Active inference as a unified model of collision avoidance behavior in human drivers: Supplementary Material

Julian F. Schumann, Johan Engström, Leif Johnson, Matthew O'Kelly,
Jens Kober, Arkady Zgonnikov

1 Active Inference Framework

To model the behavior of human drivers in different scenarios, we combine Active Inference and Evidence Accumulation. For sake of either notation, we use the shorthand of $\boldsymbol{b}_{1:B} = \{b_i | i \in \{1, \dots, B\}\}$. Additionally, while the predicted beliefs about states $\widetilde{q}_s(\boldsymbol{s}_\tau)$ and observations $\widetilde{q}_o(\boldsymbol{o}_\tau)$ for $\tau > t$ depend on the current belief $q(\boldsymbol{s}_t)$ and the chosen policy $\boldsymbol{\pi}_t$ (sequence of planned actions), so that the correct notation would be $\widetilde{q}_s(\boldsymbol{s}_\tau | \boldsymbol{\pi}_t, q(\boldsymbol{s}_t))$ and $\widetilde{q}_o(\boldsymbol{o}_\tau | \boldsymbol{\pi}_t, q(\boldsymbol{s}_t))$ respectively, we will use the shorthand expressions $\widetilde{q}_s(\boldsymbol{s}_\tau)$ and $\widetilde{q}_o(\boldsymbol{o}_\tau)$ to improve readability.

1.1 Generative process and Generative model

The basic concept of active inference is the idea that a human agent does not know the actual mechanism underlying their surrounding world (called *generative process*), but instead relies on an internal model (the so called *generative model*) approximating the true world. These describe how the true state of the world ($\boldsymbol{\eta}_t$), a state of the agents' belief about the world ($\boldsymbol{s}_t$)), the agent's actions ($\boldsymbol{a}_t$), and the observations of the true world $\boldsymbol{o}_t$ at a time t interact with each other. Here, the *generative process* consists out of two parts:

- The true transition probability $\widehat{p}(\boldsymbol{\eta}_{t+1} | \boldsymbol{\eta}_t, \boldsymbol{a}_t)$. It describes how a certain action at time t influences the future world state.
- The true observation probability $\widehat{p}(\boldsymbol{o}_t | \boldsymbol{\eta}_t)$. It describes how likely it is that a certain observation can be perceived given the current state of the world.

Meanwhile, the *generative model* consists out of two parts as well.

- The internal state transition probability $p(\boldsymbol{s}_{t+1} | \boldsymbol{s}_t, \boldsymbol{a}_t, \boldsymbol{\theta}_s)$, which might be parameterized by $\boldsymbol{\theta}_s$.
- The internal observation probability $p(\boldsymbol{o}_t | \boldsymbol{s}_t, \boldsymbol{\theta}_o)$, which might be parameterized by $\boldsymbol{\theta}_o$.

The second main idea of active inference is that an agent is uncertain about its belief about the world $\boldsymbol{s}_t$, meaning that instead of a single values, we instead assume that the agent holds a probabilistic belief, denoted by $q(\boldsymbol{s}_t)$. In general, the agent could also have some uncertainty regarding the generative model itself (i.e., there is a probabilistic belief $q(\boldsymbol{\theta})$ about the *general model*'s parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_o, \boldsymbol{\theta}_{\boldsymbol{s}}\}$), but we will not include this assumption in favor of a computationally more efficient model. We represent stochastic beliefs $q(\boldsymbol{s}_t)$ by $N = 75$ equally likely representative samples $\boldsymbol{S}_t = \boldsymbol{s}_{t,1:N}$.

Given a world with the previous state $\boldsymbol{\eta}_{t-1}$, an belief $\boldsymbol{S}_{t-1}$, and a chosen action $\boldsymbol{a}_{t-1}$, we can update those in the following way:

1. We randomly sample our future world state $\boldsymbol{\eta}_t$ using the transition probability of the *generative process*:

$$\boldsymbol{\eta}_t \sim \widehat{p}(\boldsymbol{\eta}'|\boldsymbol{\eta}_{t-1}, \boldsymbol{a}_{t-1}) \tag{1}$$

2. We generate the observations $\boldsymbol{o}_t$ that the agent makes. Under the assumption that the uncertainties in the *generative process* are negligible compared to the uncertainties in the *generative model*, we use the expected value:

$$\boldsymbol{o}_t = \mathbb{E}_{\widehat{p}(\boldsymbol{o}'|\boldsymbol{\eta}_t)} \boldsymbol{o}' \tag{2}$$

3. We lastly have to update the internal belief of the agent. Here, using variational inference, we get:

$$\begin{aligned} q(\boldsymbol{s}_t) &\propto p(\boldsymbol{o}_t|\boldsymbol{s}_t, \boldsymbol{\theta}_o)\, \mathbb{E}_{q(\boldsymbol{s}_{t-1})} p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}, \boldsymbol{a}_{t-1}, \boldsymbol{\theta}_s) \\ &\propto p(\boldsymbol{o}_t|\boldsymbol{s}_t, \boldsymbol{\theta}_o)\, q_A(\boldsymbol{s}_t) \end{aligned} \tag{3}$$

   To apply this to our sample based belief representation, we first generate the updated samples $\boldsymbol{S}_{A,t} = \boldsymbol{s}_{A,t,1:N}$ that represent $q_A(\boldsymbol{s}_t)$ and follow from the internal state transition probability $p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}, \boldsymbol{a}_{t-1}, \boldsymbol{\theta}_s)$, with $\boldsymbol{s}_{A,t,n} \sim p(\boldsymbol{s}'|\boldsymbol{s}_{t-1,n}, \boldsymbol{a}_{t-1}, \boldsymbol{\theta}_s)$. One can then get an explicit approximation for $q_A(\boldsymbol{s}_t)$ using a Kernel Density Estimate (KDE) based on $\boldsymbol{S}_{A,t}$, which can be expressed as

$$q_A(\boldsymbol{s}_t) \approx \frac{1}{N} \sum_{n=1}^{N} \mathcal{N}\left(\boldsymbol{s}_t|\boldsymbol{s}_{A,t,n}, \boldsymbol{\Sigma}_{A,t}\right) . \tag{4}$$

   At this point, with q_A and $p(\boldsymbol{o}|\boldsymbol{s})$ both known, one could generate the updated belief samples $\boldsymbol{S}_t$ with a form of the Metropolis Hastings algorithm, which is however computationally inefficient.

   Instead, a faster update is possible, as long as there exists a bijective mapping $\mathcal{L}$ under which $p(\boldsymbol{o}|\boldsymbol{s})$ can be expressed as a normal distribution:

$$\begin{aligned} p(\boldsymbol{o}_t|\boldsymbol{s}_t, \boldsymbol{\theta}_o) &= \mathcal{N}\left(\mathcal{L}(\boldsymbol{o}_t)|\mu(\mathcal{L}(\boldsymbol{o}_t)) + \mathcal{L}(\boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{b}), \boldsymbol{\Sigma}(\mathcal{L}(\boldsymbol{o}_t))\right) |\det J_{\mathcal{L}}(\boldsymbol{o}_t)| \\ &= \mathcal{N}\left(\mathcal{L}(\boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{b})|\mathcal{L}(\boldsymbol{o}_t) - \mu(\mathcal{L}(\boldsymbol{o}_t)), \boldsymbol{\Sigma}(\mathcal{L}(\boldsymbol{o}_t))\right) |\det J_{\mathcal{L}}(\boldsymbol{o}_t)| \end{aligned} \tag{5}$$

   Then, instead doing our update over $\boldsymbol{s}$ directly, we can instead do it over $\boldsymbol{s}_{\mathcal{L},t} = \mathcal{L}(\boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{b})$. Namely, we can use a KDE to instead calculate $q_{\mathcal{L},A}(\boldsymbol{s}_{\mathcal{L},t})$, where with $\boldsymbol{s}_{\mathcal{L},A,t,n} = \mathcal{L}(\boldsymbol{A}\boldsymbol{s}_{A,t,n} + \boldsymbol{b})$ we get

$$q_{\mathcal{L},A}(\boldsymbol{s}_{\mathcal{L},t}) \approx \frac{1}{N} \sum_{n=1}^{N} \mathcal{N}\left(\boldsymbol{s}_{\mathcal{L},t}|\boldsymbol{s}_{\mathcal{L},A,t,n}, \boldsymbol{\Sigma}_{\mathcal{L},A,t}\right) . \tag{6}$$

   Substituting (6) and (5) then allows us to express $q_{\mathcal{L}}(\boldsymbol{s}_{\mathcal{L},t})$ as a Gaussian multi mixture model, from which sampling $\boldsymbol{S}_{\mathcal{L},t}$ is trivial:

$$\begin{aligned} & q(\boldsymbol{s}_t) \propto p(\boldsymbol{o}_t|\boldsymbol{s}_t, \boldsymbol{\theta}_o)\, q_A(\boldsymbol{s}_t) \\ \iff\ & q_{\mathcal{L}}(\boldsymbol{s}_{\mathcal{L},t}) \propto \mathcal{N}\left(\boldsymbol{s}_{\mathcal{L},t}|\mathcal{L}(\boldsymbol{o}_t) - \mu(\mathcal{L}(\boldsymbol{o}_t)), \boldsymbol{\Sigma}(\mathcal{L}(\boldsymbol{o}_t))\right) q_{\mathcal{L},A}(\boldsymbol{s}_{\mathcal{L},t}) \\ & \qquad \mathrel{\overset{\propto}{\sim}} \sum_{n=1}^{N} \mathcal{N}\left(\boldsymbol{s}_{\mathcal{L},t}|\mathcal{L}(\boldsymbol{o}_t) - \mu(\mathcal{L}(\boldsymbol{o}_t)), \boldsymbol{\Sigma}(\mathcal{L}(\boldsymbol{o}_t))\right) \mathcal{N}\left(\boldsymbol{s}_{\mathcal{L},t}|\boldsymbol{s}_{\mathcal{L},A,t,n}, \boldsymbol{\Sigma}_{\mathcal{L},A,t}\right) \\ & \qquad \mathrel{\overset{\propto}{\sim}} \sum_{n=1}^{N} w_{t,n}\, \mathcal{N}\left(\boldsymbol{s}_{\mathcal{L},t}|\boldsymbol{\mu}_{t,n}, \boldsymbol{\Sigma}_t\right) \end{aligned} \tag{7}$$

   with $\boldsymbol{\mu}_o = \mathcal{L}(\boldsymbol{o}_t) - \mu(\mathcal{L}(\boldsymbol{o}_t))$

$$\boldsymbol{\Sigma}_t = \left(\boldsymbol{\Sigma}_{\mathcal{L},A,t}^{-1} + \boldsymbol{\Sigma}(\mathcal{L}(\boldsymbol{o}_t))^{-1}\right)^{-1}$$

$$
\begin{aligned}
\boldsymbol{\mu}_{t,n} &= \boldsymbol{\Sigma}_t \left(\boldsymbol{\Sigma}_{\mathcal{L},A,t}^{-1} \boldsymbol{s}_{\mathcal{L},A,t,n} + \boldsymbol{\Sigma}(\mathcal{L}(\boldsymbol{o}_t))^{-1} \boldsymbol{\mu}_o \right) \\
w_{t,n} &\propto \exp\left(\frac{1}{2} \left(\boldsymbol{s}_{\mathcal{L},A,t,n} - \boldsymbol{\mu}_o \right)^T \left(\boldsymbol{\Sigma}_{\mathcal{L},A,t} + \boldsymbol{\Sigma}(\mathcal{L}(\boldsymbol{o}_t)) \right)^{-1} \left(\boldsymbol{s}_{\mathcal{L},A,t,n} - \boldsymbol{\mu}_o \right) \right) .
\end{aligned}
\tag{8}
$$

After sampling $\boldsymbol{s}_{\mathcal{L},t,1:N}$ from the GMM, we can get our final sample $\boldsymbol{S}_t = \boldsymbol{s}_{t,1:N}$ with $\boldsymbol{s}_{t,n} = \boldsymbol{A}^{-1} \left(\mathcal{L}^{-1}(\boldsymbol{s}_{\mathcal{L},t,n}) - \boldsymbol{b} \right)$.

If $\boldsymbol{A}$ is not full rank, we will have to simply assume that the distribution $q(\boldsymbol{s}_t)$ orthogonally to the image of $\boldsymbol{A}$ will be identical to the one in $q_A(\boldsymbol{s}_t)$. However, it must be noted that in our simulations $\boldsymbol{A}$ will be the identity matrix.

In the update of our model, we use a time step size of $\Delta t = 0.2\,\text{s}$

1.2 Expected Free Energy

After the agent updates its internal belief $q(\boldsymbol{s})$, it then has to generate a new policy $\boldsymbol{\pi}_t = \boldsymbol{a}_{t:(t+H-1)}$ (with $\boldsymbol{a}_t = [\pi_t]_1$) over a prediction horizon of H time steps (we use $H = 30$ in our implementation). In general active inference, it is postulated that such a plan is selected based on the minimization of the *expected free energy* G, here defined using the preference function $p(\boldsymbol{o})$:

$$
\begin{aligned}
G(\boldsymbol{\pi}_t | q(\boldsymbol{s}_t)) &= \sum_{\tau=t+1}^{t+H} g(\boldsymbol{\pi}_t, q(\boldsymbol{s}_t), \tau) \\
&= \sum_{\tau=t+1}^{t+H} - \underbrace{\mathbb{E}_{\widetilde{q}_o(\boldsymbol{o}_\tau) p_{\mathrm{n}}(\boldsymbol{o}_\tau)} \ln p(\boldsymbol{o}_\tau)}_{\text{Pragmatic value } g_{\text{pragm}}} - \underbrace{\left(\mathcal{H}(\widetilde{q}_o(\boldsymbol{o}_\tau)) - \mathbb{E}_{\widetilde{q}_s(\boldsymbol{s}_\tau)} \mathcal{H}(p(\boldsymbol{o}_\tau | \boldsymbol{s}_\tau, \boldsymbol{\theta}_o)) \right)}_{\text{Epistemic value } g_{\text{epist}}}
\end{aligned}
\tag{9}
$$

Here, the normative probability p_{n} is used to implement some norm-conditioned belief about the likelihood of observations (i.e., while some observation $\boldsymbol{o}_\tau$ might be kinematically equally likely to others in $\widetilde{\boldsymbol{O}}_\tau$, it might be perceived as less likely because it violates some norms, such as driving on the wrong side of the road). We call this use of p_{n} an norm-conditioned particle filter. It must be noted that it could be argued that such beliefs are better implemented directly in the state transition function, this would require a much more detailed balancing of p_{n}, complicating the fitting of the model. Therefore, for reasons of simplicity, we chose the current approach. Meanwhile, $\widetilde{q}_s(\boldsymbol{s}_\tau)$ and $\widetilde{q}_o(\boldsymbol{o}_\tau)$ correspond to the beliefs that the agents predicts for internal states and observations when following a certain policy, with

$$
\begin{aligned}
\widetilde{q}_s(\boldsymbol{s}_\tau) &= \mathbb{E}_{\widetilde{q}_s(\boldsymbol{s}_{\tau-1})} p(\boldsymbol{s}_\tau | \boldsymbol{s}_{\tau-1}, \boldsymbol{a}_{\tau-1}, \boldsymbol{\theta}_s) \\
\widetilde{q}_o(\boldsymbol{o}_\tau) &= \mathbb{E}_{\widetilde{q}_s(\boldsymbol{s}_\tau)} p(\boldsymbol{o}_\tau | \boldsymbol{s}_\tau, \boldsymbol{\theta}_o)
\end{aligned}
\tag{10}
$$

We use again our sample based approach for belief representation, with $\widetilde{\boldsymbol{S}}_\tau = \boldsymbol{s}_{\tau,1:N}$ representing $\widetilde{q}_s(\boldsymbol{s}_t)$ (with $\boldsymbol{s}_{\tau,n} \sim p(\boldsymbol{s}' | \boldsymbol{s}_{\tau-1,n}, \boldsymbol{a}_{\tau-1}, \boldsymbol{\theta}_s)$) and $\widetilde{\boldsymbol{O}}_\tau = \boldsymbol{o}_{\tau,1:N}$ approximating $\widetilde{q}_o(\boldsymbol{o}_\tau)$ (where $\boldsymbol{o}_{\tau,n} \sim p(\boldsymbol{o}' | \boldsymbol{s}_{\tau,n}, \boldsymbol{\theta}_o)$). For the initial step of $\tau = t+1$, we can assume that $\widetilde{\boldsymbol{S}}_t = \boldsymbol{S}_t$. Based on this, one can then calculate the expected free energy g at one timestep with

$$
\begin{aligned}
g(\boldsymbol{\pi}_t, q(\boldsymbol{s}_t), \tau) \approx &- \frac{\sum_{\boldsymbol{o}_\tau \in \widetilde{\boldsymbol{O}}_\tau} p_{\mathrm{n}}(\boldsymbol{o}_\tau) \ln p(\boldsymbol{o}_\tau)}{\sum_{\boldsymbol{o}_\tau \in \widetilde{\boldsymbol{O}}_\tau} p_{\mathrm{n}}(\boldsymbol{o}_\tau)} \\
&- \left(\mathcal{H}\left(\widetilde{q}_o(\boldsymbol{o}_\tau) \right) - \frac{1}{N} \sum_{\boldsymbol{s}_\tau \in \widetilde{\boldsymbol{S}}_\tau} \mathcal{H}\left(p(\boldsymbol{o}' | \boldsymbol{s}_\tau, \boldsymbol{\theta}_o) \right) \right) ,
\end{aligned}
\tag{11}
$$

where we use the approximation

$$\begin{aligned}\mathcal{H}\left(\widetilde{q}_o(\boldsymbol{o}_\tau)\right) &= \mathcal{H}\left(\mathbb{E}_{\widetilde{q}_s(\boldsymbol{s}_\tau)} p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau, \boldsymbol{\theta}_o)\right) \\ &= -\frac{1}{N} \sum_{\boldsymbol{o}_\tau \in \widetilde{\boldsymbol{O}}_\tau} \ln \left(\frac{1}{N} \sum_{\boldsymbol{s}_\tau \in \widetilde{\boldsymbol{S}}_\tau} p(\boldsymbol{o}_\tau|\boldsymbol{s}_\tau, \boldsymbol{\theta}_o) \right)\end{aligned} \tag{12}$$

To maximize G, we use the Cross Entropy Method (CEM) for model predictive control, which is an iterative method over $k \in 1, \ldots, K$ with $K = 20$:

1. We define a distribution $p_{\pi,k}(\boldsymbol{\pi}_t) = \mathcal{N}(\boldsymbol{\pi}_t|\boldsymbol{\mu}_{\pi,t,k-1}, \mathrm{diag}(\boldsymbol{\sigma}^2_{\pi,t,k-1}))$ over the policy space, from which we sample the $M = 100$ policies $\boldsymbol{\pi}_{t,k,1:M}$ from $p_{\pi,k}(\boldsymbol{\pi}_t)$ and calculate the respective expected free energy (after adjusting for pedals with f_{real}) $G_{t,k,m} = G(f_{\text{real}}(\boldsymbol{\pi}_{t,k,m}), q(\boldsymbol{s}_t))$ (see (11)).
2. We select the $\beta = 0.1 \in [0, 1]$ percent samples $\boldsymbol{\pi}_{t,k,1:M}$ with the lowest expected free energy.
3. We update our distribution, where $\boldsymbol{\mu}_{\pi,t,k}$ and $\boldsymbol{\sigma}_{\pi,t,k}$ are the mean and standard deviation of the aforementioned βM selected best plans.

For the first iteration, we choose $\boldsymbol{\mu}_{\pi,t,0} = \mathbf{0}$, while we choose as standard deviations in $\boldsymbol{\sigma}_{\pi,t,0}$ value of $5\,\mathrm{ms}^{-2}$ for accelerations a and $0.1\,\mathrm{s}^{-1}$ for steering rates ω. The final policy is then selected as

$$\boldsymbol{\pi}_t = f_{\text{real}}\left(\underset{m \in \{1,\ldots,M\}}{\operatorname{argmin}} G(f_{\text{real}}(\boldsymbol{\pi}_{t,K,m}), q(\boldsymbol{s}_t)) \right). \tag{13}$$

Here, f_{real} is used to prevent unrealistically control inputs.

1.3 Evidence Accumulation

Commonly, the policy $\boldsymbol{\pi}_t$ is re-chosen at every timestep. However, research has shown that humans tend to make decisions (such a changing preselected policies) only if there is enough evidence supporting such a decision, in a process called *evidence accumulation*. Here, we implement this concept by having the agent accumulate evidence E_t towards the need for selecting a new policy. The agent then updates its policy in the following way:

1. We our previous policy $\boldsymbol{\pi}_{t-1}$, resulting in $\widetilde{\boldsymbol{\pi}}_t$, with $[\widetilde{\pi}_t]_{1:H-1} = [\pi_{t-1}]_{2:H}$, and only optimize the last needed time step $[\widetilde{\pi}_t]_H$ with the method described in 1.2.
2. We calculate the evidence for choosing a new plan based on the normalized, negative pragmatic value (i.e., the *surprise*), with

$$\epsilon_t = \epsilon(\widetilde{\boldsymbol{\pi}}_t|q(\boldsymbol{s}_t)) = H \max_{\boldsymbol{o}'} \ln p(\boldsymbol{o}') - \sum_{\tau=t+1}^{t+H} \mathbb{E}_{\widetilde{q}_o(\boldsymbol{o}_\tau)} \ln p(\boldsymbol{o}_\tau, \boldsymbol{a}_{\tau-1}). \tag{14}$$

   We then update our accumulated surprise with $E_t = E_{t-1} + \lambda\epsilon_t$, where we use a drift rate of $\lambda = 10^{-5.9}$.
3. If we see that $E_t \geq 1$, then we optimize the full policy $\boldsymbol{\pi}_t$ using the method described in 1.2, and set $E_t = 0$. Otherwise, we use the continued policy $\widetilde{\boldsymbol{\pi}}_t$ as our current policy $\boldsymbol{\pi}_t$

2 Specific Models

While the previous section described our general framework for using active inference, this section will detail the exact *generative process* and *generative model* we used in our scenarios.

2.1 State transition probability

When implementing the model, we use for the state transition function of both the *generative process* and the *generative model* a common bicycle model B. In this, model each vehicle can be defined by three parameters, its width d, its front length l_f and its rear length l_r. Additionally, the kinematic state of each agent then consists of its position markers x and y, its longitudinal speed v, its current heading angle θ and steering angle δ ($\boldsymbol{x} = \{x, y, v, \theta, \delta\}$). Each agent is then controlled by the acceleration $a_{\text{long}} \in [-a_{\max}, a_{\max}]$ and steering rate $\omega \in [-\omega_{\max}, \omega_{\max}]$ with $a_{\max} = 8\,\text{ms}^{-2}$ and $\omega_{\max} = 1.22\,\text{s}^{-1}$ ($\boldsymbol{u} = \{a_{\text{long}}, \omega\}$). One then can get the differential equation $\dot{\boldsymbol{x}} = B(\boldsymbol{x}, \boldsymbol{u})$:

$$
\begin{aligned}
\dot{x} &= v\cos(\theta + \beta) \\
\dot{y} &= v\sin(\theta + \beta) \\
\dot{v} &= k_{\text{tire}}\, a_{\text{long}} \\
\dot{\theta} &= \frac{v}{l_f + l_r}\tan(k_{\text{tire}}\,\delta)\cos(\beta) \\
\dot{\delta} &= \begin{cases} 0 & \frac{\text{sgn}(\omega)\text{sgn}(\delta)}{k_{\text{tire}}} > 1 \\ \omega & \text{Otherwise} \end{cases} \\
\text{with } \beta &= \arctan\left(\frac{l_r}{l_f + l_r}\tan(k_{\text{tire}}\,\delta)\right)
\end{aligned}
\tag{15}
$$

Here, we use k_{tire} and $\widehat{\delta}$ to represent the limitations imposed by the tire friction:

$$
k_{\text{tire}} = \frac{a_{\max}}{\max\left\{a_{\max}, \sqrt{a_[\text{long}]^2 + \left(\frac{v^2}{l_f + l_r}\delta\right)^2}\right\}}
\tag{16}
$$

In each scenario, where we model the ego agent in interaction with the other agents $V = \{V_1, V_2, \ldots\}$, we then can generally find the control actions $\boldsymbol{a} = \boldsymbol{u}_{\text{ego}}$ and $\boldsymbol{\eta} = \boldsymbol{o} = \boldsymbol{s} = \{\boldsymbol{x}_\nu, \boldsymbol{u}_\nu | \nu \in \{\text{ego}\} \cup V\}$. We assume that the *generative process* is deterministic, which allows us to get the following, where $f_B(\boldsymbol{x}, \boldsymbol{u})$ describes the usage of Heun's methods to propagate the state forward according to equation (15):

$$
\begin{aligned}
\widehat{p}(\boldsymbol{\eta}'|\boldsymbol{\eta}, \boldsymbol{a}) = {} & \delta\left(\boldsymbol{x}'_{\text{ego}} - f_B(\boldsymbol{x}_{\text{ego}}, \boldsymbol{a})\right)\delta\left(\boldsymbol{u}'_{\text{ego}} - \boldsymbol{a}\right) \\
& \prod_{\nu \in V}\delta\left(\boldsymbol{x}'_\nu - f_B(\boldsymbol{x}_\nu, \boldsymbol{u}_\nu)\right)\delta\left(\boldsymbol{u}'_\nu - \boldsymbol{u}_{\nu,\text{preset}}\right)
\end{aligned}
\tag{17}
$$

Here, the next control inputs $\boldsymbol{u}_{\nu,\text{preset}}$ are predefined to allow the other vehicle to follow a prescribed trajectory, which depends on the scenario (see 3.1 and 3.2). Meanwhile, some uncertainty is involved in the *generative model*:

$$
\begin{aligned}
p(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{\theta}_s) = {} & \delta\left(\boldsymbol{x}'_{\text{ego}} - f_B(\boldsymbol{x}_{\text{ego}}, \boldsymbol{a})\right)\delta\left(\boldsymbol{u}'_{\text{ego}} - \boldsymbol{a}\right) \\
& \prod_{\nu \in V}\delta\left(\boldsymbol{x}'_\nu - f_B(\boldsymbol{x}_\nu, \boldsymbol{u}_\nu)\right)\mathcal{N}\left(\boldsymbol{u}'_\nu | \boldsymbol{u}_\nu, \text{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{u}})\right)
\end{aligned}
\tag{18}
$$

In our model, we assume $\boldsymbol{\sigma}_{\boldsymbol{u}} = \boldsymbol{\sigma}_{\boldsymbol{u},0} = [3\,\text{ms}^{-2}, 0.4575\,\text{s}^{-1}]$ when updating our belief (see (3)) and

$$
\boldsymbol{\sigma}_{\boldsymbol{u}} = 0.2 f_v\left(\mathbb{E}_{\widetilde{q}_o(\boldsymbol{o})} p_{\text{n}}(\boldsymbol{o})\right)\boldsymbol{\sigma}_{\boldsymbol{u},0}
\tag{19}
$$

when predicting future states during model evaluation (see (10)). Here, p_n is the weighting used in equation (11), with

$$f_v(p) = \frac{1}{2\max\left\{\min\left\{p, 0.505\right\}, 0.01\right\} - 0.01} \; : [0,1] \to [1,10] \tag{20}$$

being used to give the agent less certainty in its belief about the future state of the other vehicle if its current state violates traffic norms.

2.2 Observation probability

For the *generative process*, we assume that observations are exact.

$$\widehat{p}(\boldsymbol{o}'|\boldsymbol{\eta}) = \delta\left(\boldsymbol{o}' - \boldsymbol{\eta}\right) \tag{21}$$

For the *generative model* meanwhile, we use a observation probability that follows the style laid out in equation (5). Here, we implement the looming based perception using the bijective mapping $\{\boldsymbol{x}_\text{ego}, \boldsymbol{\varphi}\} = \mathcal{L}(\boldsymbol{x}_\text{ego}, \boldsymbol{x}_\text{OV}, \boldsymbol{u}_\text{OV}|\boldsymbol{a})$ with $\boldsymbol{\varphi} = \{\varphi, \dot{\varphi}, \ddot{\varphi}, y_\text{OV}, \theta_\text{OV}, \delta_\text{OV}, \omega_\text{OV}\}$, where looming angle φ, looming $\dot{\varphi}$, and looming rate $\ddot{\varphi}$ are calculated as:

$$\begin{aligned}
\varphi &\approx 2\arctan\left(\frac{d}{2(x_\text{OV} - x_\text{ego})}\right) \\
\dot{\varphi} &\approx -\frac{d\left(v_\text{OV}\cos(\theta_\text{OV}) - v_\text{ego}\right)}{(x_\text{OV} - x_\text{ego})^2 + \frac{1}{4}d^2} \\
\ddot{\varphi} &\approx \frac{d}{(x_\text{OV} - x_\text{ego})^2 + \frac{1}{4}d^2}\left(a_\text{long,ego} - a_\text{OV}\cos(\theta_\text{OV}) + \frac{2(x_\text{OV} - x_\text{ego})\left(v_\text{OV}\cos(\theta_\text{OV}) - v_\text{ego}\right)^2}{(x_\text{OV} - x_\text{ego})^2 + \frac{1}{4}d^2}\right).
\end{aligned} \tag{22}$$

It must be noted that this mapping is a rough one-dimensional estimate assuming that $\theta_\text{ego} \approx 0$. However, looming based perception update is the only used if the other agent is roughly in front of it, as it unreasonable to assume that the ego vehicle would perceive the other vehicle directly with their eyes if it is not in front of them. So technically, we find that:

if in the actual world state $\boldsymbol{\eta}$ where

$$\mathcal{L}(\boldsymbol{o}|\boldsymbol{a}) = \mathcal{L}(\boldsymbol{x}_\text{ego}, \boldsymbol{x}_\text{OV}, \boldsymbol{u}_\text{OV}|\boldsymbol{a}) = \begin{cases} \{\boldsymbol{x}_\text{ego}, \boldsymbol{\varphi}\} & x_\text{OV} - x_\text{ego} > l_r + l_f \\ \boldsymbol{x}_\text{ego}, \boldsymbol{x}_\text{OV}, \boldsymbol{u}_\text{OV} & \text{Otherwise} \end{cases} \tag{23}$$

As our state and observation states $\boldsymbol{s}$ and $\boldsymbol{o}$ overlap, in Equation (5), we use $\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{b} = \boldsymbol{0}$. we also have to implement the looming threshold, for which we use the function $\mu(\boldsymbol{o}_\mathcal{L})$. Given a looming threshold of $\dot{\varphi}_0 = 0.002\,15\,\text{s}^{-1}$, we can define here:

$$\mu(\boldsymbol{o}_\mathcal{L}) = \begin{cases} \mu_\text{loom}\left(\boldsymbol{o}_\mathcal{L}\right) & (x_\text{OV} - x_\text{ego} > l_r + l_f) \wedge (|\dot{\varphi}| \leq \dot{\varphi}_0) \\ \boldsymbol{0} & \text{Otherwise} \end{cases} \tag{24}$$

with

$$\mu_{\text{loom}}\left(\boldsymbol{o}_{\mathcal{L}}\right)=\begin{Bmatrix}0\\0\\0\\0\\0\\0\\\dot{\varphi}\\\ddot{\varphi}-\frac{d}{(x_{\text{OV}}-x_{\text{ego}})^2+\frac{1}{4}d^2}\,a_{\text{ego}}\\0\\0\\0\\0\end{Bmatrix}^T. \tag{25}$$

Meanwhile, we also have to define the function $\boldsymbol{\Sigma}(\boldsymbol{o}_{\mathcal{L}})=\operatorname{diag}\left(\boldsymbol{\sigma}\left(\boldsymbol{o}_{\mathcal{L}}\right)^2\right)$, where

$$\boldsymbol{\sigma}\left(\boldsymbol{o}_{\mathcal{L}}\right)=\left\{\boldsymbol{\sigma}_{\text{ego}},\boldsymbol{\sigma}_{\text{OV}}\left(\boldsymbol{o}_{\mathcal{L}}\right)\right\} \tag{26}$$

with

$$\boldsymbol{\sigma}_{\text{ego}}=\left\{0.0002\,\text{m},0.000\,001\,\text{m},0.0002\,\text{ms}^{-1},0.000\,001,0.000\,001\right\} \tag{27}$$

and

$$\boldsymbol{\sigma}_{\text{OV}}\left(\boldsymbol{o}_{\mathcal{L}}\right)=\begin{cases}\begin{Bmatrix}0.0002\,\text{m}\\0.000\,02\,\text{m}\\0.0002\,\text{ms}^{-1}\\0.0002\\0.002\\0.000\,02\,\text{ms}^{-2}\\0.002\,\text{s}^{-1}\end{Bmatrix}^T & x_{\text{OV}}-x_{\text{ego}}<l_r+l_f\\ \begin{Bmatrix}0.000\,01\\0.000\,01\,\text{s}^{-1}\\0.000\,001\,\text{s}^{-2}\\0.000\,02\,\text{m}\\0.0002\\0.002\\0.002\,\text{s}^{-1}\end{Bmatrix}^T & (x_{\text{OV}}-x_{\text{ego}}>l_r+l_f)\wedge(|\dot{\varphi}|>\dot{\varphi}_0)\\ \begin{Bmatrix}0.000\,01\\0.0043\,\text{s}^{-1}\\0.000\,43\,\text{s}^{-2}\\0.000\,02\,\text{m}\\0.0002\\0.002\\0.002\,\text{s}^{-1}\end{Bmatrix}^T & (x_{\text{OV}}-x_{\text{ego}}>l_r+l_f)\wedge(|\dot{\varphi}|\leq\dot{\varphi}_0)\end{cases} \tag{28}$$

2.3 Control limits

We also want to limit the acceleration and deceleration patterns not achievable by actual human input. To this end, we use f_{real} (equation (13)), which in our case will use two functions, f_{pedal} and f_{jerk}. f_{pedal}

prevents unrealistically fast switching between gas and brake pedals, by setting for an acceleration $a_{\text{long},\tau}$:

$$f_{\text{pedal}}(a_{\text{long},\tau}) = \begin{cases} a_0 & (a_{\text{long},\tau-1} - a_0)(a_{\text{long},\tau} - a_0) < 0 \\ a_\tau & \text{otherwise} \end{cases} \tag{29}$$

It must be noted that in our current model we assume that the acceleration observed when releasing both pedals is $a_0 = -0.1\,\text{ms}^{-2}$, to approximate the fact that with neutral pedals, wind and roll resistances will lead to some decelerations.

Meanwhile, f_{jerk} tries to implement realistic speeds at which the pedals can be pressed and released, by limiting the jerks applied:

$$f_{\text{jerk}}(a_{\text{long},\tau}) = \min\left\{a_{\text{long},\tau-1} + j_{\min}\Delta t, \max\left\{a_{\text{long},\tau}, a_{\text{long},\tau-1} + j_{\max}\Delta t\right\}\right\} \tag{30}$$

For the jerk limits, we use:

$$\begin{aligned} j_{\min} &= \begin{cases} -5\,\text{ms}^{-3} & (a_{\text{long},\tau-1} - a_{\text{long},\tau-2}) < 0 \wedge (a_{\text{long},\tau} - a_{\text{long},\tau-1}) > 0 \\ -30\,\text{ms}^{-3} & \text{otherwise} \end{cases} \\ j_{\max} &= \begin{cases} 0\,\text{ms}^{-3} & (a_{\text{long},\tau-1} - a_{\text{long},\tau-2}) > 0 \wedge (a_{\text{long},\tau} - a_{\text{long},\tau-1}) < 0 \\ \begin{cases} 5\,\text{ms}^{-3} & a_{\text{long},\tau} \geq 0 \\ 15\,\text{ms}^{-3} & a_{\text{long},\tau} < 0 \end{cases} & \text{otherwise} \end{cases} \end{aligned} \tag{31}$$

Both those functions are applied recursively, with

$$[f_{\text{real}}(\boldsymbol{\pi})]_\tau = f_{\text{pedal}}\left(f_{\text{jerk}}\left(f_{\text{pedal}}\left([f_{\text{real}}(\boldsymbol{\pi})]_{\tau-1}\right)\right)\right) \tag{32}$$

2.4 Preference function

We use the following preference function p when minimizing the expected free energy G (see (11)):

$$p(\boldsymbol{o}) = \mathcal{N}(v_{\text{ego}}|v_0, \sigma_v)\mathcal{N}(a_{\text{long,ego}}|0, \sigma_a)\mathcal{N}(\omega_{\text{ego}}|0, \sigma_\omega)p_{\text{lat}}(y_{\text{ego}})p_{\text{coll}}(\boldsymbol{o})p_{\text{safe}}(\boldsymbol{o}) \tag{33}$$

Here,

$$p_{\text{lat}}(y_{\text{ego}}) = \mathcal{T}\left(y_{\text{rel}}\left(y_{\text{ego}}\right) \mid \frac{w-d}{2}, g_{LC}, g_{LL}\right) \tag{34}$$

with the triangular function $\mathcal{T}$:

$$\mathcal{T}(x|x_0, p_1, p_2) \propto \begin{cases} \exp\left(\frac{|x|}{x_0}p_1\right) & |x| \leq x_0 \\ \exp(p_2) & \text{otherwise} \end{cases} \tag{35}$$

We also need to define the collision preference p_{coll}:

$$p_{\text{coll}}(\boldsymbol{o}_\tau) = \min\{p_{\text{coll}}(\boldsymbol{o}_{\tau-1}), f_{\text{coll}}(\boldsymbol{o}_\tau)\} \tag{36}$$

This minimum is here so that all timesteps following upon a collision are still punished, as the model itself has no collision mechanics, allowing vehicles to phase through each other. We than get the collision preference at a single timestep f_{coll}, where we have collision condition $\mathcal{C}(\boldsymbol{o}) = |y_{\text{OV}} - y_{\text{ego}}| \leq 1.15\,d \wedge$

$|x_{\mathrm{OV}} - x_{\mathrm{ego}}| \leq 1.15\,(l_f + l_r)$:

$$f_{\mathrm{coll}}(\boldsymbol{o}) = \begin{cases} \exp\left(g_C\left(0.2 + 0.8\frac{v_{\mathrm{ego}} - v_{\mathrm{OV}}\cos(\theta_{\mathrm{ego}} - \theta_{\mathrm{OV}})}{10\,\mathrm{m\,s^{-1}}}\right)\right) & \mathcal{C}(\boldsymbol{o}) \\ \begin{cases} 1 & x_{\mathrm{OV}} - x_{\mathrm{ego}} \leq l_r + l_f \\ \mathcal{N}\left(\frac{\dot{\varphi}}{\varphi}\middle|0.2\,\mathrm{s^{-1}}, \sigma_{\tau^{-1}}\right) & \text{Otherwise} \end{cases} & \text{Otherwise} \end{cases} . \tag{37}$$

Here, the mean for the normal distribution over $\tau^{-1} = \frac{\dot{\varphi}}{\varphi}$ is taken from Markkula *et al.* [1]. In the collision cases ($\mathcal{C}(\boldsymbol{o})$), we adjust the collision cost based on the collision speed, as it is likely that human agents prefere to collide with lower impact velocities, if a collision cannot be avoided.

Lastly, we define p_{safe}, where we mainly consider the feasibility of braking when in a car following scenario:

$$p_{\mathrm{safe}}(\boldsymbol{o}_\tau) = \begin{cases} \exp\left(\frac{1}{2}g_C\left(0.2 + 0.8\frac{v_{\mathrm{ego}} - v_{\mathrm{OV}}\cos(\theta_{\mathrm{ego}} - \theta_{\mathrm{OV}})}{10\,\mathrm{m\,s^{-1}}}\right)\right) & \mathcal{C}_{\mathrm{brake}}(\boldsymbol{o}) \wedge a_{\mathrm{ego,\,req}} < -a_{\max} \\ 1 & \text{Otherwise} \end{cases} \tag{38}$$

Here, the condition $\mathcal{C}_{\mathrm{brake}}$ for being in a car following scenario is defined as:

$$\mathcal{C}_{\mathrm{brake}}(\boldsymbol{o}) = (|y_{\mathrm{OV}} - y_{\mathrm{ego}}| \leq 1.15\,d) \wedge (x_{\mathrm{OV}} - x_{\mathrm{ego}} \geq (l_f + l_r)) \wedge (\mathrm{sgn}(v_{\mathrm{ego}})\,\mathrm{sgn}\,(v_{\mathrm{OV}}\cos(\theta_{\mathrm{OV}})) \geq 0) \tag{39}$$

Meanwhile, $a_{\mathrm{ego,\,req}}$ is the required deceleration applied after a reation time of t_{react} needed to avoid a collision if the other vehicle suddenly started to accelerate towards/brake in front of the ego vehicle with $a_{\mathrm{OV,test}} = \min\{a_{\mathrm{long,OV}}, a_{\mathrm{OV,min}}\}$.

$$\begin{aligned} a_{\mathrm{ego,\,req}} &= -\frac{1}{2}\frac{\max\{v_{\mathrm{ego,react}}, 0\}^2}{\max\{d_{\mathrm{ego,react}} - 1.15\,(l_f + l_r), 0\}} \\ v_{\mathrm{ego,react}} &= v_{\mathrm{ego}} + \min\{a_{\mathrm{long,ego}}, 0\}\,t_{\mathrm{react}} \\ d_{\mathrm{ego,react}} &= \left(x_{\mathrm{OV}} - \frac{1}{2}\frac{v_{\mathrm{OV}}^2}{a_{\mathrm{OV,test}}}\right) - \left(x_{\mathrm{ego}} + v_{\mathrm{ego}}\,t_{\mathrm{react}} + \frac{1}{2}\min\{a_{\mathrm{long,ego}}, 0\}\,t_{\mathrm{react}}^2\right) \end{aligned} \tag{40}$$

The preference function can then be parameterized by the eight parameters $\sigma_v = 0.5\,\mathrm{m\,s^{-1}}$, $\sigma_a = 0.1\,\mathrm{m\,s^{-2}}$, $\sigma_\omega = 0.02\,\mathrm{s^{-1}}$, $\sigma_{\tau^{-1}} = 0.125\,\mathrm{s^{-1}}$, $g_{LC} = -1000$, $g_{LL} = -5000$, $g_C = -10000$, $t_{\mathrm{react}} = 1\,\mathrm{s}$. Meanwhile, depending on the simulation, we choose $a_{\mathrm{OV,min}}$ so that the given initial distance and speed would result in stable car following, with lower bound of $a_{\min} = -a_{\max} = -8\,\mathrm{m\,s^{-2}}$. Specifically, we simulate a one-lane front-to-rear scenario with a leading other vehicle at constant velocity v_0 for multiple values of $a_{\mathrm{OV,min}}$. For each simulation, we then extract the steady-state following distance that the agent chose for following, and the corresponding time gap. When then given a velocity v_0 and desired following distance or desired time gap, we use linear interpolation to extract the corresponding value of $a_{\mathrm{OV,min}}$ from the given data points.

While our framework is aimed to be as generalizable as possible, there are still some changes in between our two models. Namely, when calculating the y_{rel} from equation (40), we have to represent that in the front-to-rear scenario both lanes go in one direction, while in the lateral incursion scenario, the left lane is designed for oncoming traffic.

3 Specific scenarios

3.1 Front-to-rear scenario

The first scenario, which models the response of a driver to the leading other vehicle suddenly braking, contains two vehicles ($V = \{\mathrm{ego}, \mathrm{OV}\}$).

Initial state

In this scenario, there are 12 initial condition, with $\boldsymbol{x}_{\mathrm{ego},0} = \{0\,\mathrm{m}, 0\,\mathrm{m}, v_0, 0, 0\}$ and $\boldsymbol{x}_{\mathrm{OV},0} = \{v_0 \Delta t_{\mathrm{tgp},0} + l_f + l_r, 0\,\mathrm{m}, v_0, 0, 0\}$, with $\Delta t_{\mathrm{tgp},0} \in \{0.5\,\mathrm{s}, 1.0\,\mathrm{s}, 1.5\,\mathrm{s}, 2.0\,\mathrm{s}, 2.5\,\mathrm{s}, 3.0\,\mathrm{s}, 3.5\,\mathrm{s}\}$ and $v_0 \in \{10\,\mathrm{ms}^{-1}, 15\,\mathrm{ms}^{-1}, 25\,\mathrm{ms}^{-1}, 35\,\mathrm{ms}^{-1}\}$. Meanwhile, we assume a lane width $w = 3.65\,\mathrm{m}$, and vehicle sizes of $d = 1.72\,\mathrm{m}$, $l_f = 2.1\,\mathrm{m}$, and $l_r = 2.1\,\mathrm{m}$.

Other vehicles behavior

In this scenario, we set $\boldsymbol{u}_{\mathrm{OV,preset}}$ so that the other vehicle will drive straight on for exactly 5 s, after which it will start to decelerate, applying a jerk of $-10\,\mathrm{ms}^{-3}$ until reaching an acceleration value of $-6\,\mathrm{ms}^{-2}$. It will keep this acceleration until it comes to a standstill.

Lateral preference

Here, we calculate the lateral relative position y_{rel} from equation (40) as:

$$y_{\mathrm{rel}}\left(y_{\mathrm{ego}}\right) = \begin{cases} y_{\mathrm{ego}} & y_{\mathrm{ego}} \leq \frac{w-d}{2} \\ \frac{w-d}{2} & \frac{w-d}{2} < y_{\mathrm{ego}} \leq \frac{w+d}{2} \\ y_{\mathrm{ego}} - w & \frac{w+d}{2} < y_{\mathrm{ego}} \end{cases} \tag{41}$$

Norm conditioning

We define the normative probability $p_{\mathrm{n}}(\boldsymbol{o})$ (see (11)) in the following way, that punishes moving into the left lane ($p = 0.02$) or leaving the road ($p = 0.01$) :

$$p_{\mathrm{n}}(\boldsymbol{o}) = \begin{cases} 1 & -\frac{w-d}{2} \leq y_{\mathrm{OV}} \leq \frac{w-d}{2} \\ 0.02 & \frac{w-d}{2} \leq y_{\mathrm{OV}} < \frac{3w-d}{2} \\ 0.01 & \text{Otherwise} \end{cases} \tag{42}$$

Given the usage of $p_{\mathrm{n}}(\boldsymbol{o})$ as a weighing function, it is excusable to not normalize it here.

3.2 Lateral incursion scenario

The second scenario also only contains out of two vehicles ($V = \{\mathrm{ego}, \mathrm{OV}\}$).

Initial state

In this scenario, there is a single initial condition, with $\boldsymbol{x}_{\mathrm{ego},0} = \{0\,\mathrm{m}, 0\,\mathrm{m}, v_0, 0, 0\}$ and $\boldsymbol{x}_{\mathrm{OV},0} = \{300\,\mathrm{m}, 0\,\mathrm{m}, v_0, \pi, 0\}$ (the initial velocities $v_0 = 17.88\,\mathrm{ms}^{-1}$ correspond to 40 mph). Lane width and vehicle size are identical to the front-to-rear scenario (see 3.1)

Other vehicle's behavior

The other vehicle's path is preprogrammed in a manner that it start turning to the left when the time to collision $((x_{\mathrm{OV}} - x_{\mathrm{ego}})/(v_{\mathrm{OV}} + v_{\mathrm{ego}}))$ falls below 5.15 s. This turn will last for 3.3 s, at which point the other vehicle's front left corner should start crossing the central lane marker, after which the other vehicle will follow a straight path. Following [2], we run this scenario in 3 different variants, where after 5.15 s seconds, we perceive $y_{\mathrm{OV}} = -0.4w$ (Steep incursion), $y_{\mathrm{OV}} = 0\,\mathrm{m}$ (Medium incursion), or $y_{\mathrm{OV}} = 0.45w$ (Shallow incursion) (see Figure 1).

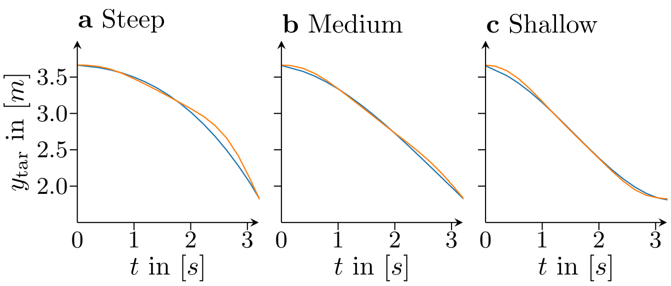

Fig. 1: The different maneuvers by the other vehicle. The orange line represents the trajectory used in our simulation, while the blue line corresponds to the original experiment [2]. $t = 0$ corresponds to the start of the maneuver, while the horizontal velocity stays constant. Afterwards, the model continues along a straight line with constant velocity.

.

Lateral preference

Here, we calculate the lateral relative position y_{rel} from equation (40) as:

$$y_{\mathrm{rel}}\left(y_{\mathrm{ego}}\right) = \begin{cases} y_{\mathrm{ego}} & y_{\mathrm{ego}} \leq \frac{w-d}{2} \\ \frac{w-d}{2} & \frac{w-d}{2} < y_{\mathrm{ego}} \leq \frac{3w-d}{2} \\ y_{\mathrm{ego}} - w & \frac{3w-d}{2} < y_{\mathrm{ego}} \end{cases} \tag{43}$$

Norm conditioning

We define the normative probability $p_{\mathrm{n}}(\boldsymbol{o})$ (see (11)) in the following way, that punishes moving into the opposite lane ($p = 0.02$), and punishes leaving the road even more ($p = 0.01$) :

$$p_{\mathrm{n}}(\boldsymbol{o}) = \begin{cases} 1 & \frac{w+d}{2} \leq y_{\mathrm{OV}} \leq \frac{3w-d}{2} \\ 0.02 & -\frac{w-d}{2} \leq y_{\mathrm{OV}} < \frac{w+d}{2} \\ 0.01 & \text{Otherwise} \end{cases} \tag{44}$$

Given the usage of $p_{\mathrm{n}}(\boldsymbol{o})$ as a weighing function, it is excusable to not normalize it here.

4 Benign scenario

This is a variation of the lateral incursion scenario, where the other vehicle just drives along its lane without any change in direction or speed. Specifically, we start the vehicles at an initial distance of 150 m, both driving with $15\,\mathrm{ms}^{-1}$. Those simulations are repeated 20 times for both the full proposed model as well as the one without the norm conditioned particle filter. As seen in Figure 2, the full model – with one very jumpy exception – does not deviate from its desired state in response to the other agent. This is much different in the model without norm conditioning, where the vehicles either move to the right, or brake, or do both. With such behavior not very realistic in everyday driving scenarios, those results highlight the need for the norm-conditioned particle filter. Of course, it must be noted that similar results without the norm-conditioned particle filter could be achieved by simply removing the prediction noise, but it was shown in the main text that that is essential for realistic collision avoidance behavior.

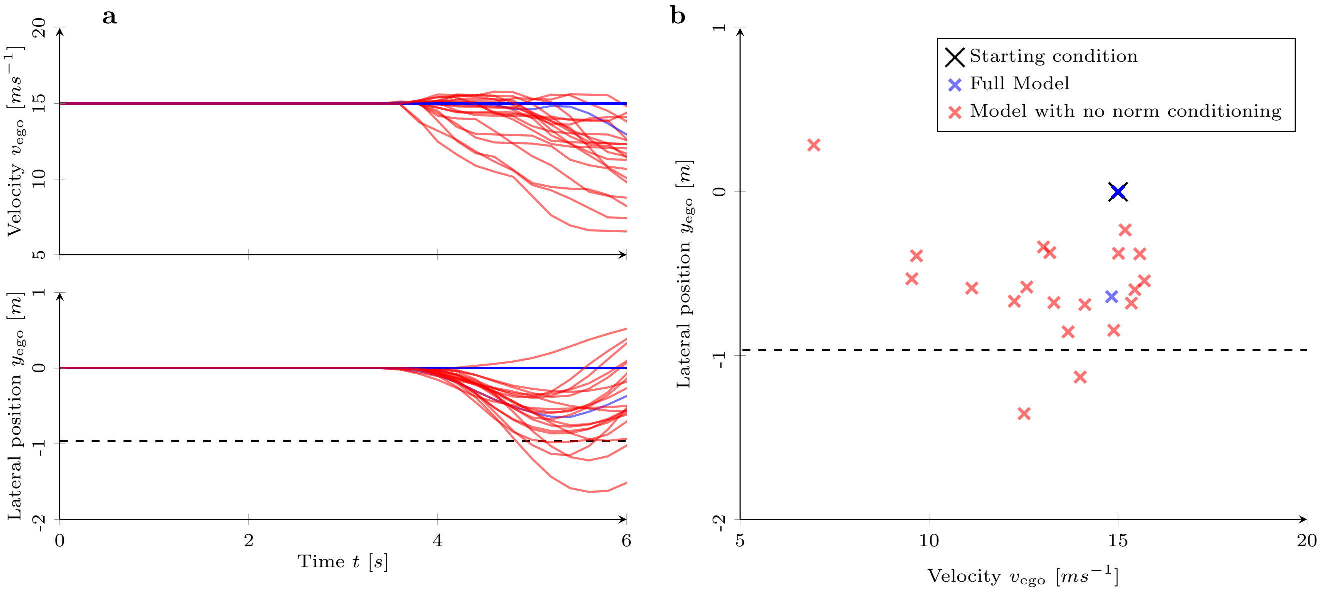

Fig. 2: Results in the 20 repetitions of the non-incursion scenario, for both the model with norm conditioned particle filter (blue) and without (red). **a**) Change of longitudinal velocity v_{ego} and lateral position y_{ego} of simulated agents over time. **b**) The kinematic state of the ego vehicle at the point in time when passing the oncoming other vehicle.

References

[1] G. Markkula, J. Engström, J. Lodin, J. Bärgman, T. Victor, A farewell to brake reaction times? Kinematics-dependent brake response in naturalistic rear-end emergencies. Accident Analysis & Prevention **95**, 209–226 (2016). https://doi.org/10.1016/j.aap.2016.07.007. URL https://www.sciencedirect.com/science/article/pii/S0001457516302366

[2] L. Johnson, J. Engström, A. Srinivasan, I. Özturk, G. Markkula, *Looking for an out: Affordances, uncertainty and collision avoidance behavior of human drivers* (2025). URL https://arxiv.org/abs/2505.14842