



Variable-Drift Diffusion Models of Pedestrian Road-Crossing Decisions

Jami Pekkanen¹ · Oscar Terence Giles² · Yee Mun Lee³ · Ruth Madigan³ · Tatsuru Daimon⁴ · Natasha Merat³ · Gustav Markkula³

Accepted: 23 August 2021 / Published online: 5 October 2021
© The Author(s) 2021

Abstract

Human behavior and interaction in road traffic is highly complex, with many open scientific questions of high applied importance, not least in relation to recent development efforts toward automated vehicles. In parallel, recent decades have seen major advances in cognitive neuroscience models of human decision-making, but these models have mainly been applied to simplified laboratory tasks. Here, we demonstrate how variable-drift extensions of drift diffusion (or evidence accumulation) models of decision-making can be adapted to the mundane yet non-trivial scenario of a pedestrian deciding if and when to cross a road with oncoming vehicle traffic. Our variable-drift diffusion models provide a mechanistic account of pedestrian road-crossing decisions, and how these are impacted by a variety of sensory cues: time and distance gaps in oncoming vehicle traffic, vehicle deceleration implicitly signaling intent to yield, as well as explicit communication of such yielding intentions. We conclude that variable-drift diffusion models not only hold great promise as mechanistic models of complex real-world decisions, but that they can also serve as applied tools for improving road traffic safety and efficiency.

Keywords Evidence accumulation · Gap acceptance · Human-robot interaction

Introduction

Human locomotion, as performed by individuals either alone or in concert with others, has been an object of scientific study for a long time (Gibson & Crooks, 1938),

and has engendered a wide range of often cross-disciplinary computational modeling research, spanning domains such as perception, motor control, decision-making, social interaction, human-robot coexistence, and more (Fajen & Warren, 2003; Hoogendoorn & Bovy, 2003; Lee, 1976; Markkula et al., 2018; Turnwald et al., 2016). In road traffic, the successes or failures of human movement and sharing of space has particularly large societal implications, in terms of mobility, productivity, and human safety, and consequently considerable effort has been invested into understanding and modeling how humans locomote both as vehicle drivers and vulnerable road users (Helbing, 2001; Markkula et al., 2012; Plöchl & Edelmann, 2007). These efforts have further intensified recently, to support development of increasingly automated vehicles (Camara et al., 2020; Sadigh et al., 2018; Schwarting et al., 2019). By many accounts, successful widespread deployment of automated vehicles will be limited by the extent to which these vehicles can encapsulate a sufficient understanding—typically in the form of computational models—of road user behavior and interaction (Brown & Laurier, 2017; Camara et al., 2020; Markkula et al., 2020; Schieben et al., 2019).

Existing approaches to computational modeling of road user behavior mirror the modeling paradigms in the

✉ Jami Pekkanen
jami.pekkanen@helsinki.fi

Oscar Giles
ogiles@turing.ac.uk

Tatsuru Daimon
daimon@ae.keio.ac.jp

Gustav Markkula
G.Markkula@leeds.ac.uk

¹ Cognitive Science, Department of Digital Humanities, University of Helsinki, P.O. box 9, Helsinki 00014, Finland

² The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, England

³ Institute for Transport Studies, University of Leeds, Leeds, UK

⁴ Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan

wider cognitive and behavioral sciences, including cognitive architectures (Salvucci, 2006), ecological psychology (Fajen, 2013), classical and optimal control theory (Plöchl & Edelmann, 2007), rational decision-making (Choudhury et al., 2007), game theory (Elvik, 2014; Hoogendoorn & Bovy, 2003), as well as data-driven modeling using machine learning approaches (Behbahani et al., 2019; Ma et al., 2016). However, most of these existing models have either emphasized detailed modeling of individual road user behavior, or more coarse-grained modeling of interactions of larger number of road users, for example, to study high-level traffic flow. Computational modeling of the subtler details of local interactions between individuals is still in its infancy (Camara et al., 2020; Markkula et al., 2020).

One type of model that has been uncommon in road user modeling, but which has over recent decades become prominent in more basic psychology and cognitive neuroscience research, is drift diffusion, or evidence accumulation, models of decision-making. Broadly speaking, these models assume that decisions are made by means of noisy integration of evidence for or against decision alternatives, up to a threshold at which the decision is made. The saliency of the available evidence (for example, the movement coherence of a set of dots on a visual display, in a paradigm where the task is to judge the overall direction of dot motion) affects evidence accumulation rate and thus overall response times, and the noise in the accumulation process introduces variability, allowing these models to predict full distributions of choices made and the corresponding response times (Gold & Shadlen, 2007; Ratcliff et al., 2016). These general ideas can take a range of more specific computational forms, some of which explicitly leverage neuroscientific concepts and modeling components (Bogacz & Gurney, 2007; Purcell et al., 2010; Usher & McClelland, 2001; Wong & Wang, 2006), for example, inhibition between competing decisions, similar to lateral inhibition in the brain. Other model formulations take a more behavioral than neural perspective, such as the well-known drift diffusion model (DDM) (Ratcliff, 1978; Ratcliff et al., 2016) or linear ballistic accumulator (LBA) (Brown & Heathcote, 2008). Overall, there is now a large literature showing that this general class of model can be highly successful at accounting for both behavioral responses as well as neural data, in both humans and non-human primates, across a range of laboratory paradigms on especially perceptual decision-making (e.g., discrimination of random dot motion direction) and value-based choice (e.g., between different food items) (Brosnan et al., 2020; Busemeyer et al., 2019; Ratcliff et al., 2016). However, it is less well known to what extent models of this nature can describe human decision-making well also in more applied and embodied contexts, for example, relating to human sensorimotor control and movement in the real world.

We and others have investigated the application of drift diffusion-type models in the road traffic context, with promising results initially for low-level locomotion decisions on applying braking or steering control (Markkula et al., 2018; Piccinini et al., 2020; Xue et al., 2018), more recently also extending to multi-agent interaction situations (Boda et al., 2020; Giles et al., 2019; Kovaceva et al., 2020; Markkula et al., 2018; Zgonnikov et al., 2020). One key step for bringing these models to bear in these contexts has been to relax the limitation to stationary or intermittently changing sensory input. This limitation has been the norm in laboratory paradigms, in part because for stationary input, model likelihood functions can be written in closed mathematical form, which simplifies model-fitting (Navarro & Fuss, 2009; Wiecki et al., 2013). However, in the context of real-world sensorimotor behavior, sensory evidence is more often than not continuously changing over time, as emphasized, for example, in the ecological psychology and perceptual control theory research traditions (Gibson, 1958; Lee, 1976; Powers, 1978). These generalizations to real-world tasks and time-varying evidence bring challenges, thus far not fully resolved, both in terms of model-fitting methodology and in terms of the more limited sample sizes that typically arise when doing controlled data collection in conditions with high external validity.

A specific type of road traffic scenario that has been the focus of increasingly intense human-automation interaction research (but which is of course relevant to traffic safety also in non-automated traffic) is pedestrian road-crossing. The majority of the research in this area has been observational (Lobjois & Cavallo, 2007; Schneemann & Gohl, 2016; Varhelyi, 1998), but some previous mathematical models exist. In the context of large-scale traffic simulation, logistic regression models have long been used to model pedestrian “gap acceptance” between vehicles in a stream of traffic (Brewer et al., 2006; Schroeder, 2008; Yannis et al., 2013), and the use of such models in automated vehicle algorithms has also been proposed (Jayaraman et al., 2021; Kapania et al., 2019). However, these models are limited to a discrete acceptance/rejection decision per gap, and do not account for the timing of road-crossing decisions, which has implications for traffic flow and acceptance of automated vehicles (Dey et al., 2020; Lee et al., 2020; Markkula et al., 2018). The existing models also do not account for how pedestrians respond to vehicles yielding to them, a process which is known to be non-trivial: Human drivers tend to communicate with pedestrians both *implicitly*, for example, using exaggerated deceleration, and *explicitly* with communicative signals (Domeyer et al., 2019; Markkula et al., 2020). Studies are currently investigating the extent to which automated vehicles should indicate their intentions in similar ways, for example, by means of external human-machine interfaces

(eHMI), either leveraging conventional signals such as headlight flashes, or using novel designs (Faas et al., 2020; Lee et al., 2019; Lee et al., 2020). However, these communicative aspects of vehicle-pedestrian interactions remain poorly understood, and have not been the subject of computational modeling.

We have previously shown that connected networks of one or more drift diffusion type models driven by time-varying sensory input can capture qualitative patterns in how pedestrian road-crossing decisions depend on time gaps and yielding deceleration (Markkula et al., 2018), and we have also reported a tentative attempt at fitting these models to quantitative human data (Giles et al., 2019). These preliminary results were promising, but we concluded that the tested models were overly complex in relation to the adopted fitting methods and the relatively small data set; the simplest model with just a single drift diffusion unit performed essentially as well as the more complex alternatives. This type of simplified model was then tested by Zgonnikov et al. (2020) on a related traffic scenario—drivers deciding to turn across oncoming traffic—and was found capable of reproducing response time distributions of “turn” versus “wait” decisions, thus confirming that existing discrete gap acceptance/rejection models of turning drivers can be generalised, using drift diffusion models with time-varying input, to model also the timing of these decisions. This study was however limited to only constant speed, non-yielding oncoming traffic.

Here, we pursue two main objectives: First, we wish to determine whether the type of variable-drift diffusion model successfully applied to gap acceptance decisions of turning drivers by Zgonnikov et al. (2020) can also account for pedestrian road-crossing decisions. Second, we attempt to go beyond pure gap acceptance, generalizing to situations where vehicles may be yielding to the decision-maker, with or without additional implicit or explicit communicative cues. To achieve these objectives, we start from our original, inconclusive pedestrian modeling efforts in Giles et al. (2019) and extend these by incorporating an additional, larger data set with higher face validity, and by leveraging more powerful model-fitting methods. In a first section below, we introduce the basic model and fitting methods. Then, we describe the first modeling study, reusing the dataset in Giles et al. (2019) to investigate and model the impact of vehicle kinematics (both non-yielding and yielding with different deceleration profiles) on pedestrian crossing decisions. Thereafter, we describe the second study, and show how our model can be extended to and validated on this second data set, which also includes explicit communication using eHMI. We then provide a general discussion and conclusions, addressing both the topic of applied use of our models in automated vehicle interaction design, and the implications of our

findings for decision-making modeling in the more basic sciences.

Computational Modeling

Model Definition

The model is based on Gaussian drift diffusion where the drift can vary over time and the drift is considered to be the momentary evidence in favor of crossing the road. A decision is made when the value of the diffusion process reaches a decision threshold. The evidence comes from sensory input, which in this work is formulated to reflect the visual information that the pedestrian receives of the approaching vehicle. The architecture of the model is illustrated in Fig. 1.

The variable drift diffusion process can be written in mathematical form as a stochastic differential equation:

$$\frac{dA}{dt} = -\alpha A(t) + s(t) + \epsilon(t) \quad (1)$$

where $A(t)$ is the accumulated evidence, α is a damping parameter, $s(t)$ is the time-varying sensory input, and $\epsilon(t)$ is the white noise process with power σ .

A decision is made at time t' when the evidence threshold A' is passed:

$$t' = \min(t) \text{ s.t. } A(t) > A' \quad (2)$$

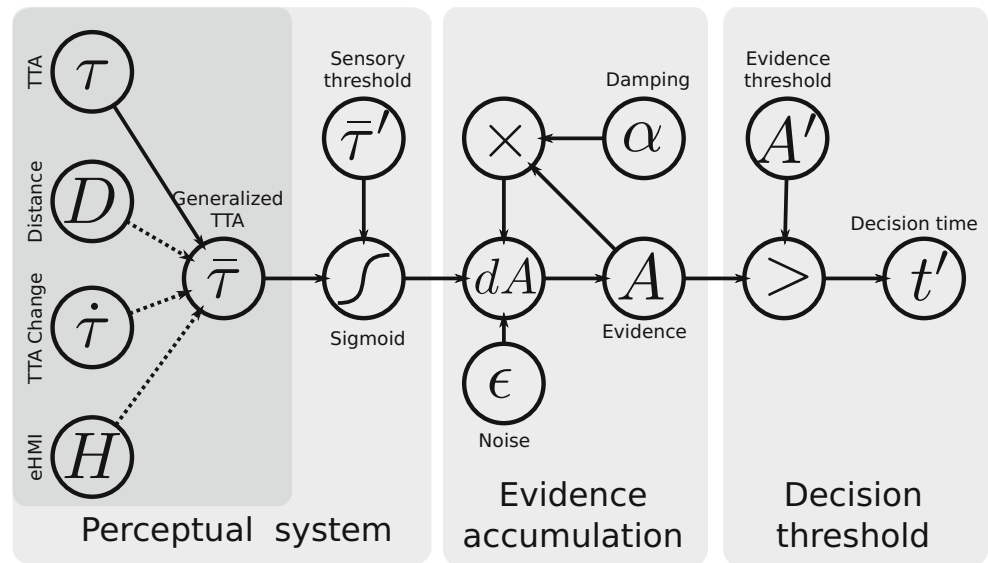
We are interested in the distribution of t' given a trajectory specified by some $s(t)$ (in our case, different kinematical trajectories of vehicle approach). Simple closed form solutions for this distribution are not known (Downes & Borovkov, 2008). To evaluate the decision time distributions, we adopt a numerical scheme presented in “Numerical Approximation”.

Generalized Time to Arrival

For the current application of pedestrian crossing, the time-varying sensory input $s(t)$ is based on a *generalized time to arrival* signal $\bar{\tau}(t)$ that comprises the apparent time to arrival (TTA) of the vehicle at the pedestrian’s location which we denote¹ $\tau(t) = D(t)/v(t)$, TTA’s first time derivative $\dot{\tau}(t)$, the distance between pedestrian and vehicle $D(t)$, the vehicle speed $v(t)$, an indicator variable for an external human-machine interface being active $H(t)$, and a passing threshold which saturates the sensory input when

¹The symbol τ is used here to refer to time to arrival, without implication whether or not the time to arrival is estimated by the participants from optical expansion information, as suggested by the tau theory of ecological psychology (Lee, 1976; Lobo et al., 2018).

Fig. 1 Schematic illustration of the variable-drift diffusion model of pedestrian crossing decisions



$\tau(t) < \tau_p$ (i.e., τ_p indicates the TTA at which the pedestrian judges the vehicles to have passed them):

$$\begin{aligned} \bar{\tau}(t) &= \tau(t) && \text{(TTA effect)} \\ &+ \beta_D(D(t)/v' - \tau(t)) && \text{(Distance effect)} \\ &+ \beta_{\dot{\tau}}(\dot{\tau}(t) + 1) && \text{(Acceleration effect)} \\ &+ \beta_H H(t) && \text{(eHMI effect)} \\ &+ \infty \text{ if } \tau(t) < \tau_p, 0 \text{ otherwise} && \text{(Vehicle passed)} \end{aligned} \quad (3)$$

The different β_* above are coefficients for the different terms and v' (fixed to 50 km/h in this study) is the prior speed, i.e., the typical speed the pedestrian assumes the vehicle is driving at before seeing it. The $\dot{\tau}$ term is formulated so that $\beta_{\dot{\tau}}$ does not affect constant speed scenarios where $\dot{\tau}(t) = -1$.

To get the sensory input $s(t)$, the generalized TTA is passed through a sigmoidal transformation:

$$s(t) = \arctan(m(\bar{\tau}(t) - \bar{\tau}')) \quad (4)$$

where m is a scaling factor and $\bar{\tau}'$ is a threshold, loosely analogous to the “critical gap” threshold considered in many existing pedestrian gap acceptance models, at which crossing and not crossing is equally likely (Schroeder, 2008). The greater the margin of the generalized TTA above (or below) this threshold, the faster the evidence A in favor of crossing increases (or decreases) in the model. The arctan transformation prevents arbitrarily high rates of evidence accumulation, which would otherwise occur when a vehicle has passed the pedestrian, and no further vehicle is approaching behind it, such that τ is infinity. The choice of arctan is somewhat arbitrary and other sigmoid functions would likely give similar results.

Overall, it may be noted that the model evidence A is determined by (i) a weighted sum of a number of different inputs (Eq. 3), (ii) a sigmoidal activation function (Eq. 4), and (iii) noise and exponential damping (Eq. 1), such that it effectively corresponds to a single neural node representing a population of neurons, as assumed in many neurally inspired models of decision-making and behavior (Schöner, 2007; Usher & McClelland, 2001).

It may also be noted that if all generalized TTA coefficients $\beta_* = 0$, the model essentially reduces to a threshold $\bar{\tau}'$ on the apparent TTA τ , such that positive evidence in favor of crossing the road will accumulate whenever the apparent TTA is above this threshold. The addition of sensory input terms in Eq. 3 to “generalize” the TTA is a formalization of the phenomenon that observed crossing decisions cannot be explained solely by TTA, but are also modulated by other factors (Petzoldt, 2014). In this paper we will introduce the different terms in Eq. 3 incrementally to study their respective contributions to the model’s behavior and ability to fit human crossing decisions.

Numerical Approximation

For computational purposes the Eq. 1 is approximated using a discrete time stochastic process:

$$\begin{aligned} \Delta A[i] &= A[i] - A[i - 1] \\ &= (-\alpha A[i - 1] + s[i])\Delta t + \epsilon[i] \end{aligned} \quad (5)$$

where Δt is the time step duration and $\epsilon[i] \sim N(0, \Delta t \sigma^2)$. The time step duration for all analyses in this study was 1/30 s.

To approximate the distribution of $A[i]$, we adopted an approximation scheme which closely resembles the forward Euler solver of the GDDM framework for solving

generalized drift diffusion models (Shinn et al., 2020) such as the presented VDDM. In brief, this method divides the time-evidence plane into a grid, and calculates a complete numerical probability distribution for the diffusing evidence at each time step, effectively until all of the evidence probability mass has been absorbed at the decision threshold. In contrast with Monte Carlo methods for decision time estimation, our method yields, for a given model parameterization and sensory input $s[i]$, a fully deterministic probability distribution of decision times, something which substantially simplifies the model fitting. For further details, please refer to Shinn et al. (2020), our provided source code at <https://github.com/jampekka/vddfit/>, and literature about stochastic differential equation approximations (Särkkä & Solin, 2019).

Model Fitting

As mentioned, the approximation described in “Numerical Approximation” yields a distribution for the crossing time t' given the input signal $s[i]$ and the parameters. We estimated the model parameters by numerically maximizing the likelihood of the observed decision times (equivalent to maximum-a-posteriori with flat priors). To remedy problems with local optima, we use the basin-hopping method with 10 iterations with Powell’s conjugate direction method as the local optimizer, as implemented in Scipy (Virtanen et al., 2020).

For Study 1, we assessed the effect of the distance coefficient β_D and acceleration coefficient β_t using nested model selection, and all other parameters were free to vary. Due to this, and likely redundancy in the parameterization (see “Results”), inferences on parameterization of Study 1 model should be restricted to these.

For Study 2, all the parameters except the passing threshold τ_p and eHMI coefficient β_H (see Eqs. 4 and 3) were fixed at values obtained in Study 1 (for rationale, see “Extending the Model to Two-Vehicle Scenarios”). Data from all participants were pooled and a single set of parameters was fitted for each study. This in contrast with most existing literature on evidence accumulation models, where models are typically fitted per participant. Typical evidence accumulation laboratory paradigms permit many repetitions per participant, and thus allow for assessing between-participants variation. Such repetitions were not feasible here, because of time constraints (each trial takes up to about 30 s to complete). It could also be argued that large numbers of repetitions could cause substantial behavior adaptation effects (Engström & Ljung Aust, 2011), reducing the external validity of these studies. For these reasons, a simplifying assumption made here is that all participants can be described with a single model parameterization. A more principled approach to account for within and between

individual variation would be a hierarchical formulation, but this was not pursued due to the additional technical and mathematical complications entailed.

The likelihood distributions (equivalent to posterior distributions with flat priors) of the parameter values were estimated using the Adaptive Metropolis method (Haario et al. 2001) using 2000 samples, with 1000 first samples discarded. The algorithm was initialized from the estimated maximum likelihood values produced by the method described above. Most of the analyses in this paper are made based on the maximum likelihood point estimates, whereas we use the likelihood distributions mainly to assess potential parameter redundancies (see “Results”).

Study 1

The aim of Study 1 was to design a minimally complex pedestrian road-crossing experiment that would allow fitting of models predicting crossing onset times, across a range of kinematical trajectories for the approaching vehicle. All procedures were approved by the relevant University of Leeds Research Ethics committee, reference AREA 18-004. Below we describe first the experiment,² and thereafter we present our methods and results for fitting our model to first constant-speed scenarios and then to scenarios with a yielding vehicle.

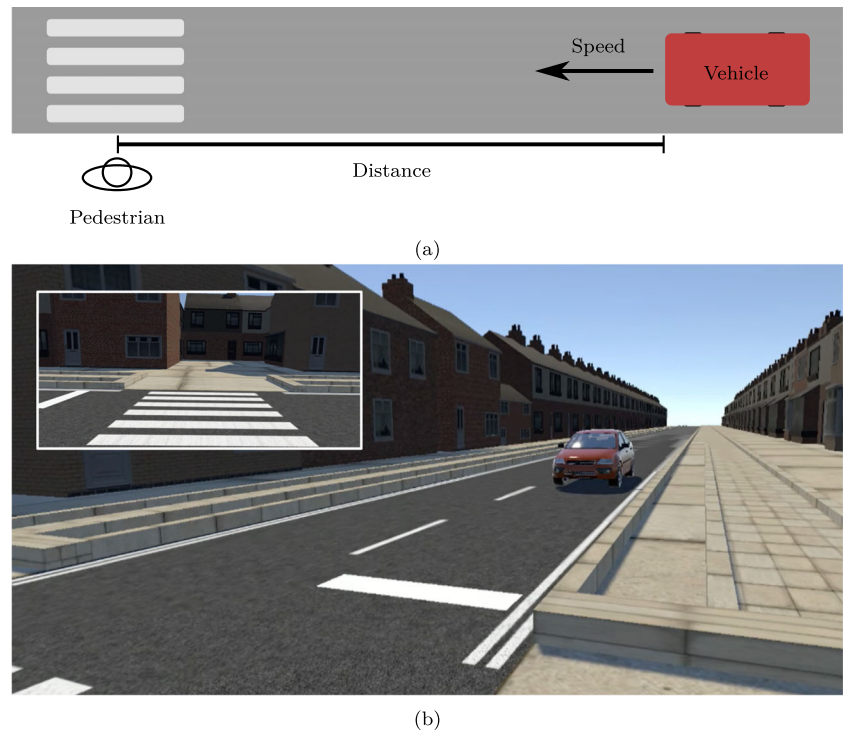
Experiment

Twenty participants (age 24–60, average 27.9 years; 11 male, 9 female) were recruited from a University participant pool, and provided informed consent before taking part in the experiment. Standing upright and wearing an HTC Vive Virtual Reality (VR) headset, the participants experienced a VR scene, created in Unity 2018, consisting of a straight two-lane road of total width 5.85 m, with a zebra crossing at the participant’s location; see Fig. 2 for an illustration.

In each trial, the participant first looked straight across the zebra crossing (see the inset in Fig. 2b), and turned their head to the right to look for oncoming (left hand driving) traffic when they felt ready to do so. Unbeknownst to the participants, their turning of the head triggered the start of a preprogrammed scenario, whereby a car was positioned in the virtual world at a certain initial distance and speed (see the large image in Fig. 2b). The participants’ task was to cross the road as soon as they felt safe to do so, either before or after the car had passed them, by pressing a button on the HTC Vive’s controller. Upon this button press, the

²Initial work on the dataset collected in this experiment has been previously reported in a conference paper (Giles et al., 2019), but we provide a full description of the experiment here nonetheless.

Fig. 2 Study 1. (a) Schematic bird's eye view of the pedestrian crossing scenario (not to scale). (b) Example views of the virtual scene as shown in the head-mounted display, at the start of each trial (inset) and once the participant turned their head to look for oncoming traffic



location of their viewpoint in the VR world moved in a straight line across the road at a speed of 1.31 m/s (a typical average walking speed (Chandra & Bharti, 2013)); during this time their head rotation still controlled the rotation of the VR viewpoint. An alternative approach would have been to let the pedestrians physically walk to cross the virtual road, but in Study 1 we opted for this button-pressing approach instead, to make it easier to identify the timing of the crossing decision, and to minimise the impact of different preferred walking speeds between participants as a possible source of variability in their crossing decisions. Once the participant reached the other side of the road, the trial concluded, an all-white VR scene was shown with an instruction on where to look (in the direction straight across the zebra crossing) before pressing the controller button again to start a new trial.

The participants were allowed to practice this task until they were comfortable with it, and then followed a sequence of 16 trials per participant. In each of these trials, the vehicle approach behavior was different, following one of three scenario types: In six *constant-speed* scenarios, the vehicle appeared at a distance D_0 (all distances measured longitudinally along the road from the participant's location to the front of the car) and maintained a constant speed v_0 while approaching and passing the zebra crossing. In eight *yielding* scenarios, the vehicle appeared at initial distance and speed D_0 and v_0 , and immediately decelerated at a constant rate to stop at a distance D_{stop} from the participant. There were also two scenarios where the vehicle only decelerated down to a speed of 5 km/h before passing the

zebra crossing, but these were excluded from analysis here due to a scenario programming error corrupting some of the collected data. Full details about all of the included scenarios are provided in Table 1, where also the initial time to arrival $\tau_0 = D_0/v_0$ is listed.

We chose these scenarios and scenario parameters to allow us to model the impact on crossing decisions of (i) TTA and distance, previously observed to both separately influence gap acceptance decisions in constant-speed scenarios (Petzoldt, 2014), and (ii) yielding decelerations of different magnitudes.

Table 1 Vehicle approach scenarios in Study 1

Scenario type	v_0 (m/s)	D_0 (m)	τ_0 (s)	d_{stop} (m)
Constant speed	6.94	15.90	2.29	N/A
	13.89	31.81	2.29	N/A
	6.94	31.81	4.58	N/A
	13.89	63.61	4.58	N/A
	6.94	47.71	6.87	N/A
	13.89	95.42	6.87	N/A
Yielding	6.94	15.90	2.29	4
	13.89	31.81	2.29	4
	13.89	31.81	2.29	8
	6.94	31.81	4.58	4
	13.89	63.61	4.58	4
	13.89	63.61	4.58	8
	6.94	47.71	6.87	4
	13.89	95.42	6.87	4

Table 2 Model fits for Study 1

σ	α	m	τ'	A'	τ_p	β_D	$\beta_{\dot{\tau}}$	loglik
0.64	1.84	0.59	1.64	0.84	−0.14	0.75	0.59	−400.9
0.71	1.90	0.43	2.57	0.84	−0.17	0	0.38	−416.3
0.28	4.34	0.74	0.23	0.45	−0.18	0.32	0	−417.6
0.30	4.22	0.62	0.42	0.47	−0.17	0	0	−420.7
1	0	1	2	1	0	0	0	−595.8

Fixed parameters (used also as initial values) in bold

Results

The parameters estimated from Study 1 data are listed in Table 2 and their marginal likelihood distributions are illustrated in Fig. 3. The nested model selection for distance coefficient β_D and acceleration coefficient $\beta_{\dot{\tau}}$ shows considerable improvement in the likelihood, especially

when both parameters are free to vary (log likelihood improvement 19.8, AIC improvement 35.6), indicating that under the model vehicle distance and acceleration cues had a significant impact on pedestrian crossing, beyond the impact of the pure TTA cue.

The marginal likelihood ranges (equivalent to posterior credible intervals with flat priors) of the parameters

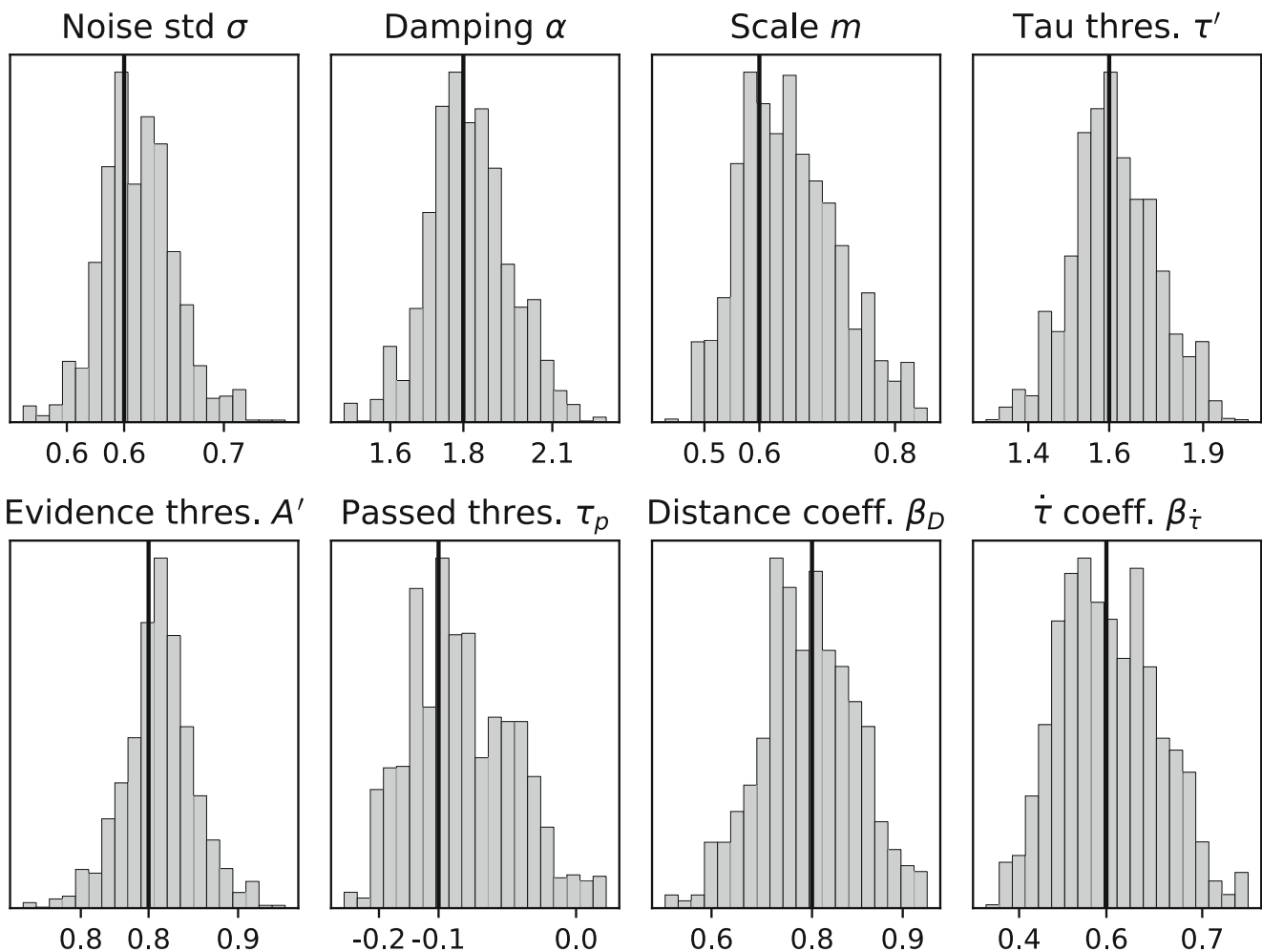


Fig. 3 Optimized parameter values (black vertical lines) and their estimated marginal likelihoods (histograms). Left and rightmost ticks mark the 95% highest density range. The distributions should be interpreted with caution, as strong pairwise correlations can be seen in

some parameters, indicating likely redundancy in the parameterization. Figure 14 in the Appendix shows pairwise covariation in the distribution

are also relatively tight, and notably both generalized TTA coefficients for distance β_D and acceleration β_i significantly differ from zero. However, the parameter posterior distributions exhibit high (Pearson ρ up to 0.92) correlation among parameters, likely redundancy in parameterization, and thus limiting the interpretation of the marginal distributions. Pairwise covariation of the posterior estimates can be assessed in the Appendix Fig. 14.

Below, we provide further insight into the model fits, separately for the constant-speed and yielding scenarios.

Constant-Speed Scenarios

To begin with, we consider the constant-speed scenarios, only including the time-to-arrival τ and distance D terms of the generalized TTA expression in Eq 3. As seen in Figs. 4 and 5, the model clearly captures the bimodal pattern of early vs late crossers (crossing before and after the vehicle, respectively) and its strong dependency on initial time to arrival, as well as the general shape of the crossing onset time distributions.

Furthermore, Fig. 5 shows that given vehicles approaching at identical TTAs, those with higher speeds, or conversely longer distances, exhibit earlier crossing decisions. This behavior is not possible to explain with models relying solely on TTA, which is a common assumption in ecological psychology for object avoidance (Merchant & Georgopoulos, 2006), but which the current model is able to capture due to the distance term. Without a nonzero distance coefficient β_D the model would produce identical distributions for the two scenarios with different speeds but same initial TTA (the differently colored lines in Fig. 5), while the observed distributions are clearly not identical in the higher initial TTA cases (see also Fig. 10). As mentioned, this is reflected also in the 95 % marginal likelihood range of the β_D coefficient, from 0.6 to 0.9, i.e., excluding zero (Fig. 3).

Yielding Scenarios

As has been previously observed, pedestrian crossing behavior in situations where a vehicle yields also tends to be bimodal, with some crossings seen early on, and a second mode of crossings once the vehicle is close to standstill, with few crossings in between (Dey et al., 2020; Schneemann & Gohl, 2016). As can be seen in Figs. 6 and 7 this general pattern of behavior was present also in our data, and was captured rather well by the model.

Again in line with previous reports (Dietrich et al., 2020), increased deceleration magnitudes (higher D_{stop} distances) led to a shift toward earlier crossing onset times. This can be seen in Fig. 7 as the difference of the green and black

lines in both measured and model-predicted data. However, this observation in itself does not permit the conclusion that the participants made use of deceleration cues to make their crossing decision, since the apparent TTA is also affected by the deceleration. Qualitatively, a model relying on TTA only would also predict earlier crossing onsets for increased deceleration. However, since our model considers both cues separately we can tease apart their model-estimated contributions, and our results do indicate that participants took the crossing decision more readily in the presence of deceleration cues than what would be predicted from the deceleration-induced change in TTA alone. Visually this effect can be seen in Fig. 6 as the difference between the blue and orange curves (the latter showing the behavior of the model with β_i set to zero) and quantitatively in the model selection in Table 2 and in the parameter estimates in Fig. 3, with a 95% marginal likelihood interval from 0.4 to 0.7 for coefficient β_i .

The overall predictive power of the mean crossing times across different scenarios can be seen in Fig. 8. Note that mean crossing time is by no means a perfect summary of the observed and model-predicted crossing times, given the multimodality and strong skew of the distributions, but not least from an applied perspective we still find this aggregate view of model performance useful.

Study 2

The aim of Study 2 was twofold. First, to test the models fitted as part of Study 1 on a separate, larger dataset, collected in a setting with higher face validity. Second, to attempt extension of the model to scenarios with eHMI indications of yielding intention. Study 2 is somewhat akin to an out-of-sample predictive test for the model developed and fitted for Study 1. However, due to clear differences in the experimental environment and vehicle scenarios (see next section), we allow for one free parameter (namely the generalized TTA coefficient τ_p to account for these changes (see “[Extending the Model to Two-Vehicle Scenarios](#)” below). While this means that we don’t have a pure out-of-sample replication, it should be noted that τ_p only controls a very limited aspect of the model’s behavior.

Experiment

This experiment was originally developed to study the combined impact of vehicle approach kinematics and different types of eHMI on human road-crossing behavior; full methodological details and non-modeling analyses of the same dataset are reported by Lee et al. (2020). All procedures were approved by the relevant University of Leeds Research Ethics committee, reference LTTRAN-

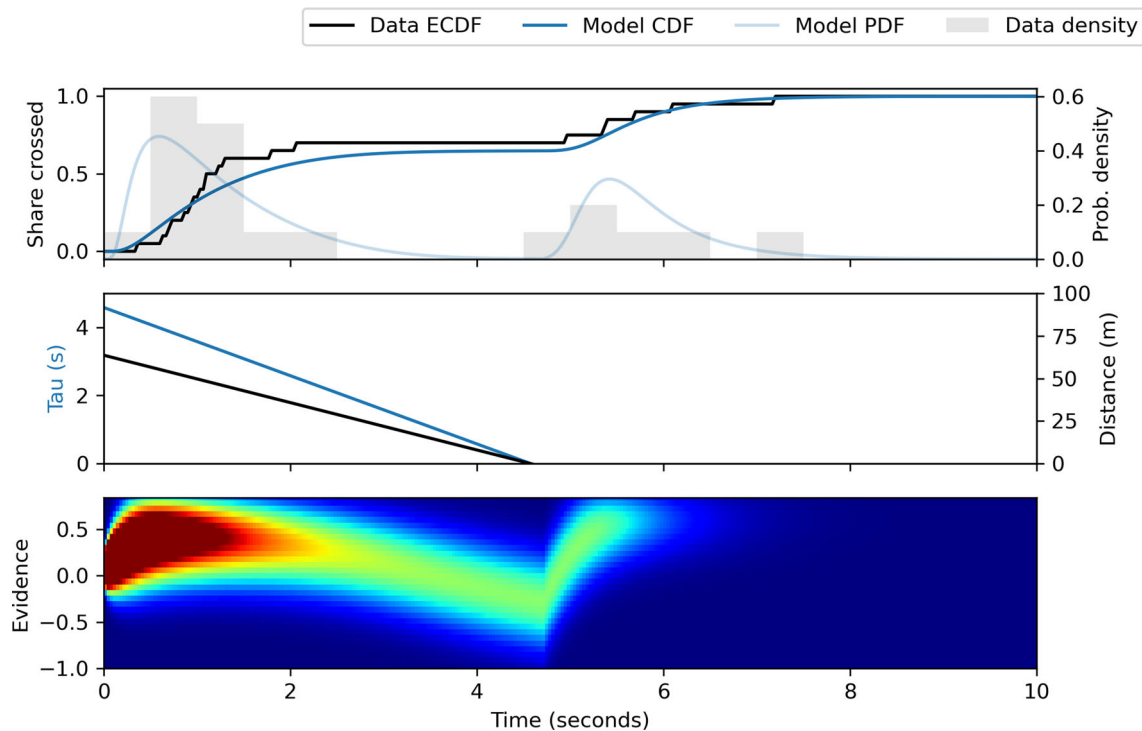


Fig. 4 Model response to an example constant-speed scenario. Top panel shows cumulative density functions and probability densities for observed (black) and model predicted (blue) crossing decision times. Mid-panel shows the TTA (blue) and distance (black) of the

approaching vehicle over time. Bottom panel visualizes the evidence distribution (approximately the distribution of $A(t)$ of Fig. 1 and Eq. 1) and its evolution over time

107. Forty participants (age 19–36, average 28.5 years; 23 male, 17 female) were recruited via a participant pool, message board notices, and social media posts, and provided informed consent before taking part. The experiment took place in the University of Leeds HIKER (Highly Immersive Kinematic Experimental Research) lab, a CAVE-based pedestrian simulator with projection on three walls and a $9\text{ m} \times 4\text{ m}$ floor space on which participants could walk

freely, while their head position and orientation was being tracked, for perspective correction and movement recording. The experimental paradigm was adapted from (Lobjois & Cavallo, 2007): As shown in Fig. 9a, the participant started each trial standing at a specified location next to a 3.5 m wide single-lane street, along which two vehicles approached. The instruction to the participants was to “cross (or decide not to cross) between the two approaching cars

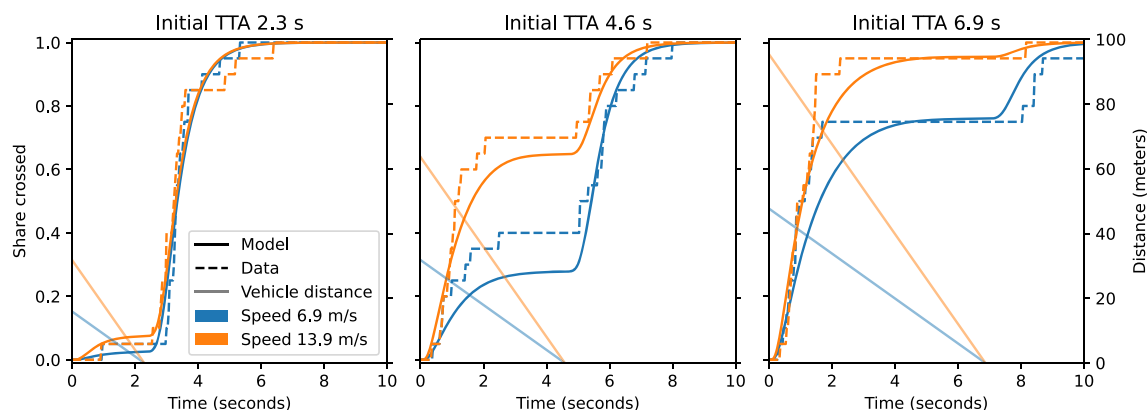
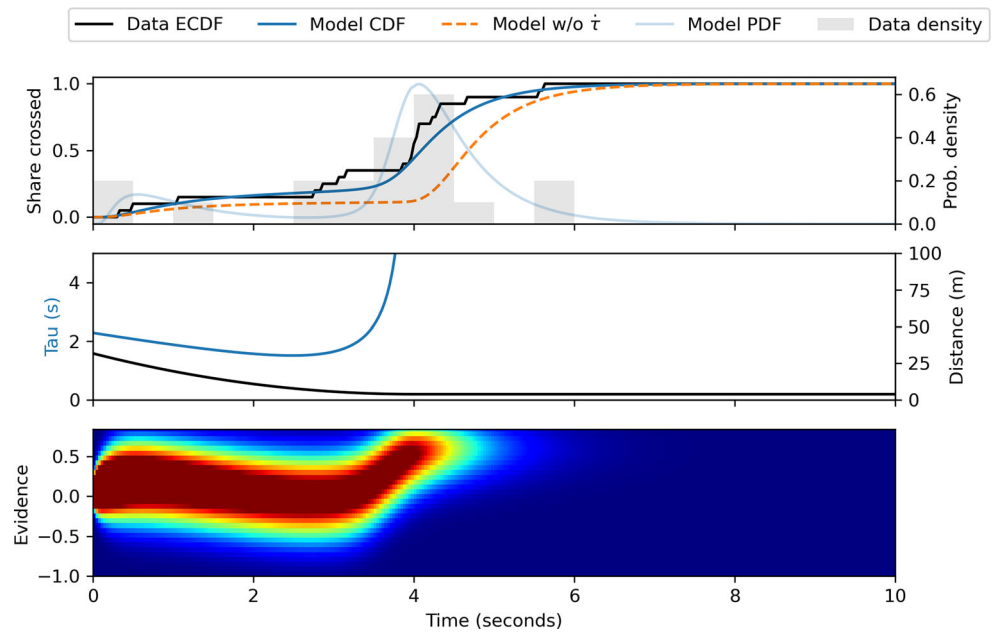


Fig. 5 Cumulative probability functions for decision times in the constant speed scenarios. The three panels show scenarios with identical TTA and different speeds within these scenarios are displayed as lines of different colors. As we assume, for simplicity, that all individuals

have the same parameterization, these can be interpreted as the model predicted crossing times for any pedestrian crossing in an identical kinematic setting

Fig. 6 Model response to an example yielding scenario. To highlight the independent effect of deceleration in the model, the dashed orange line shows predicted cumulative distribution if the TTA change coefficient $\beta_i = 0$ (see “Generalized Time to Arrival”). For further figure explanation, see Fig. 4



[...] when you feel comfortable to do so, such as you would in real traffic.”

The participants were allowed to practice the task until they were familiar with it, and then followed three experimental blocks with a short break between each. Each block included 48 trials, which were identical between blocks, but with the order randomised per participant and block. As in Study 1, the trials followed two main scenario types: constant-speed scenarios and yielding scenarios, but the kinematics were different from Study 1. For both scenario types, there were 12 kinematic variations, across all combinations of three initial vehicle speeds $v_0 \in$

{11.11, 13.33, 15.56} m/s and four initial time gaps between the vehicles $\tau_0 \in \{2, 3, 4, 5\}$ s. In constant-speed scenarios, both vehicles maintained their initial speed throughout, whereas in the yielding scenarios, the second vehicle started decelerating when 38.5 m from the participant, with a constant deceleration so as to stop 2.5 m from the participant. Each kinematic variation was repeated twice in each block. Table 3 provides a summary overview of the trials in each block. A between-participant factor was also included: For twenty of the participants (Group 2 in Table 3), for half of the yielding trials an eHMI on the second vehicle was activated at deceleration onset, as

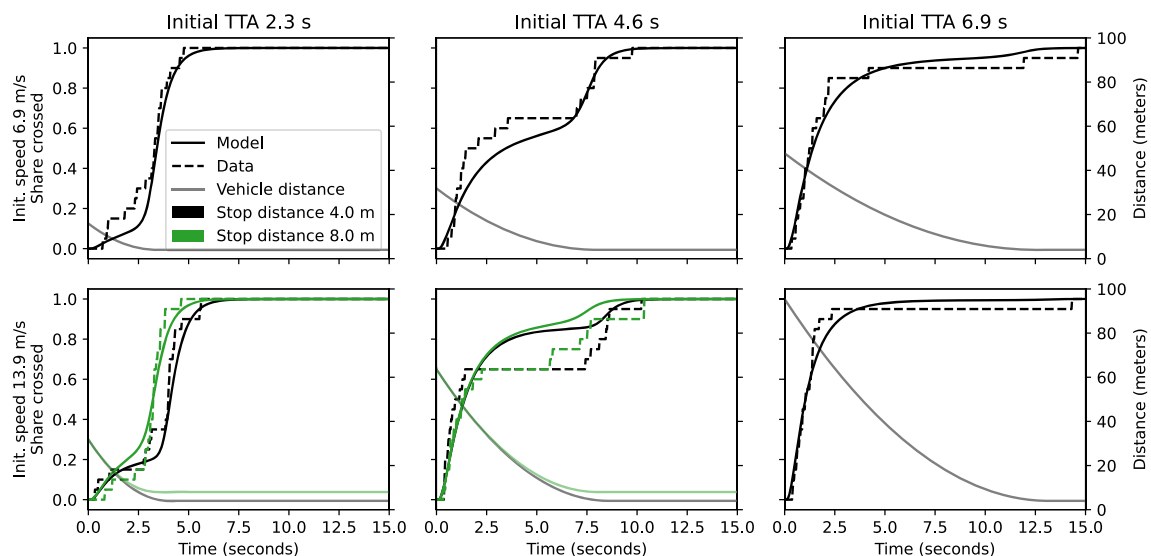
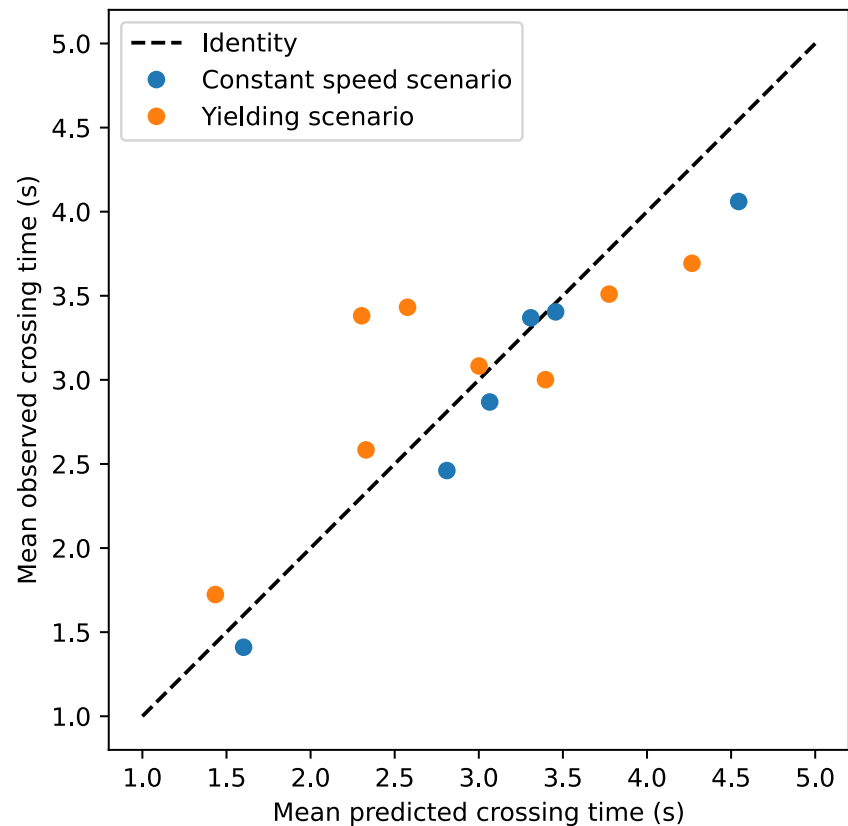


Fig. 7 Cumulative density functions for decision times in the deceleration scenarios. The rows have identical initial speed and the columns have identical initial TTA. Different distances of the vehicle to the pedestrian at full stop (D_{stop}) are represented in black and green (4 m and 8 m respectively)

Fig. 8 Predicted vs observed average crossing times across the different scenarios in Study 1. Mean absolute deviation (MAD) of the predictions across scenarios is 0.37 s. MAD for constant speed scenarios is 0.22 s and for yielding scenarios it is 0.47 s



an explicit communication of yielding intent. The original study included three experimental groups, and two different eHMI; here we are only using the groups who experienced either no eHMI (Group 1) or a *flashing headlights* eHMI (Group 2). The latter took the form of three quick flashes of the front headlights, chosen for being a commonly used signal for yielding intentions in the UK. As described by Lee et al. (2020), a third group of participants experienced a *slow pulsing light band* eHMI, but since the effect of this eHMI on participant behavior was considerably smaller, we did not include these participants in our analyses here. The eHMIs were not mentioned at all in the information provided to the participants before the experiment, since a goal of the original study was to investigate how quickly participants would deduce the meaning of the eHMI signals. Lee et al. (2020) found that the impact of the flashing headlights eHMI on crossing decisions was already present from the first experimental block, hence we are not considering learning effects here.

Extending the Model to Two-Vehicle Scenarios

To extend the model to accommodate two vehicles, the crossing decision for both is modeled using the same single evidence accumulation, and in line with the experimental instructions the simulated pedestrian is only allowed to

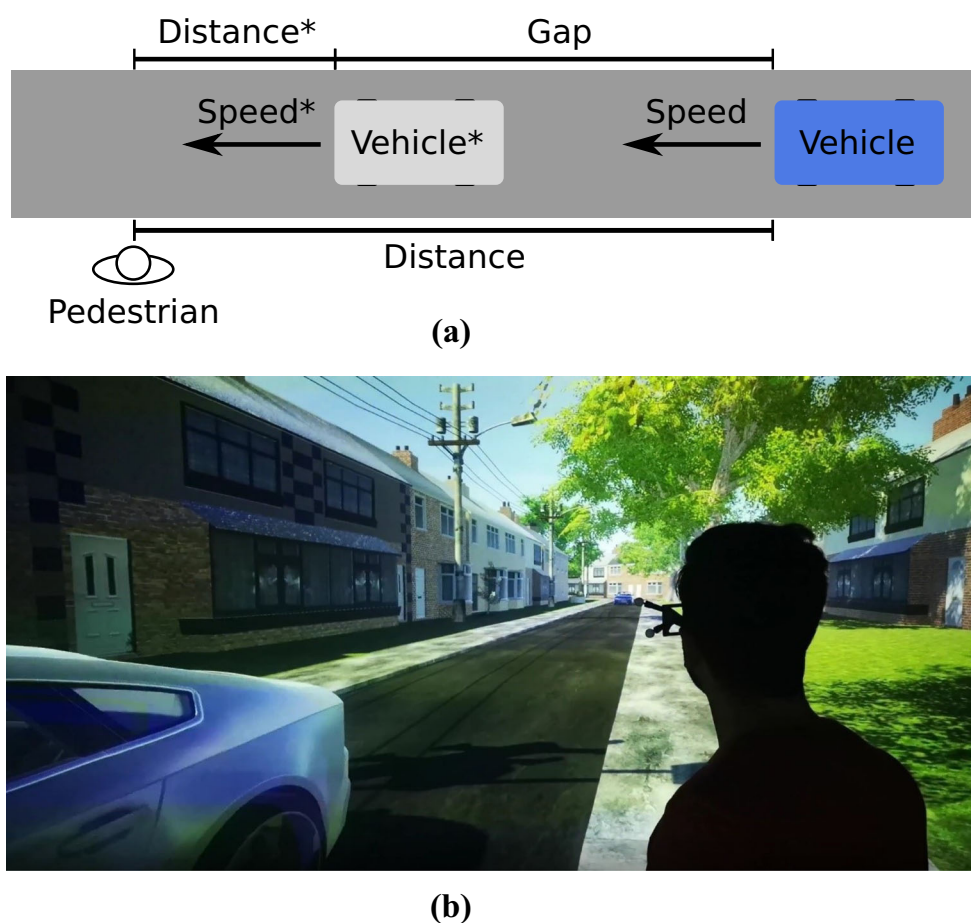
cross between the two vehicles; otherwise, the pedestrian does not cross at all. In the model, this is enforced by altering the parameterization so that for the first vehicle τ' is fixed at ∞ (such that the model will never cross before the first vehicle), and for the second vehicle τ_p is set to $-\infty$ (such that the model will never pass after the second vehicle). All the other parameters are shared for both vehicles.

To account for the differences in the experimental settings, the passed threshold τ_p was optimized separately (in conjunction with a new eHMI parameter τ_H discussed in the next section), arriving at value 0.33 s, whereas all other parameters were fixed to values fitted in Study 1 (see Fig. 3).

The nested model selection, listed in Table 4, shows considerable increase in model likelihood when both the pass threshold τ_p and eHMI coefficient β_H are free to vary (log likelihood improvement 3668.9, AIC improvement 7333.8). For τ_p this indicates that the model parameterization needs at least this adjustment for the new scenario of Study 2. Improvement due to incorporating the β_H (log likelihood improvement 175.2 and AIC improvement 348.4 after optimized τ_p) implies that eHMI has an additional effect on crossing times under this model.

The motivation for adapting the τ_p parameter was the observation that in Study 2, participants sometimes began crossing the road slightly before the first vehicle had passed

Fig. 9 Study 2. (a) Schematic bird's eye view of the pedestrian crossing scenario. (b) An example view of the pedestrian crossing scenario in the HIKER CAVE environment. In this photo, the first car is just passing the participant, and the second car is visible in the distance



them (negative crossing times in Fig. 10), presumably due to a perceived time pressure from the approaching second vehicle. This type of time pressure was not present in Study 1 when passing after the (sole) vehicle.

Figure 10 shows that the model generalizes quite well to the Study 2 data. The overall share of participants crossing between the two vehicles is on average predicted well (rightmost ends of lines in the figure). The effect of the initial distance is at least qualitatively correctly predicted, but with clear overestimation of the effect in the case of initial TTA of 3.0 s.

A systematic lack of model fit can be seen in the “elbow” of the CDF, where the model seems to have more variation

in the latencies of the crossing decision (seen in the CDF as more gradual increases over time of the share crossed). This tendency can to some extent also be seen in the Study 1 results (especially the TTA 6.9 s panel in Fig. 10); in the Discussion we consider possible reasons for this lack of fit.

Yielding and eHMI

For half of the participants in Study 2, the vehicle activated eHMI (flashed headlights) to the participant when it started decelerating. As described in “Generalized Time to Arrival”, the eHMI was incorporated in the model by adding it as an indicator variable $H(t)$ (0 when no eHMI, 1 when eHMI

Table 3 Trials per block in Study 2

Scenario type	eHMI present (Group 1)	eHMI present (Group 2)	Number of trials
Constant speed	N/A	N/A	12 kinematic variations $\times 2$
Yielding	No	Yes	12 kinematic variations
	No	No	12 kinematic variations

See the text for a description of the kinematic scenario variations

Table 4 Model fits for Study 2

Pass threshold τ_p	eHMI coefficient β_t	loglik
0.33	0.94	−7151.2
0.34	0	−7326.4
− 0.14	1.05	−10607.1
− 0.14	0	−10820.1

Fixed parameters (used also as initial values) in bold. Other parameters from Study 1 fit

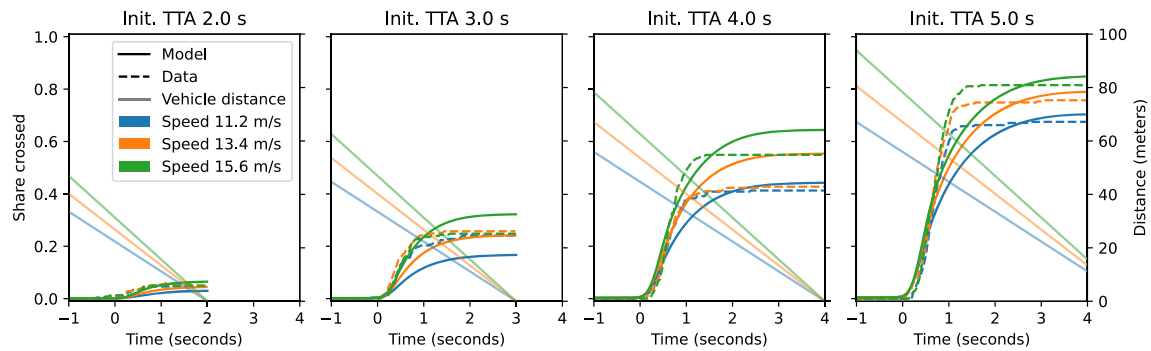


Fig. 10 Cumulative probability functions for decision times in the constant speed scenarios in Study 2. The four panels show scenarios with identical TTA and different speeds within these scenarios are displayed as lines of different colors. The time axis is zero when the first car passes the pedestrian position

was active) to the generalized time to arrival (see Fig. 1 and Eq. 3). As can be seen in Fig. 11, the model was able to reproduce relatively well the crossing onset distributions both in eHMI and non-eHMI scenarios. The maximum likelihood value for the coefficient β_H was found to be

0.94, suggesting that flashing headlights is (marginally) equivalent to an increase of 0.94 s of TTA for pedestrian crossing decisions.

The significant contribution of eHMI to the crossing time distribution can be seen in Fig. 11 in both observed and

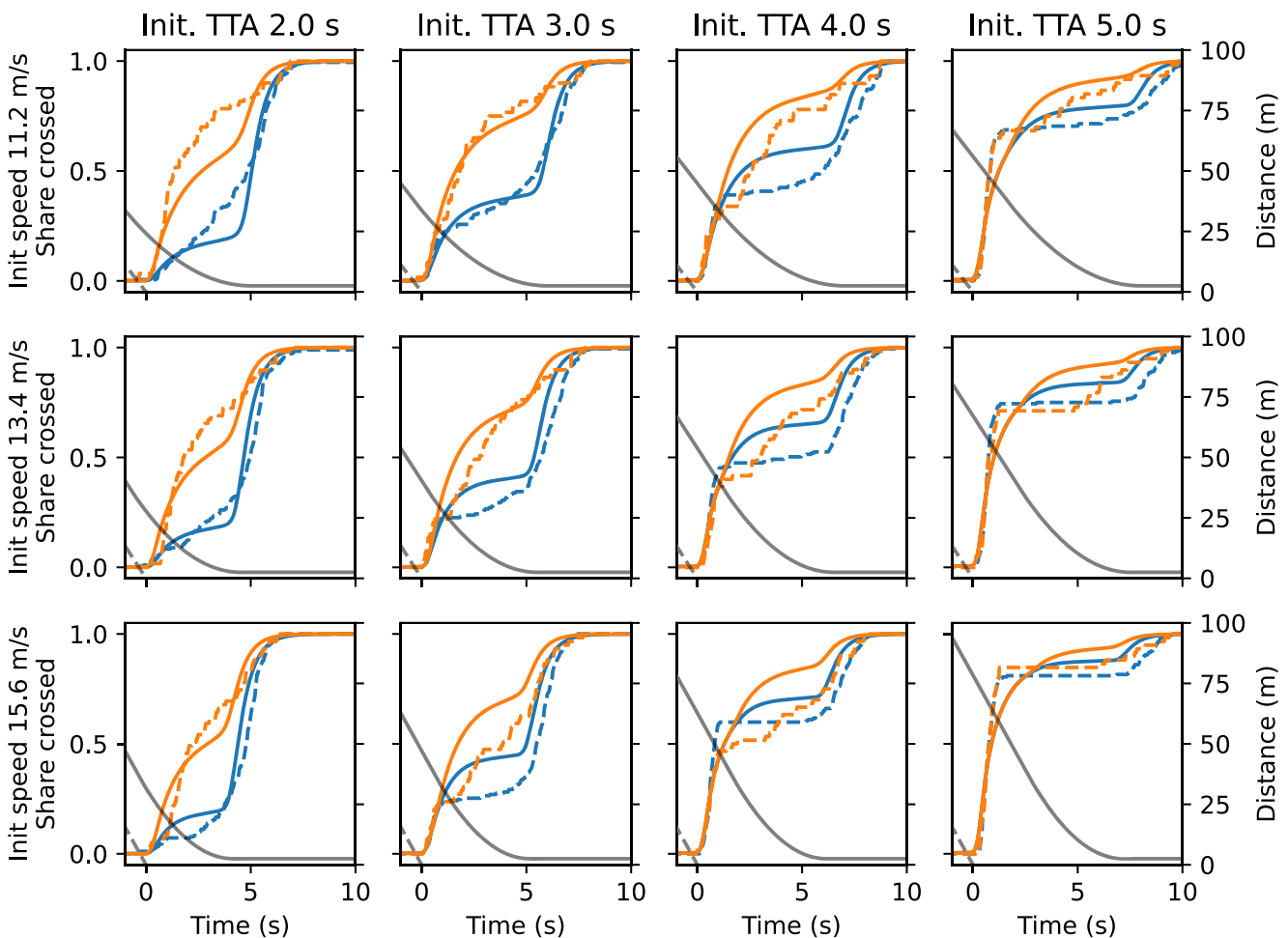
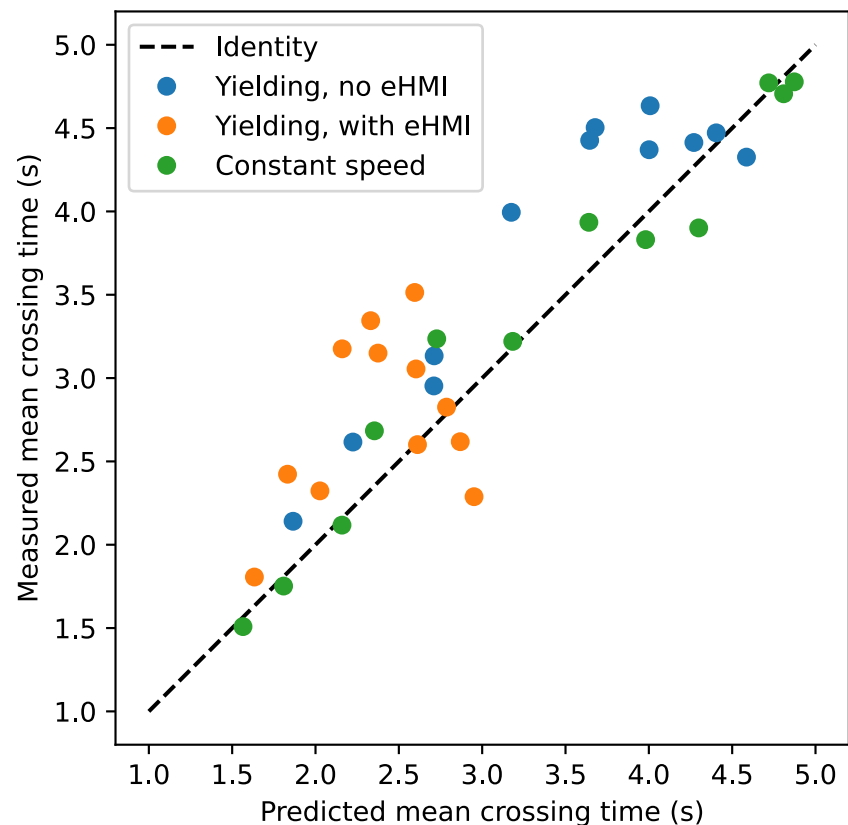


Fig. 11 Cumulative probability functions for decision times in the deceleration scenarios of Study 2. The rows have identical initial speed and the columns have identical initial TTA. Blue lines are trials with no eHMI and orange lines are trials with eHMI

model-predicted crossing times, where lower crossing times are systematically more likely when eHMI is present. It should be noted that the magnitude of the impact of eHMI is dependent on other features of the vehicle's trajectory. For example, in very high initial TTA situations (rightmost row of Fig. 11) pedestrians tend to cross even before eHMI is enacted, whereas the eHMI's contribution is higher for low-to-mid TTA scenarios (leftmost rows). This interaction is qualitatively predicted by the model, although some quantitative differences can be seen; the “elbow” exhibits some lack of fit, as already seen for the constant speed scenarios and in Study 1, and the quantitative crossing times are also off for some scenarios.

Figure 12 provides a summary overview of the prediction performance for mean crossing times. The accuracy in predicting the mean crossing times across scenarios is similar to Study 1 (see Fig. 12). Most directly comparable are the yielding scenarios with no eHMI and yielding scenarios of Study 1, for which the mean absolute deviations are essentially identical (Study 2 0.44 s vs Study 1 0.47 s). The prediction error increases with eHMI scenarios slightly to 0.52 s. The error for constant speed trials is somewhat lower for Study 2 (0.18 s vs 0.22 s), but the constant speed means for Study 2 are somewhat arbitrary due to imputation (see caption of Fig. 12).

Fig. 12 Predicted vs observed average crossing times across the different scenarios in Study 2. For non-eHMI yielding trials the Mean absolute deviation (MAD) of the predictions is 0.44 s. For the with-eHMI trials the MAD is 0.52 s. In the experiment the pedestrian could not pass after the vehicle, so for constant speed scenarios all crossing times greater than the initial TTA are censored. To compute a mean for the censored scenarios, a crossing time of 5.0 s (the longest possible constant speed trial) was input. Note that this makes the MAD for constant speed trials somewhat arbitrary, but as computed it is 0.18 s, and the overall MAD across all scenarios is 0.38 s.



Discussion

Below we provide a discussion of what new insights our results bring regarding pedestrian road-crossing decisions and how to model them, as well as of the limitations of our work. Then, we discuss possible implications and future directions for computational modeling of cognition and behavior in general, as well for applied work on automated vehicles and traffic safety.

Computational Modeling of Road-crossing Decisions

The work by Zgonnikov et al. (2020) demonstrated that, for car drivers turning across oncoming non-yielding traffic, variable-drift diffusion models allow modeling of not only the frequency of gap acceptance as a function of vehicle TTA and distance, but also the distributions of timing of these decisions. Our results replicate this finding for a pedestrian road-crossing scenario. We parameterised our model using one dataset and tested its performance on another, finding that our model accounted rather well for the substantial variations of crossing onset distributions across a wide range of kinematical conditions.

We extend beyond the constant-speed scenarios considered by Zgonnikov et al. (2020) by also considering

scenarios with yielding deceleration, again across the two separate experiments. The presence of yielding makes the situation under study significantly more complex from a decision-making point of view, since now the oncoming car may be considered not only as an object moving through space, but as (controlled by) an agent who can have one of several different intentions with respect to oneself (Pezzulo et al., 2019). It has been previously established that pedestrian crossing in front of yielding vehicles tends to follow a bimodal pattern, with one early mode and one mode which occurs later, once the vehicle is approaching a full stop, but with few occurrences of crossing in between (Dey et al., 2020; Schneemann & Gohl, 2016). Our model was able to capture this pattern well, again across a wide range of kinematical scenarios over the two datasets, and our model also provides a mechanistic explanation for why this pattern arises: According to our model, the early crossing decisions are effectively equivalent to those which are observed in constant-speed scenarios, and occur because the pedestrian judges the apparent gap to be large enough to cross, regardless of whether the car yields or not. However, as the car approaches further, even though it is decelerating the apparent TTA will further decrease for a while (see Fig. 6, and see Lee (1976) for a mathematical analysis of $\dot{\tau}$), such that pedestrians who did not cross early are even less likely to cross in this period of time. The apparent TTA then starts to increase dramatically a short while before the car comes to a full stop, giving rise, in our model, to the late mode of crossing onsets.

There are interesting nuances, however, to the exact timing of this second mode. Past work has shown that larger deceleration magnitudes lead to faster crossing decisions (Dietrich et al., 2020), but it has not been clarified whether this is simply because the car comes to a full stop earlier when the deceleration is greater, or whether it is also due to some form of intent recognition on the part of the pedestrian, facilitated by more obvious deceleration cues. Indeed, our model-based analyses permit the conclusion that crossing decisions to decelerating vehicles are faster than one would expect from a pedestrian who solely responds to the deceleration-induced change in the apparent gap. By including the time derivative of TTA ($\dot{\tau}$) as a source of decision evidence, our model was able to capture the timing of the late mode of crossing onsets. In other words, we provide evidence that pedestrian road-crossing decisions do involve a process akin to intent recognition (or acceleration estimation, which to some extent is equivalent). That this would be the case seems intuitively true from everyday experience, but has not been previously conclusively demonstrated, as far as we are aware.

Since yielding intent can be recognised from deceleration behavior, deceleration has been referred to as a form of implicit communication from drivers to surrounding road users (Dey et al., 2020; Domeyer et al., 2019; Markkula et al., 2020). In Study 2, we also considered

explicit communication, in the form of headlight flashes that, on some trials, indicated the onset of yielding. Previous empirical work has shown that explicit indications of yielding (either by headlight flashes or more novel eHMIs) increase the tendency for crossing in front of an approaching vehicle, and speeds up the decision to do so, but again this phenomenon has not been subject to computational modeling. Our model assumes that explicit eHMI indications of yielding are considered by the pedestrian as an additional piece of evidence in favour of a road-crossing decision. The evidence “boost” that was added at eHMI activation was kinematics-independent, but it can be noted in Fig. 11 (and indirectly in Fig. 13) that the model nevertheless managed to capture the clear interaction between kinematics and eHMI presence, with largest impact of eHMI for smaller TTAs. This pattern arises naturally from the model, because at larger TTAs there is already strong kinematic evidence in favour of crossing, such that added evidence from eHMI will have a relatively minor impact (even more so given the saturating transfer function constraining $s(t)$).

Limitations

Although the model was able to capture all of the main qualitative effects of the studied factors on road-crossing

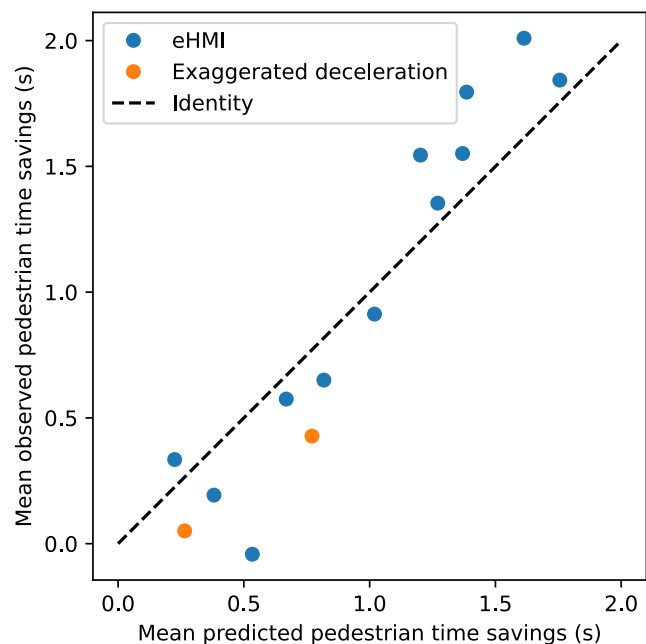


Fig. 13 Observed and predicted average pedestrian time savings due to implicit and explicit communication of vehicle yielding intentions. Each data point is the difference in average crossing decision time between two scenarios, where the only difference between scenarios was the absence or presence of an exaggerated yielding deceleration (orange; Study 1) or an eHMI indication of yielding (blue; Study 2). Overall $R^2 = 0.90$ and mean absolute deviation 0.28 s

decisions, it is clear that the quantitative model fits were not perfect for all scenarios. There seems to be at least two systematic shortcomings in the model's predictions:

First, a specific shortcoming of the model is a tendency to exhibit a non-negligible crossing probability at low apparent TTAs in some situations where close to no human participants initiated crossing. This is visible as smoother model CDFs versus more stepwise empirical CDFs in especially the constant-speed scenarios (Figs. 5 and 10), particularly those with high initial TTAs, and particularly so in Study 2 (Fig. 10).

Second, while the distance term in the model allowed capturing the overall increased tendency to cross with increased vehicle distance (or differently put, with increased speed for a fixed TTA), again the quantitative fits were not perfect. As can be seen in Figs. 5 and 10, the observed impact of the distance manipulation was greatest at intermediate initial TTAs around 4–5 s, and the model captures this nonlinear interaction qualitatively, but exaggerates it quantitatively (see the same figures, and also Fig. 7). Our formulation of the distance term in the model was mathematically equivalent to the one used by Zgonnikov et al. (2020), but was reformulated here to more clearly express one possible explanation for the phenomenon as such: Our formulation suggests that the distance-dependency arises due to pedestrians' prior expectations of vehicles driving at a certain speed, such that higher than expected speeds lead to over-estimates of generalized TTA, and vice versa.

Some of the shortcomings mentioned above can likely be attributed to the simplifying assumptions made: one parameterization for all individuals, constant accumulation noise variance and decision boundaries, and independent linear summation of the different sources of evidence.

The choice of using a single parameterization across participants was made mainly for technical purposes. A hierarchical estimation of parameters would be conceptually straightforward using well known methods, such as partial pooling and MCMC, but would likely require more per-participant repetitions, especially compared to Study 1. The other simplifying assumptions we made affect also the structure of the model, and changing them would require further theoretical development.

Inferences of the model's parameter values are hindered by high correlations in the posterior estimates. The correlations are likely due to redundant parameterization in the presented formulation. Future work should try to find a more parsimonious model formulation which would give clearer interpretation for the parameters and more quantitatively link more naturalistic tasks to what is known from evidence accumulation models in simpler task environments.

The generalizability of the parameter values or distributions to new data is not fully understood, and especially due to the relatively large amount of parameters (8 in Study 1) to a limited dataset (320 crossings in Study 1), there is a risk of overfitting. This is addressed somewhat by Study 2, although due to the re-estimation of the passed threshold τ_p this doesn't serve as a pure replication. Future work should address the generalizability with replication data or with computational methods such as cross validation with a larger data set.

A relatively straightforward extension to the model would be to allow a collapsing decision boundary for the accumulated evidence, to capture potential time pressure effects. This is a somewhat standard extension in drift diffusion models with existing implementations available (Shinn et al., 2020), and has been recently applied also to road-crossing modeling by Zgonnikov et al. (2020), in their case to account for a small but statistically significant decrease of response times with decreasing TTAs. Similar approaches could be used to study, e.g., time-varying accumulation noise.

The general approach we adopted for modeling the input sensory evidence to be accumulated, describing it as a generalised TTA constructed from separate sources of evidence in favor of crossing, worked relatively well overall, but there are many possible variations to Eq. 3 that we did not investigate here. Not least our assumption that the various factors affecting the generalised TTA are independent and additive seems likely to be an oversimplification. For example, as shown by Lee et al. (2019), the visibility of eHMI varies with distance, suggesting inclusion of the distance $D(t)$ also in the $H(t)$ term in Eq. 3, and possibly something similar could be the case also for perception of acceleration (the $\dot{\tau}$ term).

The sensory model could also be improved by incorporating more details of what is known about the limitations of human sensory system. Similar discussion and some development is happening within car following models used for traffic simulation (Saifuzzaman & Zheng, 2014), e.g., using visually projected angles of an object rather than its distance from the observer. However, such models typically cause non-linearities within the model, which would require also further deviation from the Gaussian-linear assumptions of classical drift-diffusion models. On the other hand for example a Bayesian observer based modeling would establish further links to recent developments in computational neuroscience (Friston, 2012), as has been proposed for car following (Pekkanen et al., 2018).

More generally, it is worth pointing out that our results do not permit conclusions about the exact sensory information made use of by our participants while deciding on their road-crossing. It is possible that some of the success of our model is due to our basing it on sensory quantities

that are visually accessible to human observers, such as τ and $\hat{\tau}$, but it may just as well be that the participants used some other sensory information that roughly covaries with these quantities. To draw more specific conclusions in this respect, one would need experiments and model comparisons targeting these questions specifically.

Implications for Wider Computational Modeling of Cognition and Behavior

Besides the task-specific future modeling directions discussed above, there are also possible implications from our work for computational modeling of decision-making more in general.

As mentioned in the introduction, most evidence accumulation modeling work in the literature focuses on abstract tasks of perceptual discrimination or value-based choice. Our results instead add to a more limited (but growing) body of support for drift diffusion type models in locomotion and general sensorimotor interaction with the world (Boda et al., 2020; Giles et al., 2019; Kovaceva et al., 2020; Markkula et al., 2018; Markkula et al., 2018; Piccinini et al., 2020; Xue et al., 2018; Zgonnikov et al., 2020). In this type of context, decisions are less purely “perceptual” or “cognitive,” and instead arguably more embodied in nature, yet interestingly the same type of decision mechanisms seem to apply. This is a largely unexplored research area, with many potentially fruitful directions to pursue. One aspect of decision-making that becomes more obvious in this context is that sensorimotor tasks are often highly dynamic, such that the associated decision evidence tends to vary over time, often continuously so. This highlights the similarity of evidence accumulation models with frameworks like dynamic field theory, where dynamic sensory input can also be accumulated up to thresholds where new behaviors occur (Schöner (2007); making the link between these modeling approaches more explicit could be one interesting direction. Another could be to extend the existing battery of laboratory paradigms used for evidence accumulation modeling, by developing paradigms providing continuously time-varying decision information, rather than static or intermittently changing decision evidence as in most existing paradigms. New paradigms emphasising continuously time-varying evidence could be devised both in tasks with some external validity, such as we have done here, or tasks that may be more artificial in nature but which may allow collection of larger datasets and thus allow fitting of models to individual participants.

Addressing more complicated tasks and models brings about some technical considerations. For conventional drift diffusion models, and various extensions, closed form solutions or efficient numerical approximations are available (Navarro & Fuss, 2009; Wiecki et al., 2013).

For some more general models, such as supporting an arbitrarily varying drift rate, however, such solutions or approximations are typically not known or even necessarily possible.

Drift diffusion type-models are mathematically stochastic difference or differential equations (SDE) and are relatively easy to sample from (at least in discrete time formulations or approximations), but parameter estimation and statistical inference becomes complicated, as evidenced by our previous inconclusive attempts (Giles et al., 2019). To simplify estimation and inference, we opted here to instead use discretized approximations of the underlying distributions and relatively simple Euler stepping solution of the SDE. More sophisticated methods for solving and approximating SDEs are known and actively studied (Smith, 2000; Särkkä & Solin, 2019) and applying such could increase the accuracy of the approximation and curb computational complexity.

Applied Implications

The model we propose here may be useful in applied settings in a few different ways, for example, to predict and optimize the impact of automated vehicle design on traffic flow efficiency. As summarized in Fig. 13, our model was rather successful at predicting the average time savings for the pedestrians achieved in our studies when the vehicle communicated its yielding intentions implicitly (via exaggerated deceleration; Study 1) or explicitly (via eHMI; Study 2). It is worth noting that the time savings for the vehicle passengers would typically be larger, due to the acceleration dynamics of vehicles (both deceleration and subsequent acceleration take time) (Markkula et al., 2018).

The model may also be useful as a component in algorithms for real-time sensor data interpretation. Making use of models of pedestrian behavior in automated vehicle algorithms is an active area of research, but so far the models used have been relatively simplistic (Camara et al., 2020; Jayaraman et al., 2021; Kapania et al., 2019). Another important role of human behavior models in vehicle development is as agents in simulation environments for virtual testing (Behbahani et al., 2019; Camara et al., 2020; Markkula et al., 2018). For both of these applications, since the current model only considers the specific case of a pedestrian who is stationary at the kerb, one would likely want to adapt the model into a larger framework for behavior prediction, to allow modeling of a richer variety of scenarios.

Also beyond the context of vehicle automation, the model may be of use for studying and improving safety in conventional, non-automated traffic, where car-pedestrian collisions account for a substantial fraction of casualties (Schneider, 2020; Organization, 2018). The model, or future

improvements of it, could, for example, be used to study the probability of unsafe pedestrian crossing behavior as a function of road design decisions affecting, for example, vehicle speed and visibility.

Conclusion

We have demonstrated that variable-drift diffusion models can be used to account for timing of pedestrian road-crossing decisions, and how these are affected by a number of different factors: (i) The impact of vehicle kinematics (distance, speed, and deceleration) of the approaching vehicle was established across two independent—and methodologically rather different—studies, with the second study serving to validate the model fits obtained from the first study. (ii) It was known since before that vehicle deceleration magnitudes affect pedestrian crossing decisions, but our model analyses permit the conclusion that this effect does not arise solely due to the impact of the deceleration on the apparent time gap, but rather due to a separate process whereby the pedestrians recognize the yielding intent of the vehicle. (iii) We also show how the impact of explicit communication of

yielding intent (e.g., eHMI) can be modelled as providing an extra source of evidence in favour of initiating road-crossing.

One central feature of our model was our assumption that different sources of decision evidence could be seen as independent additions to a generalised time to arrival quantity, which was then thresholded to yield the momentary evidence accumulation, and we found that this approach worked rather well in general. However, the quantitative fits between observed and model-predicted behavior were far from perfect for all scenarios, and future work may refine the exact formulation of the model, for example, in terms of the translation from sensory input to decision evidence, to further improve the model's predictions.

We have illustrated and discussed how our model can be put to applied use in research and development work on vehicle automation and road safety. From a perspective of computational cognitive models more in general, we conclude that variable-drift diffusion models provide a promising framework for describing human decision-making also in complex real-world situations, where continuous integration of multiple sources of time-varying evidence is necessary for successful behavior.

Appendix

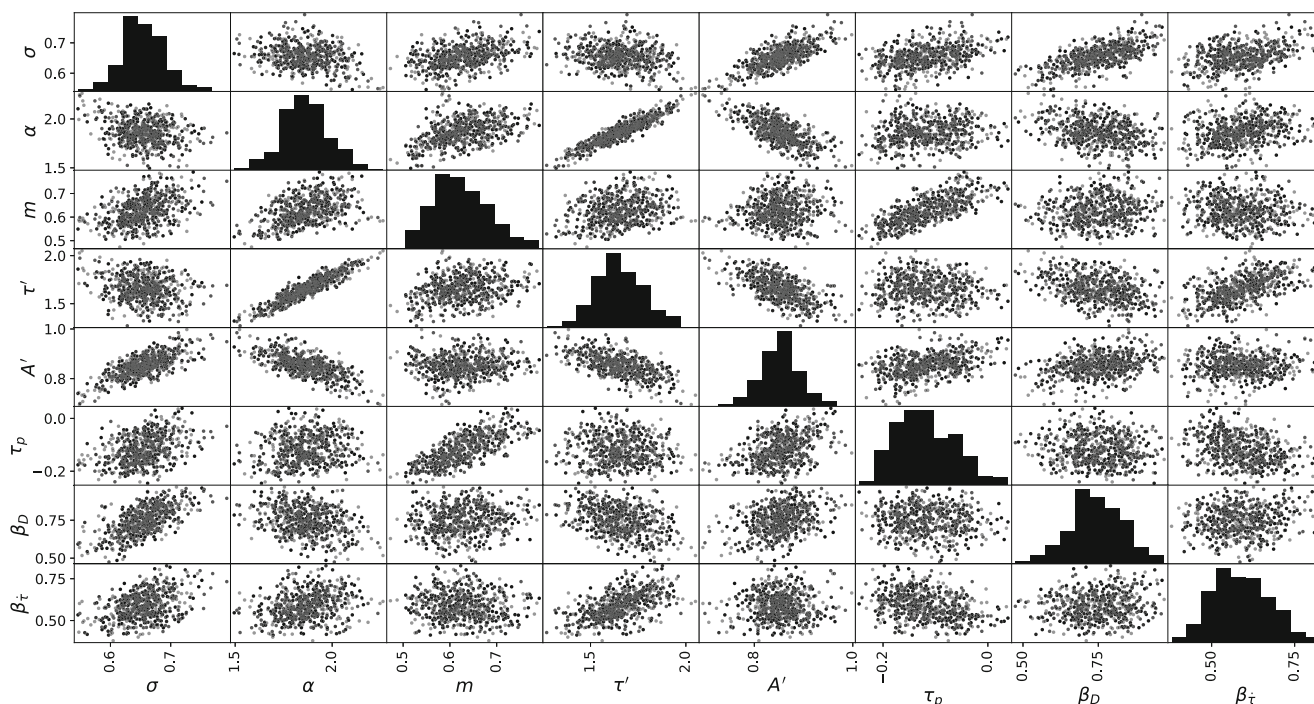


Fig. 14 Marginal and pairwise posterior distributions of parameters for Study 1. High correlations (up to 0.92 for α and τ') are observed for some variables, indicating that the parameterization is likely redundant and marginal distributions should be interpreted with caution

Author Contribution Funding acquisition: NM, GM, TD. Conceptualization and methodology: all authors. Investigation: OG, YML, RM. Data curation and formal analysis: JP, OG. Writing—original draft: JP, GM. Writing—review and editing: all authors

Funding Open access funding provided by University of Helsinki including Helsinki University Central Hospital. This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement 723395 and by the UK Engineering and Physical Sciences Research Council under grant EP/S005056/1.

Availability of Data and Material The data and model, analysis and figure code is available at <https://github.com/jampekka/vddfif/>.

Declarations

Conflict of Interests The authors declare that there are no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Behbahani, F., Shiarlis, K., Chen, X., Kurin, V., Kasewa, S., Stirbu, C., Gomes, J., Paul, S., Oliehoek, F. A., Messias, J., & Whiteson, S. (2019). Learning from demonstration in the wild. In *IEEE international conference on robotics and automation (ICRA)* 2019.
- Boda, C. N., Lehtonen, E., & Dozza, M. (2020). A computational driver model to predict driver control at unsignalised intersections. *IEEE Access*, 8, 104619–104631. <https://doi.org/10.1109/ACCESS.2020.2999851>, <https://ieeexplore.ieee.org/document/9107141/>.
- Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, 19(2), 442–477. <https://doi.org/10.1162/neco.2007.19.2.442>.
- Brewer, M. A., Fitzpatrick, K., Whitacre, J. A., & Lord, D. (2006). Exploration of pedestrian gap-acceptance behavior at selected locations. *Transportation Research Record*, 1982(1), 132–140.
- Brosnan, M. B., Sabarodien, K., Silk, T., Genc, S., Newman, D. P., Loughnane, G. M., Fornito, A., O'Connell, R. G., & Bellgrove, M. A. (2020). Evidence accumulation during perceptual decisions in humans varies as a function of dorsal frontoparietal organization. *Nature Human Behaviour*, 4(8), 844–855. <https://doi.org/10.1038/s41562-020-0863-4>.
- Brown, B., & Laurier, E. (2017). The trouble with autopilots: Assisted and autonomous driving on the social road. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 416–429).
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178. <https://doi.org/10.1016/j.cogpsych.2007.12.002>.
- Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in Cognitive Sciences*, 23(3), 251–263. <https://doi.org/10.1016/j.tics.2018.12.003>.
- Camara, F., Bellotto, N., Cosar, S., Weber, F., Nathanael, D., Althoff, M., Wu, J., Ruenz, J., Dietrich, A., Markkula, G., Schieben, A., Tango, F., Merat, N., & Fox, C. (2020). Pedestrian models for autonomous driving part II: High-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems* pp. 1–20. <https://doi.org/10.1109/TITS.2020.3006767>.
- Chandra, S., & Bharti, A. K. (2013). Speed distribution curves for pedestrians during walking and crossing. *Procedia - Social and Behavioral Sciences*, 104, 660–667. <https://doi.org/10.1016/j.sbspro.2013.11.160>.
- Choudhury, C. F., Ben-Akiva, M. E., Toledo, T., Lee, G., & Rao, A. (2007). Modeling cooperative lane changing and forced merging behavior. In *86th annual meeting of the transportation research board*. Washington.
- Dey, D., Matvienko, A., Berger, M., Pflöging, B., Martens, M., & Terken, J. (2020). Communicating the intention of an automated vehicle to pedestrians: The contributions of eHMI and vehicle behavior. it - Information Technology p. in press.
- Dietrich, A., Maruhn P., Schwarze, L., & Bengler, K. (2020). Implicit communication of automated vehicles in urban scenarios: effects of pitch and deceleration on pedestrian crossing behavior. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar (Eds.) *Human systems engineering and design II, advances in intelligent systems and computing* (pp. 176–181). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_27.
- Domeyer, J., Dinparastdjadid, A., Lee, J. D., Douglas, G., Alsaïd, A., & Price, M. (2019). Proxemics and kinesics in automated vehicle-pedestrian communication: Representing ethnographic observations. *Transportation Research Record*, 2673(10), 70–81.
- Downes, A. N., & Borovkov, K. (2008). First passage densities and boundary crossing probabilities for diffusion processes. *Methodology and Computing in Applied Probability*, 10(4), 621–644.
- Elvik, R. (2014). A review of game-theoretic models of road user behaviour. *Accident Analysis & Prevention*, 62, 388–396. <https://doi.org/10.1016/j.aap.2013.06.016>.
- Engström, J., & Ljung Aust, M. (2011). Adaptive behavior in the simulator: Implications for active safety system evaluation. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.) *Handbook of driving simulation for engineering, medicine, and psychology* (pp. 41–115).
- Faas, S. M., Mathis, L. A., & Baumann, M. (2020). External HMI for self-driving vehicles: Which information shall be displayed?. *Transportation Research Part F: Traffic Psychology and Behaviour*, 68, 171–186. <https://doi.org/10.1016/j.trf.2019.12.009>, <https://linkinghub.elsevier.com/retrieve/pii/S1369847819305479>.
- Fajen, B. R. (2013). Guiding locomotion in complex, dynamic environments. *Frontiers in Behavioral Neuroscience* 7. <https://doi.org/10.3389/fnbeh.2013.00085>.
- Fajen, B. R., & Warren, W. H. (2003). Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 343.
- Friston, K. (2012). The history of the future of the bayesian brain. *NeuroImage*, 62(2), 1230–1233.
- Gibson, J. J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology*, 49(3), 182–194. <https://doi.org/10.1111/j.2044-8295.1958.tb00656.x>.

- Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. *The American Journal of Psychology*, 51(3), 453. <https://doi.org/10.2307/1416145>.
- Giles, O., Markkula, G., Pekkanen, J., Yokota, N., Matsunaga, N., Merat, N., & Daimon, T. (2019). At the zebra crossing: Modelling complex decision processes with variable-drift diffusion models. In *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 366–372). Cognitive Science Society.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535–574.
- Haario, H., Saksman, E., Tamminen, J., et al. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2), 223–242.
- Helbing, D. (2001). Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73(4), 75.
- Hoogendoorn, S., & Bovy, H. L. P. (2003). Simulation of pedestrian flows by optimal control and differential games. *Optimal Control Applications and Methods*, 24(3), 153–172. <https://doi.org/10.1002/oca.727>.
- Jayaraman, S. K., Robert, L. P., Yang, X. J., & Tilbury, D.M. (2021). Multimodal hybrid pedestrian: A hybrid automaton model of urban pedestrian behavior for automated driving applications. *IEEE Access*, 9, 27708–27722. <https://doi.org/10.1109/ACCESS.2021.3058307>.
- Kapania, N. R., Govindarajan, V., Borrelli, F., & Gerdes, J.C. (2019). A hybrid control design for autonomous vehicles at uncontrolled crosswalks. In *2019 IEEE intelligent vehicles symposium (IV)* (pp. 1604–1611). France: IEEE. <https://doi.org/10.1109/IVS.2019.8814116>, <https://ieeexplore.ieee.org/document/8814116/>.
- Kovaceva, J., Bärgrman, J., & Dozza, M. (2020). A comparison of computational driver models using naturalistic and test-track data from cyclist-overtaking manoeuvres. *Transportation research part F: traffic psychology and behaviour*, 75, 87–105.
- Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, 5(4), 437–459.
- Lee, Y. M., Madigan, R., Garcia, J., Tomlinson, A., Solernou, A., Romano, R., Markkula, G., Merat, N., & Uttley, J. (2019). Understanding the messages conveyed by automated vehicles. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications* (pp. 134–143).
- Lee, Y. M., Madigan, R., Uzundu, C., Garcia, J., Romano, R., Markkula, G., & Merat, N. (2020). Learning to interpret novel eHMI: The effect of communicative indications and vehicle kinematics on pedestrians' crossing behaviour. Manuscript in preparation.
- Lobjois, R., & Cavallo, V. (2007). Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task. *Accident Analysis & Prevention*, 39(5), 934–943. <https://doi.org/10.1016/j.aap.2006.12.013>.
- Lobo, L., Heras-Escribano, M., & Travieso, D. (2018). The history and philosophy of ecological psychology. *Frontiers in Psychology*, 9, 2228. <https://doi.org/10.3389/fpsyg.2018.02228>.
- Ma, Y., Lee, E. W. M., & Yuen, R.K.K. (2016). An artificial intelligence-based approach for simulating pedestrian movement. *IEEE Transactions on Intelligent Transportation Systems*, 17(11), 3159–3170. <https://doi.org/10.1109/TITS.2016.2542843>.
- Markkula, G., Benderius, O., Wolff, K., & Wahde, M. (2012). A review of near-collision driver behavior models. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(6), 1117–1143. <https://doi.org/10.1177/0018720812448474>.
- Markkula, G., Boer, E., Romano, R., & Merat, N. (2018). Sustained sensorimotor control as intermittent decisions about prediction errors: Computational framework and application to ground vehicle steering. *Biological Cybernetics*, 112(3), 181–207. <https://doi.org/10.1007/s00422-017-0743-9>.
- Markkula, G., Madigan, R., Nathanael, D., Portouli, E., Lee, Y. M., Dietrich, A., Billington, J., Schieben, A., & Merat, N. (2020). Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science* pp. 1–24. <https://doi.org/10.1080/1463922X.2020.1736686>.
- Markkula, G., Romano, R., Madigan, R., Fox, C. W., Giles, O. T., & Merat, N. (2018). Models of human decision-making as tools for estimating and optimizing impacts of vehicle automation. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(37), 153–163. <https://doi.org/10.1177/0361198118792131>.
- Merchant, H., & Georgopoulos, A. P. (2006). Neurophysiology of perceptual and motor aspects of interception. *Journal of Neurophysiology*, 95(1), 1–13.
- Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in wiener diffusion models. *Journal of Mathematical Psychology*, 53(4), 222–230.
- Organization, W. H. (2018). Global status report on road safety 2018. Tech. rep., World Health Organization.
- Pekkanen, J., Lappi, O., Rinkkala, P., Tuhkanen, S., Frantsi, R., & Summala, H. (2018). A computational model for driver's cognitive state, visual perception and intermittent attention in a distracted car following task. *Royal Society Open Science*, 5(9), 180194.
- Petzoldt, T. (2014). On the relationship between pedestrian gap acceptance and time to arrival estimates. *Accident Analysis & Prevention*, 72, 127–133. <https://doi.org/10.1016/j.aap.2014.06.019>, <https://linkinghub.elsevier.com/retrieve/pii/S0001457514001924>.
- Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., & Castelfranchi, C. (2019). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of Life Reviews*, 28, 1–21. <https://doi.org/10.1016/j.plrev.2018.06.014>.
- Piccinini, G. B., Lehtonen, E., Forcolin, F., Engström, J., Albers, D., Markkula, G., Lodin, J., & Sandin, J. (2020). How do drivers respond to silent automation failures? Driving simulator study and comparison of computational driver 2 braking models 3. *Human Factors*, 62(7), 1212–1229. <https://doi.org/10.1177/0018720819875347>.
- Plöchl, M., & Edelmann, J. (2007). Driver models in automobile dynamics application. *Vehicle System Dynamics*, 45(7–8), 699–741.
- Powers, W. T. (1978). Quantitative analysis of purposive systems: Some spade-work at the foundations of scientific psychology. *Psychological Review*, 85(5), 417–435.
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T.J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, 117(4), 1113–1143. <https://doi.org/10.1037/a0020311>.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4), 260–281. <https://doi.org/10.1016/j.tics.2016.01.007>.
- Sadigh, D., Landolfi, N., Sastry, S. S., Seshia, S. A., & 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. *Autonomous Robots*, 42(7), 1405–1426. <https://doi.org/10.1007/s10514-018-9746-1>.
- Saifuzzaman, M., & Zheng, Z. (2014). Incorporating human-factors in car-following models: A review of recent developments and research needs. *Transportation Research Part C: Emerging Technologies*, 48, 379–403.

- Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors*, 48(2), 362–380.
- Särkkä, S., & Solin, A. (2019). *Applied stochastic differential equations* Vol. 10. Cambridge: Cambridge University Press.
- Schieben, A., Wilbrink, M., Kettwich, C., Madigan, R., Louw, T., & Merat, N. (2019). Designing the interaction of automated vehicles with other traffic participants: Design considerations based on human needs and expectations. *Cognition Technology & Work*, 21(1), 69–85.
- Schneemann, F., & Gohl, I. (2016). Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. In *Proceedings of the 2016 IEEE intelligent vehicles symposium (IV)* (pp. 38–43). Sweden.
- Schneider, R. J. (2020). United States pedestrian fatality trends, 1977 to 2016. *Transportation Research Record*, 2674(9), 1069–1083.
- Schöner, G. (2007). Dynamical systems approaches to cognition. In R. Sun (Ed.) *Cambridge handbook of computational cognitive modeling*. Cambridge University Press.
- Schroeder, B. J. (2008). A behavior-based methodology for evaluating pedestrian-vehicle interaction at crosswalks. Ph.D. thesis, North Carolina State University.
- Schwartz, W., Pierson, A., Alonso-Mora, J., Karaman, S., & Rus, D. (2019). Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50), 24972–24978. <https://doi.org/10.1073/pnas.1820676116>.
- Shinn, M., Lam, N. H., & Murray, J.D. (2020). A flexible framework for simulating and fitting generalized drift-diffusion models. *eLife*, 9, e56938. <https://doi.org/10.7554/eLife.56938>.
- Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, 44(3), 408–463.
- Turnwald, A., Althoff, D., Wollherr, D., & Buss, M. (2016). Understanding human avoidance behavior: Interaction-aware decision making based on game theory. *International Journal of Social Robotics*, 8(2), 331–351.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592.
- Varhelyi, A. (1998). Drivers' speed behaviour at a zebra crossing: A case study. *Accident Analysis & Prevention*, 30(6), 731–743.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... van Mulbregt, P. (2020). SciPy 1.0 contributors: SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- Wiecki, T. V., Sofer, I., & Frank, M.J. (2013). Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, 7, 14.
- Wong, K. F., & Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4), 1314–1328. <https://doi.org/10.1523/JNEUROSCI.3733-05.2006>.
- Xue, Q., Markkula, G., Yan, X., & Merat, N. (2018). Using perceptual cues for brake response to a lead vehicle: Comparing threshold and accumulator models of visual looming. *Accident Analysis & Prevention*, 118, 114–124. <https://doi.org/10.1016/j.aap.2018.06.006>.
- Yannis, G., Papadimitriou, E., & Theofilatos, A. (2013). Pedestrian gap acceptance for mid-block street crossing. *Transportation Planning and Technology*, 36(5), 450–462. <https://doi.org/10.1080/03081060.2013.818274>.
- Zgonnikov, A., Abbink, D., & Markkula, G. (2020). Should I stay or should I go? Evidence accumulation drives decision making in human drivers. <https://doi.org/10.31234/osf.io/p8dxn>. Publisher: PsyArXiv.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.