0.800	0.847	0.824	0.548
rounD	$0.953^{\pm 0.011}$	$0.951^{\pm 0.011}$	$0.955^{\pm 0.011}$	$0.951^{\pm 0.016}$	$0.953^{\pm 0.012}$	$0.954^{\pm 0.010}$	$0.940^{\pm 0.044}$	$0.956^{\pm 0.014}$	$0.976^{\pm 0.007}$	$0.976^{\pm 0.011}$	$0.981^{\pm 0.008}$
(V_E and V_T)	0.695	0.669	0.661	0.650	0.683	0.690	0.686	0.683	0.855	0.865	$\underline{0.934}$
Leeds	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$0.989^{\pm 0.022}$	$0.999^{\pm 0.003}$	$\underline{1.000}$	$\underline{1.000}$
(V_E and V_T)	0.938	0.935	$\underline{1.000}$	0.938	$\underline{1.000}$	$\underline{1.000}$	0.812	$\underline{1.000}$	$\underline{1.000}$	0.991	0.738

TABLE V

AVERAGE DISPLACEMENT ERROR ADE_1 FOR PREDICTIONS AT CONSTANT-SIZED GAPS ($t_0 = \min\{t \mid t_{\underline{C}}(t) - t = \Delta t\}$)

Dataset	**Models**										
	$T++$	LR_{2D}	LR_{1D}	$CM_{IIJ1,\mathcal{L}_1}$	$CM_{INJ1,\mathcal{L}_1}$	$CM_{IIJ2,\mathcal{L}_1}$	$CM_{INJ2,\mathcal{L}_1}$	$CM_{AIJ1,\mathcal{L}_1}$	$CM_{ANJ1,\mathcal{L}_1}$	$CM_{AIJ2,\mathcal{L}_1}$	$CM_{ANJ2,\mathcal{L}_1}$
L-GAP	$1.429^{\pm 0.135}$	$1.258^{\pm 0.088}$	$1.332^{\pm 0.065}$	$1.143^{\pm 0.081}$	$1.146^{\pm 0.084}$	$1.139^{\pm 0.064}$	$1.151^{\pm 0.086}$	$1.217^{\pm 0.064}$	$1.160^{\pm 0.104}$	$1.176^{\pm 0.127}$	$1.172^{\pm 0.120}$
	2.189	2.353	2.481	2.758	2.767	2.764	2.664	2.752	2.758	2.766	2.767
rounD	$1.213^{\pm 0.129}$	$1.195^{\pm 0.109}$	$1.185^{\pm 0.090}$	$1.154^{\pm 0.089}$	$1.158^{\pm 0.083}$	$1.156^{\pm 0.093}$	$\underline{1.150}^{\pm 0.089}$	$1.153^{\pm 0.086}$	$1.160^{\pm 0.085}$	$1.153^{\pm 0.088}$	$1.156^{\pm 0.086}$
(V_E and V_T)	1.743	$\underline{1.697}$	1.749	1.747	1.764	1.743	1.748	1.804	1.907	1.829	1.744
Leeds	$2.234^{\pm 0.330}$	$1.940^{\pm 0.265}$	$1.933^{\pm 0.259}$	$1.852^{\pm 0.205}$	$1.911^{\pm 0.336}$	$1.834^{\pm 0.205}$	$1.849^{\pm 0.196}$	$1.858^{\pm 0.222}$	$1.856^{\pm 0.217}$	$1.843^{\pm 0.210}$	$1.846^{\pm 0.206}$
(V_E and V_T)	3.150	5.732	5.972	3.758	3.714	4.618	4.618	3.709	3.814	4.463	3.881

	$CM_{IIA1,\mathcal{L}_1}$	$CM_{INA1,\mathcal{L}_1}$	$CM_{IIA2,\mathcal{L}_1}$	$CM_{INA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_1}$	$CM_{ANA1,\mathcal{L}_1}$	$CM_{AIA2,\mathcal{L}_1}$	$CM_{ANA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_2}$	$CM_{ANA1,\mathcal{L}_2}$	$CM_{AIJ2,\mathcal{L}_2}$
L-GAP	$1.080^{\pm 0.139}$	$1.081^{\pm 0.112}$	$1.086^{\pm 0.124}$	$\underline{1.079}^{\pm 0.125}$	$1.194^{\pm 0.116}$	$1.202^{\pm 0.089}$	$1.252^{\pm 0.172}$	$1.240^{\pm 0.191}$	$1.261^{\pm 0.120}$	$1.253^{\pm 0.067}$	$1.235^{\pm 0.116}$
	2.193	2.182	2.300	2.201	$\underline{2.177}$	2.241	2.179	2.185	2.445	2.337	3.453
rounD	$1.153^{\pm 0.088}$	$1.172^{\pm 0.091}$	$1.155^{\pm 0.090}$	$1.176^{\pm 0.075}$	$1.161^{\pm 0.086}$	$1.161^{\pm 0.093}$	$1.216^{\pm 0.208}$	$1.163^{\pm 0.091}$	$1.252^{\pm 0.090}$	$1.248^{\pm 0.085}$	$1.212^{\pm 0.085}$
(V_E and V_T)	1.972	1.877	2.045	1.856	1.880	1.799	2.030	1.789	1.859	1.938	1.732
Leeds	$1.784^{\pm 0.186}$	$1.776^{\pm 0.186}$	$1.797^{\pm 0.218}$	$1.867^{\pm 0.373}$	$\underline{1.772}^{\pm 0.183}$	$1.843^{\pm 0.356}$	$1.792^{\pm 0.193}$	$1.937^{\pm 0.452}$	$2.032^{\pm 0.352}$	$1.898^{\pm 0.342}$	$1.943^{\pm 0.194}$
(V_E and V_T)	2.899	2.924	$\underline{2.095}$	3.779	2.159	2.223	4.690	2.164	6.854	6.617	6.253

TABLE VI

TNR-PR FOR PREDICTIONS AT CRITICAL GAPS ($t_0 = t_{\text{CRIT}} - t_\epsilon$)

Dataset	**Models**										
	$T++$	LR_{2D}	LR_{1D}	$CM_{IIJ1,\mathcal{L}_1}$	$CM_{INJ1,\mathcal{L}_1}$	$CM_{IIJ2,\mathcal{L}_1}$	$CM_{INJ2,\mathcal{L}_1}$	$CM_{AIJ1,\mathcal{L}_1}$	$CM_{ANJ1,\mathcal{L}_1}$	$CM_{AIJ2,\mathcal{L}_1}$	$CM_{ANJ2,\mathcal{L}_1}$
L-GAP	$0.955^{\pm 0.043}$	$0.861^{\pm 0.118}$	$0.910^{\pm 0.099}$	$0.494^{\pm 0.494}$	$0.395^{\pm 0.484}$	$0.494^{\pm 0.494}$	$0.299^{\pm 0.456}$	$0.395^{\pm 0.484}$	$0.393^{\pm 0.481}$	$0.393^{\pm 0.481}$	$0.393^{\pm 0.481}$
	0.861	0.101	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
rounD	$0.699^{\pm 0.458}$	$\underline{0.994}^{\pm 0.008}$	$0.993^{\pm 0.011}$	$0.699^{\pm 0.458}$	$0.599^{\pm 0.489}$	$0.699^{\pm 0.458}$	$0.699^{\pm 0.458}$	$0.699^{\pm 0.458}$	$0.699^{\pm 0.458}$	$0.699^{\pm 0.458}$	$0.599^{\pm 0.489}$
(V_E and V_T)	0.000	$\underline{0.936}$	0.929	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	$CM_{IIA1,\mathcal{L}_1}$	$CM_{INA1,\mathcal{L}_1}$	$CM_{IIA2,\mathcal{L}_1}$	$CM_{INA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_1}$	$CM_{ANA1,\mathcal{L}_1}$	$CM_{AIA2,\mathcal{L}_1}$	$CM_{ANA2,\mathcal{L}_1}$	$CM_{AIA1,\mathcal{L}_2}$	$CM_{ANA1,\mathcal{L}_2}$	$CM_{AIJ2,\mathcal{L}_2}$
L-GAP	$0.779^{\pm 0.390}$	$0.788^{\pm 0.394}$	$0.784^{\pm 0.392}$	$0.785^{\pm 0.393}$	$\underline{0.983}^{\pm 0.016}$	$0.980^{\pm 0.020}$	$0.783^{\pm 0.392}$	$0.786^{\pm 0.393}$	$0.884^{\pm 0.295}$	$0.884^{\pm 0.295}$	$0.879^{\pm 0.293}$
	0.000	0.000	0.000	0.000	$\underline{0.949}$	0.000	0.000	0.000	0.000	0.000	0.000
rounD	$0.400^{\pm 0.490}$	$0.599^{\pm 0.489}$	$0.400^{\pm 0.490}$	$0.400^{\pm 0.490}$	$0.500^{\pm 0.500}$	$0.500^{\pm 0.500}$	$0.400^{\pm 0.490}$	$0.499^{\pm 0.499}$	$0.564^{\pm 0.464}$	$0.959^{\pm 0.055}$	$0.881^{\pm 0.295}$
(V_E and V_T)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.922

C. *Evaluating the promise of the* Commotions *model*

In Tab. VII, VIII, IX, and X, one uses $n_I \in \{2, 3, 8\}$. Similarly to the previous section, the upper values in each of the fields represent the average result as well its standard deviation on the ten different random test sets (for each of the different n_I, from $n_I = 2$ in the left, and $n_I = 8$ in the right). Meanwhile, the lower value depicts the result on the test set that includes the most unintuitive samples for the respective n_I.

TABLE VII
AUC FOR PREDICTIONS AT INITIAL GAPS ($t_0 = t_S$)

Dataset	**Models**								
	$T++$			LR_{2D}			LR_{1D}		
rounD	$0.990^{\pm 0.004}$ $\underline{0.916}$	$0.995^{\pm 0.003}$ $\underline{0.953}$	$0.995^{\pm 0.003}$ $\underline{0.954}$	$\underline{0.991}^{\pm 0.002}$ 0.645	$\underline{0.996}^{\pm 0.001}$ 0.658	$\underline{0.995}^{\pm 0.002}$ 0.659	$0.988^{\pm 0.003}$ 0.772	$0.991^{\pm 0.003}$ 0.747	$0.989^{\pm 0.004}$ 0.780
rounD (V_E and V_T)	$0.990^{\pm 0.004}$ $\underline{0.855}$	$0.995^{\pm 0.002}$ $\underline{0.962}$	$0.995^{\pm 0.002}$ $\underline{0.956}$	$\underline{0.991}^{\pm 0.002}$ 0.570	$\underline{0.997}^{\pm 0.001}$ 0.615	$\underline{0.997}^{\pm 0.001}$ 0.633	$0.988^{\pm 0.004}$ 0.770	$0.992^{\pm 0.003}$ 0.756	$0.992^{\pm 0.004}$ 0.763
Leeds	$\underline{0.989}^{\pm 0.014}$ $\underline{0.887}$	$\underline{0.997}^{\pm 0.008}$ $\underline{0.943}$	$\underline{0.996}^{\pm 0.007}$ $\underline{0.990}$	$0.981^{\pm 0.021}$ 0.775	$0.981^{\pm 0.021}$ 0.773	$0.982^{\pm 0.020}$ 0.776	$0.979^{\pm 0.027}$ 0.750	$0.981^{\pm 0.021}$ 0.750	$0.982^{\pm 0.026}$ 0.750
Leeds (V_E and V_T)	$\underline{0.987}^{\pm 0.017}$ $\underline{0.990}$	$\underline{0.994}^{\pm 0.011}$ $\underline{0.971}$	$\underline{0.991}^{\pm 0.017}$ $\underline{0.990}$	$0.986^{\pm 0.017}$ 0.777	$0.984^{\pm 0.018}$ 0.777	$0.989^{\pm 0.017}$ 0.776	$0.986^{\pm 0.017}$ 0.777	$0.984^{\pm 0.018}$ 0.777	$0.989^{\pm 0.017}$ 0.775

TABLE VIII
AUC FOR PREDICTIONS AT CONSTANT-SIZED GAPS ($t_0 = \min\{t \,|\, t_{\underline{C}}(t) - t = \Delta t\}$)

Dataset	**Models**								
	$T++$			LR_{2D}			LR_{1D}		
L-GAP	$0.978^{\pm 0.014}$ $\underline{0.885}$	$\underline{0.982}^{\pm 0.013}$ $\underline{0.876}$	$\underline{0.979}^{\pm 0.014}$ $\underline{0.885}$	$\underline{0.978}^{\pm 0.013}$ 0.838	$0.980^{\pm 0.012}$ 0.799	$0.978^{\pm 0.013}$ 0.745	$0.971^{\pm 0.015}$ 0.811	$0.970^{\pm 0.014}$ 0.802	$0.963^{\pm 0.017}$ 0.745
rounD	$0.986^{\pm 0.007}$ $\underline{0.901}$	$\underline{0.997}^{\pm 0.003}$ $\underline{0.973}$	$\underline{0.996}^{\pm 0.004}$ $\underline{0.975}$	$\underline{0.991}^{\pm 0.005}$ 0.861	$0.993^{\pm 0.005}$ 0.835	$0.986^{\pm 0.012}$ 0.871	$0.991^{\pm 0.004}$ 0.825	$0.993^{\pm 0.004}$ 0.799	$0.994^{\pm 0.006}$ 0.816
rounD (V_E and V_T)	$0.986^{\pm 0.007}$ 0.917	$0.995^{\pm 0.003}$ $\underline{0.975}$	$0.995^{\pm 0.004}$ $\underline{0.979}$	$\underline{0.992}^{\pm 0.005}$ $\underline{0.933}$	$\underline{0.996}^{\pm 0.004}$ 0.940	$\underline{0.997}^{\pm 0.003}$ 0.924	$0.990^{\pm 0.004}$ 0.835	$0.994^{\pm 0.004}$ 0.833	$0.995^{\pm 0.004}$ 0.841
Leeds	$\underline{1.000}$ $\underline{1.000}$	$\underline{1.000}$ $\underline{1.000}$	$\underline{1.000}$ $\underline{1.000}$	$\underline{1.000}$ 0.929	$\underline{1.000}$ 0.973	$\underline{1.000}$ 0.875	$\underline{1.000}$ 0.929	$\underline{1.000}$ 0.973	$\underline{1.000}$ 0.875
Leeds (V_E and V_T)	$\underline{1.000}$ $\underline{1.000}$	$\underline{1.000}$ $\underline{1.000}$	$\underline{1.000}$ $\underline{1.000}$	$\underline{1.000}$ 0.938	$\underline{1.000}$ 0.973	$\underline{1.000}$ 0.875	$\underline{1.000}$ 0.938	$\underline{1.000}$ 0.973	$\underline{1.000}$ 0.875

TABLE IX

AVERAGE DISPLACEMENT ERROR ADE_1 FOR PREDICTIONS AT CONSTANT-SIZED GAPS ($t_0 = \min\{t \,|\, t_{\underline{C}}(t) - t = \Delta t\}$)

Dataset	**Models**								
	$T++$			LR_{2D}			LR_{1D}		
L-GAP	$1.429^{\pm 0.135}$	$\underline{1.127}^{\pm 0.137}$	$\underline{1.087}^{\pm 0.146}$	$\underline{1.258}^{\pm 0.088}$	$1.164^{\pm 0.105}$	$1.122^{\pm 0.111}$	$1.332^{\pm 0.065}$	$1.247^{\pm 0.089}$	$1.199^{\pm 0.097}$
	$\underline{2.189}$	$\underline{1.863}$	$\underline{1.684}$	2.353	2.421	2.506	2.481	2.455	2.484
rounD	$1.220^{\pm 0.130}$	$0.904^{\pm 0.123}$	$0.843^{\pm 0.081}$	$1.102^{\pm 0.089}$	$\underline{0.812}^{\pm 0.081}$	$0.777^{\pm 0.078}$	$\underline{1.093}^{\pm 0.085}$	$0.813^{\pm 0.083}$	$\underline{0.764}^{\pm 0.057}$
	$\underline{1.520}$	$\underline{1.203}$	$\underline{1.051}$	1.760	1.524	1.361	1.706	1.561	1.311
rounD	$1.213^{\pm 0.129}$	$0.906^{\pm 0.140}$	$0.856^{\pm 0.086}$	$1.195^{\pm 0.109}$	$\underline{0.871}^{\pm 0.089}$	$\underline{0.788}^{\pm 0.073}$	$\underline{1.185}^{\pm 0.090}$	$0.877^{\pm 0.080}$	$0.808^{\pm 0.062}$
(V_E and V_T)	1.743	$\underline{1.143}$	$\underline{0.992}$	$\underline{1.697}$	1.305	1.198	1.749	1.294	1.177
Leeds	$\underline{2.018}^{\pm 0.294}$	$1.894^{\pm 0.277}$	$1.897^{\pm 0.300}$	$2.041^{\pm 0.294}$	$1.414^{\pm 0.235}$	$1.450^{\pm 0.141}$	$2.026^{\pm 0.292}$	$\underline{1.414}^{\pm 0.231}$	$\underline{1.446}^{\pm 0.148}$
	$\underline{3.543}$	$\underline{2.674}$	$\underline{2.449}$	7.618	4.864	5.091	7.576	4.939	5.115
Leeds	$2.234^{\pm 0.330}$	$1.879^{\pm 0.200}$	$\underline{1.843}^{\pm 0.292}$	$1.940^{\pm 0.265}$	$\underline{1.521}^{\pm 0.239}$	$1.992^{\pm 0.173}$	$\underline{1.933}^{\pm 0.259}$	$1.524^{\pm 0.236}$	$1.996^{\pm 0.171}$
(V_E and V_T)	$\underline{3.150}$	$\underline{2.926}$	$\underline{2.678}$	5.732	4.737	5.931	5.972	4.713	5.936

TABLE X

TNR-PR FOR PREDICTIONS AT CRITICAL GAPS ($t_0 = t_{\text{CRIT}} - t_\epsilon$)

Dataset	**Models**								
	$T++$			LR_{2D}			LR_{1D}		
L-GAP	$\underline{0.955}^{\pm 0.043}$	$\underline{0.993}^{\pm 0.013}$	$\underline{0.891}^{\pm 0.297}$	$0.861^{\pm 0.118}$	$0.896^{\pm 0.102}$	$0.869^{\pm 0.121}$	$0.910^{\pm 0.099}$	$0.919^{\pm 0.087}$	$0.566^{\pm 0.453}$
	$\underline{0.861}$	0.000	$\underline{0.000}$	0.101	$\underline{0.127}$	$\underline{0.000}$	0.000	0.000	$\underline{0.000}$
rounD	$0.600^{\pm 0.490}$	$0.900^{\pm 0.300}$	$0.900^{\pm 0.300}$	$\underline{0.992}^{\pm 0.006}$	$\underline{0.988}^{\pm 0.010}$	$\underline{0.995}^{\pm 0.004}$	$0.989^{\pm 0.017}$	$0.986^{\pm 0.019}$	$0.981^{\pm 0.025}$
	0.000	0.000	$\underline{0.996}$	$\underline{0.968}$	$\underline{0.945}$	0.966	0.932	0.931	0.902
rounD	$0.699^{\pm 0.458}$	$0.700^{\pm 0.458}$	$\underline{1.000}^{\pm 0.001}$	$\underline{0.994}^{\pm 0.008}$	$\underline{0.994}^{\pm 0.009}$	$0.997^{\pm 0.005}$	$0.993^{\pm 0.011}$	$0.992^{\pm 0.012}$	$0.993^{\pm 0.011}$
(V_E and V_T)	0.000	0.000	$\underline{0.998}$	$\underline{0.936}$	0.925	0.968	0.929	$\underline{0.929}$	0.915