At the Zebra Crossing: Modelling Complex Decision Processes with Variable-Drift Diffusion Models

Oscar Giles (o.t.giles@leeds.ac.uk)
Institute for Transport Studies and School of Psychology, University of Leeds, LS2 9JT, Leeds, United Kingdom

Gustav Markkula (g.markkula@leeds.ac.uk)
Institute for Transport Studies, University of Leeds, LS2 9JT, Leeds, United Kingdom

Jami Pekkanen (j.j.o.pekkanen@leeds.ac.uk)
School of Psychology and Institute for Transport Studies, University of Leeds, LS2 9JT, Leeds, United Kingdom

Naoki Yokota (ujaeu-gfuir4@a5.keio.jp)
Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan

Naoto Matsunaga (pinen07mn@a2.keio.jp)
Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan

Natasha Merat (n.merat@its.leeds.ac.uk)
Institute for Transport Studies, University of Leeds, LS2 9JT, Leeds, United Kingdom

Tatsuru Daimon (daimon@keio.jp)
Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan

Abstract

Drift diffusion (or evidence accumulation) models have found widespread use in the modelling of simple decision tasks. Extensions of these models, in which the model's instantaneous drift rate is not fixed but instead allowed to vary over time as a function of a stream of perceptual inputs, have allowed these models to account for more complex sensorimotor decision tasks. However, many real-world tasks seemingly rely on a myriad of even more complex underlying processes. One interesting example is the task of deciding whether to cross a road with an approaching vehicle. This action decision seemingly depends on sensory information both about own affordances (whether one can make it across before the vehicle) and action intention of others (whether the vehicle is yielding to oneself). Here, we compared three extensions of a standard drift diffusion model, with regards to their ability to capture timing of pedestrian crossing decisions in a virtual reality environment. We find that a single variable-drift diffusion model (S-VDDM) in which the varying drift rate is determined by visual quantities describing vehicle approach and deceleration, saturated at an upper and lower bound, can explain multimodal distributions of crossing times well across a broad range vehicle approach scenarios. More complex models, which attempt to partition the final crossing decision into constituent perceptual decisions, improve the fit to the human data but further work is needed before firm conclusions can be drawn from this finding.

Keywords: complex decision making; road crossing; variable-drift diffusion models

Introduction

Sensorimotor decision making, how people decide what motor actions to take and when, has been a key object of research over the past hundred years in the psychological sciences. One area of particular progress has been in the development of mathematical models which predict action choices and reaction times. In particular, drift diffusion models (DDMs) and various related models, which describe the decision making process as a noisy accumulation of sensory information to a bound, have been found to very successfully capture behavioral data across a plethora of experimental tasks (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Ratcliff, Smith, Brown, & McKoon, 2016) and have shown success in bridging the gap between neurophysiological and behavioral data (Purcell et al., 2010).

DDMs and related models have most commonly been applied to two alternative force choice (2-AFC) tasks, in which people make a decision between two alternative choices based on perceptual information. Quintessential among these is the kinematogram task in which people decide the direction of a random flow of dots (Ratcliff et al., 2016). DDMs have also been successfully applied to more complex sensorimotor tasks, such as determining the action intentions of other people (Koul, Soriano, Tversky, Becchio, & Cavallo, 2019). However, standard DDMs and related models of the evidence accumulation type typically assume that the drift rate (i.e., the rate at which evidence accumulates to a bound), is set to a fixed value. Yet many sensorimotor decisions take place in the context of a continuous stream of varying sensory information. Models with variable drift rate, which we will refer to here as variable-drift diffusion models (VDDMs), have been successful in the vehicle driving context, accounting well for driver brake responses to the time varying visual looming of an approaching vehicle (Xue, Markkula, Yan, & Merat, 2018) as well as for steering responses during lane-keeping (Markkula, Boer, Romano, & Merat, 2018).

However, further generalization to more complex real world decisions brings additional challenges. Firstly, more

complex decisions may depend on multiple types of sensory cues, raising the question of how different cues should contribute to the drift rate. In this paper, we will consider a pedestrian's decision of when to cross at a zebra crossing with an approaching vehicle, a decision relying on at least two types of cues (Rasouli, Kotseruba, & Tsotsos, 2017): (1) Cues regarding own affordances, for example in terms of the time to arrival (TTA) of the approaching vehicle, in relation to the width of road to be crossed. (2) Cues regarding the action intention of the vehicle driver, in the form of kinematic cues (e.g., vehicle deceleration) and/or communicative cues (e.g., flashing headlights).

Secondly, when the sensory inputs to the model vary over a large magnitude, this may result in undesirable model behavior. For example, when a vehicle decelerates to a stop, its perceptually estimated TTA will go to infinity. If this is used as a model input then the accumulator will be guaranteed to reach its threshold (and initiate a crossing) immediately when the vehicle stops, when in fact people show a probabilistic delay in crossing times.

Finally, it remains unclear how complex decisions, like the zebra crossing decisions, are structured in practice. Is the overt behavior the result of only a single action decision ("I am crossing now"), or is that action decision underpinned by separate, purely perceptual decisions about the affordances and action intentions mentioned above (e.g., "I can make it across before the car"; "The car is stopping for me")? There are many examples in the broader literature of psychological, cognitive, and robotics models where multiple parallel units of activation dynamics akin to evidence accumulators have been interconnected to produce more complex emergent behavior (e.g., Cooper & Shallice, 2000; Sandamirskaya, Richter, & Schöner, 2011), but DDM type decision models have seemingly not been previously generalized in this direction.

In the current study we wished to test three novel VDDMs, which aim to address the above three challenges. Firstly we wished to test a model recently proposed by Markkula, Romano, et al., (2018), which we refer to as the connected variable-drift diffusion model (C-VDDM). The C-VDDM models action decisions and perceptual decisions as separate but interconnected accumulator units as discussed above (see Figure 1, top), where the drift rate of each perceptual unit is a function of a time varying sensory input. In turn, the drift rate of the action unit is a function of the current activation levels of each of the two perceptual units. The activation of each perceptual unit is bounded to ± 1 which ensures that large perceptual inputs do not immediately lead to the action unit reaching threshold. Markkula, Romano, et al., (2018) showed that this model could qualitatively account for bimodal distributions of crossing decision times, as reported for human pedestrians, but did not formally test or fit the model with human data.

We also wished to test a simplification of the C-VDDM model, in which a single perceptual unit has a drift rate which varies as a function of a linear combination of multiple sensory cues (see Figure 1, middle), in turn

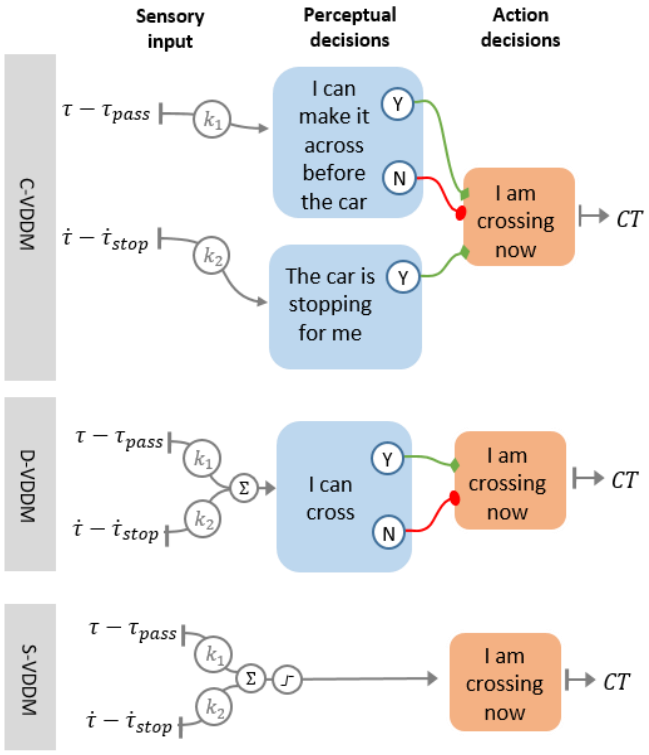

Figure 1: The three variable-drift diffusion models (VDDMs).

modulating the activation level of an action unit. We refer to this as the dual variable-drift diffusion model (D-VDDM). Like the C-VDDM, this model ensures that large sensory inputs do not result in the reaction time distribution collapsing to a spike. However, unlike the C-VDDM, this model does not independently represent different underlying decision processes.

Finally, we also wanted to test a model consisting of a single accumulator unit with a drift rate that varies as a function of a linear combination of sensory cues (see Figure 1, bottom). Instead of the D-VDDM's intermediate accumulator unit, to ensure that high input values did not result in rapid termination of the accumulation process, we limited the drift rate of this model by saturating the perceptual input, such that it could not have a magnitude greater than a certain value. We refer to this model as the S-VDDM.

To compare the three models we collected data on pedestrian crossing times, using a virtual reality (VR) headset. VR allowed us to carefully control the experimental stimulus (i.e., vehicle approach trajectories) and avoid confounding variables that may be present when observing crossing behavior in the real world (e.g., effects of other pedestrians or additional vehicles on crossing behavior). We used a large range of vehicle approach trajectories, which were specifically chosen with the aim of creating different types of situations with respect to pedestrian affordances and vehicle action intentions.

Virtual Reality Road Crossing Task

Participants

Twenty participants (age 24-60, average 27.9 years; 11 male) took part in the study and were recruited from a University participant pool. All participants provided informed consent, and the study was approved by the University Research Ethics committee.

Materials and Design

Participants wore an HTC Vive Virtual Reality headset while standing. All stimuli were created in Unity 2018. The stimuli consisted of a straight two lane road (width: 5.85 m) with a zebra crossing and pavements on either side. The trial started with the participant standing at the edge of the zebra crossing, looking directly across it. To start the trial the participant turned their head to the right, which (unbeknownst to the participants) instantiated the approaching car at its initial position and speed for the scenario in question. For increased experimental control and simplicity, the participants did not physically walk across the VR pedestrian crossing, but instead pressed a trigger button on an HTC Vive controller when they decided it was safe to cross, and the participant's view point in the virtual world then translated across the crossing at 1.31 $\mathrm{m\,s^{-1}}$. Once the participant had crossed the road in VR, the trial ended. The time at which the participant initiated the crossing, measured from the point at which the vehicle began moving, was the primary outcome measure.

Scenarios

To preserve as much as possible a natural road-crossing behavior, the number of trials per participant was limited to 16. Each of these trials used a different vehicle approach scenario, presented in a pseudo-randomized order to the participants. The scenarios were defined so as to elicit a broad range of different crossing situations, and were of three general types, with parameters as listed in Table 1:

"*Constant velocity*" (6 scenarios): The vehicle appeared at distance D_{init} from the pedestrian, and maintained a constant velocity v_{init}, i.e., it had an initial time to arrival TTA_{init}.

"*Decelerate to a stop*" (8 scenarios): The vehicle appeared at distance D_{init} from the pedestrian, with initial speed v_{init}, and immediately decelerated at a constant rate so as to reach zero speed at distance D_{stop}.

"*Decelerate without stopping*" (2 scenarios): The vehicle appeared at distance D_{init} at speed v_{init} and immediately decelerated at a constant rate until distance D_{stop}, where it continued to travel at a final speed of 5 km/h.

Table 1: Scenario parameters

Scenario type	v_{init} (km/h)	D_{init} (m)	TTA_{init} (s)	D_{stop} (m)
Constant velocity	25	15.90	2.29	N/A
	50	31.81	2.29	N/A
	25	31.81	4.58	N/A
	50	63.61	4.58	N/A
	25	47.71	6.87	N/A
	50	95.42	6.87	N/A
Decelerate to a stop	25	15.90	2.29	4
	50	31.81	2.29	4
	50	31.81	2.29	8
	25	31.81	4.58	4
	50	63.61	4.58	4
	50	63.61	4.58	8
	25	47.71	6.87	4
	50	95.42	6.87	4
Decelerate w/o stopping	50	27.78	2	8
	50	41.67	3	8

Variable-Drift Diffusion Models

We developed three models to capture the road crossing times (CT) of pedestrians in the VR study, as illustrated in Figure 1. All models received the same perceptual inputs. As in Markkula, Romano et al., (2018) the first input was based on the instantaneous apparent time to arrival (TTA) of the vehicle, disregarding any deceleration. This apparent TTA is visually available, as the relative rate of optical expansion τ (Lee, 1976). The model input was given by $\tau - \tau_{pass}$, where $\tau_{pass} = 2.46$ (the time it took to cross the VR road). Thus the model input was positive when it was possible to make it across the road before the vehicle (based on apparent TTA), and negative when it was not. The second model input was based on the derivative of the vehicle's apparent TTA, $\dot{\tau}$. The input was defined as $\dot{\tau} - \dot{\tau}_{pass}$, with $\dot{\tau}_{pass} = -0.5$, corresponding to the vehicle stopping to just exactly touch the participant (Lee, 1976). Thus, the input was positive when the vehicle was decelerating so as to stop before the participant, and negative when not.

For the C-VDDM model these inputs were fed into two separate "perceptual decision" units. For the D-VDDM and S-VDDM model these were linearly combined and fed into a single accumulator unit. For the S-VDDM this combined weighted input was also limited such that it could not exceed a certain magnitude.

Model Specification

The models were all specified on the same general form, following Markkula, Romano et al., (2018), of which a brief summary is provided here. At any point in time t, the activation level of each of the model's accumulator units is described by the vector, $\boldsymbol{A}_t = \left[A_{1,t}, A_{2,t}, \ldots, A_{U,t}\right]^T$, where U is the number of accumulator units, and each unit's activation is limited to $-1 \leq A_{i,t} \leq 1$, with 1 and -1 signifying "yes" and "no" decision states, respectively. At each simulation time step, the activation levels are updated according to,

$$\frac{d}{dt}\boldsymbol{A}_t = -\frac{1}{T}\boldsymbol{A}_t + f_c(\boldsymbol{W}_I D(\boldsymbol{K})\boldsymbol{I}_t, \eta) + \boldsymbol{W}_Y D(\boldsymbol{Y}) f_Y(\boldsymbol{A}_t) + \boldsymbol{W}_N D(\boldsymbol{N}) f_N(\boldsymbol{A}_t)$$

$$\boldsymbol{A}_{t+dt} \sim MultiNorm\left(\boldsymbol{A}_t + d\boldsymbol{A}_t, \boldsymbol{\Sigma}\sqrt{dt}\right),$$

where $\boldsymbol{I}_t = \left[\tau_t - \tau_{pass}, \dot{\tau}_t - \dot{\tau}_{stop}\right]^T$, is a vector of perceptual inputs. $\boldsymbol{K} = [k_1, k_2]^T$ is a vector of relative weights for these two perceptual inputs, $\boldsymbol{Y}$ and $\boldsymbol{N}$ are vectors of connection weights for the "yes" and "no" accumulator output connections respectively and $D(\boldsymbol{x})$ is a diagonal matrix with diagonal $\boldsymbol{x}$. $\boldsymbol{W}_I$, $\boldsymbol{W}_N$ and $\boldsymbol{W}_Y$ are design matrices which specify accumulator inputs and connections, with elements $\boldsymbol{W}_{[j,k]} \in \{0, 1\}$. The function $f_c(\boldsymbol{W}_I D(\boldsymbol{K})\boldsymbol{I}_t, \eta)$ limits the perceptual inputs to the accumulators between $\pm\eta$. In the C-VDDM and D-VDDM η was fixed at infinity (and so had no effect), while in the S-VDDM it was a free parameter. This allowed the S-VDDM's activation to gradually rise to 1, even when the inputs were at large values. The function $f_Y(x)$ limits the input between 0 and 1, thus returning $f_Y(A_{i,t}) = 1$ for an accumulator activation $A_{i,t} = 1$ (a "yes" state), while $f_N(x) = f_Y(-x)$, such that $f_N(A_{i,t}) = 1$ for $A_{i,t} = -1$ (a "no" state). $\boldsymbol{\Sigma}$ is a covariance matrix with all off diagonal elements set to 0, and all diagonal elements sharing the same value, σ^2, representing noise in the decision process.

When the activation of the action decision accumulator reaches a value of 1, a decision to cross the road is made, and the time at which this occurs is the crossing time, CT_m.

Model Fitting

To simplify notation, here we denote all the parameters of a given VDDM model as θ. Fitting to the VR dataset is made challenging as calculating the likelihood function, $P(CT|\theta)$, involves computing a high dimension integral.

Instead we estimated the likelihood function using a large number of data simulations, referred to as the pseudo-likelihood estimation, $\hat{P}(CT|\theta)$. For each trial scenario we generated 5000 simulated crossing times, CT_m, from the model being fitted. We then calculated a numerical probability distribution $\boldsymbol{b}$ over 80 bins equally spaced between 0 and 20 seconds, where $\boldsymbol{b}$ is a vector where each element, b_i, is the relative frequency of CT_m falling into the ith bin. $\hat{P}(CT|\theta)$ was then estimated as the value of $\boldsymbol{b}$ for the bin corresponding to CT.

Due to the finite number of model simulations, with this method it is possible that a bin is assigned zero probability (no values of CT_m fell within that bin), despite the model having support over this region. If CT falls within such a bin then $\hat{P}(CT|\theta) = 0$, which can cause issues for the model fitting. To avoid this, we ensured that all bins had a non-zero probability by adjusting $\boldsymbol{b}$ by a constant $\boldsymbol{z}$, to $\boldsymbol{b}\lambda + \boldsymbol{z}(1-\lambda)$, where $\lambda = .98$. $\boldsymbol{z}$ was set as the probability of drawing a value from any given bin when sampling from a uniform distribution with bounds 0 and 20. In practice this had almost no discernible effect on the estimate of $\hat{P}(CT|\theta)$, but ensured non-zero support over all values of CT. Finally we removed the first "decelerate without stopping" trial from the analysis. This was because many participants began crossing while the vehicle was still in front of them, which the models were not designed to capture.

We used PSO (Wahde, 2008) to fit the models using the pseudo-likelihood estimation method described above. A swarm of 50 particles was used and optimized for 50 iterations. In all cases the algorithm appeared to converge to some local optimum (pseudo log-likelihood estimates stopped increasing) before the 50th iteration.

Table 2: Log likelihood and Akaike information criterion (AIC) for each of the models. *indicates the model with highest log likelihood estimate

Model	$log\ P(CT\|\hat{\theta})$	N param	AIC
C-VDDM	-953.72	7	1921.4
D-VDDM*	-871.90	6	1755.8
S-VDDM	-882.04	5	1774.0

Table 3: Estimated parameter values for each model. Fixed parameters are shown in italics.

Param	C-VDDM	D-VDDM	S-VDDM
T	0.67	0.26	0.34
$\boldsymbol{K}$	[4.35, 0.46]	[0.66, 0.42]	[0.47, 0.19]
$\boldsymbol{Y}$	[*0*, 0.44, 1.83]	[*0*, 3.25]	N/A
$\boldsymbol{N}$	[*0*, 0.76, *0*]	[*0*, 10.0]	N/A
σ	0.87	1.03	1.05
η	N/A	N/A	2.5

Results

Table 2 shows the pseudo log-likelihood estimate and AIC of the VR crossing time data for each of the three models. We can see that the D-VDDM captured the data the best (highest log likelihood) and had the lowest AIC value. The S-VDDM performed slightly worse, while the C-VDDM performed poorer than both. The parameters returned by the PSO algorithm are shown in Table 3.

To explore the model fits in more detail we simulated 5000 crossing times (CT_m) for each vehicle approach scenario and each fitted models. The left panel of Figure 2 shows the real CT (top panel), and simulated CT_m (bottom panel) for one of the "constant velocity" scenarios. We also plot the model activations for the S-VDDM (black traces, bottom panel). In this trial the vehicle starts far enough away that the participant has time to successfully cross the road, if this decision is made relatively quickly. However, the vehicle soon comes too close for a successful crossing to take place. Some participants crossed early in the vehicle's trajectory, while some waited for the vehicle to pass. All of the models were able to capture this trend. However, it appears that the S-VDDM (blue line; bottom panel) and

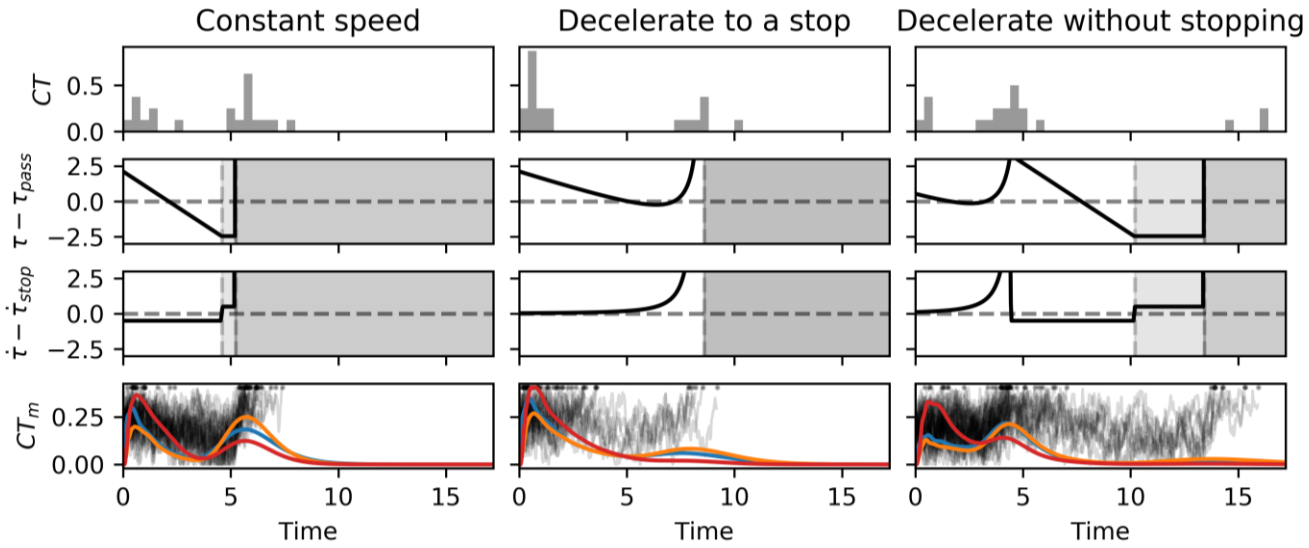

Figure 2: Human and model behavior in three example scenarios. Top panels show the observed human crossing times (CT) in the virtual reality experiment, and the second and third rows of panels show the sensory input cues to the models. Dark grey regions indicate that the vehicle has come to a stop or has passed the participant. Light grey regions indicate that the vehicle is passing the participant. The bottom panels show the simulated crossing time (CT_m) for the C-VDDM (red lines), D-VDDM (orange lines), and S-VDDM (blue lines). The black traces show example activations of the S-VDDM accumulator unit.

especially the C-VDDM (red line; bottom panel) showed a larger peak around the early crossings, while the D-VDDM (orange line; bottom panel) showed a larger peak after the vehicle had crossed, which better matched the participants' behavior.

The middle panels of Figure 2 shows the same plots for one of the "decelerate to a stop" scenarios. Again, we observed a bimodal distribution of crossing times (top panel), with some participants crossing early in the vehicle's trajectory, and others waiting until the vehicle had nearly or completely stopped. Here both the D-VDDM and S-VDDM captured this trend rather well, with a larger mode at the early crossing times and a smaller mode after the vehicle stopped. However, the C-VDDM was not able to capture the later crossing mode.

The right panels of Figure 2 show the same plots for one of the "decelerate without stopping" trials. Here, beyond the bimodal pattern already described for the "decelerate to a stop" scenario, a small number of participants also waited for the vehicle to completely pass before crossing. Thus the observed CT showed a tri-modal distribution. Both the D-VDDM and S-VDDM models reproduced these three modes well (the third mode is rather flat, but its presence can be seen from the black activation traces in bottom panel), while again the C-VDDM appeared to place too much weight over the initial mode, and predicted close to zero participants crossing after the vehicle had passed. Figure 3 shows the observed CT and model simulations, CT_m, for all scenarios.

However, we were concerned that the more complex C-VDDM's poor performance might be caused by the PSO algorithm getting stuck in a local optimum. Indeed, rerunning the fitting of the different VDDMs with new initial random seeds, and/or additional constraints on the parameter search range, we obtained slightly different parameterizations, but for the C-VDDM these never performed better than either the D-VDDM or S-VDDM.

Assuming that the C-VDDM's poor relative performance is not the result of challenges in finding the global optimum, we wondered whether one issue might be that the connected accumulator models all share a single σ parameter. Thus we refit the C-VDDM model with a separate σ parameter for each accumulator unit. This improved the model fit, achieving a log likelihood of -874.17 (AIC 1766.3), a better fit than for the S-VDDM and approaching the performance of the D-VDDM. For completeness, we also tested a version of the D-VDDM with separate σ parameters for its two accumulator units, achieving a log likelihood of -924.66 (AIC 1863.3), i.e., a worse fit than the single-σ D-VDDM. This is clearly a local optimum, since the better-performing single-σ D-VDDM is actually present in the parameter search space of the dual-σ VDDM (along the line where both σ are equal).

Discussion

Here we explored the ability of variable-drift diffusion models (VDDMs) to capture complex sensorimotor decisions based on a continuous stream of multiple sensory cues. Our initial hypothesis was that a complex model consisting of several parallel VDDM processes (the C-VDDM) would be needed to capture the multimodal decision time distributions exhibited by humans in the zebra crossing situation. Instead, we found that a relatively simple model with just a single VDDM unit (the S-VDDM) and five free parameters was able to reproduce multimodal probability distributions of human crossing times, across 15 separate scenarios with a diverse range of vehicle approach

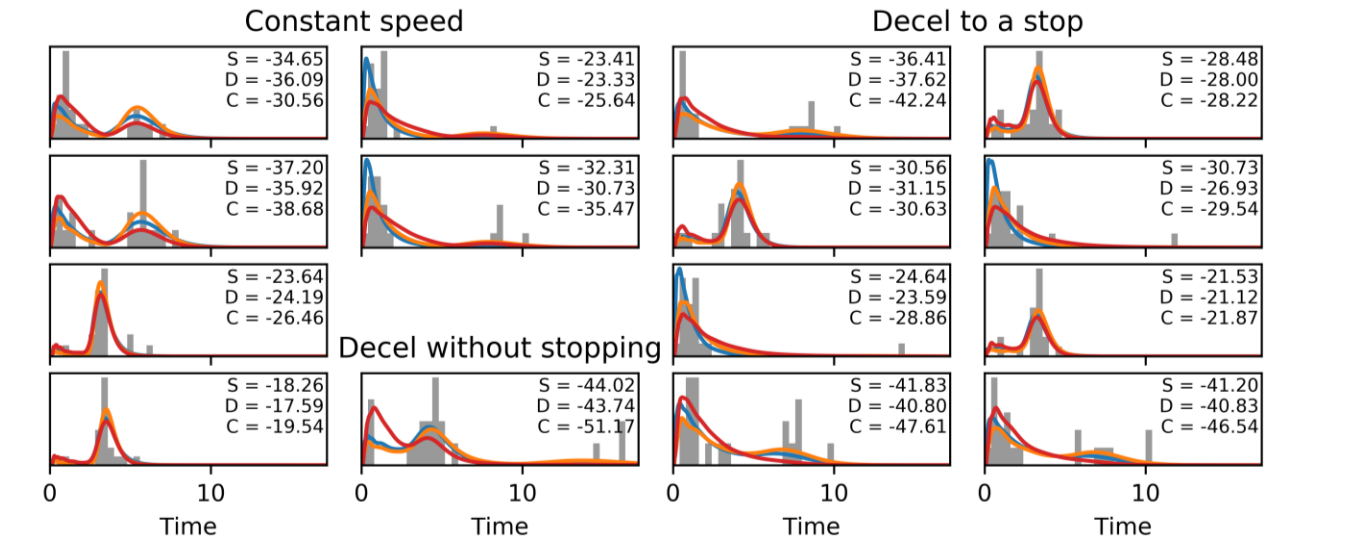

Figure 3: Observed crossing times versus predictions by the C-VDDM (red lines), D-VDDM (orange), and S-VDDM (blue) for all scenarios. Y-axis scale varies between panels. The text shows density estimate log-likelihoods for the three models.

trajectories. This is arguably the most striking finding from this work.

One important insight here, and seemingly a main reason behind the good performance of the S-VDDM, is that the τ variable (the apparent time to arrival; TTA), can in itself help explain the observed human behavior to a large extent. As seen in Figure 2, with more positive $\tau - \tau_{pass}$, participants became more likely to initiate crossing, and the non-trivial variation of τ over time during each scenario seemed to drive the number and location of peaks in the crossing time distribution. The VDDM provides a potential mechanistic explanation for how the observed crossing time distribution arises from this time-varying perceptual input.

Another critical aspect of the S-VDDM model was that while the drift rate was allowed to vary as a function of the perceptual inputs, we also limited its magnitude with a saturation threshold parameter. This ensured that large inputs, arising when the vehicle decelerated to a stop, did not result in the drift rate immediately trending to a very large value. This enabled the model to capture the distribution of crossing times that are observed after a vehicle comes to a stop or passes.

With respect to the more complex model variants, it is difficult to draw firm conclusions from the present results. If the C-VDDM model had been able to capture qualitative aspects of the human data that the S-VDDM was unable to, this could have been taken as tentative evidence for the C-VDDM's hypothesized partition of the decision process into constituent perceptual and action decisions. However, since the best version of the C-VDDM, with three separate σ parameters, simply improved the goodness of fit without changing the qualitative nature of the model behavior, it cannot be excluded that the added model complexity simply led to overfitting to the present data. To further investigate whether there is some merit to the hypotheses behind the C-VDDM, larger datasets with even more diverse scenarios would be useful, and more stringent methods than AIC for controlling for overfitting, such as hold-out validation on parts of the dataset.

Exactly the same argument applies to the D-VDDM, which was the model for which the overall best fit was obtained. The D-VDDM was adopted here as an intermediate-complexity model, in practice replacing the static input saturation step of the S-VDDM with a time-dynamic accumulator. Again, for the same reasons as mentioned above, further work is needed to shed light on whether the improved fits for this model over the S-VDDM have some theoretical relevance.

These difficulties in drawing conclusions from the fits of the more complex models are exacerbated by the apparent tendency of the PSO algorithm to get stuck in local optima. This was evidenced clearly when the PSO found a provably suboptimal parameterization for the two-σ D-VDDM, but may also be part of the reason for the somewhat surprising finding that the relatively complex single-σ C-VDDM yielded the poorest goodness of fit across all tested models. Existing methods for efficient DDM fitting are based on the conventional assumption of constant drift rate (e.g., Vandekerckhove, Tuerlinckx, & Lee, 2011); good methods for fitting also VDDMs would be a valuable future pursuit.

In summary, we demonstrate that already simple VDDMs are able to capture sensorimotor decision making behavior in a task that is more complex, and arguably of higher applied relevance, than the laboratory decision-making tasks typically modelled with DDMs. We suspect that VDDMs could be applied to a wide range of non-trivial real world sensorimotor decision making tasks, but methodological developments are needed to efficiently and reliably fit these models to data.

Acknowledgments

This work is part of the interACT project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement 723395.

References

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765.

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*(4), 297-338.

Koul, A., Soriano, M., Tversky, B., Becchio, C., & Cavallo, A. (2019). The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. *Cognition*, *182*(October 2018), 213–219.

Lee, D. N. (1976). A theory of visual control of braking based on information about time to collision. *Perception*, *5*(4), 437–459.

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.

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*, *2672*(37), 153–163.

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.

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2017). Agreeing to cross: How drivers and pedestrians communicate. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 264–269). IEEE.

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.

Sandamirskaya, Y., Richter, M., & Schöner, G. (2011, August). A neural-dynamic architecture for behavioral organization of an embodied agent. In *2011 IEEE International Conference on Development and Learning (ICDL)* (Vol. 2, pp. 1-7). IEEE.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*(1), 44–62.

Wahde, M. (2008). *Biologically inspired optimization methods: an introduction.* WIT Press.

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*(March), 114–124.