



## OPEN ACCESS

## EDITED BY

Tom Ziemke,  
Linköping University, Sweden

## REVIEWED BY

Fabio Tango,  
Centro Ricerche FIAT, Italy  
Alice Plebe,  
University of Trento, Italy

## \*CORRESPONDENCE

Johan Engström  
✉ jengstrom@waymo.com

RECEIVED 20 November 2023

ACCEPTED 07 March 2024

PUBLISHED 21 March 2024

## CITATION

Engström J, Wei R, McDonald AD, Garcia A,  
O'Kelly M and Johnson L (2024) Resolving  
uncertainty on the fly: modeling adaptive  
driving behavior as active inference.  
*Front. Neurobot.* 18:1341750.  
doi: 10.3389/fnbot.2024.1341750

## COPYRIGHT

© 2024 Engström, Wei, McDonald, Garcia,  
O'Kelly and Johnson. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC  
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Resolving uncertainty on the fly: modeling adaptive driving behavior as active inference

Johan Engström<sup>1\*</sup>, Ran Wei<sup>2</sup>, Anthony D. McDonald<sup>3</sup>,  
Alfredo Garcia<sup>2</sup>, Matthew O'Kelly<sup>1</sup> and Leif Johnson<sup>1</sup>

<sup>1</sup>Waymo LLC, Mountain View, CA, United States, <sup>2</sup>Department of Industrial and Systems Engineering, Texas A&M, College Station, TX, United States, <sup>3</sup>Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States

Understanding adaptive human driving behavior, in particular how drivers manage uncertainty, is of key importance for developing simulated human driver models that can be used in the evaluation and development of autonomous vehicles. However, existing traffic psychology models of adaptive driving behavior either lack computational rigor or only address specific scenarios and/or behavioral phenomena. While models developed in the fields of machine learning and robotics can effectively learn adaptive driving behavior from data, due to their black box nature, they offer little or no explanation of the mechanisms underlying the adaptive behavior. Thus, generalizable, interpretable, computational models of adaptive human driving behavior are still rare. This paper proposes such a model based on active inference, a behavioral modeling framework originating in computational neuroscience. The model offers a principled solution to how humans trade progress against caution through policy selection based on the single mandate to minimize expected free energy. This casts goal-seeking and information-seeking (uncertainty-resolving) behavior under a single objective function, allowing the model to seamlessly resolve uncertainty as a means to obtain its goals. We apply the model in two apparently disparate driving scenarios that require managing uncertainty, (1) driving past an occluding object and (2) visual time-sharing between driving and a secondary task, and show how human-like adaptive driving behavior emerges from the single principle of expected free energy minimization.

## KEYWORDS

active inference, driving behavior, epistemic action, driver model, uncertainty, visual time-sharing, driver distraction, pedestrian

## 1 Introduction

A fundamental aspect of motor vehicle driving, and locomotion in general, is to find an adequate balance between progress and caution. The main purpose of driving is typically to get to the destination as efficiently as possible. However, one also needs to make sure to get there without crashing and avoid other undesired consequences such as getting ticketed. A key challenge here is to manage the uncertainty inherent in most traffic situations. For example, will the vehicle in front brake and, if so, how hard? Is there a risk of a pedestrian suddenly appearing behind the stopped bus ahead? Being too pessimistic about how a traffic situation may play out in uncertain situations could lead to overcautious behavior or even a complete lack of progress. Moreover, non-human-like, overcautious, behavior may be surprising to other road users with potential negative safety consequences (Dinparastdjadid et al., 2023). Yet, being too assertive in an uncertain situation may incur a significant risk of collision.

Human drivers are most often able to manage this tradeoff, even in highly complex and uncertain traffic situations. Thus, understanding how human drivers manage uncertainty is critical for developing realistic and explainable computational models of their behavior. Such models have a range of applications such as explaining crash causation (Summala, 1996), representing other road users in traffic simulation (Treiber and Kesting, 2013; Igl et al., 2018; Suo et al., 2021; Markkula et al., 2023), establishing behavior benchmarks for autonomous vehicles (Engström et al., 2024), or as part of the AV software itself (Sadigh et al., 2018).

Contemporary generative AI models can learn complex driving behaviors from large quantities of data (e.g., Suo et al., 2021; Igl et al., 2022). There has also been extensive work in machine learning and robotics on developing models able to manage uncertainty through concepts like artificial curiosity and intrinsic motivation (e.g., Schmidhuber, 1991; Sun et al., 2011; Hester and Stone, 2012). However, due to the black-box nature of these models, they do not lend themselves to explaining the cognitive mechanisms underlying human adaptive behavior, which is also typically not their purpose.

There is a long-standing tradition in traffic psychology in modeling adaptive road user behavior, such as the regulation of speed and headway. Models in this tradition, often referred to as motivational models, include the risk homeostasis theory (Wilde, 1982), the zero-risk theory (Näätänen and Summala, 1976; Summala, 1988), and the task capability interface (TCI) model (Fuller, 2005) (see Lewis-Evans, 2012 for a review). In these models, excitatory “forces” such as the motivation to make progress toward the destination are balanced against inhibitory “forces” where uncertainty is typically a key component. However, these traditional motivational driving models are typically conceptual in nature, lacking mathematical rigor.

One example of an early computational model of adaptive driving behavior, of particular relevance for this paper, is the pioneering work by Senders (1967) based on the visual occlusion paradigm. In visual occlusion, subjects wear a helmet featuring a visor that, while driving, intermittently occludes the subject's view. The occlusion and viewing times may be fixed, or the subject could be given control over the occlusion time by means of a switch that opens the visor for fixed viewing time period (typically 0.25–1 s). Senders et al. found that, when occlusion times were controlled at different levels, subjects adapted their speed such that shorter occlusion times resulted in higher speed and vice versa. Conversely, when subjects had voluntary control over the occlusion time, and speed was controlled, higher speeds led to shorter occlusion times and vice versa. To explain these results, the authors developed an information theoretic model based on the idea that uncertainty builds up during the occlusion intervals and that the observed adaptive behaviors (speed and occlusion interval adjustments) reflects the attempt of the driver to control this uncertainty. Senders et al. further proposed two main sources of uncertainty in driving that need to be controlled: (1) the uncertainty about the traffic situation ahead due to loss of relevant visual information and (2) uncertainty about the vehicle's position on the road due to random disturbances in vehicle lateral control.

These early occlusion results have since been replicated for other types of driving scenarios. For example, in a recent

study, Pekkanen et al. (2017) found reduced voluntary occlusion times with reduced headway. Pekkanen et al. (2018) developed a computational model of drivers' visual sampling where uncertainty about the consequences of action (acceleration), resulting from uncertainty in state estimation, was used to directly control attention (operationalized as a voluntary opening of the occlusion visor).

Victor et al. (2005) analyzed drivers' visual time-sharing between driving and a secondary task and found, in line with the occlusion studies, that the visual demand of driving during curve negotiation led to reduced off-road glance durations. Johnson et al. (2014), based on earlier computational modeling in non-driving domains (Sprague and Ballard, 2003), proposed a model of visual time-sharing during car following based on a tradeoff between task priority and uncertainty.

Most computational models of human adaptive driving behavior have focused on visual sampling and models with a more general behavioral scope are rare. One notable exception is the model by Kolekar et al. (2020), based on zero-risk theory and the field of safe travel concept from Gibson and Crooks (1938). The model represents uncertainty about potential collisions in terms of a dynamical risk field, and was demonstrated to account for a wide range of empirical adaptive driving behavior results reported in the literature. However, the current version of the model is limited to static scenarios with no other road users present.

Another example of a generic computational model of adaptive behavior, related in some ways to Kolekar et al.'s (2020) field model, is the model developed by da Lio et al. (2023) based on the affordance competition hypothesis originating in neuroscience (Cisek, 2007). The concept of affordances were introduced by Gibson (2014), based on the earlier field of safe travel model (Gibson and Crooks, 1938). Affordances refer to opportunities for action offered to an agent by its environment. For example, a chair affords sitting and an empty adjacent lane affords overtaking the car ahead. The model by da Lio et al. implements a control architecture with a set of affordances (e.g., stay in the current lane or move to the next lane) which compete for action control, where the selection of affordances is based on a reward function which can be set to represent desired driving characteristics (e.g., avoid collisions, follow the road rules etc.). It was demonstrated that a wide range of relative complex human-like driving behaviors emerges from this general architecture, such as merging onto a highway, overtaking a lead vehicle, responding to a cut-in event and interacting with a pedestrian at a crosswalk. The affordance concept has also been used as a basis for action-based representations in machine learning-based driver models (e.g., Xiao et al., 2021).

A common denominator in most of these existing models is that adaptive driving behavior is, on the one hand, driven by a motivation to achieve goals and, on the other, by inhibitory motives such as the need to control uncertainty. The computational models reviewed above typically represent specific aspects of this phenomenon (e.g., visual sampling and time-sharing), except for a few notable developments of models with a more general scope (Kolekar et al., 2020; da Lio et al., 2023). However, a generic computational model of adaptive driving behavior, applicable across all types of scenarios and behaviors, is still lacking.

In this paper, we propose such a model based on active inference, a behavior modeling framework originating from computational neuroscience (Friston et al., 2017; Parr and Friston, 2017; Parr et al., 2022). The application of active inference, and the closely related predictive processing framework (Clark, 2013, 2015, 2023), in the automotive domain were explored in Engström et al. (2018), but only conceptually. In a series of recent papers (Wei et al., 2022, 2023a,b), we have demonstrated how computational driver models based on active inference can be implemented and learned from data, thus offering a potential “middle ground” between traditional “black box” machine learning models and mechanistic human behavior models. In this paper, we focus specifically on how active inference can provide a conceptual and computational basis for modeling human adaptive driving behavior and, in particular, how a (Bayes) optimal balance between goal-directed action and the resolution of uncertainty emerges “automatically” from the minimization of expected free energy.

In active inference, the agent estimates the free energy associated with alternative future policies,  $\pi$ , defined as sequences of actions  $\pi = a_{1:H}$  within a predefined planning time horizon  $H$ , and, at each time step, selects the action associated with the policy that has the lowest *expected free energy* (EFE). Expected free energy can be formulated in several different ways (see Friston et al., 2015, 2017; Parr et al., 2022). For present purposes we choose the formulation in Equation (1) which defines the expected free energy as the (negative) sum of a pragmatic value and an epistemic value, where the pragmatic value relates to goal seeking behavior and epistemic value to information-seeking (uncertainty-resolving) behavior, mapping conceptually to progress vs. caution or exploitation vs. exploration.

$$\begin{aligned} \text{EFE} &= G(\pi) \\ &= - \underbrace{\mathbb{E}_{Q(\mathbf{o}|\pi)}[\log P(\mathbf{o})]}_{\text{Pragmatic value}} - \underbrace{\mathbb{E}_{Q(\mathbf{s},\mathbf{o}|\pi)} D_{KL}[Q(\mathbf{s}|\mathbf{o},\pi)||Q(\mathbf{s}|\pi)]}_{\text{Epistemic value}} \quad (1) \end{aligned}$$

In Equation (1),  $\mathbf{s} = s_{1:H}$  and  $\mathbf{o} = o_{1:H}$  are sequences of future states and observations within the lookahead time horizon  $H$ ,  $\mathbb{E}$  denotes expectations, and  $D_{KL}$  denotes the Kulback-Leibler (KL) divergence, a statistical measure of the distance between two distributions.  $Q(\mathbf{o}|\pi)$  and  $Q(\mathbf{o},\mathbf{s}|\pi)$  are the agent’s belief or prediction about future observation and state-observation sequences, respectively.

In the formulation of expected free energy in Equation (1), the pragmatic value is defined based on a prior probability distribution over observations  $P(o)$  that is biased such that observations preferred by the agent have the highest probability and, hence, the highest pragmatic value. Thus, selecting policies that generate preferred observations will maximize the pragmatic value and contribute to minimizing the expected free energy. This hence implements a mechanism that generates goal-directed (or aversive) behavior, in a similar way as optimizing against a reward or cost function in optimal control or reinforcement learning (Sutton and Barto, 2018).

The epistemic value represents the value of obtaining new information that may help to resolve uncertainty in the belief about future states. This may, in turn, enable (or “open up”) new policies that maximize pragmatic value and thus realize the agent’s

preferred observations. For example, when planning to overtake a car ahead, there is typically uncertainty about whether this will lead to a conflict with a potential vehicle approaching from behind in the adjacent lane. The uncertainty can be resolved by checking the rearview mirror to verify that the lane is clear. Epistemic value scores such information-seeking actions, contributing to the overall expected free energy. In Equation (1), the epistemic value of a policy is defined as the expected KL divergence between the agent’s prior belief  $Q(\mathbf{s}|\pi)$  and posterior belief  $Q(\mathbf{s}|\mathbf{o},\pi)$  about external states associated with that policy, corresponding to (expected) Bayesian Surprise (Itti and Baldi, 2009; Dinparastdjadid et al., 2023). Intuitively, this means that epistemic value is maximized for observations that lead to a maximum change in beliefs. Epistemic value can also be expressed as in Equation (2), as the difference between the posterior predictive entropy and the expected ambiguity (Parr et al., 2022, p. 135):

$$\mathbb{E}_{Q(\mathbf{o}|\pi)} D_{KL}[Q(\mathbf{s}|\mathbf{o},\pi)||Q(\mathbf{s}|\pi)] = \mathcal{H}[Q(\mathbf{o}|\pi)] - \mathbb{E}_{Q(\mathbf{s}|\pi)} \mathcal{H}[P(\mathbf{o}|\mathbf{s})] \quad (2)$$

where  $\mathcal{H}$  denotes Shannon entropy.

In Equation (2), the posterior predictive entropy (first term;  $\mathcal{H}[Q(\mathbf{o}|\pi)]$ ) represents the uncertainty about future observations associated with a given policy. That is, a policy with a high posterior predictive entropy may yield a variety of different observations with a strong potential for gaining new information. The expected ambiguity (second term;  $\mathbb{E}_{Q(\mathbf{s}|\pi)} \mathcal{H}[P(\mathbf{o}|\mathbf{s})]$ ) represents the expected diversity of observations for a given state. Intuitively, this means that the epistemic value of a policy is discounted if the state to be visited does not generate reliable observations (e.g., due to darkness or otherwise reduced visibility). Thus, epistemic value is maximized when the expected ambiguity is zero, that is, when the observation generated by the policy is expected to completely resolve the uncertainty.

Thus, in uncertain situations, minimizing expected free energy “automatically” promotes policies with high epistemic value, generating observations expected to resolve the uncertainty. As we will see below, a single action (e.g., moving forward) often carries both pragmatic (moving closer to the goal or away from danger) and epistemic value (getting a better view to resolve uncertainty). This leads to a key distinguishing feature of active inference: Goal directed (pragmatic) and information-seeking (epistemic) value are defined in a common currency and an optimal balance between them (given the agent’s beliefs and preferences) can be established by minimizing the expected free energy.

The key objective of this paper is to demonstrate how active inference can provide a novel conceptual and computational basis for modeling adaptive driving behavior. We explore how uncertainty can be resolved “on the fly” as an integral part of the general planning objective to minimize expected free energy. Specifically, we demonstrate how a model based on the single mandate to minimize expected free energy can account for two apparently disparate adaptive driving behaviors: (1) safely passing an occluding object and (2) visual time-sharing behavior, for example when texting on a cell phone.

## 2 Materials and methods

### 2.1 Overview

A conceptual overview of our model is given in [Figure 1](#). Control actions (e.g., acceleration and steering inputs) are generated as the result of a planning process where policies  $\pi$ , constituting sequences of future actions up to the planning horizon, are selected based on identifying the policy with the minimum expected free energy (i.e.,  $\min_{\pi} G(\pi)$ ). At each time step, the first action of the selected policy is executed. The action planning is based on the driver's beliefs over hidden states  $Q(s)$  and preferences defined as priors over observations  $P(o)$ . The beliefs are continuously updated into posterior beliefs  $Q(s|o)$  based on new observations. The precision (inverse variance) of the beliefs represents the certainty of these beliefs over states. The preferences  $P(o)$  define the observations that the driver is seeking to realize through action (e.g., maintain a speed near the speed limit) and their precision represents the “strength”, or priority, of the preference (i.e., how motivated the driver is to keep the speed near the speed limit).

In action planning, a set of alternative counterfactual policies are sampled at each time step  $t$  and rolled out up to a time horizon representing the planning window. In all simulations reported here, a planning horizon of 4 s is used. For each candidate policy, the beliefs over states are propagated forward in time from the current belief based on a state transition model and the propagated beliefs are used to calculate the counterfactual pragmatic and epistemic value at each future time step,  $\tau$ , in the planning window. The prior preferences are used to score the pragmatic value based on the counterfactual observations generated, where the pragmatic value at a given future time step,  $\tau$ , is computed based on the probability of the counterfactual observation made at that time step under the preference distribution. Similarly, epistemic value is computed based on counterfactual beliefs about future states and observations. The overall expected free energy of each policy is then scored based on [Equation \(1\)](#) and [\(2\)](#) (as the negative sum of the epistemic and pragmatic values over the planning horizon), and the action at the next time step is sampled from a distribution of the first actions in the highest scoring policies. These general principles can be implemented in many different ways, where the current implementation is based on a particle filter, as described in more detail below.

The driver's generative model is represented as a discrete-time Partially Observable Markov Decision Process (POMDP) which describes how hidden environment states (e.g., pose of the ego vehicle) evolve over time depending on the ego vehicle driver's chosen policies and resulting control actions, and generate signals observed by the driver, for example the new pose of the ego vehicle or the presence of a pedestrian. Importantly, some of these state variables are not always directly observable to the driver, for example, the driver cannot observe a pedestrian when they are occluded by an object and cannot observe the road ahead while looking away from the road.

The model uses a mixture of discrete and continuous state, observation, and action variables with highly structured dependencies, which makes exact computation of the belief update and action selection intractable. We thus perform approximate

belief update and policy/action selection using a particle filter and a particle planner. At a high level, this means that the model represents uncertainty about the hidden states using an ensemble of hypothetical states, where each ensemble member represents a different possibility. The model then selects the best actions by simulating future state-action-observation trajectories under different policies using a state transition model and an observation model and scores each policy for expected free energy using [Equation \(1\)](#) and [\(2\)](#). Such a simulation-based inference method is known to approach the theoretically optimal (or exact) solution with a large number of particles ([Murphy, 2012](#)).

### 2.2 Implementation

In this section, we first describe the perception and action process of active inference agents. We then describe our particle-based implementation.

#### 2.2.1 Active inference and expected free energy

We use the standard notation for POMDP ([Kaelbling et al., 1998](#)), where  $S = \{s\}$  denotes a set of states,  $A = \{a\}$  denotes a set of actions,  $O = \{o\}$  denotes a set of observations. The active inference agent has a generative model of the environment which consists of a state transition distribution  $P(s'|s, a)$  and an observation distribution  $P(o|s)$ . We denote the environment (also known as the generative process in the active inference literature; [Parr et al., 2022](#)) that the agent interacts with as  $\mathbb{P}(o'|h, a)$ , where the time-indexed  $h_t = (o_{1:t}, a_{1:t-1})$  is the interaction history.

Upon receiving an observation  $o_t \sim \mathbb{P}(\cdot|h_{t-1}, a_{t-1})$  from the environment, the agent updates its belief, defined as a probability distribution over the hidden environment state  $Q(s_t)$ , by minimizing the variational free energy of its generative model (see [Parr et al., 2022](#)). The optimal belief distribution is known to have the form given by [Equation 5](#) ([Da Costa et al., 2020](#)):

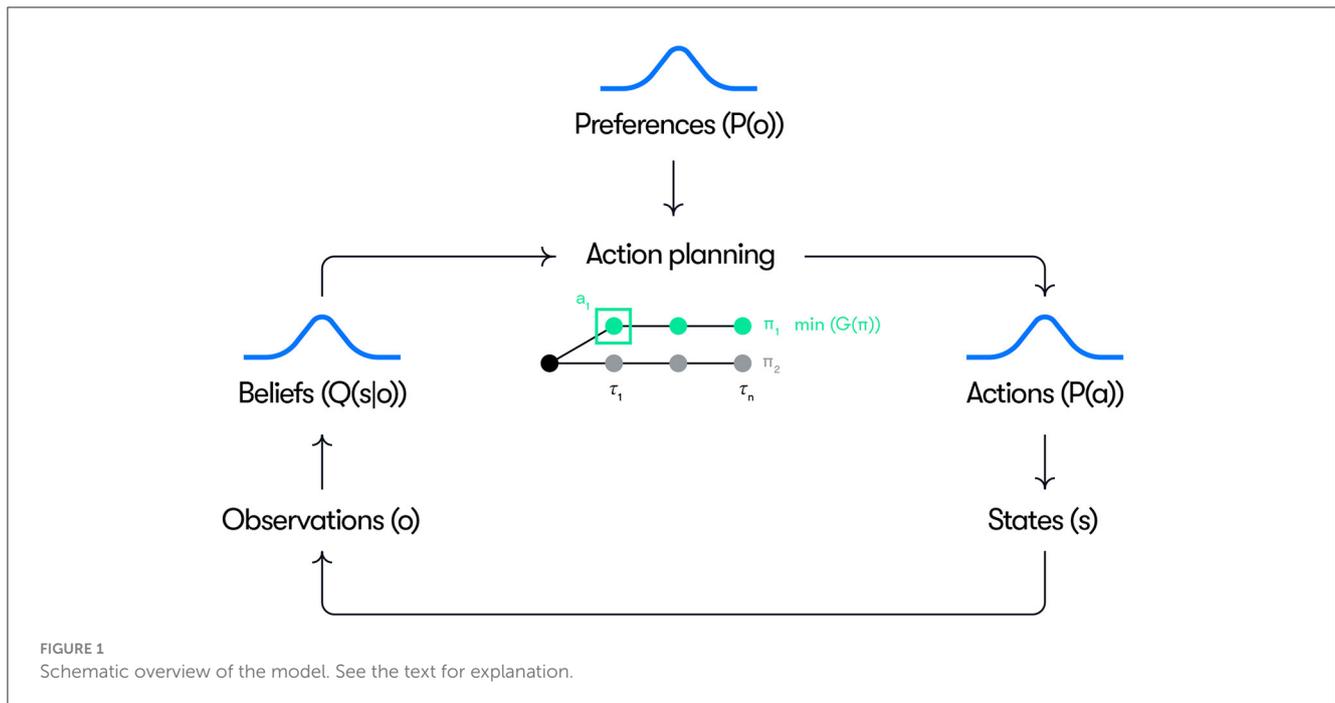
$$Q(s_t) \propto \exp(\log P(o_t|s_t) + \mathbb{E}_{Q(s_{t-1})}[\log P(s_t|s_{t-1}, a_{t-1})]) \quad (3)$$

Starting from the updated belief, the agent constructs predictions over future state-observation trajectories ( $s_{t+1:t+H}, o_{t+1:t+H}$ ) for a lookahead horizon of  $H$  time steps given a policy  $\pi = a_{t:t+H-1}$ . These predictions, defined over the lookahead time steps  $\tau \in \{t+1, \dots, t+H\}$  in the form of probability distributions, can be constructed sequentially (i.e., via rollout) as follow:

$$Q(o_{\tau}, s_{\tau}|\pi) = \mathbb{E}_{Q(s_{\tau-1}|\pi)}[P(o_{\tau}|s_{\tau})P(s_{\tau}|s_{\tau-1}, a_{\tau-1})] \quad (4)$$

The quality of each policy is scored by the expected free energy function defined in [Equation \(1\)](#) as:

$$G(\pi) = \sum_{\tau=t+1}^{t+H} \underbrace{-\mathbb{E}_{Q(o_{\tau}|\pi)}[\log P(o_{\tau})]}_{\text{Pragmatic value}} - \underbrace{\mathbb{E}_{Q(s_{\tau}, o_{\tau}|\pi)} D_{KL}[Q(s_{\tau}|o_{\tau}, \pi) || Q(s_{\tau}|\pi)]}_{\text{Epistemic value}}$$



## 2.2.2 Particle-based algorithm

We use a particle-based approach to belief representation, inference, evaluation, and planning. We describe these components below and summarize the entire process as pseudocode in [Algorithm 1](#).

Using a particle filter, we represent the ego's belief  $b_t$  over all state variables using a set of  $N$  particles  $\{\delta_1, \dots, \delta_N\}$ ,  $\delta_n \in R^{|S|}$ , where each particle consists of a vector corresponding to a realization of the state variables and a weight  $w_n \geq 0$ ,  $\sum_n w_n = 1$  representing the likeliness of the realization under the ego's belief distribution. Under this representation, the ego belief has high certainty (high precision) if all particles with high weights are similar, for example, all particles correspond to the pedestrian being present, and low certainty if all particles have even weights and are substantially different, for example, the element representing the pedestrian's position in each particle is evenly distributed on the road map.

Upon executing an action  $a_t$  and receiving a new observation  $o_{t+1}$ , we update the set of particles, including both the state vectors and weights, using a Sequential Importance Resampling (SIR) filter ([Murphy, 2012](#)). The SIR filter first samples a next state conditioned on the current state and action and updates the weights using [Equation 5](#):

$$w_{t+1,n} \propto w_{t,n} P(o_{t+1} | \delta_{t+1,n}) \quad (5)$$

where  $\delta_{t+1,n} \sim P(\cdot | \delta_{t,n}, a_t)$ .

To address the mode collapse problem common in particle filters, we use a large number of samples and apply systematic resampling so that the sampled particle weights are equidistant. Resampling of the current set of particles is only performed if the effective sample size  $N_{\text{eff}} \approx N / (1 + \sum_n w_n^2)$  is less than  $N/2$ . In this way, low weight particles are still represented.

## 2.2.3 EFE computation

Evaluating the EFE of an action sequence under a particular belief requires propagating the belief particles to compute the state and observation distributions at each counterfactual future time step and using the propagated particles to evaluate the pragmatic and epistemic values.

The particles can be easily propagated by recursively sampling the next state from the transition distribution conditioned on the previous sampled state and policy action and then sampling the next observation from the observation distribution conditioned on the sampled next state, i.e., [Equation \(4\)](#). To evaluate the pragmatic value [the first term in [Equation \(1\)](#)], we compute the average log likelihood of each observation sample under the preference distribution.

To compute the epistemic value [term 2 in [Equation \(1\)](#)], we use the decomposition of the expected information gain in [Equation \(2\)](#), which is the difference between the posterior predictive entropy and the expected ambiguity:

$$\mathbb{E}_{Q(o|\pi)} D_{KL}[Q(s|o, \pi) || Q(s|\pi)] = \mathcal{H}[Q(o|\pi)] - \mathbb{E}_{Q(s|\pi)} \mathcal{H}[P(o|s)]$$

We approximate the intractable posterior predictive entropy using a Kernel Density Estimator [similar to [Fischer and Tas \(2020\)](#)].

## 2.2.4 Model predictive control

Given each updated belief, we compute the approximately optimal EFE-minimizing actions using the Cross Entropy Method (CEM) for model predictive control ([De Boer et al., 2005](#)). CEM iteratively refines a distribution over action sequences (i.e., policies) by sampling a batch of action sequences, simulating their trajectories forward, selecting the top  $r$  percent scoring trajectories, and refitting the action distribution to the selected

**Require:** Environment model  $\mathbb{P}(o'|h,a)$ , agent transition model  $P(s'|s,a)$ , agent observation model  $P(o|s)$ , agent preference model  $P(o)$ , number of belief particles  $N$ , number of planning particles  $\tilde{N}$ , number of plans to evaluate  $M$ , percent of top plans to retain  $r$ , number of planning iterations  $K$ .

- 1: Initialize environment observation  $o_0$
- 2: Initialize belief particles  $b_0 = (\delta_{0,1:N}, w_{0,1:N})$
- 3: **for**  $t = 0:T$  **do**
- 4:   # Agent planning
- 5:   Initialize planning particles  $\tilde{\delta}_{1:\tilde{N}}$  using multinomial resampling from  $b_t$
- 6:   **for**  $k = 1:K$  **do**
- 7:     Sample  $M$  action sequences  $\{a_{m,1:H}\}_{m=1}^M$  from distribution  $P(a_{1:H})$
- 8:     **for**  $m = 1:M$  **do**
- 9:      # Evaluate plan
- 10:      Sample a sequence of state and observation particles  $(\tilde{\delta}_{1:\tilde{N},1:H}, \tilde{o}_{1:\tilde{N},1:H})$  using forward rollouts of  $P(s'|s,a)$  and  $P(o|s)$
- 11:      Compute the EFE of  $(\tilde{\delta}_{1:\tilde{N},1:H}, \tilde{o}_{1:\tilde{N},1:H})$
- 12:      Refit distribution  $P(a_{1:H})$  to the top  $r$  percent of the  $M$  action sequences by EFE
- 13:     **end for**
- 14:   **end for**
- 15:   # Choose action and update environment
- 16:   Choose  $a_t$  as the mean of  $P(a_0)$
- 17:   Sample observation  $o_{t+1} \sim \mathbb{P}(\cdot|h_t, a_t)$
- 18:   # Update belief using particle filter
- 19:   Propagate belief particles forward:  $\delta_{t+1,n} \sim P(\cdot|\delta_{t,n}, a_t)$
- 20:   Update weights:  $w_{t+1,n} \propto w_{t,n} P(o_{t+1}|\delta_{t+1,n})$
- 21:   Apply systematic resampling on particles and weights if  $N/(1 + \sum_n w_n^2) \leq N/2$
- 22: **end for**

Algorithm 1. Simulation of particle-based active inference agent.

action sequences. This process can be understood as sampling from a distribution of action sequences in proportion to their EFE values similar to relative entropy policy search (Peters et al., 2010).

Our application of CEM has two major differences from its normal use in optimal control and trajectory optimization. First, the optimal decision for an active inference agent is based on its belief as opposed to a known state. Thus, we adapt the default CEM by treating a set of belief particles as the state. Specifically, we draw  $\tilde{N}$  particles from the belief set according to their weights as the input to the CEM planner. This allows us to compute EFE using sampled-based averaging during policy search. Second, we use a mixture of discrete (e.g., gaze direction) and continuous actions (e.g., vehicle control) in the visual time-sharing scenario, whereas the default CEM typically only optimizes continuous actions. We solve this by separately fitting the discrete and continuous variables once the best trajectory samples are selected in each iteration.

## 3 Results

This section describes the simulation results from applying our model to two different driving scenarios that require the control of uncertainty: (1) passing an occluding object and (2) visually time-sharing gaze between driving and a secondary task. In the first scenario, the uncertainty concerns the potential presence of a pedestrian hidden behind the object who may step into the ego vehicle's path. In the second scenario, uncertainty about the lateral position of the vehicle in the lane builds up during glances away from the road due to disturbances such as wind gusts and an uneven road surface (Senders, 1967).

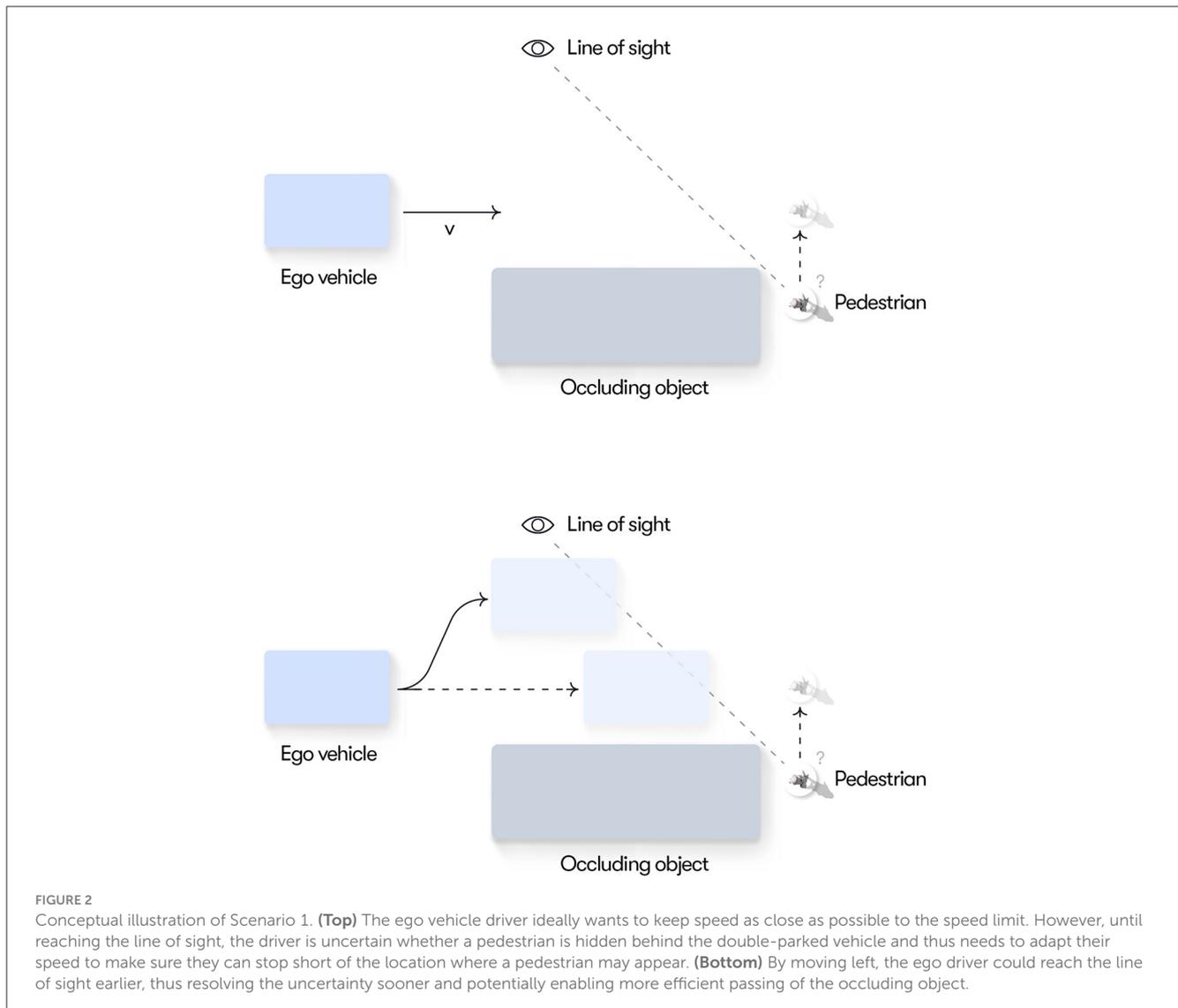
### 3.1 Scenario 1: passing an occluding object

In this scenario, the ego vehicle approaches a large occluding object (e.g., a stopped bus) and there is uncertainty about whether a pedestrian, potentially hidden behind the object, will encroach into the ego vehicle's path (see Figure 2). Managing uncertainties around occlusions is a key behavior that must be mastered both by human drivers and autonomous vehicles and thus an interesting first use case for our model. We make the simplifying assumptions that a single pedestrian is the only possible obstacle that could be hidden behind the object and that, if a pedestrian is present, it will always step out in the road and cause a potential conflict with the ego vehicle. We further assume that the pedestrian can only be present at a given point along the horizontal ( $x$ ) dimension so that it, if present, always becomes visible when it gets into the line of sight of the ego vehicle driver (see Figure 2). These assumptions simplify the current model implementation but they do not impose any fundamental limitations on the general modeling framework.

We further assume that, as a default, the ego vehicle driver prefers to keep the speed at the speed limit to maximize progress while respecting the rules of the road. However, if the driver believes that there is a risk for a hidden pedestrian encroaching into their path, they need to adapt their speed to be able to stop well ahead of the pedestrian (to meet the preference of conflict avoidance) without harsh braking (to meet their preferences for deceleration comfort). When the ego vehicle reaches the line of sight, the driver's uncertainty about the presence of the pedestrian is resolved and, if no pedestrian is present, they can speed up again to the preferred speed. Furthermore, as shown in Figure 2, bottom, by moving left in its lane the ego driver can reach the line of sight earlier, thus resolving the uncertainty that hinders progress and enabling a potentially faster trajectory past the occluding object. The general goal of the current simulation is to demonstrate how these adaptive driving behaviors emerge solely on the basis of minimizing expected free energy.

#### 3.1.1 Model specification

The state, observation and action variables in the model are listed in Table 1. The ego vehicle kinematic states are denoted as  $s^{ego} = [x, v_x, a_x, y, v_y, a_y]$  and are assumed to follow linear dynamics. In this scenario, we model the ego vehicle simply as a



point mass which was motivated by the desire to keep the model as simple as possible. Since the source of uncertainty here is the presence or absence of the pedestrian, a point mass dynamics model is sufficient to illustrate the key model principles of interest. By contrast, in Scenario 2, where the main source of uncertainty is the vehicle position on the road, we replaced the point mass model with a bicycle model, as further described below. The pedestrian's position  $s^{ped} = [x, y]$  is determined by a context variable,  $I$ , denoting whether the pedestrian is present or not present. If the pedestrian is present, its  $x$  and  $y$  positions will be set next to the occluding object. Otherwise, its  $x$  and  $y$  position will be set to a null value far away from the ego vehicle (e.g.,  $-1,000$ ).  $C$  represents whether there is a conflict between the ego and the pedestrian or if the vehicle exits the lane, in which case  $C$  is set to 1, and 0 otherwise. A conflict is defined as the pedestrian being present and the longitudinal distance between the ego and the pedestrian is less than a safe distance (2 m).

As described above, the main source of uncertainty in this scenario is that the presence or absence of the pedestrian cannot

be determined before the vehicle reaches the line of sight. Geometrically, as shown in Figure 2, the pedestrian is occluded if it is behind (i.e., with smaller  $x$  coordinate) the line of sight connecting the ego vehicle and the upper right tip of the occluding object (since we are modeling the ego vehicle as a point mass, this is also the reference point from which the observation is made). For the observation space, we introduce a variable  $o^I$  to represent the context observed by the ego.  $o^I$  is set to "observed" when the ego vehicle enters the line of sight, otherwise, it is set to "not observed". The rest of the observation variables share the same semantics and dimensionality as the underlying state variables, except when the pedestrian is occluded. In this case, the observations of the pedestrian's position,  $o^{ped}$ , are set to null. Hence, as long as there is no occlusion, we assume no uncertainty over the observations and that, hence, the ego can observe both its own and the pedestrian's kinematics states exactly.

As described above, the model's preferences are defined as priors over observations. The preferences of the model used in Scenario 1 and their default values are given in Table 2. The

TABLE 1 State, observation and action variables in the POMDP for Scenario 1.

Variable name	Symbol	Values	Type
<b>States</b>			
Pedestrian context	$I$	Not present (0), present (1)	Discrete
Is conflict or lane exit	$C$	Yes (1), no (0)	Discrete
Pedestrian position	$s^{ped}$	xy position	Continuous
Ego kinematics	$s^{ego}$	xy position, xy speed, xy acceleration	Continuous
<b>Observations</b>			
Pedestrian context observation	$o^I$	Not observed (1), observed (2)	Discrete
Conflict observation	$o^C$	Yes (1), no (0)	Discrete
Pedestrian position observation	$o^{ped}$	xy position, null	Continuous
Ego kinematics observation	$o^{ego}$	xy position, xy speed, xy acceleration	Continuous
<b>Action</b>			
Ego control	$a^{ego}$	xy acceleration	Continuous

TABLE 2 Preference priors for Scenario 1.

Preference prior	Specification	Default values
Speed keeping	Gaussian distribution centered at the speed limit	$\mu = 10m/s$ $\sigma = 1m/s$
Lane keeping	Triangular distribution centered at 0 and bounded at the lane boundaries	N/A
Acceleration	Gaussian distribution centered at zero for accelerations (x & y)	$\mu = 0m/s^2$ $\sigma = 0.5m/s^2$
Conflict	Categorical distribution representing an absolute preference over “no-conflict”	N/A

default parameter values were set by hand and no systematic model optimization was performed. Again, the key purpose was to demonstrate the basic model principles rather than to achieve optimal performance. All simulations were run with a 200 ms time step (i.e., an update frequency of 5 Hz). Due to the stochasticity of the model, each simulation run yields a unique trace. However, for clarity, we only plot randomly selected single simulation traces.

### 3.1.2 Simulations

This section presents the results of different permutations of the occlusion scenario with the purpose to illustrate the key aspects of our model described above. We begin with the case where the ego driver can only drive straight (i.e., not move laterally, e.g., due to a narrow lane) and then extend this to allow for lateral movement that enables the driver to resolve uncertainty earlier through epistemic action (moving left to reach the line of sight earlier). In all simulations, we assume that no pedestrian is *actually* present while the ego driver model may, or may not, initially *believe* that a pedestrian could be present with a given probability.

#### 3.1.2.1 Simulation 1a: safely passing an occluding object

The purpose of this initial simulation is to show how our active inference model generates successful adaptive behavior in the occlusion scenario, finding an optimal balance between progress and caution given its set beliefs and preferences. Since, in this first simulation, the driver can only drive straight and not move laterally, the ego driver's preferences reduce to prior distributions on preferred speed, comfortable accelerations and avoiding conflicts.

We initiate the model at the speed limit (10 m/s) and a 20% belief that the pedestrian is present. We assume that this belief matches the true probability that a pedestrian is present (in other words, the generative model matches the generative process). As shown in [Figure 3A](#) (top panel), the model initially slows down and then speeds up again after the uncertainty about the presence of the pedestrian has been resolved (bottom panel) when reaching the line of sight and it can be observed that no pedestrian is present (the panel third from top).

In our model, the initial slowing down behavior results from selecting a policy that maximizes the probability of preferred observations (maintaining the preferred speed, avoiding harsh deceleration and avoiding conflicts) taking into account the initial

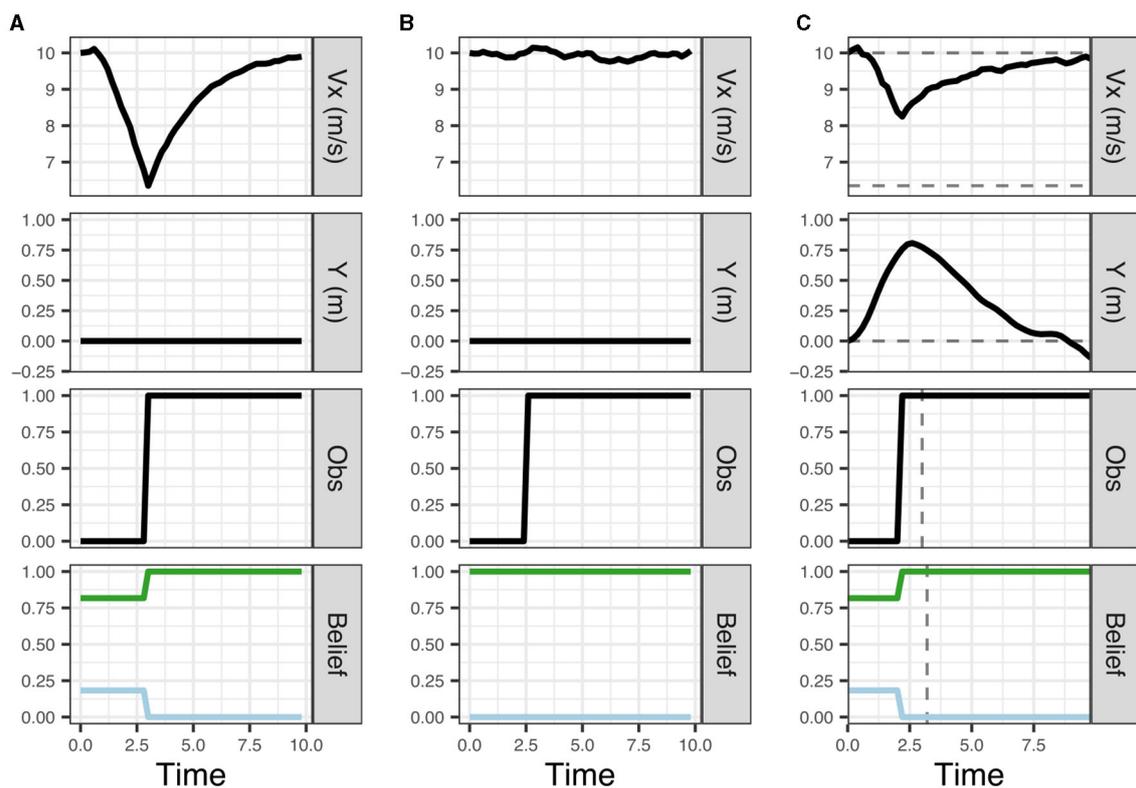


FIGURE 3

Results from simulations (A) left, (B) middle, and (C) right. In (A), the model initially believes there is a 20% chance of a pedestrian being present behind the occluding object. In (B), the model has a false certainty that no pedestrian is present. In (C), the driver takes epistemic action (change in  $Y$ ) to view the possible pedestrian sooner. Each column shows the vehicle speed profile (top;  $V_x$ ), the vehicle's lateral position (second from the top;  $Y$ ), the driver's observations of the area behind the occluding object (third from the top; Obs), and the driver's predicted beliefs about the pedestrian presence (bottom; Belief). On the belief charts, the blue line represents belief that the pedestrian is present and green the belief that the pedestrian is absent. The dashed lines in (C) show the deviations from scenarios (A, B). Note, that by taking epistemic action the driver observes the pedestrian earlier and is able to maintain a higher speed than in (A).

uncertainty in the belief about the presence of the pedestrian. Uncertainty about the presence of the pedestrian is represented by the dispersion of belief particles on the pedestrian context and position variables, and these variables are not updated by the particle filter until the ego vehicle reaches the line of sight. Once the uncertainty about the presence of the pedestrian has been resolved (when the ego reaches the line of sight) the policies with the highest pragmatic value are those that generate speed observations close to the preferred speed (i.e., the 10 m/s speed limit) and thus the model speeds up again. In this case, the behavior (policy selection) is driven solely by maximizing pragmatic value, thus minimizing expected free energy (Equation 1). The model's behavior (moving forward) also carries epistemic value as it eventually brings the pedestrian site within its line of sight, but this epistemic value here “comes for free” with the move-forward policy that maximizes the pragmatic value. Thus, in this scenario, pragmatic value does not have to be traded against epistemic value. This illustrates the key point that a given behavior often carries both pragmatic and epistemic value. The simulation in Figure 3A provides a simple first illustration for how minimizing expected free energy (in this case by maximizing expected pragmatic value) leads to adaptive behavior that is optimal given the ego driver's preferences and beliefs.

### 3.1.2.2 Simulation 1b: false certainty

We now change the ego driver model's prior belief from a 20% to a 0% belief that a pedestrian is present. Since we assume that there is a true 20% chance that a pedestrian may appear, the model now has a false (overly certain) belief to the contrary. In other words, this represents a mismatch between the prior belief and the true statistics of the situation (i.e., the generative model does not match the generative process). As shown in Figure 3B, as a result of the false belief (bottom panel) the ego maintains a constant speed near the speed limit (top panel) as it drives through the line of sight thus ignoring the potential presence of the pedestrian (the panel third from top).

In terms of our model, due to the lack of uncertainty in the beliefs about the pedestrian (and the resulting false certainty about the lack of risk for a conflict), the policy that minimizes the expected free energy is that which maximizes progress, in this case maintaining the preferred speed past the occlusion. This can be seen as an example of “optimal behavior with a sub-optimal model” (Schwartenbeck et al., 2016). Behavior is optimal given the ego driver's beliefs, but the beliefs do not match the true statistics of the environment, that is, the ego driver has a false certainty about the absence of pedestrians which leads to overly assertive behavior. Real world examples of this phenomenon include novice drivers

that have not yet learned the true statistics (generative process) of all traffic situations or a drunk driver with a biased generative model (generating overly certain beliefs), resulting in reduced risk aversion.

### 3.1.2.3 Simulation 1c: harvesting epistemic value through lateral movement

In Simulations 1a and 1b, policy and action selection was driven by pragmatic value only. However, in many real world situations, actions differ in their pragmatic and epistemic value and the optimal policy may, for example, involve an initial action yielding mainly epistemic value (e.g., turning on the light in a dark room, checking the rear view mirror) which resolves the uncertainty and enables pragmatic actions that realize the agent's goal (exit the room, overtake the vehicle ahead). To illustrate how such policies, where uncertainty is resolved "on the fly", can be found by minimizing expected free energy, we introduce the possibility for the ego vehicle to move laterally (along the  $y$ -axis). Due to the geometry of the situation (see Figure 2), the lateral movement allows the ego driver to resolve the uncertainty about the potential pedestrian earlier (by reaching the line of sight earlier; see Figure 2, bottom). This may lead to a more efficient overall path past the occluding object; even if the initial lateral movement temporarily reduces the pragmatic value (due to the agent's prior preference to stay in the middle of the lane) this may be offset by the advantage of being able to speed up earlier (thus satisfying the preference of maintaining a speed close to the speed limit).

The driver's preferences are the same as in 1a and 1b with the addition of a lane keeping preference defined by a triangular distribution centered at  $x = 0$  and bounded at the lane edges, with a lane width of 3 m. The driver's prior belief is the same as in Simulation 1a, that is, the driver believes with 20% probability that a pedestrian would appear.

The simulation results are shown in Figure 3A. The model initially slows down as in Simulation 1a (top panel) but now also moves to the left, reaching the line of sight earlier. This allows the model to resolve the uncertainty one second earlier than in Simulation 1a (around 2 s compared to 3 s in Simulation 1a; see the bottom panels of Figure 3) and is thus also able to speed up earlier than in Simulation 1a.

In our model, the leftward movement is generated by the possibility of the driver to harvest epistemic value or, equivalently, minimize the epistemic component of the expected free energy by obtaining information about the pedestrian's presence or absence when reaching the line of sight. As described above in Equation (2), the epistemic value can be decomposed as the difference between the posterior predictive entropy and the expected ambiguity (Parr et al., 2022). In the current simulation, we assume that the ego vehicle's observations of the pedestrian's presence are precise: the pedestrian is precisely unobserved before the ego vehicle reaches the line of sight and precisely observed afterwards. Thus, the expected ambiguity, is zero and does not change the epistemic value associated with different actions. On the other hand, the posterior predictive entropy, which corresponds to the diversity of the ego vehicle's future observations of the presence or absence of the pedestrian, depends on the ego vehicle's current belief uncertainty as well as the ego's position. Before the ego vehicle reaches the line of sight, it can only observe the "null" pedestrian position value;

however, when it passes the line of sight, it may observe either the "null" value if the pedestrian is not present or the pedestrian's actual position. This observation diversity increases the epistemic value of road positions beyond the line of sight.

To further illustrate this mechanism, we fixed the ego vehicle's belief to have 20% probability over the pedestrian's presence and placed the ego vehicle at different positions on the road map to calculate the epistemic value for each ego position. Figure 4 shows a plot of these ego vehicle positions, colored by the corresponding epistemic values (yellow indicates high epistemic value and purple indicates low epistemic value). The figure shows that lateral positions above and beyond the line of sight had about 75 more units of epistemic value than positions below the line of sight. This shows that there is high epistemic value to be gained by moving to the left in the lane, which temporarily trumps the pragmatic value of continuing forward in the center of the lane, leading to the selection of a policy that aims to reach the line of sight sooner. The pattern in Figure 4 can be seen as a *saliency map* representing *epistemic affordances* in terms of future locations from which observations may yield valuable, uncertainty-resolving, information. Such epistemic affordances can be seen as a specific type of traditional (Gibsonian) affordances, representing opportunities for actions that yield new valuable information. Once the vehicle passes the line of sight (i.e., physically rather than during the counterfactual planning operation), and uncertainty is resolved, the saliency map will change such that there is no longer a difference in epistemic value along the  $y$  dimension, and the driver model will shift back to the lane center, driven by the pragmatic value of maintaining the preferred central lane position.

This simulation illustrates one of the key takeaways of this paper: *Since pragmatic and epistemic value are scored in the same currency and optimized under the same objective, selecting policies based on expected free energy minimization allows the driver to find an optimal balance between assertiveness and caution* (conditioned on the driver's beliefs and preferences). This allows the driver to resolve uncertainty "on the fly" simply by moving to a location that provides a better view, an epistemic affordance, which unlocks pragmatic affordances for maintaining efficient progress.

### 3.1.2.4 Simulation 1d: epistemic value depends on the driver's beliefs

Figure 5 shows the results of a simulation where, similar to Simulation 1b, we set the model's prior belief such that it is (falsely) certain (believes with 100% probability) that no pedestrian is present. In contrast to Simulation 1c the ego vehicle no longer moves to the left but rather proceeds straight as in 1b. This happens because, from the model's (false) perspective, there is no uncertainty about the pedestrian and hence no epistemic value to be gained by moving left. Thus, the model's behavior is, again, driven by pragmatic value only. This is akin to a situation where a human driver is (falsely) certain that no other road users will enter into their path and thus fails to visually scan the road scene (or take other epistemic actions) to the extent that is warranted by the true uncertainty of the situation.

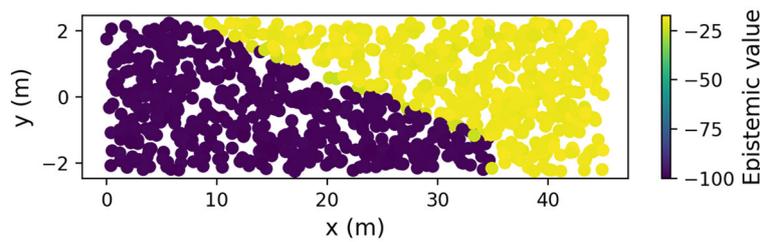


FIGURE 4 Epistemic affordance for moving left (see the text for explanation).

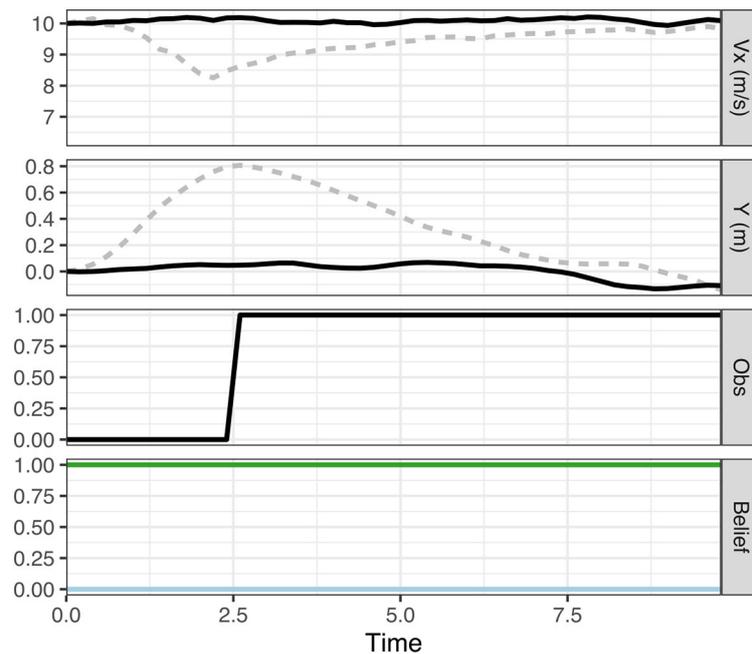


FIGURE 5 Simulation results for 1d illustrating a scenario where the driver has a strong prior belief that the pedestrian is not present. The vehicle kinematics for this scenario are illustrated in black lines. Gray dashed reference lines illustrate the comparison to the results from Simulation 1c, where uncertainty is present.

### 3.2 Scenario 2: visual time-sharing

In this scenario, we apply a slightly expanded version of the model applied in Scenario 1 to a scenario where the ego vehicle driver is performing a visually demanding secondary task that requires looking away from the road, such as texting on a cellphone. Visual distraction and time sharing with secondary tasks while driving have been extensively studied in the traffic safety and driving behavior literature, in particular due to its well-demonstrated relation to crash risk (Klauer et al., 2006, 2010; Victor et al., 2015). This means that there are plenty of empirical results available in the literature, both on visual behavior during time sharing under different driving conditions, and on the effect of visual time sharing on driving performance, which can be used to validate the model. Another reason for choosing this scenario is that it is substantially different from the occlusion scenario

in Simulation 1, thus testing the generality of the proposed modeling framework.

For simplicity, the specifics of the visual task is not modeled, only its effect on glance behavior. That is, we are assuming that the modeled glances away from the road are associated with a visual secondary task that the driver is motivated to perform. Furthermore, we assume that during the off-road glances, no visual information of the road ahead is available to the driver (i.e., ignoring peripheral vision). The driving task involves maintaining longitudinal and lateral vehicle control on a straight road with no other traffic, with a default lane width of 3 m. This set-up is akin to the visual occlusion paradigm (Senders, 1967; Pekkanen et al., 2017, 2018). However, a key difference is that we are here explicitly modeling the motivation to perform the secondary task in terms of a preference (prior distribution) for on-road vs. off-road glances.

During off-road glances, the uncertainty about the situation on the road ahead and the vehicle's position in the lane (i.e., the

two sources of uncertainty identified by Senders (1967) discussed above) builds up and the driver needs to intermittently look back to the road to update their beliefs and reset the uncertainty. In the active inference framework, such on-road glances can be seen as actions carrying epistemic value, similar to looking behind the occluding object in Scenario 1 to check for pedestrians. The drive to reduce uncertainty through epistemic actions (on-road glances) is offset by the motivation to look away from the road to perform the secondary task. At the same time, the driver is motivated to keep up speed, maintaining a central lane position and avoid exiting the lane. All Scenario 2 simulations were performed for a segment of 30 s. Each scenario was run for 10 times and the time series plots below show one randomly selected run. However, the results from all 10 runs were used to compute summary statistic metrics for comparison with human data.

### 3.2.1 Model specification

The visual time-sharing model builds on the model described under Scenario 1 and, unless otherwise stated, uses the same parameters and values. In this scenario, we replaced the point mass vehicle dynamics model in the previous scenario with a kinematic bicycle model to more accurately capture the vehicle motion constraints (Polack et al., 2017). To generate uncertainty in the ego vehicle's actual position on the road we introduce a small steering noise (0.001 rad/s) to represent the disturbance caused by uncontrollable environment effects such as uneven road surface or wind gusts.

We model the state of the driver-vehicle system using the kinematic state of the vehicle in lane coordinates  $s^{ego} = [x, y, \theta, \delta, v, a, w]$ , two binary variables  $Cl$  and  $Cr$  for whether the vehicle has crossed the left or right lane boundary, and a binary variable  $I$  for the driver's gaze direction (i.e., off-road or on-road). The lane crossing variables can be seen as representing rumble strips that generate vibrations and sound when crossed, thus conveying to the driver their lane positions even when the driver is looking off-road. The model makes decisions about two types of actions: the kinematic control of the vehicle  $a^{control}$ , and the gaze action,  $a^I$ , which represents a deterministic transition of the corresponding gaze state.

To represent the buildup of uncertainty in the model's belief about the vehicle's position and heading angle during off-road glances, we added steering noise also to the model's counterfactual steering actions, matching the noise added to the actual steering (0.001 rad/s). Thus, the driver's generative model matched the generative process. Since the road ahead was always empty, we here only address the second source of uncertainty proposed by Senders (1967) related to vehicle position in the lane. However, the model could be extended to incorporate other sources of uncertainty, for example, related to the behavior of other road users [the first type of uncertainty in Senders (1967)].

We define the observations in the same space as the state variables. However, when the driver glances off-road, we assume that the vehicle's lateral ( $y$ ) and longitudinal ( $x$ ) position and the heading angle ( $\theta$ ) cannot be observed and they are thus set to null values. This means that, due to the noise in the propagation of the beliefs about these states, the uncertainty in the beliefs continue to

grow until the driver looks back to the road and the uncertainty is reset. For simplicity, we assume that the belief update when looking back occurs instantaneously during a single time step. As a consequence, on-road glances in our simulations typically have a duration of a single time step (200 ms) which is clearly not a realistic representation of visual behavior but sufficient for demonstrating the key principles of the model. This is similar to the occlusion setup where the viewing (visor opening) times also have a fixed duration (250–1,000 ms in Senders (1967)). The state, observation and action variables used in Scenario 2 are listed in Table 3.

The model's default preferences are generally the same as in Scenario 1, but the lateral acceleration preference is replaced by a preference on steering rate and, as mentioned above, we include two categorical distributions ( $Cl$  and  $Cr$ ) to represent a strong preference for not exiting the lane. Since preferences are defined over observations, during off-road glances we do not evaluate the pragmatic value over the variables that cannot be observed when glancing off-road (i.e., vehicle lateral and longitudinal position and the heading angle, as mentioned above).

We also specify a preference over gaze direction using the log probability of on-road glances based on which we can also compute the log probability of off-road glances to ensure a normalized Bernoulli distribution. The preferences and their default values are defined in Table 4.

### 3.2.2 Simulations

#### 3.2.2.1 Simulation 2a: effects of visual time-sharing on vehicle control

The goal of this first simulation was to establish that our model could generate realistic visual time-sharing behavior and reproduce effects of visual time-sharing on vehicle control established in the literature. These effects include (1) increased lateral control variability (e.g., Zwahlen et al., 1988; Greenberg et al., 2003; Östlund et al., 2004; Engström et al., 2005; Horrey et al., 2006; Merat and Jamson, 2008; McDonald et al., 2020), (2) an increased frequency and magnitude of steering corrections (Macdonald and Hoffmann, 1980; Engström et al., 2005; Markkula and Engström, 2006), and (3) a reduction in speed (e.g., Antin et al., 1990; Engström et al., 2005; Merat and Jamson, 2008).

To establish a baseline condition representing vehicle control only (i.e., with no visual time-sharing), the preference value for on-road gaze was set to 0 (i.e., the prior probability of on-road glances equals 100%), resulting in a simulation with no eyes off-road glances. We then added a visual secondary task to be performed concurrently with the vehicle control task by lowering the on-road gaze preference from 0 to  $-7$  (hence in effect increasing the preference for looking off-road). This can be seen as endowing the model with a motivation to glance off-road to perform the visual secondary task.

Simulation results for the baseline and visual time-sharing conditions are shown in Figure 6. As can be seen, the model finds a way to visually timeshare between the driving and the secondary task. When the model is looking off-road, the uncertainty in the belief about lateral position [the standard deviation of the belief particles,  $\sigma(b_y)$ ] increases until it is reset by an on-road glance. By visual inspection of the plot, it can also be observed that the

TABLE 3 State, observation and action variables in the POMDP for Scenario 2.

Variable name	Symbol	Values	Type
<b>States</b>			
Gaze	$I$	Off-road (0), on road (1)	Discrete
Left bound crossed	$Cl$	Yes (1), no (0)	Discrete
Right bound crossed	$Cr$	Yes (1), no (0)	Discrete
Ego kinematics	$s^{ego}$	$x, y, \theta, \delta, v, a, w$	Continuous
<b>Observations</b>			
Gaze observation	$o^I$	Off road (0), on road (1)	Discrete
Left bound crossed observation	$o^{Cl}$	Yes (1), no (0)	Discrete
Right bound crossed observation	$o^{Cr}$	Yes (1), no (0)	Discrete
Ego kinematics observation	$o^{ego}$	$x, y, \theta, \delta, v, a, w$	Continuous
<b>Action</b>			
Ego kinematic control	$a^{control}$	Acceleration (a), steering rate (w)	Continuous
Gaze action	$a^I$	Off road (0), on road (1)	Discrete

TABLE 4 Preferences specifications for Scenario 2.

Preference	Specification	Default values
Speed keeping	Gaussian distribution centered at the speed limit	$\mu = 10 \text{ m/s}$ $\sigma = 1 \text{ m/s}$
Lane keeping	Triangular distribution centered at 0 and bounded at the lane boundaries	N/A
Longitudinal acceleration	Gaussian distribution centered at zero	$\mu = 0 \text{ m/s}^2$ $\sigma = 0.5 \text{ m/s}^2$
Steering rate	Gaussian distribution centered at zero	$\mu = 0 \text{ rad/s}$ $\sigma = 0.005 \text{ rad/s}$
Gaze preference	The log probability of an on-road glance based on a Bernoulli distribution of on-road/off glances.	-7
Lane exit left/right	Categorical distribution representing an absolute preference for not exiting the lane	N/A

model reduces speed and the variation in lateral vehicle position, and the amount of steering control activity increases during visual time-sharing.

Figure 7 compares summary statistic metrics of speed, lane position and steering reversals. The metrics are mean speed, standard deviation of lane position and the number of large steering reversals, which were computed for 10 simulation runs, for the baseline and visual time-sharing conditions, respectively. Steering reversal rate was defined as the number of time steps the front wheel angle exceeds 0.0025 rad/0.14 deg. As shown in the plots, the model reproduces the general effects of visual time-sharing on vehicle control in the human data reviewed above, showing a pattern of increased standard deviation of lane position (SDLP), increased rate of large steering reversals and reduced speed.

The increased lane keeping variability (SDLP) can be generally explained as the result of increased lane drifts that occur during off-road glances driven by the secondary task preference. The increasing number of steering corrections can be explained as

corrections to the lane drifts that the driver performs to align the observed lane position with their prior lane keeping preference. Finally, the speed reduction can be understood as a way to settle on an optimal balance between the preference to maximize time on the secondary task and the competing preference to keep the speed close to the speed limit. By reducing speed, the model can “buy” more time for the secondary task (since it takes longer to reach the lane boundary at lower speed), but the model cannot slow down more than mandated by the speed prior preference distribution.

To help explain the underlying visual time-sharing mechanism implemented by our expected free energy minimizing model, we visualize the tradeoff between pragmatic and epistemic value in Figure 8. As in Scenario 1, the pragmatic value is determined by the deviation of observations from the modes of the preference prior distributions and the epistemic value is solely driven by the posterior predictive entropy (first term in Equation 2) since, again, there is no ambiguity in the state-observation mapping (i.e., the second term in Equation 2 equals zero).

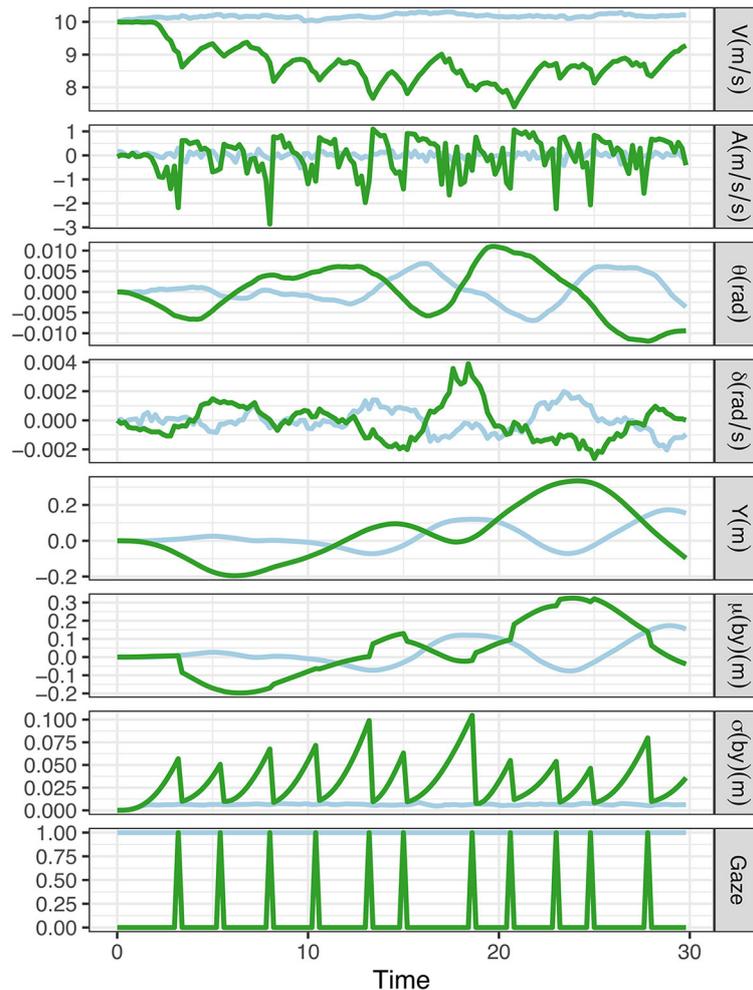


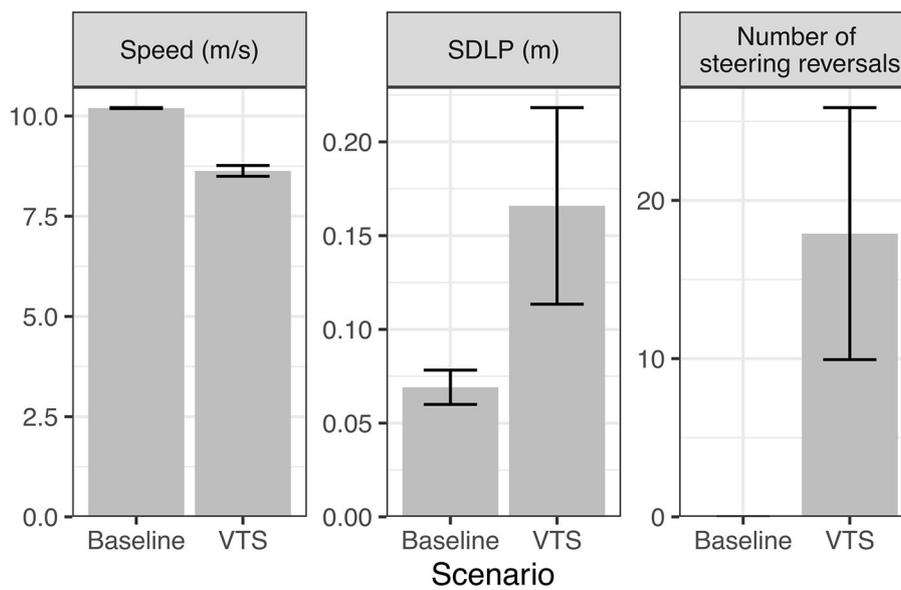
FIGURE 6

Simulation of baseline (on-road gaze preference = 0; light blue lines) and visual time-sharing between the vehicle control and the secondary task (on-road gaze preference = -7; green lines). The charts correspond to the following variables:  $V$ , speed;  $A$ , longitudinal acceleration;  $\theta$ , heading angle;  $\delta$ , steering angle;  $y$ , lateral position.

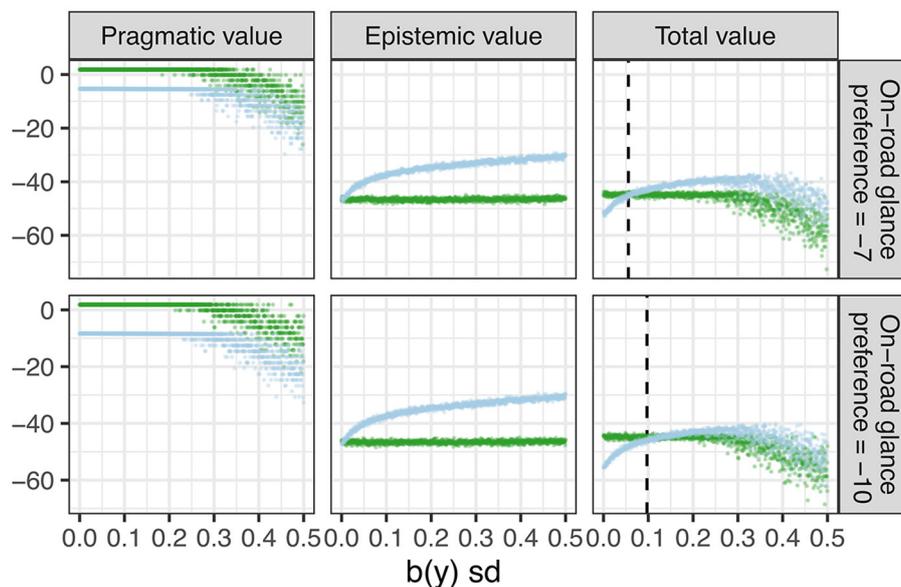
Figure 8 illustrates how the pragmatic, epistemic and total value (negative expected free energy) when gazing off-road (green) vs. on-road (blue) vary with increasing standard deviation of artificially generated belief particles representing uncertainty in lateral position. Here, the expected free energy was evaluated for each individual time step (and not for the 4 s policy as a whole). The top and bottom panels differ only in the value of the on-road preference. In the top panels we use the default value of -7 and in the bottom panels, we reduced it to -10 (representing a stronger motivation to look off road). In the plots for pragmatic and total value, we truncate the log preference probability of lane exits (a very large negative value) to -100 in order to visualize critical decision points along the vertical axis.

The left panels show that on-road glances generally have lower pragmatic value because they do not fulfill the preference of engaging with the secondary task. The total pragmatic value during on-road glances is higher for the higher on-road glance preference (-7) compared to the lower on-road gaze preference (-10). With increasing dispersion in the belief about lateral position ( $b(y)$ ), the

pragmatic values during on-road and off-road glances initially do not change significantly (it is constant for off-road glances and decreases slightly for on-road glances but this is not visible in the figure), but both start decreasing after a cut-off point, reflecting the increased counterfactual risk of exiting the lane. The middle panels show that the epistemic value of an on-road glance increases with increasing uncertainty in the belief about lateral position, which is unaffected by gaze preference (i.e., the top and bottom plots are very similar). Combining the trend in pragmatic and epistemic value, the right panels shows that the total value (negative expected free energy) during an on-road glance exceeds that of an off-road glance at a certain value of  $b(y)$  which depends on the gaze preference. At this crossover point (dashed lines in Figure 8), a policy involving a glance back to road will thus be selected. Thus, due to the stronger prior preference for on-road glances, the model in the top panel requires less epistemic value to cross over than the model in the bottom panel with weaker on-road glance preference. Hence, the former model is more prone to look back to the road than the latter when the uncertainty about lateral position increases.



**FIGURE 7** Effects of visual time-sharing (VTS; on-road gaze preference = -7) on vehicle control compared to baseline (on-road gaze preference = 0) in the current simulation. SDLP, standard deviation of lane position.



**FIGURE 8** Pragmatic value (left), epistemic value (middle), and total value (right) during eyes on road (light blue) and eyes off road (green) glances. Top panels: On-road glance preference = -7. Bottom panels: On-road glance preference -10. The dashed reference lines on the total value charts indicate the standard deviation of the belief about lateral position where the total expected free energy of the off road glance exceeds that of the on road glances (or equivalently, the total value associated with an on-road glance exceeds that of an off-road glance).

The on-road visual sampling in our model is driven by epistemic value in the same principal way as in the occluded pedestrian scenario in Simulation 1 above. The posterior predictive entropy in Equation (1) (which in our model fully determines the epistemic value) of a policy will be highest when the belief is dispersed (high uncertainty) and the model chooses to look back to the road, since this policy is expected to generate a greater

variety of possible observations compared to when continuing to look away (and observe nothing) or looking back when the belief is certain (and no new information expected is from looking back). Hence, looking back to the road in this scenario is analogous to reaching the line of sight in the occluded pedestrian scenario.

To summarize, this simulation demonstrates how visual time-sharing behavior and its effect on vehicle control emerges from

selecting policies that minimize expected free energy. Based on this sole objective, the model strikes a balance between different prior preferences (motivations for performing the secondary task and maintaining vehicle control), and visual sampling of the road ahead is driven by the epistemic value of resolving uncertainty about the current vehicle state (here primarily lateral position). It was also shown that the model explains and reproduces well-established effects of visual time-sharing on vehicle control in the driver behavior literature. In the following simulations, we further explore how the model accounts for effects of driving demand and prior preferences on visual time-sharing.

### 3.2.2.2 Simulation 2b: effects of driving demand and prior preferences on visual time sharing

As reviewed above, empirical studies has established a strong relationship between the demand (difficulty) of the vehicle control task and visual sampling of the forward roadway. In particular, Senders (1967) found that increasing driving demand by increasing the set speed led drivers to choose shorter voluntary occlusion intervals (i.e., increased viewing of the road ahead). Similarly, Victor et al. (2005) found a significantly shorter mean off-road glance duration when driving in curves compared to straight road sections, as well as a trend for more frequent off-road glances in curves.

To explore if our model was able to reproduce and explain these effects, we manipulated driving demand in terms of lane width. It is well-established that (all other things being equal) reducing lane width leads to a reduction in speed (Yagar and Van Aerde, 1983; Fitzpatrick et al., 2001).

As shown in Figure 9, reducing the lane width from 3 to 2.5 m in our simulation leads to a reduction in speed as well as shorter and more frequent off-road glances compared to the wider road, in line with the empirical results from human drivers. This is also accompanied by a reduction in lane keeping variation (SDLP), and less frequent large steering reversals which can be interpreted as a response to the need for tighter lateral control on the narrower road.

In terms of our model, the speed reduction can be explained as an attempt to limit the loss of expected pragmatic value during off-road glances, due to the increased risk of exiting the lane, by slowing down the dispersion of beliefs about future lane position. The reduced duration and increased frequency of off-road glances reflects a need for the model to sample the road more often on the narrower road as driven by epistemic value.

Figure 10 further explores the dynamics of the model's visual behavior and vehicle control and plots the glance behavior along with the overall epistemic and pragmatic value of the optimal (selected) policy at each time step (where the optimal policy is that which minimizes the expected free energy, i.e., maximizes the pragmatic plus epistemic value) during the 4 s planning horizon. The plot zooms in on the last 15 s of the simulation segment. It can be seen that, for the wider lane (left panel), each on-road glance is preceded by the selection of policies with increasing epistemic value. Eventually, due to the build-up of uncertainty in lateral position and the corresponding increasing epistemic value of an off-road glance, a policy involving an on-road glance scores the maximum total value and is selected (based on the mechanism illustrated in Figure 8). The uncertainty (and the epistemic value)

is then reset as the result of the on road glance, leading to the model performing a new off-road glance at the next time step (to maximize pragmatic value in the temporary absence of uncertainty).

For the narrow lane (left panel), the pragmatic value is generally lower than on the wider lane and is also intermittently reduced during the off-road glances due to the increased proportion of lane exit risk in the selected policies (the large dips in pragmatic value). Thus, less epistemic value is needed to trigger the selection of an on-road glance policy, leading to shorter and more frequent on-road glances (highlighted by the red points in the top charts).

Figure 11 explores the tradeoff between speed preference and glance preference by systematically varying on-road glance and speed preference precision on the wide (3 m) lane. Reducing the on-road glance preference (increasing the off-road glance preference) leads to reduced speed but only if the speed preference precision is sufficiently low to allow for speed compensation. This also leads to longer mean off-road single glance durations where, again, the effect is stronger when the speed preference precision is low. Thus, as long as the model is not strongly motivated to keep up the speed, the model adopts a strategy of slowing down to allow for longer off-road glances while still maintaining sufficient lateral control (i.e., avoid exiting the lane). By contrast, if the speed keeping cannot be sacrificed due to a high precision of the speed preference prior, the model only looks away briefly before having to look back to the road again. This shows how the visual time-sharing strategies adopted by the model depend on the relative motivations for performing the driving and secondary tasks and, more generally, how such strategies emerge naturally from the single mandate of minimizing expected free energy.

## 4 Discussion

Adaptive driving behavior has been extensively studied for the past sixty years and there exists an influential tradition of (mostly) conceptual models offering explanations for the key underlying mechanisms (e.g., Senders, 1967; Näätänen and Summala, 1976; Summala, 1988; Fuller, 2005), which has been followed by more recent computational models addressing different adaptive driving behavior phenomena (Johnson et al., 2014; Pekkanen et al., 2018; Kolekar et al., 2020; da Lio et al., 2023).

A common theme in the existing models of adaptive driving behavior is that drivers manage the progress vs. caution tradeoff by balancing the motivation to achieve goals against the control of uncertainty. Our active inference-based model implements motivations (preferences) as prior distributions over observations that the driver seeks to realize through behavior (yielding pragmatic value) and the control of uncertainty as information-seeking behavior (yielding epistemic value). Adaptive driving behavior, finding an optimal (given the agent's beliefs and preferences) balance between progress and caution, then simply emerges by selecting policies and actions that minimize the expected free energy, which can be computed as the (negative) sum of the pragmatic and epistemic value expected in the future.

Pragmatic value can be seen as corresponding to the notion of reward in standard reinforcement learning (Sutton and Barto, 2018). However, pragmatic value in the active inference context is conceptually different from the classical notion of (externally

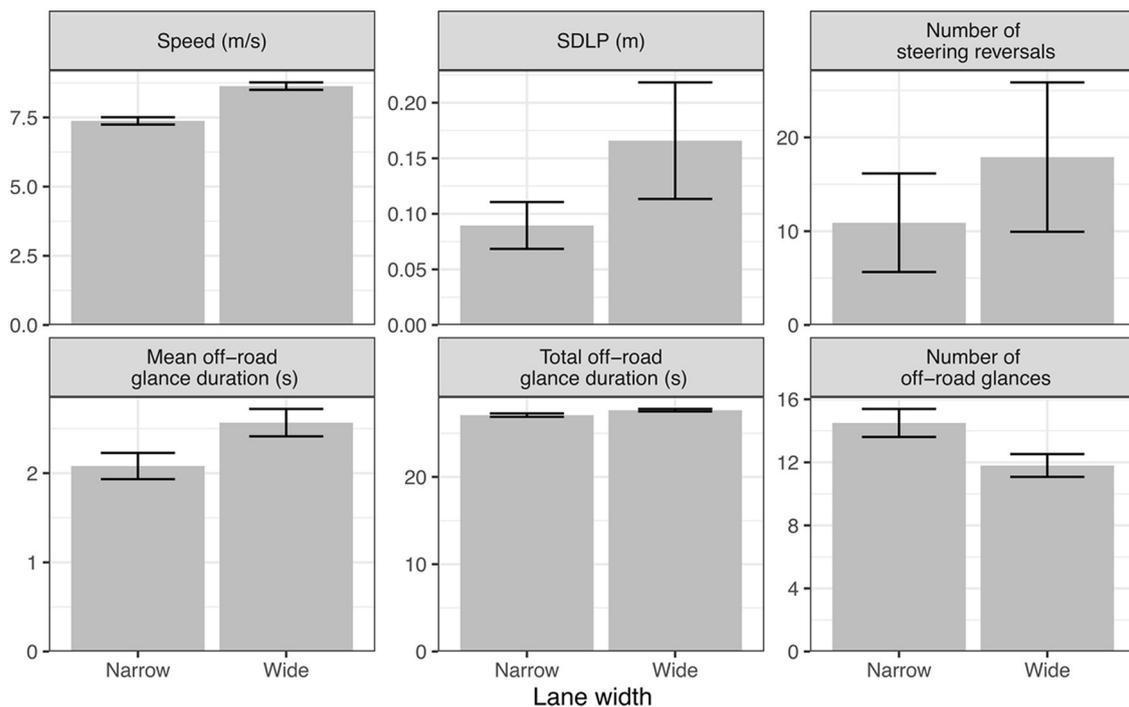


FIGURE 9  
Effects of lane width on vehicle control and glance metrics (wide = 3 m, narrow = 2.5 m).

imposed) reward in that it is internal to the agent and rests on the idea of self-evidencing, emphasizing the key role of agency: In order to maintain their existence, organisms need to seek evidence for their own generative model of the world (Friston et al., 2012), where the preferred outcomes, defined as prior beliefs over observations, are those expected to be realized through action. This aspect of active inference makes close contact with classical cybernetics models (Miller et al., 1960) and, in particular, Perceptual Control Theory (PCT), which suggests that behavior can be explained as the control of sensory input signals relative to a reference, or goal, value (Powers, 1973) (see Stephan et al., 2016 for a discussion about the similarities between PCT and active inference). The self-evidencing concept in active inference also makes close contact with ecological psychology (Gibson, 2014) and embodied and enactive approaches to cognitive science (e.g., Clark, 2013, 2015, 2023; Kiverstein et al., 2022).

Our model also has similarities with the affordance competition hypothesis (Cisek, 2007) and its implementation in the computational driver behavior model by da Lio et al. (2023). Specifically, both models are based on the continuous selection of action representations biased by value. The policies in our model are action sequences that get selected if (1) the environment offers the agent an opportunity to perform them (i.e., an affordance) and (2) they generate preferred (valuable) observations. One key difference is that, in our model, policy and action selection is additionally driven by epistemic affordances, that is, opportunities for actions that generate new information which could unlock new pragmatic (traditional) affordances.

A key feature in our model is that the strength of the preferences is controlled by the precision (inverse variance) of the preference

priors. Thus, when a preference is tuned to high precision, only a narrow range of observations yield high pragmatic value and the model will tend to select policies that do not deviate from this range. This offers a mechanism for protecting a given preference (e.g., keeping the speed near the speed limit) against other conflicting goals (e.g., looking away from the road), and the current model demonstrates how these factors interact in producing different tradeoffs in adaptive behavior (Figure 11). This can be seen as a potential mechanism underlying *cognitive control*, that is, the ability to adaptively prioritize and focus behavior on the currently most relevant goals and protect against distractions (Miller and Cohen, 2001; Engström et al., 2017).

Another key aspect of our model is the resolution of uncertainty in beliefs through epistemic actions. This leads to actions seeking novel, newsworthy, information yielding epistemic value. In the model simulations described above, such epistemic actions included taking a wider turn to get a better view behind an occluding object to resolve uncertainty about conflicts with a potentially hidden pedestrian (Scenario 1), and looking back to the road to resolve uncertainty about one's position in the lane which has accumulated during the off-road glance (Scenario 2). As shown in this paper, these apparently disparate behaviors both emerge from the same expected free energy minimizing objective.

Epistemic actions may unlock more efficient realizations of the preferred observations (pragmatic value), in our case, being able to safely speed up past the occluding object earlier (Scenario 1) and enabling a new off-road glance after the uncertainty about the lane position has been reset (Scenario 2). This seamless interplay between pragmatic and epistemic value can be seen as the core feature of our model (and the active inference framework

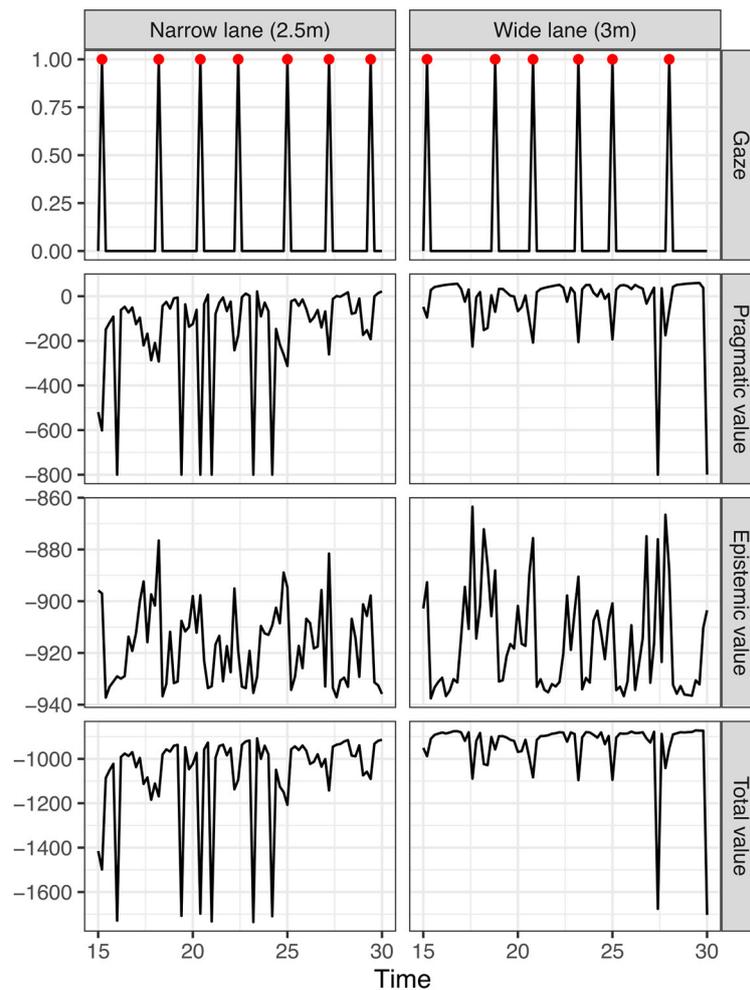


FIGURE 10

Dynamics of expected free energy minimization during visual time-sharing. Large dips in pragmatic value due to lane exits are truncated by setting the lane exit log preference probability to  $-100$ , similar to Figure 8. The red points in the gaze plot indicate on-road glances.

in general). In active inference, this is enabled by casting the two quantities in the common currency of expected free energy (Equation 1). By selecting policies that minimize expected (future) free energy, uncertainty can be resolved “on the fly” as an integral part of the generated behavior. As we demonstrate in this paper, a single action often carries both pragmatic and epistemic value, for example, in Simulation 1a where the ego vehicle was moving forward toward the goal (yielding pragmatic value) which also brought about a better view of the area behind the occluding object (epistemic value).

Epistemic value has also been explored in machine learning where it is typically referred to as artificial curiosity and intrinsic motivation (Schmidhuber, 1991; Sun et al., 2011; Hester and Stone, 2012). In this context, the purpose of curiosity and intrinsic motivation is generally to facilitate model learning by promoting a wider exploration of the state space. By contrast, the focus of the present paper is on the role of epistemic action in resolving uncertainty during inference. However, active inference also extends to learning, which can also be cast in terms of expected free energy minimization and facilitated by epistemic

action (Friston et al., 2015). While classical machine learning approaches to exploration often involve *ad-hoc* features and/or separate mechanisms for exploitation and exploration, active inference has the advantage of offering a more principled approach for both inference and learning solely based on expected free minimization.

Minimizing expected free energy “automatically” yields an optimal combination of epistemic and pragmatic action, but the ensuing behavior is only optimal given the agent’s subjective preferences and beliefs. This allows for conceptualizing and modeling situations where a driver’s understanding of the situation is incorrect, that is, the driver’s generative model does not match the actual state of affairs (the generative process) but the behavior is still optimal given the agent’s subjective beliefs and preferences. In this paper, we demonstrated how an overly certain (high precision) belief may expose the driver to an increased risk of collision. In our Simulations 1b and 1d, the model falsely believed with full certainty that no pedestrian could be hidden behind the occluding object, in which case it did not slow down or moved left to get a better view. Based on the model’s own beliefs and preferences, this

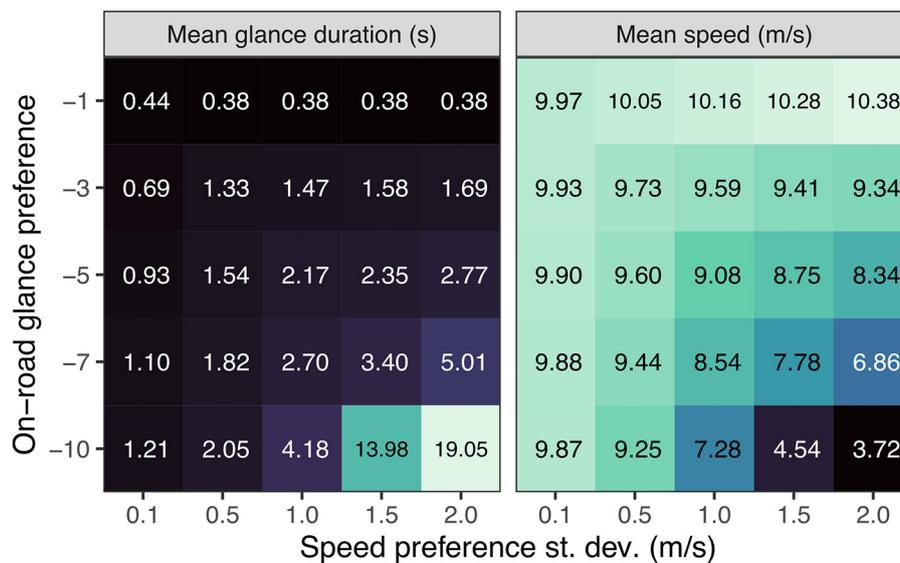


FIGURE 11

Interaction of on-road glance preference and speed preference on mean speed and mean off-road single glance duration.

represented the optimal behavior in this situation, but the model's beliefs were not well-calibrated to the situation, representing the general phenomenon of optimal inference with suboptimal models studied in computational psychiatry (Schwartenbeck et al., 2016). It is widely believed that failures to properly adapt behavior to the traffic situation due to false beliefs is a leading factor behind road crashes (Summala, 1996, 2007).

In Scenario 2, we demonstrated that our model was able to reproduce and explain the underlying mechanisms behind key human behavioral patterns in the literature on visual sampling and visual time-sharing in driving. Specifically, the model explains how visual time-sharing emerges as a result of pragmatic value motivating off road glances to perform a secondary task, counteracted by the epistemic value of resolving uncertainty about one's position in the lane, all governed by selecting policies that minimize expected free energy. At the high level, our model shares many of the key concepts behind the existing computational visual sampling models by Senders (1967), Pekkanen et al. (2018), and Johnson et al. (2014), in particular regarding the explicit modeling of uncertainty and the key role of visual sampling in resolving uncertainty. However, whereas the existing models are specifically applicable to visual sampling, our model is, in principle, applicable to any form of adaptive driving behavior. While the current relatively simple scenarios were chosen to illustrate the key principles of the model, we believe that our model, thanks to its generality, can in principle be applied to any traffic scenario. Even though the specific implementation details may differ, the current framework suggests that all that is needed to model a certain behavior is to define the preference distribution, belief distribution, the observation and action variables (as probability distributions), and the generative model for how they evolve. Adaptive driving behavior will then fall out "automatically" from the selection of policies that minimize expected free energy.

While the current implementation of the model is based on active inference principles (specifically expected free energy minimization), it does not represent a "pure" implementation of the type typically found in the active inference literature (e.g., Parr et al., 2022). Rather the model makes use of standard engineering methods such as particle filtering (for representing beliefs about future states) and the cross entropy method for policy selection. A related approach is presented in Fountas et al. (2020) which used Monte Carlo Tree Search to generate policies evaluated through expected free energy minimization. This yields a modular architecture where the different components can be substituted for other models and methods. For example, in order to represent more sophisticated probabilistic beliefs (e.g., about the future behavior of other road users), the particle filter in the current model could be replaced by a more advanced machine-learned behavior prediction model (e.g., Chai et al., 2019), while still retaining the key principles of active inference discussed above. Such alternative implementations could also help increase computational efficiency and make the model more amenable to real time implementation.

Whereas, in the current model, the parameter values were set by hand, the model naturally lends itself to learning the parameters from data (see Wei et al., 2022, 2023b for our earlier work in this direction). The design space for incorporating techniques from contemporary generative AI in active inference models is large and the exploration of these possibilities has only begun (see e.g., Fountas et al., 2020; Tschantz et al., 2020; Lanillos et al., 2021; Friston et al., 2022; Mazzaglia et al., 2022). However, regardless of the implementation, behavioral models developed based on the active inference principles outlined above are fundamentally explainable and interpretable which is one of their key potential advantages compared to existing black-box approaches for agent modeling (Albarracin et al., 2023).

The modeling framework outlined in this paper has many potential applications in road traffic research beyond driver agent

modeling. For example, it can be used as the basis for defining and modeling human road user failure modes behind road crashes, and for designing effective countermeasures. As discussed above, one such mechanism is how too high precision (false certainty) in one's belief in how a situation will play out leads to overly assertive behavior (e.g., not accounting for potential occlusions, and increased crash risk). While such crash causation mechanisms have been previously outlined conceptually (Engström et al., 2018), the current model offers a precise computational formulation of these principles. The framework can also be used to develop models that generate reference trajectories for the evaluation of autonomous vehicles which are optimal given preferences that reflect the societal norms in the community where the AI is deployed and reasonably foreseeable assumptions (beliefs) about the behavior of other road users, as formalized in the IEEE p2846 standard (IEEE, 2022).

The model presented in this paper has a number of specific limitations related to the simplifying assumptions of the current implementation described above (e.g., assuming only a single pedestrian in Scenario 1, ignoring peripheral vision in Scenario 2 etc.). In general we do not think these simplifying assumptions imply any fundamental limitations for the proposed behavior modeling framework. However, there are certain aspects of the current model set-up that warrants further consideration. For example, the current model assumes that some variables (vehicle  $x$  and  $y$  position and heading) are not observable when looking off-road, and that this also applies during the counterfactual evaluation of policies. During these counterfactual off-road glances, the pragmatic value related to vehicle  $x$  and  $y$  position and heading cannot be evaluated. This is a consequence of pragmatic value being defined over observations and necessitated the addition of the lane exit variable which is assumed to be always observable. Clearly, this is somewhat arbitrary and an alternative approach would be to define pragmatic value during counterfactual reasoning over states instead of observations, in which case only the lane position variable would be needed (and the lane exit variable would be redundant). Such alternative model variants can be explored in further work.

More broadly, there are many ways in which the current model can be further developed. First, as already mentioned, using more advanced generative models for representing beliefs and learning parameters from data will likely be needed to scale up to more realistic scenarios, for example including other vehicles and beliefs about their future trajectories. Second, it would be interesting to explore further the extent to which the model can generalize across different types of traffic scenarios. Of particular interest is to model collision avoidance behaviors in critical traffic conflicts. Such behaviors are difficult to learn solely from data (due to the sparsity of such long-tail events in human driving data), while the present model could be better suited for such scenarios thanks to the possibility to explicitly specify the behavioral mechanisms. This would require a detailed model of the belief updating process to model response timing, a feature which is lacking in the current model (see Engström et al., 2024). Another very interesting avenue of future model development is in the context of road user interactions, for example at a crosswalk (Markkula et al., 2023). In such scenarios, active inference models can represent the situation understanding shared between interacting road users in terms of

shared generative models (see Friston and Frith, 2015; Friston et al., 2022) and, conversely, failing interactions can be understood in terms of mismatching generative models.

Another interesting question not addressed in the present paper concerns how to model the difference between explicit planning and more automatized habitual behavioral selection. The current model is based on sampling and rolling out a large number of policies (plans) where the optimal plan is identified based on expected free energy minimization. However, human action selection in everyday situations is at least partly governed by habits. In skilled performance, humans do not evaluate all possible policies but rather learn to “see” the best policy given the information currently available (e.g., in terms of Gibsonian affordances; Gibson, 2014). This is particularly true in the context of driving which is often to a large extent automatized for experienced drivers (Engström et al., 2017). In current ML-based active inference models, such habit formation has been modeled in terms of “habit networks” which are able to learn successful policies through amortization (e.g., Mazzaglia et al., 2022) and this is an interesting topic for future development of the present framework.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

JE: Conceptualization, Investigation, Methodology, Project administration, Supervision, Visualization, Writing—original draft, Writing—review & editing. RW: Conceptualization, Investigation, Methodology, Software, Validation, Writing—original draft, Writing—review & editing. AM: Conceptualization, Funding acquisition, Project administration, Supervision, Visualization, Writing—original draft, Writing—review & editing, Methodology. AG: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing—original draft, Writing—review & editing. MO'K: Investigation, Methodology, Software, Writing—original draft, Writing—review & editing. LJ: Investigation, Methodology, Software, Writing—original draft, Writing—review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. AG and RW's contributions were supported by a contract from Waymo LLC.

## Acknowledgments

This article was previously submitted to the arXiv preprint server.

## Conflict of interest

JE, MO'K, and LJ were employed by Waymo LLC and conducted the research without any external funding from third-parties. Techniques discussed in this paper may be described in U.S. Patent Application No. 63/454,012.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Albarracín, M., Hipólito, I., Tremblay, S. E., Fox, J. G., René, G., Friston, K., et al. (2023). Designing explainable artificial intelligence with active inference: A framework for transparent introspection and decision-making. *arXiv [preprint]*. doi: 10.1007/978-3-031-47958-8\_9
- Antin, J. F., Dingus, T. A., Hulse, M. C., and Wierwille, W. W. (1990). An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. *Int. J. Man Mach. Stud.* 33, 581–594. doi: 10.1016/S0020-7373(05)80054-9
- Chai, Y., Sapp, B., Bansal, M., and Anguelov, D. (2019). Multipath: multiple probabilistic anchor trajectory hypotheses for behavior prediction. *arXiv [preprint]*. doi: 10.48550/arXiv.1910.05449
- Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos. Transact. R. Soc. B Biol. Sci.* 362, 1585–1599. doi: 10.1098/rstb.2007.2054
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- Clark, A. (2023). *The Experience Machine: How Our Minds Predict and Shape Reality*. New York, NY: Pantheon Books.
- Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., and Friston, K. (2020). Active inference on discrete state-spaces: a synthesis. *J. Math. Psychol.* 99:102447. doi: 10.1016/j.jmp.2020.102447
- da Lio, M., Cherubini, A., Papini, G. P. R., and Plebe, A. (2023). Complex self-driving behaviors emerging from affordance competition in layered control architectures. *Cogn. Syst. Res.* 79, 4–14. doi: 10.1016/j.cogsys.2022.12.007
- De Boer, P.-T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Ann. Operat. Res.* 134, 19–67. doi: 10.1007/s10479-005-5724-z
- Dinparastdjadid, A., Supeene, I., and Engstrom, J. (2023). Measuring surprise in the wild. *arXiv [preprint]*. doi: 10.48550/arXiv.2305.07733
- Engström, J., Bärngman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G. B., et al. (2018). Great expectations: a predictive processing account of automobile driving. *Theoret. Iss. Ergon. Sci.* 19, 156–194. doi: 10.1080/1463922X.2017.1306148
- Engström, J., Johansson, E., and Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transport. Res. Part F* 8, 97–120. doi: 10.1016/j.trf.2005.04.012
- Engström, J., Liu, S.-Y., Dinparastdjadid, A., and Simoiu, C. (2024). Engström, J., Liu, S.-Y., Dinparastdjadid, A., and Simoiu, C. (2024). Modeling road user response timing in naturalistic traffic conflicts: a surprise-based framework. *Accid. Anal. Prevent.* 198. Available online at: <https://www.sciencedirect.com/science/article/pii/S0001457524000058>
- Engström, J., Markkula, G., Victor, T., and Merat, N. (2017). Effects of cognitive load on driving performance: the cognitive control hypothesis. *Hum. Fact.* 59, 734–764. doi: 10.1177/0018720817690639
- Fischer, J., and Tas, Ö. S. (2020). “Information particle filter tree: an online algorithm for pomdps with belief-based rewards on continuous domains,” in *International Conference on Machine Learning* (PMLR), 3177–3187.
- Fitzpatrick, K., Carlson, P., Brewer, M., and Wooldridge, M. (2001). Design factors that affect driver speed on suburban streets. *Transp. Res. Rec.* 1751, 18–25. doi: 10.3141/1751-03
- Fountas, Z., Sajid, N., Mediano, P., and Friston, K. (2020). Deep active inference agents using monte-carlo methods. *Adv. Neural Inf. Process. Syst.* 33, 11662–11675.
- Friston, K., Adams, R., and Montague, R. (2012). What is value—accumulated reward or evidence? *Front. Neurobot.* 6:11. doi: 10.3389/fnbot.2012.00011
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017). Active inference: a process theory. *Neural Comput.* 29, 1–49. doi: 10.1162/NECO\_a\_00912
- Friston, K., and Frith, C. (2015). A duet for one. *Conscious. Cogn.* 36, 390–405.
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214. doi: 10.1080/17588928.2015.1020053
- Friston, K. J., Ramstead, M. J., Kiefer, A. B., Tschantz, A., Buckley, C. L., Albarracín, M., et al. (2022). Designing ecosystems of intelligence from first principles. *arXiv [preprint]*. doi: 10.48550/arXiv.2212.01354
- Fuller, R. (2005). Towards a general theory of driver behaviour. *Accid. Anal. Prev.* 37, 461–472. doi: 10.1016/j.aap.2004.11.003
- Gibson, J. J. (2014). *The Ecological Approach to Visual Perception: Classic Edition*. New York, NY: Psychology Press.
- Gibson, J. J., and Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. *Am. J. Psychol.* 51, 453–471. doi: 10.2307/1416145
- Greenberg, J., Tijerina, L., Curry, R., Artz, B., Cathey, L., Kochhar, D., et al. (2003). Driver distraction: evaluation with event detection paradigm. *Transp. Res. Rec.* 1843, 1–9. doi: 10.3141/1843-01
- Hester, T., and Stone, P. (2012). “Intrinsically motivated model learning for a developing curious agent,” in *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (San Diego, CA: IEEE), 1–6.
- Horrey, W. J., Wickens, C. D., and Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *J. Exp. Psychol. Appl.* 12:67. doi: 10.1037/1076-898X.12.2.67
- IEEE (2022). *Standard for Assumptions in Safety-Related Models for Automated Driving Systems: Ieee*. Technical Report. doi: 10.1109/IEEESTD.2022.9761121
- Igl, M., Kim, D., Kuefler, A., Mouglin, P., Shah, P., Shiarlis, K., et al. (2022). “Symphony: learning realistic and diverse agents for autonomous driving simulation,” in *2022 International Conference on Robotics and Automation (ICRA)* (Philadelphia, PA: IEEE), 2445–2451.
- Igl, M., Zintgraf, L., Le, T. A., Wood, F., and Whiteson, S. (2018). “Deep variational reinforcement learning for pomdps,” in *International Conference on Machine Learning* (Stockholm: PMLR), 117–2126.
- Itti, L., and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Res.* 49, 1295–1306. doi: 10.1016/j.visres.2008.09.007
- Johnson, L., Sullivan, B., Hayhoe, M., and Ballard, D. (2014). Predicting human visuomotor behaviour in a driving task. *Philos. Transact. R. Soc. B* 369:20130044. doi: 10.1098/rstb.2013.0044
- Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artif. Intell.* 101, 99–134. doi: 10.1016/S0004-3702(98)00023-X
- Kiverstein, J., Kirchhoff, M. D., and Froese, T. (2022). The problem of meaning: the free energy principle and artificial agency. *Front. Neurobot.* 16:844773. doi: 10.3389/fnbot.2022.844773
- Klauer, C., Dingus, T. A., Neale, V. L., Sudweeks, J. D., Ramsey, D. J., et al. (2006). The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data. Technical report, United States. *Nati. Highway Traffic Saf. Admin.* doi: 10.1037/e729262011-001

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Klauer, S. G., Guo, F., Sudweeks, J., and Dingus, T. A. (2010). *An Analysis of Driver Inattention Using a Case-Crossover Approach on 100-Car Data*. Technical Report, National Highway Traffic Safety Administration. Washington, DC.
- Kolekar, S., de Winter, J., and Abbink, D. (2020). Human-like driving behaviour emerges from a risk-based driver model. *Nat. Commun.* 11:4850. doi: 10.1038/s41467-020-18353-4
- Lanillos, P., Meo, C., Pezzato, C., Meera, A. A., Baioumy, M., Ohata, W., et al. (2021). Active inference in robotics and artificial agents: survey and challenges. *arXiv [preprint]*. doi: 10.48550/arXiv.2112.01871
- Lewis-Evans, B. (2012). *Testing models of driver behaviour*. Groningen: University Library Groningen.
- Macdonald, W. A., and Hoffmann, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Hum. Fact.* 22, 733–739. doi: 10.1177/001872088002200609
- Markkula, G., and Engström, J. (2006). “A steering wheel reversal rate metric for assessing effects of visual and cognitive secondary task load,” in *Proceedings of the 13th ITS World Congress* (London), 1–12.
- Markkula, G., Lin, Y.-S., Srinivasan, A. R., Billington, J., Leonetti, M., Kalantari, A. H., et al. (2023). Explaining human interactions on the road by large-scale integration of computational psychological theory. *PNAS Nexus* 2:pgad163. doi: 10.1093/pnasnexus/pgad163
- Mazzaglia, P., Verbelen, T., Çatal, O., and Dhoedt, B. (2022). The free energy principle for perception and action: a deep learning perspective. *Entropy* 24:301. doi: 10.3390/e24020301
- McDonald, A. D., Ferris, T. K., and Wiener, T. A. (2020). Classification of driver distraction: a comprehensive analysis of feature generation, machine learning, and input measures. *Hum. Fact.* 62, 1019–1035. doi: 10.1177/0018720819856454
- Merat, N., and Jamson, A. H. (2008). The effect of stimulus modality on signal detection: implications for assessing the safety of in-vehicle technology. *Hum. Fact.* 50, 145–158. doi: 10.1518/001872008X250656
- Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167
- Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the Structure of Behavior*. New York, NY: Holt.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.
- Näätänen, R., and Summala, H. (1976). *Road-User Behaviour and Traffic Accidents*. Amsterdam: North-Holland Publishing Company.
- Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, H., Jamson, S., et al. (2004). *Haste Deliverable 2: Hmi and Safety-Related Driver Performance*. Technical Report, European Commission EC. Brussels.
- Parr, T., and Friston, K. J. (2017). Uncertainty, epistemics and active inference. *J. R. Soc. Interf.* 14:20170376. doi: 10.1098/rsif.2017.0376
- Parr, T., Pezzulo, G., and Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. Cambridge, MA: MIT Press.
- Pekkanen, J., Lappi, O., Itkonen, T. H., and Summala, H. (2017). Task-difficulty homeostasis in car following models: experimental validation using self-paced visual occlusion. *PLoS ONE* 12:e0169704. doi: 10.1371/journal.pone.0169704
- Pekkanen, J., Lappi, O., Rinkkala, P., Tuhkanen, S., Frantsi, R., and Summala, H. (2018). A computational model for driver's cognitive state, visual perception and intermittent attention in a distracted car following task. *R. Soc. Open Sci.* 5:180194. doi: 10.1098/rsos.180194
- Peters, J., Mulling, K., and Altun, Y. (2010). Relative entropy policy search. *Proc. AAAI Conf. Artif. Intell.* 24, 1607–1612. doi: 10.1609/aaai.v24i1.7727
- Polack, P., Alché, F., d'Andréa Novel, B., and de La Fortelle, A. (2017). “The kinematic bicycle model: a consistent model for planning feasible trajectories for autonomous vehicles?,” in *2017 IEEE Intelligent Vehicles Symposium (IV)* (Redondo Beach, CA: IEEE), 812–818.
- Powers, W. T. (1973). Feedback: Beyond behaviorism: stimulus-response laws are wholly predictable within a control-system model of behavioral organization. *Science* 179, 351–356. doi: 10.1126/science.179.4071.351
- Sadigh, D., Landolfi, N., Sastry, S. S., Seshia, S. A., and Dragan, A. D. (2018). Planning for cars that coordinate with people: leveraging effects on human actions for planning and active information gathering over human internal state. *Auton. Robots* 42, 1405–1426. doi: 10.1007/s10514-018-9746-1
- Schmidhuber, J. (1991). “Curious model-building control systems,” in *Proc. International Joint Conference on Neural Networks* (Singapore), 1458–1463.
- Schwartenbeck, P., FitzGerald, T., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., et al. (2016). Optimal inference with suboptimal models: addiction and active bayesian inference. *Med. Hypotheses* 84, 109–117. doi: 10.1016/j.mehy.2014.12.007
- Senders, J. (1967). On the distribution of attention in a dynamic environment. *Acta Psychol.* 27, 349–354. doi: 10.1016/0001-6918(67)90079-0
- Sprague, N., and Ballard, D. (2003). “Eye movements for reward maximization,” in *Advances in Neural Information Processing Systems Vol. 16* (MIT Press).
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., et al. (2016). Allostatic self-efficacy: a metacognitive theory of dyshomeostasis-induced fatigue and depression. *Front. Hum. Neurosci.* 10:550. doi: 10.3389/fnhum.2016.00550
- Summala, H. (1988). Risk control is not risk adjustment: the zero-risk theory of driver behaviour and its implications. *Ergonomics* 31, 491–506. doi: 10.1080/00140138808966694
- Summala, H. (1996). Accident risk and driver behaviour. *Saf. Sci.* 22, 103–117. doi: 10.1016/0925-7535(96)00009-4
- Summala, H. (2007). “Towards understanding motivational and emotional factors in driver behaviour: comfort through satisficing,” in *Modelling Driver Behaviour in Automotive Environments* (London: Springer), 189–207. doi: 10.1007/978-1-84628-618-6
- Sun, Y., Gomez, F., and Schmidhuber, J. (2011). “Planning to be surprised: optimal bayesian exploration in dynamic environments,” in *Artificial General Intelligence: 4th International Conference, AGI 2011, Mountain View, CA, USA, August 3-6, 2011. Proceedings 4* (Springer), 41–51.
- Suo, S., Regalado, S., Casas, S., and Urtasun, R. (2021). “TrafficSim: learning to simulate realistic multi-agent behaviors,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Nashville, TN), 10400–10409.
- Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- Treiber, M., and Kesting, A. (2013). Microscopic calibration and validation of car-following models—a systematic approach. *Proc. Soc. Behav. Sci.* 80, 922–939. doi: 10.1016/j.sbspro.2013.05.050
- Tschantz, A., Baltieri, M., Seth, A. K., and Buckley, C. L. (2020). “Scaling active inference,” in *2020 International Joint Conference on Neural Networks (IJCNN)* (Glasgow: IEEE), 1–8.
- Victor, T., Dozza, M., Bärghman, J., Boda, C.-N., Engström, J., Flannagan, C., et al. (2015). *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk*. Technical Report, Transportation Research Board. Washington, DC.
- Victor, T. W., Harbluk, J. L., and Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transport. Res. Part F* 8, 167–190. doi: 10.1016/j.trf.2005.04.014
- Wei, R., Garcia, A., McDonald, A., Markkula, G., Engström, J., Supeene, I., et al. (2023a). “World model learning from demonstrations with active inference: application to driving behavior,” in *Active Inference: Third International Workshop, IWAI 2022, Grenoble, France, September 19, 2022, Revised Selected Papers* (Springer), 130–142.
- Wei, R., McDonald, A. D., Garcia, A., and Alambeigi, H. (2022). Modeling driver responses to automation failures with active inference. *IEEE Transact. Intell. Transport. Syst.* 23, 18064–18075. doi: 10.1109/TITS.2022.3155381
- Wei, R., McDonald, A. D., Garcia, A., Markkula, G., Engstrom, J., and O'Kelly, M. (2023b). An active inference model of car following: advantages and applications. *arXiv*. doi: 10.48550/arXiv.2303.15201
- Wilde, G. J. (1982). The theory of risk homeostasis: implications for safety and health. *Risk Anal.* 2, 209–225. doi: 10.1111/j.1539-6924.1982.tb01384.x
- Xiao, Y., Codevilla, F., Pal, C., and Lopez, A. (2021). “Action-based representation learning for autonomous driving,” in *Conference on Robot Learning* (London: PMLR), 232–246.
- Yagar, S., and Van Aerde, M. (1983). Geometric and environmental effects on speeds of 2-lane highways. *Transport. Res. Part A* 17, 315–325. doi: 10.1016/0191-2607(83)90094-8
- Zwahlen, H., Adams, C., and DeBald, D. (1988). *Vision in Vehicles II*. Amstredam: Elsevier, 335–344.