Predicting Mispredictions: A Model of Human Misjudgment About Vulnerable Road Users' Trajectories

Alessandro Colombo, *Senior Member, IEEE*, Matteo Depaola, Francesco Ferrise, Nicolò Dozio, and Gabriel Rodrigues de Campos

***Abstract*— This paper presents a cognitive model designed to reproduce human drivers' errors in predicting the motion of nearby vulnerable road users. We aim to define a computational model that, given both the trajectory of the eye gaze of a human driver and the trajectory of a bicycle, can compute the probability distribution of where the human driver believes the bicycle will be in the near future. For the design and validation of the proposed cognitive model, we tested 51 subjects in immersive virtual reality scenarios. The results indicate that the proposed model can generate probability distributions of the human drivers' beliefs about the future bicycle position that are very similar, though not statistically equivalent, to those obtained experimentally. Such models could easily be generalized to describe how drivers misjudge the motion of other road users. This may enable ADAS to evaluate and improve drivers' situational awareness. In the future, these models could also be used by autonomous cars to evaluate situational awareness of nearby humans, enabling a safer coexistence of autonomous vehicles and vulnerable road users.**

***Index Terms*— ADAS, cognitive model, motion prediction, situation awareness, vulnerable road users.**

I. Introduction

EVERY year, road traffic accidents cause over 1.19 million deaths; more than half are Vulnerable Road Users (VRU) [1]. Many of these accidents are related to human mistakes and are driven by incorrect situation assessment, inattentiveness, or poor situation awareness. To tackle these issues, several generations of Advanced Driver Assistance Systems (ADAS) have been deployed over the last decades, with a relevant impact on real-world traffic and traffic-related injuries [2], [3]. ADAS provide support in cases where the driver is likely to fail the driving task, by increasing driver awareness and engagement (e.g., through audio/visual alerts), by evaluating the driver's awareness and actions for threat assessment purposes, and ultimately by enforcing automatic corrective actions whenever appropriate. Such systems include, for example, automated braking systems, lane-keeping assistance, and, more recently, head-on collision avoidance or advanced cruising features.

Recently, there has been increasing interest in the development of Driver Monitoring Systems (DMS), which are expected to have a high potential in terms of driver's state/behavior estimation and situation awareness assessment [4], [5]. Situation awareness, an overarching research field covering multiple scientific domains (e.g., human factors, system engineering, dynamic systems) and industrial fields (e.g. aviation, ground transportation, nuclear power), refers to the cognitive process through which individuals perceive the elements of their environment, comprehend their meaning, and project their status in the near future [6], [7], [8]. DMS concepts are diverse and include a large body of techniques and tools, such as heart rate variability measurements by electrocardiography [9], brain activity estimation using electroencephalograms [10], as well as camera-based solutions, frequently employed for eye gaze or body posture detection [11], [12], [13]. Leveraging these new information sources, many works in the literature focus on detecting driver inattentiveness, distraction [14], [15], [16], drowsiness [17], [18] as well as estimating driver's intent and behavior [19], [20], [21], [22]. Driver state information can then be used in the design of new support/safety systems for improving existing ADAS, which are traditionally not based on in-vehicle sensing. Examples of this integration are some implementations of the Lane Keeping Assistance [21], [23] and Lane Change Assistance [24], [25], [26] systems.

Most DMS provide a binary classification of driver attention or awareness. Reality can however be more subtle. For instance, statistics from a large dataset of accidents reported in Sweden [27] show that, in almost 30% of car-bicycle accidents where car and bike paths intersect, the car driver had seen the bike before the crash. This raises to almost 50% in scenarios where car and bicycle paths are parallel. These statistics suggest that a simple binary classification of driver awareness is not sufficient to predict a relevant portion of

Received 19 December 2023; revised 22 April 2024 and 24 September 2024; accepted 16 October 2024. The Associate Editor for this article was M. Brackstone. *(Corresponding author: Alessandro Colombo.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of Politecnico di Milano under Application No. 4/2021.

Alessandro Colombo is with DEIB, Politecnico di Milano, 20133 Milan, Italy (e-mail: alessandro.colombo@polimi.it).

Matteo Depaola is with Siemens Digital Industries Software, 3001 Leuven, Belgium (e-mail: matteo.depaola@siemens.com).

Francesco Ferrise and Nicolò Dozio are with the Department of Mechanical Engineering, Politecnico di Milano, 20133 Milan, Italy (e-mail: francesco.ferrise@polimi.it; nicolo.dozio@polimi.it).

Gabriel Rodrigues de Campos is with Zenseact, 417 56 Gothenburg, Sweden (e-mail: gabriel@decampos.eu).

Digital Object Identifier 10.1109/TITS.2024.3484004

collisions: misunderstanding of each other's intentions plays an important role.

Based on the above observation, we explore in this work the extent to which information about the driver's eye gaze and the position of nearby VRUs can be used to quantify, statistically, the misunderstanding by the driver of the VRUs' trajectory. We propose for this purpose a novel cognitive model, reproducing how the driver perceives and understands the motion of nearby objects. In this objective, our work shares some similarities with the cognitive architectures developed in computational cognitive sciences, such as Adaptive Control of Thought—Rational (ACT-R) [28] and Soar [29], which aspire to model and simulate general cognition while, at the same time, reproducing more or less faithfully the cognitive structures postulated by psychological theory. The reader can refer to [30], where more recent ideas about embodiment (perception and action) and subsymbolic processes are discussed, and to [31], [32], [33], and [34], where both general-purpose and domain-specific cognitive architectures are employed to model driver cognition. In very general terms, our objective is positioned in the same area and is founded on the same principles: we aim to model the cognitive processes that allow a subject (in our case, the driver) to form cognition of a perceived object's motion (which we call a *percept*), and we do so by assuming that the subject utilizes a *mental model* (or internal model, to use a term more common in Control Theory) of the percept's dynamics, learned through prior experience, to form such a cognition.

Rather than explicitly modeling the procedural and declarative memory functions that realize such a process, we take a higher-level view of the phenomenon. We use a minimal (i.e., as simple as possible) mathematical model to reproduce the experimental evidence gathered from a set of experiments. In this sense, our approach is closer to that used in dynamic cognitive sciences (see, e.g., [35]), and follows in the path of recent attempts to model human perception through Bayesian observers [36], [37], [38], [39]. Moreover, our model does not address the complex learning processes that lead to the formation of the above-mentioned mental model. Rather, we assume the mental model is set, and for this purpose, we employ percepts whose dynamics are well-known to the test subjects. The experiments that we used to identify the model's parameters were designed to minimize the effects of learning, and we tested the ability of our model to reproduce the imprecision with which human subjects perceive and then predict the motion of the percept. In a nutshell, we aim to reproduce, with as simple a model as possible, the imprecision with which the average human subject understands and predicts the short-term motion of a percept of a well-known nature.

Our results show that the proposed cognitive model can reproduce the mismatch between the expected (by the human driver) motion of the bicycle and the actual cyclist motion. By exploiting both in-cabin sensing (driver gaze) and surrounding sensing, our approach could improve the ADAS ability to detect the driver's unawareness of potential risks (see Fig. 1). This would enable the design of a new generation of ADAS implementing stronger precautionary measures to avoid unreasonable risks, in line with the principles described in [40].

The remainder of the paper is organized as follows. Section II presents the proposed cognitive model. Section III outlines the experimental design and data processing. Finally, results and discussion are presented in Section IV, and conclusions and future research perspectives are outlined in Section V.

II. Model Description

In the design of our cognitive model, we consider in particular cyclists as the VRUs. Furthermore, and without loss of generality, we consider interactions at traffic intersections, a paramount element of today's traffic infrastructure known for their complexity, diverse topologies, as well for their accident-prone nature [41].

The cognitive model incorporates two submodels (see Fig. 2):

- a *mental model* describing prior knowledge about the percept's laws of motion (see Section II-A).
- a *perceptive model* characterizing the inaccuracy of human vision (see Section II-B).

The full cognitive model then represents how the subjects update their prior knowledge about the percept using visual information, and how they predict its future movement (see Section II).

We assume that the process through which the human mind forms cognition of the percept is similar to a Bayesian observer that iteratively gathers information (chiefly through vision) and integrates it into the current understanding of the percept's state. We use the Kalman filter as the foundation of this mathematical structure. We take the position and heading of the bicycle, perceived by the driver, as the Kalman filter observed variable. The state of the filter represents the cognition that an average driver has of the bicycle's state at each time instant, while the filter covariance matrix encodes the probability distribution with which a population of drivers understands a given bicycle's state. The model's parameters are then identified from data collected through a set of experiments using immersive Virtual Reality (VR) scenarios, which involved a total of 51 subjects and are described in Sec. III.

In the following, lowercase letters indicate scalar quantities (s), bold lowercase letters are used for vectors ($\mathbf{v}$), and uppercase letters are matrices (M). We use subscripts (v_i or $m_{i,j}$) to denote the i-th element of a vector $\mathbf{v}$ or the element of matrix M of i-th row and j-th column. The superscript ^{S} denotes that a vector v is an element of vector space S, i.e., $\mathbf{v}^S \in S$. The dimensions of a matrix M with m rows and n columns are indicated as $M_{m\times n}$. Moreover, according to the terminology already introduced in the previous section, we call *subject* the human driver, and *percept* the perceived object.

A. The Mental Model

The mental model encodes the prior knowledge of the subject about the percept. We deal in particular with two types of percept: a stationary object, and a bicycle. The state $\mathbf{x}$ of

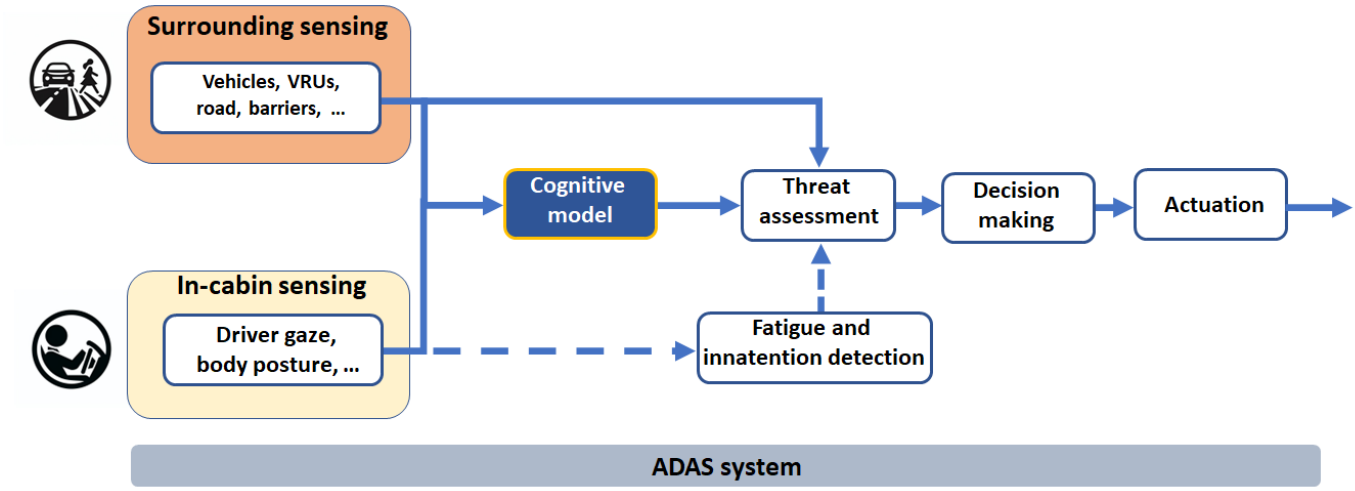

Fig. 1. Standard ADAS structure (dashed arrows) VS our cognitive-model-equipped ADAS (solid arrows).

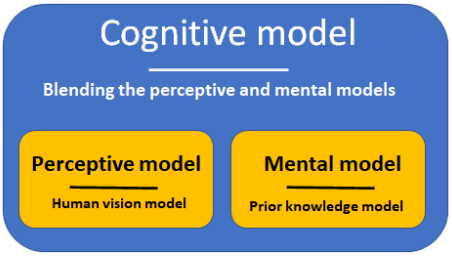

Fig. 2. Building blocks of the cognitive model. The mental model describes the prior knowledge about the percept's dynamics; the perceptive model describes the uncertainty with which the percept is perceived. These two models, together with the equations necessary to represent how prior knowledge is blended with newly perceived information, form the cognitive model.

the mental model of the percept is the set of quantities that characterize the subject's understanding of its state of motion. We assume it is a vector with $n_{\mathbf{x}}$ elements. Given $\mathbf{x}$, the mental model is the dynamical system

$$\mathbf{x}(t+1) = f(\mathbf{x}(t)) + \mathbf{e}_Q,$$

where $f{:}\mathbb{R}^{n_{\mathbf{x}}} \to \mathbb{R}^{n_{\mathbf{x}}}$ models the subject's prior knowledge about the percept's dynamics, while $\mathbf{e}_Q \sim \mathcal{N}(\mathbf{0}, Q)$, with $Q \in \mathbb{R}^{n_{\mathbf{x}} \times n_{\mathbf{x}}}$, models the uncertainty about these dynamics.

In the case of the stationary object, the state coincides with the position of the ground projection of the percept's center of mass, hence $n_{\mathbf{x}} = 2$. Moreover, since the subject knows (is told before the experiment) that the object is stationary, the mental model describes a constant position, with no uncertainty about the dynamics:

$$f(\mathbf{x}(t)) = \mathbf{x}(t), \quad \mathbf{e}_Q = 0. \tag{1}$$

In the case of the bicycle, we considered for the mental model both a kinematic unicycle model and a kinematic bicycle model. We observed a better model fit with the latter, so for the sake of conciseness, we discuss the latter only. In this case, $n_{\mathbf{x}} = 5$. States x_1 and x_2 represent the position of the ground projection of the bicycle's center of mass; x_3 is its heading; x_4 is its steering angle; and x_5 is its velocity. We then have

$$f(\mathbf{x}(t)) = \begin{bmatrix} x_1(t) + \tau_s x_5(t)\cos(x_3(t) + \beta(t)) \\ x_2(t) + \tau_s x_5(t)\sin(x_3(t) + \beta(t)) \\ x_3(t) + \tau_s x_5(t)\dfrac{\tan(x_4(t))\cos(\beta(t))}{L} \\ x_4(t) \\ \alpha x_5(t) \end{bmatrix}, \tag{2}$$

and we assume the covariance matrix Q to be diagonal and nonzero, with $q_{1,1} = q_{2,2}$ since these both model uncertainty in the horizontal position. Parameter L is the bicycle wheelbase ($L = 1.15$m in this work), $\beta = \arctan(l_R \tan(x_4)/L)$ is the sideslip angle, while τ_s is the sampling time, and was set to 0.01s. Notice that the fifth element of (2) models the (mental model of the) bicycle velocity as accelerated, according to a parameter α which will be identified from the experimental data.

B. The Perceptive Model

Borrowing from the computer vision jargon [42], we call *pose* the set of quantities that define the percept's perceivable state. When the percept is a stationary object, we identify its pose with the coordinates of its center of gravity, i.e., the coordinates of the ground projection of its center of mass. When the percept is a bicycle, we initially attempted to fit the cognitive model to the experimental data using the coordinates of the bicycle's center of gravity as the solely observed quantity. Such a model failed to reproduce the experimental data to any reasonable degree of accuracy, suggesting that the perception of the heading is as important as the perception of the center of gravity in forming the subject's understanding of the percept's state. Hence, when the percept is a bicycle, the pose consists of the bicycle's center of gravity and its heading. The role of the perceptive model is to give a mathematical representation of the errors affecting the subject's perception of the percept's pose. According to recent results [43], such information is employed by working memory to construct cognition of the percept.

Before discussing our perceptive model for the two kinds of percepts, we shall introduce an error model for the perception of a generic point in space. In our experiments, the subject sees percepts lying on the ground plane of a 3-dimensional virtual space. We represent a location in this space as a vector $\mathbf{z}^{\mathcal{W}}$ in the *World Reference Frame* $\mathcal{W} \subseteq \mathbb{R}^2$, where the two components z_1 and z_2 represent coordinates on the ground plane. We assume that $\mathcal{W}$ has the origin of the axes at the subject's feet. The World Reference Frame is the mathematical representation of the inertial frame in which the subjects place the positional coordinates of their mental model, e.g., the x_1 and x_2 coordinates of the bicycle's center of gravity.

Let us define a coordinate plane $\mathcal{G} \subseteq \mathbb{R}^2$ as a time-dependent rotation of $\mathcal{W}$ such that a vector $\mathbf{z}^{\mathcal{G}} \in \mathcal{G}$ has component $z_1^{\mathcal{G}}$ aligned with the subject's gaze, and $z_2^{\mathcal{G}}$ orthogonal to this direction. We have

$$\mathbf{z}^{\mathcal{G}} := T(\theta)\mathbf{z}^{\mathcal{W}} \tag{3}$$

where

$$T(\theta) := \begin{bmatrix} \cos\theta(t) & -\sin\theta(t) \\ \sin\theta(t) & \cos\theta(t) \end{bmatrix}$$

is the time-dependent rotation by the angle θ.

We assume a monocular pinhole model of vision, modeling the subject's retina as the plane $\mathcal{R} \subseteq \mathbb{R}^2$ with axis $z_1^{\mathcal{R}}$ pointing laterally, and $z_2^{\mathcal{R}}$ pointing down: a point located at $\mathbf{z}^{\mathcal{G}}$, at a vertical distance v below the eyes of the subject ($v = 1$ m in the experiments), is mapped onto

$$\mathbf{z}^{\mathcal{R}} = n(\mathbf{z}^{\mathcal{G}}) := \begin{bmatrix} \frac{z_2^{\mathcal{G}}}{z_1^{\mathcal{G}}} \\ \frac{v}{z_1^{\mathcal{G}}} \end{bmatrix} \tag{4}$$

on the retina plane. Let us call $\mathbf{z}_f^{\mathcal{R}}$ the location of the fovea on the retina. According to the above mapping, and letting $\mathbf{z}_f^{\mathcal{G}}$ be the center of the subject's gaze, we have

$$\mathbf{z}_f^{\mathcal{R}} = \begin{bmatrix} 0 \\ \frac{v_f}{z_{f1}^{\mathcal{G}}} \end{bmatrix}.$$

Through mapping (4), and with the assumption that perception error variances depend on the distance between $\mathbf{z}^{\mathcal{R}}$ and the fovea, we can construct a model of the error with which a subject perceives a point located at $\mathbf{z}^{\mathcal{G}}$, at a vertical distance v below the eyes. We assume that the perception of the point is affected by a Gaussian noise $\mathbf{e}^{\mathcal{G}} \sim \mathcal{N}(\mathbf{b}^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}), R^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}))$, where $\mathbf{b}^{\mathcal{G}}$ is a bias, and $R^{\mathcal{G}}$ is a covariance matrix. To construct the covariance matrix $R^{\mathcal{G}}$, we assume that perception error covariance on the retina is characterized by a diagonal covariance matrix

$$R^{\mathcal{R}}(\mathbf{z}^{\mathcal{R}}, \mathbf{z}_f^{\mathcal{R}}) := \begin{bmatrix} r_{1,1}^{\mathcal{R}}(\mathbf{z}^{\mathcal{R}}, c_1, s_1) & 0 \\ 0 & r_{2,2}^{\mathcal{R}}(\mathbf{z}^{\mathcal{R}}, \mathbf{z}_f^{\mathcal{R}}, c_2, s_2) \end{bmatrix}$$

with

$$r_{1,1}^{\mathcal{R}}(\mathbf{z}^{\mathcal{R}}, c_1, s_1) := ((1 + c_1(z_1^{\mathcal{R}})^2)s_1)^2,$$
$$r_{2,2}^{\mathcal{R}}(\mathbf{z}^{\mathcal{R}}, \mathbf{z}_f^{\mathcal{R}}, c_2, s_2) := ((1 + c_2(z_2^{\mathcal{R}} - z_{f2}^{\mathcal{R}})^2)s_2)^2.$$

Here c_1 and c_2 are coefficients, while s_1 and s_2 are the standard deviations of the perception error on the retina at the fovea. Calling $\mathbf{e}^{\mathcal{R}} \sim \mathcal{N}(0, R^{\mathcal{R}}(\mathbf{z}^{\mathcal{R}}, \mathbf{z}_f^{\mathcal{R}}))$, using (4) we can write the Taylor expansion of the image of point $\mathbf{z}^{\mathcal{G}}$ on the retina subject to error as

$$\mathbf{z}^{\mathcal{R}} + \mathbf{e}^{\mathcal{R}} = n(\mathbf{z}^{\mathcal{G}} + \mathbf{e}^{\mathcal{G}}) = n(\mathbf{z}^{\mathcal{G}}) + J_n(\mathbf{z}^{\mathcal{G}})\mathbf{e}^{\mathcal{G}} + \ldots,$$

where J_n is the Jacobian matrix of n. Knowing that

$$J_n^{-1}(\mathbf{z}^{\mathcal{G}}) = \begin{bmatrix} 0 & -\frac{(z_1^{\mathcal{G}})^2}{v} \\ z_1^{\mathcal{G}} & -\frac{z_1^{\mathcal{G}} z_2^{\mathcal{G}}}{v} \end{bmatrix},$$

and using function (4) to relate $\mathbf{z}^{\mathcal{G}}$ and $\mathbf{z}^{\mathcal{R}}$, we can now write the following lowest-order approximation of $R^{\mathcal{G}}$ as a function of $R^{\mathcal{R}}$:

$$R^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}) = J_n^{-1}(\mathbf{z}^{\mathcal{G}}) R^{\mathcal{R}}(n(\mathbf{z}^{\mathcal{G}}), n(\mathbf{z}_f^{\mathcal{G}})) J_n^{-1}(\mathbf{z}^{\mathcal{G}})^{\top}.$$

To obtain a model of the bias $\mathbf{b}^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}})$, starting from experimental data we postulated the following functional form, where parameters $k_1, \ldots, k_4$ are identified from data.

$$\mathbf{b}^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}) = \begin{bmatrix} k_2 \left(z_2^{\mathcal{G}}\right)^2 \left(z_1^{\mathcal{G}} - z_{f1}^{\mathcal{G}} - k_3\right) - \left(z_1^{\mathcal{G}} - z_{f1}^{\mathcal{G}}\right) e^{-k_4\left(z_1^{\mathcal{G}} - z_{f1}^{\mathcal{G}}\right)^2} \\ k_1 \arctan\left(\frac{z_2^{\mathcal{G}}}{z_1^{\mathcal{G}}}\right) \end{bmatrix} \tag{5}$$

We are now ready to discuss the perceptive model for the two kinds of percept. Let us denote by $\mathbf{y}$ the pose of the percept, as perceived by the subjects in their field of view. As we stated above, in the case of a stationary object the pose $\mathbf{y}$ coincides with its center of gravity $\mathbf{z}^{\mathcal{G}}$. We can express it in terms of the inertial coordinates $\mathbf{z}^{\mathcal{W}}$ of the object's center of gravity, and of the angle θ between frames $\mathcal{W}$ and $\mathcal{G}$, as

$$\mathbf{y} = h(\theta, \mathbf{z}^{\mathcal{W}}) := T(\theta)\mathbf{z}^{\mathcal{W}}. \tag{6}$$

When the percept is the bicycle, the pose consists of its center of gravity in the field of view, i.e., $\mathbf{z}^{\mathcal{G}}$, and its heading $\rho^{\mathcal{G}}$. Calling $\rho^{\mathcal{W}}$ the bicycle's heading expressed in the inertial frame $\mathcal{W}$ we have

$$\mathbf{y} = h(\theta, \mathbf{z}^{\mathcal{W}}, \rho^{\mathcal{W}}) := \begin{bmatrix} T(\theta)\mathbf{z}^{\mathcal{W}} \\ \rho^{\mathcal{W}} - \theta \end{bmatrix}. \tag{7}$$

Let now

$$\tilde{\mathbf{y}} \sim \mathcal{N}(\mathbf{y} + \mathbf{b}(\mathbf{y}, \mathbf{z}_f^{\mathcal{G}}), R(\mathbf{y}, \mathbf{z}_f^{\mathcal{G}})) \tag{8}$$

be the pose that is perceived by the subject. This is a normal distribution with the mean given by the percept's pose affected by bias $\mathbf{b}$, and covariance matrix R. In the case of a stationary object, we can simply define $\mathbf{b} := \mathbf{b}^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}})$, and $R := R^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}})$. When the percept is the bicycle, the bias on the perception of the center of gravity is described once again

by (5), and we assume that the heading is perceived without any bias, so that

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}) \\ 0 \end{bmatrix}.$$

In order to define the covariance matrix R, we assume that the subject perceives the heading as the relative position of two points located at an unknown position along the bicycle's longitudinal axis. We can define a further point $\mathbf{z}_p^{\mathcal{G}}$ along the heading direction, at an unknown distance d from $\mathbf{z}^{\mathcal{G}}$, such that

$$\rho^{\mathcal{G}} = \arctan(\mathbf{z}^{\mathcal{G}} - \mathbf{z}_p^{\mathcal{G}}),$$

where $\mathbf{z}^{\mathcal{G}}$ is the bicycle's center of gravity, and $\arctan(\cdot)$ is the 2-argument inverse tangent. We have that

$$\mathbf{z}_p^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \rho^{\mathcal{G}}) = \mathbf{z}^{\mathcal{G}} + d \begin{bmatrix} \cos(\rho^{\mathcal{G}}) \\ \sin(\rho^{\mathcal{G}}) \end{bmatrix}$$

and

$$\mathbf{y} = g(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_p^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \rho^{\mathcal{G}})) := \begin{bmatrix} \mathbf{z}^{\mathcal{G}} \\ \arctan(\mathbf{z}^{\mathcal{G}} - \mathbf{z}_p^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \rho^{\mathcal{G}})) \end{bmatrix}. \quad (9)$$

Calling $J_g(\rho^{\mathcal{G}}) \in \mathbb{R}^{3\times 4}$ the Jacobian of g in (9) in the variables $\mathbf{z}^{\mathcal{G}}$ and $\mathbf{z}_p^{\mathcal{G}}$, and assuming that the covariances of the perception error of both $\mathbf{z}^{\mathcal{G}}$ and $\mathbf{z}_p^{\mathcal{G}}$ are $R^{\mathcal{G}}(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}) \simeq R^{\mathcal{G}}(\mathbf{z}_p^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}) = R^{\mathcal{G}}$, we obtain

$$\begin{aligned} &R(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}}, \rho^{\mathcal{G}}) \\ &= J_g(\rho^{\mathcal{G}}) \begin{bmatrix} R^{\mathcal{G}} & 0_{2\times 2} \\ 0_{2\times 2} & R^{\mathcal{G}} \end{bmatrix} J_g^\top(\rho^{\mathcal{G}}) \\ &= \begin{bmatrix} R^{\mathcal{G}} & \frac{1}{d} R^{\mathcal{G}} \begin{bmatrix} -\sin(\rho^{\mathcal{G}}) \\ \cos(\rho^{\mathcal{G}}) \end{bmatrix} \\ \frac{1}{d} \begin{bmatrix} -\sin(\rho^{\mathcal{G}}) \\ \cos(\rho^{\mathcal{G}}) \end{bmatrix}^\top R^{\mathcal{G}} & \frac{2}{d^2} \begin{bmatrix} -\sin(\rho^{\mathcal{G}}) \\ \cos(\rho^{\mathcal{G}}) \end{bmatrix}^\top R^{\mathcal{G}} \begin{bmatrix} -\sin(\rho^{\mathcal{G}}) \\ \cos(\rho^{\mathcal{G}}) \end{bmatrix} \end{bmatrix}, \end{aligned}$$

where we omitted the dependence of $R^{\mathcal{G}}$ on $(\mathbf{z}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}})$ for the sake of conciseness.

C. The Cognitive Model

The cognitive model blends the prior knowledge about the percept's motion, described by the mental model given in Section II-A, with newly perceived information, characterized by the perceptive model introduced in Section II-B (see Fig 2). The task of the cognitive model is to describe the probability with which a subject's cognition of a percept is described by state $\mathbf{x}$, given the history of observations of the percept by the subject. We assume, in particular, that $\mathbf{x} \sim \mathcal{N}(\hat{\mathbf{x}}, P)$, where $\hat{\mathbf{x}}$ is the expected value of the cognition $\mathbf{x}$ and P is its covariance matrix. Both quantities depend on the trajectories of $\mathbf{y}$, the perceived pose, and of $\mathbf{z}_f^{\mathcal{G}}$, the subject's gaze.

The mathematical structure of the following model is that of the Extended Kalman filter (EKF), therefore we use the standard notation for the EKF: $\hat{\mathbf{x}}(t|t)$ and $P(t|t)$ denote the expected value and covariance matrix of the cognition state $\mathbf{x}$ at time t, while $\hat{\mathbf{x}}(t+1|t)$ and $P(t+1|t)$ are the *a priori* estimates at time $t+1$, before being corrected according to the perceived information at time $t+1$. Let us call

$$F := \left.\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}(t|t)}$$

the Jacobian of the function f in (1) or (2), depending on the nature of the percept. We also need to define the Jacobians of functions h in (6) and (7), once the percept's pose (position and, for the bicycle, heading) is identified with the corresponding variables in the cognitive state $\mathbf{x}$. For such a purpose, in the case of the stationary object (when $\mathbf{x} \in \mathbb{R}^2$), let us define

$$h(\theta, \mathbf{x}) := h(\theta, \mathbf{z}^{\mathcal{W}})|_{z_1^{\mathcal{W}}=x_1, z_2^{\mathcal{W}}=x_2}.$$

Similarly, in the case of the bicycle (when $\mathbf{x} \in \mathbb{R}^5$), let us define

$$h(\theta, \mathbf{x}) := h(\theta, \mathbf{z}^{\mathcal{W}}, \rho^{\mathcal{W}})|_{z_1^{\mathcal{W}}=x_1, z_2^{\mathcal{W}}=x_2, \rho^{\mathcal{W}}=x_3}.$$

We then define

$$H(\theta) := \frac{\partial h(\theta, \mathbf{x})}{\partial \mathbf{x}}.$$

The cognitive model is composed of the following equations:

$$\begin{aligned} \hat{\mathbf{x}}(t+1|t) &= f(\hat{\mathbf{x}}(t|t)), \\ P(t+1|t) &= FP(t|t)F^\top + Q, \\ S &= H(\theta(t+1))P(t+1|t)H^\top(\theta(t+1)) \\ &\quad + R(\mathbf{y}(t+1), \mathbf{z}_f^{\mathcal{G}}(t+1)), \\ K &= P(t+1|t)H^\top(\theta(t+1))S^{-1}, \\ \hat{\mathbf{y}}(t+1) &= \mathbf{y}(t+1) + \mathbf{b}(\mathbf{y}(t+1), \mathbf{z}_f^{\mathcal{G}}(t+1)) \\ \hat{\mathbf{x}}(t+1|t+1) &= \hat{\mathbf{x}}(t+1|t) \\ &\quad + \gamma(t+1)K\hat{\mathbf{y}}(t+1) \\ &\quad - \gamma(t+1)Kh(\theta(t+1), \hat{\mathbf{x}}(t+1|t)), \\ P(t+1|t+1) &= P(t+1|t) - \gamma(t+1)KSK^\top. \end{aligned} \quad (10)$$

In the above equations, the binary variable $\gamma(t) \in \{0, 1\}$ encodes the condition that at time t, the percept is within the subject's field of view (when $\gamma = 1$) or not ($\gamma = 0$), and $\hat{\mathbf{y}}$ is the expected value of the perceived pose of the percept. This according to (8) is equal to the actual pose plus a pose-dependent bias.

As a final remark, note that a trajectory of the above model depends both on the trajectory $\mathbf{y}(t)$ of the percept's pose, and on the initial conditions $\hat{\mathbf{x}}(0|0)$ and $P(0,0)$. While the trajectory of the percept's pose is an experimental datum, the initial conditions are, in principle, unknown. We model the initial state of ignorance of the subject about the percept by setting

$$P(0|0) = \lim_{p\to\infty} pI_{n_x \times n_x}. \quad (11)$$

In the case of the stationary object, this (together with the trajectory of $\mathbf{y}(t)$) fully determines the trajectory of $\hat{\mathbf{x}}(t|t)$, since (11) implies that $\hat{\mathbf{x}}(1|1) = \hat{\mathbf{y}}(1)$ and $P(1,1) = R$. The choice of $\hat{\mathbf{x}}(0|0)$ is therefore irrelevant.

In the case of the bicycle, we further set $(\hat{x}_4(0|0), \hat{x}_5(0|0)) = (0,0)$, while due to (11) the trajectory of $\hat{\mathbf{x}}(t|t)$ is independent of the initial value of $\hat{x}_1, \ldots, \hat{x}_3$.

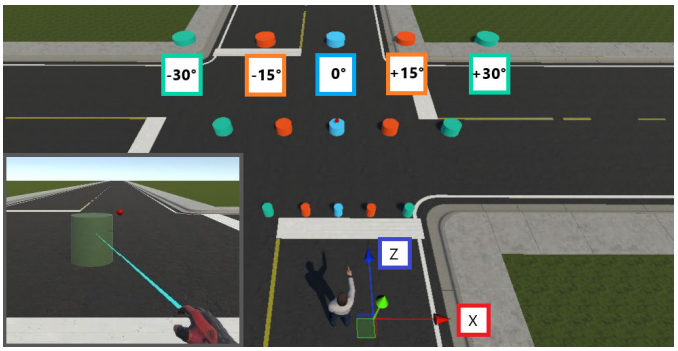

Fig. 3. Overview of the percept locations and the participant's point of view in the PXP study.

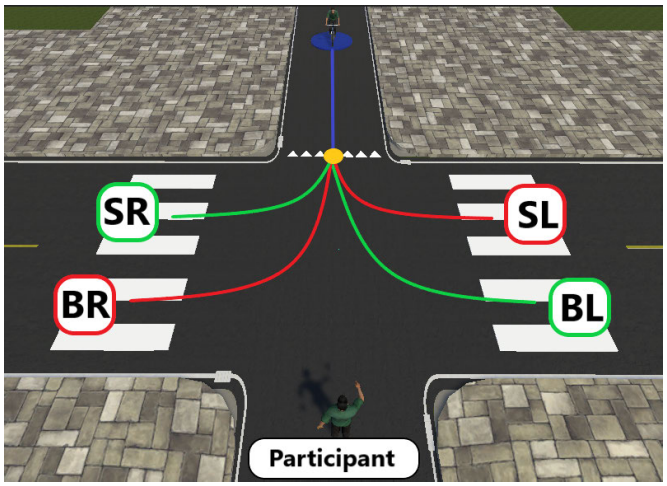

Fig. 4. Overview of the simplified road intersection and the four maneuvers considered in the CXP study.

III. Experimental Design

The model described above depends on 8 parameters when the percept is a stationary object, and 15 parameters when the percept is a bicycle. We ran two experiments to identify them. The first one, named *perception experiment* (PXP), focused on identifying the parameters of the perceptive model, namely $c_1, c_2, s_1, s_2, k_1, \ldots, k_4$. The second one, called *cognition experiment* (CXP), aimed at identifying the remaining parameters, namely Q, α, and d of the bicycle mental model. A detailed report and data analysis of the two experiments is found in [44]. In this section, we describe the experimental setup and the data-cleaning process.

A. Data Acquisition

Both experiments were held in an immersive virtual environment based on the Unity engine (version 2020.3.23) using a Meta Quest 2 visor (resolution 1832×1920 per eye; refresh rate 72Hz; horizontal field of view of approximately 90°). The virtual environment reproduced a highly simplified road intersection (see Fig. 4). The objects within the virtual environment were scaled so that 1 VR unit corresponded to 1 meter, and the subject's point of view was placed approximately 2m above ground. The subjects viewed the scene in first person and were stationary during both experiments. They could interact with the virtual environment using a Meta Touch Controller with their dominant hand (hand dominance was self-reported).

The participants were 51 students of Politecnico di Milano (28 male, 23 female) aged between 20 and 34 years ($M = 25.6$, $SD = 2.8$). All participants had normal or corrected to normal vision, and 90.2% were right-handed. Inclusion criteria were to have a driving license and to be used to right-hand traffic. Each participant provided written informed consent before taking part in the study. All data was collected anonymously. The study was approved by the Politecnico di Milano Ethical Committee.

In PXP, subjects were asked to keep their gaze on a red fixation point located 11m from them and 1m above ground, in order to standardize their field of view. The percept, in this case, a figure of a pedestrian, appeared near one of 15 possible locations within the field of view. These locations were aligned at distances of 5, 11, and 18 meters, at an angle (*eccentricity*) of 0, ±15, ±30 degrees from the subject's gaze (see Fig. 3). The percept was visible for 0.5s at a *spawning position* equal to one of the 15 above-mentioned locations, plus a random perturbation uniformly distributed in the circle of radius 1m. This served to minimize the learning effect. After this time the percept disappeared and the subject had to wait a further 0.5s, keeping the gaze on the fixation point. To ensure that participants' gazes remained at the fixation point, they were asked to aim a laser ray originating from the index finger of their avatar onto the fixation point. After this time interval, the fixation point turned green, a cylinder appeared at the end of the laser ray (see bottom-left insert in Fig. 3), and subjects were asked to position the cylinder at the location where they had perceived the percept. Each subject was tested 54 times, 6 for each location with eccentricity 0, and 3 for the other locations.

TABLE I

The MOI Names (S/B Stand for Small/Big, Respectively, LR Stand for Left/Right, Respectively) With Their Corresponding Steering Angle

MOI	SL	BL	BR	SR
steering angle (rad)	$+0.1297$	$+0.1034$	-0.1034	-0.1297

In CXP the subjects were required to look at the percept (a bicycle moving at a constant speed of 4m/s) and follow it with the virtual indicator, which was shaped like the cylinder in PXP. The bicycle with its rider was 1.6m tall and had a 1.15m wheelbase, as we mentioned before, and appeared at a randomized spawning position.

Let us denote, without loss of generality, $t = 0$ the time when the percept appeared. The percept moved straight until $t = 3$s towards the subject (the blue trajectory in Fig. 4), then it initiated one of four maneuvers of interest (MOI) depicted in red or green Fig. 4, consisting in a turn with constant steering angle from Table I. The percept remained visible until a time instant $t = T_v$, then it became invisible. T_v ranged from 3.25s to 4.25s, in steps of 0.25s. After T_v, subjects were asked to continue following what they predicted to be the trajectory of the percept with the virtual pointer. Finally, at $t = T_t$ ($T_t \in \{5, 6\}$s), the location of the virtual pointer (i.e., the subject's prediction of the percept's location) was

recorded. The corresponding *prediction time*, i.e., the interval between the last moment the bicycle was visible and the time where the prediction is recorded, therefore ranged from 0.75s (for $T_t = 5$s and $T_v = 4.25$s) to 2.75s (for $T_t = 6$s and $T_v = 3.25$s). Each subject was tested once in each of the 40 combinations of the above conditions. The percept's spawning position for MOI BL and BR was randomly selected (uniform distribution) within a circle of radius 1m around point (24.9, −0.35)m; for MOI SL and SR it was randomly selected (uniform distribution) within a circle of radius 1.7m around point (27.4, −0.35)m, all expressed in the $\mathcal{W}$ reference frame.

Notice that, both in PXP and in CXP, the percept's spawning position was equal to a nominal value plus a random perturbation δ, while the result of each experiment was a predicted position $\mathbf{z}^{\mathcal{W}}$. Before processing, the predictions from PXP and CXP were de-randomized by computing $\mathbf{z}^{\mathcal{W}} - \delta$, and the resulting distributions (for each set of experimental conditions) were cleaned from outliers using the MATLAB function *robustcov*.

IV. Results and Discussion

As we mentioned before, the model parameters were grouped in two sets: a first set

$$\rho_1 := \{c_1, c_2, s_1, s_2, k_1, \ldots, k_4\}$$

was identified from data collected in PXP, while a second set

$$\rho_2 := \{q_{11}, q_{22}, q_{33}, q_{44}, q_{55}, \alpha, d\}$$

was identified from data collected in CXP. Notice that parameters c_1, c_2, s_1, and s_2 affect the cost function through matrix R in (10), parameters $k_1, \ldots, k_4$ affect the bias $\mathbf{b}$ (see (5)), while the parameters in set ρ_2 do not appear in equations (10) when the percept is the static object, i.e., in the experimental setup of PXP.

Using subscript j to denote any one of the 15 nominal spawning positions of PXP, let $\mathbf{y}_j(t)$ be the corresponding pose of the percept (constant for all t), and let $\mu_j \in \mathbb{R}^2$ and $\Sigma_j \in \mathbb{R}^{2\times 2}$ be the mean and covariance matrix of the corresponding experimental distribution of predictions $\mathbf{z}^{\mathcal{W}}$. We estimated the parameters set ρ_1 as the optimizer of the following problem

$$\min_{\rho_1} \sum_j C_j(\hat{\mathbf{x}}_j(T|T)), P_j(T|T))$$

s.t. $\hat{\mathbf{x}}_j(t|t)$ and $P_j(t|t)$ are solutions of (10), with
f specified in (1) and with:
$\gamma(t) = 1, \quad \forall\, t \le 50$
$\gamma(t) = 0, \quad \forall\, t > 50$
$\mathbf{y}(t) = \mathbf{y}_j(t), \quad \forall\, t \ge 0,$
$\mathbf{z}_f^{\mathcal{G}}(t)$ on the fixation point, $\quad \forall\, t \ge 0,$

where

$$C_j(\hat{\mathbf{x}}_j(T|T)), P_j(T|T)), := \|\hat{\mathbf{x}}_j(T|T) - \mu_j\|_2 + \|P_j(T|T) - \Sigma_j\|_F. \quad (12)$$

In the above formula, $\|\cdot\|_2$ is the Euclidean norm, and $\|\cdot\|_F$ is the Frobenius norm, while the value of T is set to 100. Given that we assumed a discretization time step $\tau_s = 0.01$s, this corresponds to $1s$, i.e., the time elapsed from the moment the percept was first visible to the subject, to the moment the subject was required to indicate its perceived position. For the same reason, the conditions on $\gamma(t)$ encode the fact that the percept was visible for the first 0.5s, then became invisible.

The remaining parameters set ρ_2 was identified from data collected in CXP, using for the parameters in ρ_1 the values determined through the previous experiment. Let subscript j indicate any one of the 40 experimental conditions of CXP (4 MOIs, 2 possible values of T_t, 5 possible values of T_v). Like before, let $\mu_j \in \mathbb{R}^2$ and $\Sigma_j \in \mathbb{R}^{2\times 2}$ be the mean and covariance matrix of the corresponding experimental distribution of predictions $\mathbf{z}^{\mathcal{W}}$, $\mathbf{y}_j(t)$ be the corresponding pose of the percept (center of gravity and heading of the bicycle), and $T_j = T_t/\tau_s$, i.e., the duration in time steps of the experiment in experimental condition j. We estimated parameters ρ_2 as the minimizer of the following problem

$$\min_{\rho_2} \sum_j C_j(\hat{\mathbf{x}}_j(T_j|T_j)), P_j(T_j|T_j))$$

s.t. $\hat{\mathbf{x}}_j(t|t)$ and $P_j(t|t)$ are solutions of (10), with
f specified in (2), and with
$\gamma(t) = 1, \quad \forall\, t \le T_v,$
$\gamma(t) = 0, \quad \forall\, t > T_v,$
$\mathbf{y}(t) = \mathbf{y}_j(t), \quad \forall\, t \ge 0,$
$\mathbf{z}_f^{\mathcal{G}}(t)$ in the centre of gravity of the bicycle, $\quad \forall\, t \ge 0.$

Unlike in (12), the cost function C_j now depends only on a subset of the elements of $\hat{\mathbf{x}}$ and P, namely those that correspond to spatial coordinates in $\mathcal{W}$. Let us call

$$m(\hat{\mathbf{x}}_j) := (\hat{x}_{j,1}, \hat{x}_{j,2}),$$
$$S(P_j) := \begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}.$$

We have

$$C_j(\hat{\mathbf{x}}_j(T_j|T_j)), P_j(T_j|T_j), w) := 0.01\|m(\hat{\mathbf{x}}_j(T_j|T_j)) - \mu_j\|_2 + 0.99\|S(P_j(T_j|T_j)) - \Sigma_j\|_F.$$

Note that the above cost functions depend on a difference between the perception/cognition model mean and covariance matrices, and the experimental mean and covariance matrices. We are therefore fitting the parameters to optimally approximate the experimental distributions, i.e., to reproduce the inaccuracy with which subjects perceived and predicted the positions of the percepts.

To assess the quality of the resulting fitted model, we compared the model and experimental distributions obtained from each experimental condition (i.e., the 15 spawning positions in PXP, and the 40 combinations of MOI, T_t, and T_v in CXP) through a one-sample Kolmogorov-Smirnov (KS) test.

A. Results of PXP

The experimental and model distributions of the perception errors in PXP are represented in Fig. 5, with fitted parameters

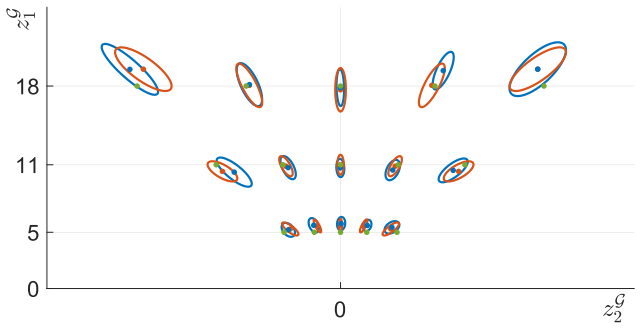

Fig. 5. Experimental (blue) and fitted-model (red) distributions from PXP.

TABLE II
THE VALUES OF THE FITTED PARAMETERS ρ_1

s_1	s_2	c_1	c_2
0.015	0.012	7.092	30.701
k_1	k_2	k_3	k_4
0.011	0.005	3.228	0.062

reported in Table II. In the figure, percept locations are aligned along 5 directions with an angle of 0, ± 15, and ± 30 degrees from the subject's gaze. The ellipses are the equidensity contours at one standard deviation of the respective distributions, while the dots are the mean values of the corresponding distributions. The green dots represent the actual spawning position of the percept, after the removal of the random perturbation.

The overlap of the experimental and model distributions suggests that the model reproduces rather faithfully the perception error distributions in all 15 of the tested locations. We further tested the 15 model distributions for statistical equivalence with the experimental ones, using a one-sample KS test with a significance level of 0.05. Given that the distributions are bivariate, while the KS test applies to univariate distributions, the test was repeated 10 times with Bonferroni correction, for 10 projections of the distributions along directions equally spaced in the round angle, i.e., with a spacing of 18°. A model distribution is then considered statistically equivalent to the experimental one if all 10 projections of the model distributions are simultaneously statistically equivalent to the corresponding projections of the experimental distribution. According to this test, only 3 out of the 15 model distributions (namely the one with eccentricity $-15°$ and distance 18m, and the ones with eccentricity 30° and distances 5m and 18m) are statistically equivalent to the experimental ones. The experimental distributions are, however, visibly non-symmetric with respect to the $z_1^{\mathcal{G}}$ axis, thus violating one of our modeling assumptions. Given the large sample size (153 samples for each location with nonzero eccentricity, 306 for locations with 0 eccentricity), we may conclude that this is a feature of the process. The presence of some degree of left/right asymmetry in perception accuracy is indeed well documented [45], [46] and is therefore not surprising. The gist of our modeling approach, however, is to minimize the number of parameters and root the mathematical representation on first principles as much as possible, in order to minimize the risk of overfitting and preserve interpretability. It is therefore unclear, at this stage, how the model could be improved to account for these asymmetries without denaturing the approach. Despite this limitation, as we will see in the next section, the significant overlap between model and experimental distributions is sufficient to support the functionality of the cognitive model.

TABLE III
THE VALUES OF THE FITTED PARAMETERS IN ρ_2

q_{11}, q_{22}	q_{33}	q_{44}
3.11×10^{-3}	4.45×10^{-8}	9.01×10^{-6}
q_{55}	α	d
2.80×10^{-3}	9.96×10^{-1}	3.98×10^{11}

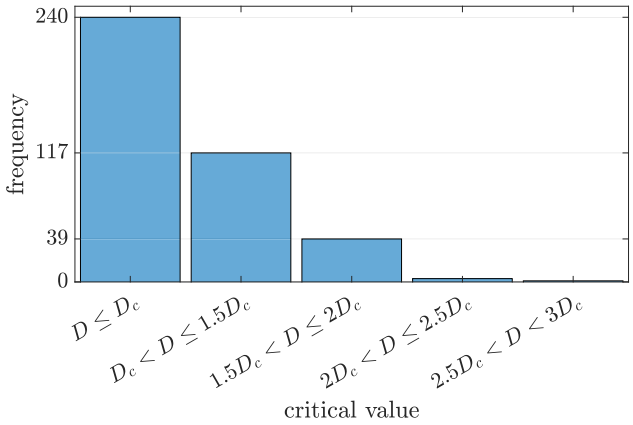

Fig. 6. Distribution of the KS test statistics D of the model distributions from CXP (40 conditions, 10 projections per condition), relative to the tests' critical values D_c.

B. Results of CXP

The fitted parameters are reported in Tab. III. The model distributions were once again tested for statistical equivalence with the experimental ones with a one-sample KS test on 10 equispaced projections, with a significance level of 0.05 and Bonferroni correction, finding that 9 of the 40 distributions are statistically equivalent. Statistical fit is therefore far from perfect. We may more precisely quantify *how far* the model is from reproducing the experimental distributions, by comparing the KS statistics D of each of the 400 KS tests (40 conditions and 10 projections per condition) with the corresponding KS critical value D_c [47], [48]. We remind that a model distribution projection is statistically equivalent to the corresponding experimental distribution projection if its KS test statistic satisfies $D \leq D_c$. We see in Fig. 6 that 240 out of the 400 model distribution projections are statistically equivalent to the corresponding experimental distribution projections, and in 117 of the remaining tests the KS statistic falls within 1.5 times the critical value. In other words, our model reproduces the experimental data with statistics below or near the critical value in almost 90% of the projections: the model distributions are not always statistically equivalent to, but are very close to the experimental ones under most

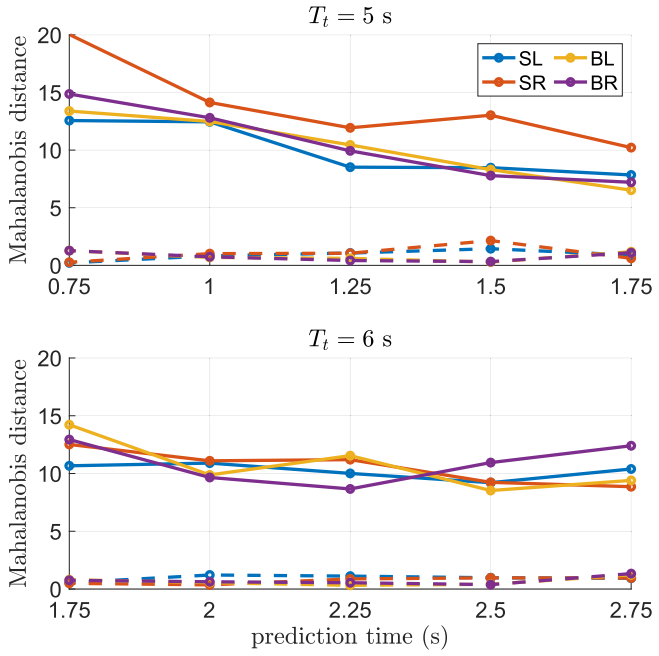

Fig. 7. Mahalanobis distances between the mean value of trivial predictor (solid) and cognitive model (dashed) distributions, and the distribution of experimental predictions, for the four maneuvers described in Table I.

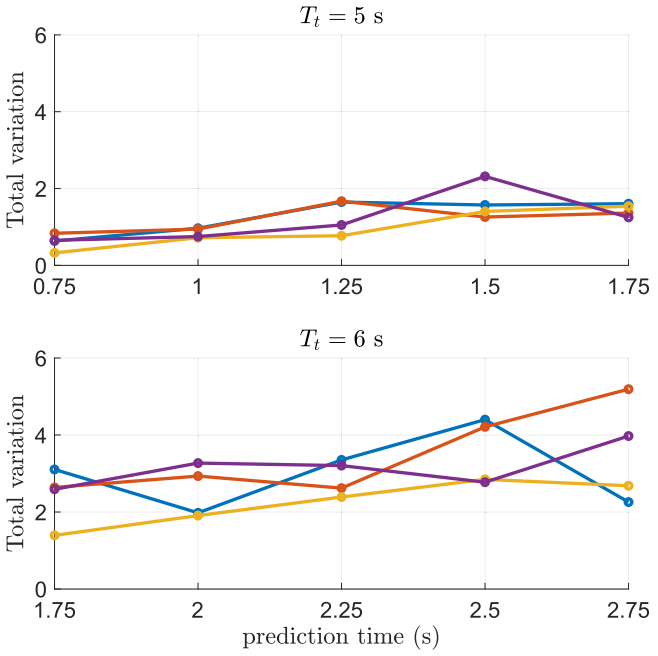

Fig. 8. Total variation of the distributions of experimental predictions, color-coded as in Fig. 7.

projections. This remarkable performance is further supported if we compare the model with the *trivial predictor*, obtained by assuming that the subject's prediction coincides with the actual position of the bicycle, plus a zero-mean Gaussian noise. We compare the two predictors through the Mahalanobis distance of their mean predictions from the experimental distributions in Fig. 7. Note that in this and the following figures we display data for fixed values of T_t, and for increasing prediction time. Given that T_v is the difference between T_t and the prediction time, the same figures can be interpreted as showing data for *decreasing* T_v: reading the figures left to right, the experimental condition has subjects seeing a shorter portion of the MOI, and predicting over a longer time horizon. Under all experimental conditions, the mean values of the cognitive model's predictions are significantly closer to the experimental distributions than those of the trivial predictor. Our cognitive model, in other words, does a much better job at approximating the (mean) prediction errors made by the subjects of our experiment, despite not reproducing all experimental distributions of such errors with statistical significance. Furthermore, the accuracy with which our cognitive model reproduces the experimental prediction error is quite independent of T_v and of the prediction time in Mahalanobis distance. Interestingly, the distance between the mean values of the trivial prediction and the experimental distribution exhibits a visible decreasing trend when $T_t = 5$s for all MOIs. One might expect the opposite, i.e., a larger discrepancy between the trivial model and experimental data at longer prediction times and shorter T_v. This is, however, an artifact of the increasing total variation (i.e., the trace of the covariance matrix) of the experimental distributions (Fig. 8), coupled with the fact that the Mahalanobis distance is an

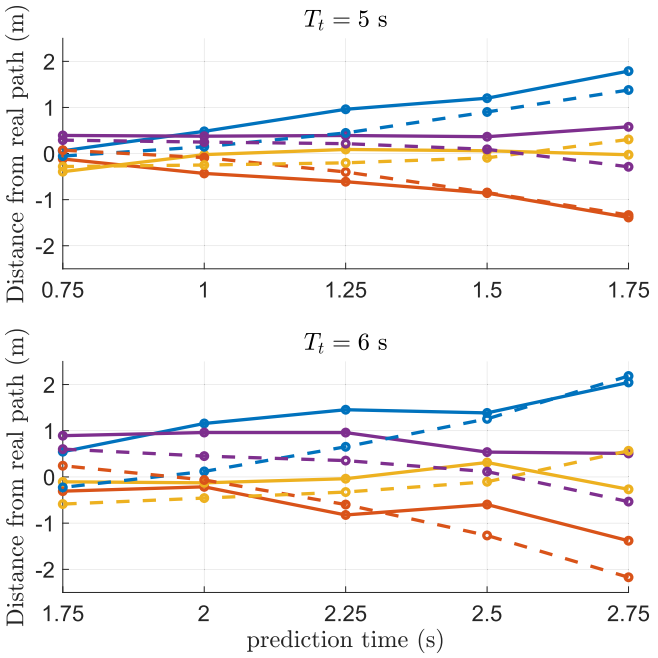

Fig. 9. Signed distance between the mean value of the cognitive model distribution and the real path of the bicycle (dashed) and between the mean value of the experimental distributions and the real path of the bicycle (solid). Color-coding as in Fig. 7. The signed distance is positive if the mean value of the distribution lies to the right of the bicycle path (in the direction of motion), negative otherwise.

Euclidean distance weighted on the distribution's covariance matrix.

The experimental distributions also display an interesting trend in underestimating the curvature radius for the tighter turns. This is clearly visible in Fig. 9: the mean experimental prediction for the tight left turn (solid blue line, MOI: SL)

lies to the right of the real bicycle path, increasingly more for longer prediction time and shorter T_v; similarly, the mean experimental prediction for the tight right turn (solid red line, MOI: SR) lies to the left of the real bicycle path, increasingly more for longer prediction time and shorter T_v. The same trend is not as clearly visible for the larger turns (MOI: BL and BR). Our model (dashed lines in Fig. 9) reproduces these trends faithfully over all prediction horizons, thus correctly capturing, at least qualitatively, the tendency of the tested to underestimate the curvature of tighter turns.

Interestingly, parameter d in Table III is extremely large. Remembering that, in our first-principles modeling framework, d represents the distance between the endpoints of the imaginary segment utilized by the subject to perceive the percept's heading, this implies that the subject is perceiving heading as purely directional information, and not through the processing of the relative positions of the two endpoints. From a computational perspective, it corresponds to setting to (almost) 0 the last row and column of the matrix $R(\mathbf{y}^{\mathcal{G}}, \mathbf{z}_f^{\mathcal{G}})$ in (10), thus stating that the error on the perceived heading is negligible with respect to that on position.

V. Conclusion

We have proposed a novel cognitive model capable of reproducing the error with which human subjects estimate the short-term trajectory of a moving bicycle, using information about the location of the bicycle within the human's field of view and a model of human perception.

Our primary objective in exploring the feasibility of such a model is to develop a computational means to statistically characterize human misunderstanding of the short-term trajectory of a bicycle (and, in the future, of other VRUs). A vehicle equipped with an eye-tracking device could use such a model to acquire information regarding the probability that its driver is significantly misjudging the near-future behavior of the bicycle. It could use this information, together with information collected through its onboard sensors, to decide whether an ADAS intervention is appropriate.

The model is inspired by the similarity between human cognition of moving objects and the Kalman filter. The model depends on 15 parameters, which were identified from experimental data involving 51 subjects. We followed a postdictive approach [30], therefore aiming to prove that the chosen model can reproduce the experimental data. Our objective was, in other words, to test whether the model, with suitably tuned parameters, could reproduce the errors in the human prediction of the position of the bicycle in all 40 experimental conditions described in Sec. III.

The cognition error distributions generated by the model are, in many cases, not statistically equivalent to the experimental ones. However, we have shown in Sec. IV that the distributions generated by the model emulate the experimental ones significantly well, and can even reproduce the tested subject's tendency to misjudge the bicycle turning radius systematically. This is a particularly significant outcome since this trend is an emergent property, not explicitly encoded in the model structure. Therefore, it suggests a strong similarity between the model functioning and that of the modeled cognitive process. As a consequence, our result strongly supports the opportunity of utilizing models such as the one proposed here, to approximate the likely cognition error of a human driver in predicting the trajectory of VRUs. The simplistic assumptions of a bike moving at a constant speed and piecewise constant curvature were justified by our aim to predict misjudgments over a time horizon of a few seconds. On such a short time interval, the velocity and curvature changes of a real bicycle are not very large. Additionally, the choice of keeping our subjects still during the experiments (rather than, for instance, simulating a driver sitting in a moving car) was taken to simplify the visual scenario and minimize the potential impact of visual references and occlusions on the subjects' predictions. It was, in other words, dictated by the need to obtain more clearly interpretable results, at the cost of some realism. We recognize, however, that the effect of these simplifications should be assessed, by testing our model on more complex scenarios, as well as by assessing the effects of ego-motion on the cognitive model. An important future extension will therefore be to test this model on a larger variety of percept trajectories, for instance on naturalistic trajectories, with imperfect geometry, variable speed profile, and viewed from different angles, and to test its ability to generalize to trajectories of a different nature than those used to fit the model. Note, however, that the use of a relatively simple model, with a very small number of parameters, makes this family of models much less prone to overfitting than, for instance, a neural network.

Another interesting extension will be to train and test the model with VRUs of a different nature, for example, pedestrians. Note that the application of our model to a different type of percept requires the redesign of the mental model, but not of the whole cognitive model. While this is not a trivial task, we believe that the framework that we presented in this work provides a very valuable headstart towards such extensions.

In principle, the model could also be extended to scenarios where multiple percepts are simultaneously present, by running multiple Kalman filters in parallel. There is however some debate in the literature regarding the ability (or lack of) of the human mind to process multiple data sources simultaneously, as well as about the mechanisms underlying these limits [49], [50]. In order to address a complex multi-percept scenario, the processing bottleneck should likely be modeled explicitly, and validated with experimental data.

It is also worth mentioning that the proposed method could be extended to model other subject-percept configurations. For example, the cognition by a human cyclist (as the subject) of the motion of an automated vehicle (as the percept). This would allow the autonomous vehicle to assess the cyclist's situation awareness and adapt its driving policy to the potentially irrational behavior of the cyclist. While the considered experimental data does not allow the validation of such cases, the authors believe that the theoretical framework of this work would be suitable to handle them, provided that the necessary data is available.

References

[1] *Global Status Report on Road Safety 2023*, World Health Org., Geneva, Switzerland, 2023.

[2] *Target Crash Population for Crash Avoidance Technologies in Passenger Vehicles*, Nat. Highway Traffic Saf. Admin., Washington, DC, USA, Mar. 2019.

[3] *Advanced Driver Assistance Systems*, Eur. Commission, Brussels, Belgium, 2018.

[4] S. Bouhsissin, N. Sael, and F. Benabbou, "Driver behavior classification: A systematic literature review," *IEEE Access*, vol. 11, pp. 14128–14153, 2023.

[5] M. N. Azadani and A. Boukerche, "Driving behavior analysis guidelines for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6027–6045, Jul. 2022.

[6] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 37, no. 1, pp. 32–64, Mar. 1995.

[7] M. R. Endsley, B. Bolte, and D. G. Jones, *Designing for Situation Awareness*. New York, NY, USA: Taylor & Francis, 2003.

[8] A. Eskandarian, *Handbook of Intelligent Vehicles*. Cham, Switzerland: Springer, 2012.

[9] R. Buendia, F. Forcolin, J. Karlsson, B. Arne Sjöqvist, A. Anund, and S. Candefjord, "Deriving heart rate variability indices from cardiac monitoring—An indicator of driver sleepiness," *Traffic Injury Prevention*, vol. 20, no. 3, pp. 249–254, Mar. 2019.

[10] Y. Lu and L. Bi, "Combined lateral and longitudinal control of EEG signals-based brain-controlled vehicles," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1732–1742, Sep. 2019.

[11] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, and M. Fathy, "A review on driver face monitoring systems for fatigue and distraction detection," *Int. J. Adv. Sci. Technol.*, vol. 64, pp. 73–100, Mar. 2014.

[12] A. Kashevnik, I. Lashkov, A. Ponomarev, N. Teslya, and A. Gurtov, "Cloud-based driver monitoring system using a smartphone," *IEEE Sensors J.*, vol. 20, no. 12, pp. 6701–6715, Jun. 2020.

[13] A. Kashevnik, I. Lashkov, and A. Gurtov, "Methodology and mobile application for driver behavior analysis and accident prevention," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2427–2436, Jun. 2020.

[14] C. Ahlström and K. Kircher, "Review of real-time visual driver distraction detection algorithms," in *Proc. 7th Int. Conf. Methods Techn. Behav. Res.*, Aug. 2010, pp. 1–4.

[15] A. Fernández, R. Usamentiaga, J. Carús, and R. Casado, "Driver distraction using visual-based sensors and algorithms," *Sensors*, vol. 16, no. 11, p. 1805, Oct. 2016.

[16] A. Kashevnik, R. Shchedrin, C. Kaiser, and A. Stocker, "Driver distraction detection methods: A literature review and framework," *IEEE Access*, vol. 9, pp. 60063–60076, 2021.

[17] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

[18] A. Quddus, A. Shahidi Zandi, L. Prest, and F. J. E. Comeau, "Using long short term memory and convolutional neural networks for driver drowsiness detection," *Accident Anal. Prevention*, vol. 156, Jun. 2021, Art. no. 106107.

[19] A. Doshi and M. Trivedi, "A comparative exploration of eye gaze and head motion cues for lane change intent prediction," in *Proc. IEEE Intell. Vehicles Symp.*, vol. 14, Jun. 2008, pp. 49–54.

[20] N. Kuge, T. Yamamura, O. Shimoyama, and A. Liu, "A driver behavior recognition method based on a driver model framework," SAE Tech. Paper 2000-01-0349, 2000.

[21] J. Dahl, G. R. de Campos, and J. Fredriksson, "Intention-aware lane keeping assist using driver gaze information," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2023, pp. 1–7.

[22] A. Doshi and M. Trivedi, "Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2009, pp. 887–892.

[23] J. Pohl, W. Birk, and L. Westervall, "A driver-distraction-based lane-keeping assistance system," *Proc. Inst. Mech. Eng. I, J. Syst. Control Eng.*, vol. 221, no. 4, pp. 541–552, Jun. 2007.

[24] Y. Xing, C. Lv, H. Wang, D. Cao, and E. Velenis, "An ensemble deep learning approach for driver lane change intention inference," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102615.

[25] X. Li, W. Wang, and M. Roetting, "Estimating driver's lane-change intent considering driving style and contextual traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3258–3271, Sep. 2019.

[26] S. E. Lee, E. C. Olsen, and W. W. Wierwille, "A comprehensive examination of naturalistic lane-changes," Nat. Highway Traffic Saf. Admin., Washington, DC, USA, Tech. Rep. DOT-HS-809-702, 2004.

[27] I. Isaksson-Hellman and J. Werneke, "Detailed description of bicycle and passenger car collisions based on insurance claims," *Saf. Sci.*, vol. 92, pp. 330–337, Feb. 2017.

[28] J. Anderson and C. Lebiere, *The Atomic Components of Thought*. London, U.K.: Psychology Press, 1998.

[29] J. E. Laird, A. Newell, and P. S. Rosenbloom, "SOAR: An architecture for general intelligence," *Artif. Intell.*, vol. 33, no. 1, pp. 1–64, Sep. 1987.

[30] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of mind," *Psychol. Rev.*, vol. 111, no. 4, pp. 1036–1060, 2004.

[31] K. S. Haring, "A cognitive model of drivers attention," in *Proc. 11th Int. Conf. Cognit. Model.*, 2012, pp. 275–280.

[32] B. Song, D. Delorme, and J. VanderWerf, "Cognitive and hybrid model of human driver," in *Proc. IEEE Intell. Vehicles Symp.*, Aug. 2000, pp. 1–6.

[33] A. Fries, F. Fahrenkrog, K. Donauer, M. Mai, and F. Raisch, "Driver behavior model for the safety assessment of automated driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2022, pp. 1669–1674.

[34] A. D. Dumbuya and R. L. Wood, "Visual perception modelling for intelligent virtual driver agents in synthetic driving simulation," *J. Experim. Theor. Artif. Intell.*, vol. 15, no. 1, pp. 73–102, Jan. 2003.

[35] R. D. Beer, "Dynamical approaches to cognitive science," *Trends Cognit. Sci.*, vol. 4, no. 3, pp. 91–99, Mar. 2000.

[36] L. Acerbi, W. J. Ma, and S. Vijayakumar, "A framework for testing identifiability of Bayesian models of perception," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[37] M. Fritsche, E. Spaak, and F. P. de Lange, "A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception," *eLife*, vol. 9, Jun. 2020, Art. no. e55389.

[38] K. J. Lakshminarasimhan, M. Petsalis, H. Park, G. C. DeAngelis, X. Pitkow, and D. E. Angelaki, "A dynamic Bayesian observer model reveals origins of bias in visual path integration," *Neuron*, vol. 99, no. 1, pp. 194–206.e5, Jul. 2018.

[39] T. S. Manning, B. N. Naecker, I. R. McLean, B. Rokers, J. W. Pillow, and E. A. Cooper, "A general framework for inferring Bayesian ideal observer models from psychophysical data," *eNeuro*, vol. 10, no. 1, pp. 1–17, Jan. 2023, doi: 10.1523/ENEURO.0144-22.2022.

[40] G. R. De Campos, R. Kianfar, and M. Brännström, "Precautionary safety for autonomous driving systems: Adapting driving policies to satisfy quantitative risk norms," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 645–652.

[41] J. Anderson, "Driver interaction, informal rules, irritation and aggressive behavior," Ph.D. thesis, Dept. Psychol., Uppsala Univ., Uppsala, Sweden, 2005.

[42] P. Corke, *Robotics, Vision & Control*. Cham, Switzerland: Springer, 2007.

[43] A. H. Yoo, L. Acerbi, and W. J. Ma, "Uncertainty is maintained and used in working memory," *J. Vis.*, vol. 21, no. 8, p. 13, Aug. 2021.

[44] N. Dozio, L. Rozza, M. S. Lukasiewicz, A. Colombo, and F. Ferrise, "Localization and prediction of visual targets' position in immersive virtual reality," *PRESENCE, Virtual Augmented Reality*, vol. 31, pp. 5–21, Dec. 2022.

[45] G. Jewell and M. E. McCourt, "Pseudoneglect: A review and meta-analysis of performance factors in line bisection tasks," *Neuropsychologia*, vol. 38, no. 1, pp. 93–110, Jan. 2000.

[46] A. Nuthmann and C. N. L. Clark, "Pseudoneglect during object search in naturalistic scenes," *Exp. Brain Res.*, vol. 241, no. 9, pp. 2345–2360, Sep. 2023.

[47] W. Feller, "On the Kolmogorov–Smirnov limit theorems for empirical distributions," *Ann. Math. Statist.*, vol. 19, no. 2, pp. 177–189, Jun. 1948.

[48] F. J. Massey Jr., "The Kolmogorov–Smirnov test for goodness of fit," *J. Amer. Statist. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.

[49] J. M. Scimeca and S. L. Franconeri, "Selecting and tracking multiple objects," *WIREs Cognit. Sci.*, vol. 6, no. 2, pp. 109–118, Mar. 2015.

[50] A. Holcombe, *Attending to Moving Objects* (Elements in Perception). Cambridge, U.K.: Cambridge Univ. Press, 2023.

Alessandro Colombo (Senior Member, IEEE) received the Diplôme d'Ingénieur degree from ENSTA in 2005 and the Ph.D. degree from the Politecnico di Milano in 2009. He was a Post-Doctoral Associate with Massachusetts Institute of Technology from 2010 to 2012 and is currently an Associate Professor with the Department of Electronics, Information, and Bioengineering, Politecnico di Milano. His research interests include analysis and control of hybrid systems and in applications of control theory to the modeling and understanding of different human activities.

Matteo Depaola received the M.Sc. degree in automation and control engineering from the Politecnico di Milano, Italy, in 2023, with a thesis on the development of a cognitive model. He is currently pursuing the Ph.D. degree with KU Leuven, Belgium. He is also a Researcher with Siemens Digital Industries Software, Leuven, Belgium. His research interests include the applications of estimation techniques, in particular, in the field of microelectronics.

Francesco Ferrise received the Ph.D. degree in design and methods for product development from the Politecnico di Milano, Italy, in 2010. He is currently a Full Professor with the Department of Mechanical Engineering, Politecnico di Milano. His research interests include the use of virtual and augmented reality technologies to support product development. He is an Associate Editor of the *ASME Journal of Computing and Information Science in Engineering* (JCISE) and the *IEEE Computer Graphics and Applications*.

Nicolò Dozio received the Ph.D. degree in mechanical engineering from the Politecnico di Milano, Italy, in 2023. He is currently a Post-Doctoral Researcher with the Department of Mechanical Engineering, Politecnico di Milano. His research interests include the use of eXtended reality to explore cognitive and emotional aspects involved in human–machine interaction.

Gabriel Rodrigues de Campos received the Ph.D. degree in automatic control from Grenoble University/Grenoble INP, France, in 2012. He is currently a Researcher with Zenseact, Sweden. Prior to joining Zenseact, he was a Post-Doctoral Fellow with the Department of Signals and Systems, Chalmers University of Technology, Sweden, and DEIB, Politecnico di Milano, Italy. His research interests include cooperative and distributed control, safety assurance, driver behavior, and human factors, as well as threat assessment and decision-making.

Open Access funding provided by 'Politecnico di Milano' within the CRUI CARE Agreement