Nudging human drivers via implicit communication by automated vehicles
Empirical evidence and computational cognitive modeling

Zgonnikov, Arkady; Beckers, Niek; George, Ashwin; Abbink, David; Jonker, Catholijn

DOI
10.1016/j.ijhcs.2024.103224

Publication date
2024

Document Version
Final published version

Published in
International Journal of Human-Computer Studies

Int. J. Human–Computer Studies 185 (2024) 103224

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Nudging human drivers via implicit communication by automated vehicles: Empirical evidence and computational cognitive modeling

Arkady Zgonnikov [a,*], Niek Beckers [a], Ashwin George [a], David Abbink [a], Catholijn Jonker [b]

[a] Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, Mekelweg 2, Delft, 2628CD, The Netherlands
[b] Department of Intelligent Systems, Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Van Mourik Broekmanweg 6, Delft, 2628XE, The Netherlands

ARTICLE INFO

Dataset link: https://osf.io/5hu7e/

Keywords:
Driver behavior
Automated vehicles
Implicit communication
Decision making
Cognitive modeling
Evidence accumulation

ABSTRACT

Understanding behavior of human drivers in interactions with automated vehicles (AV) can aid the development of future AVs. Existing investigations of such behavior have predominantly focused on situations in which an AV a priori needs to take action because the human has the right of way. However, future AVs might need to proactively manage interactions even if they have the right of way over humans, e.g., a human driver taking a left turn in front of the approaching AV. Yet it remains unclear how AVs could behave in such interactions and how humans would react to them. To address this issue, here we investigated behavior of human drivers (N = 19) when interacting with an oncoming AV during unprotected left turns in a driving simulator experiment. We measured the outcomes (Go or Stay) and timing of participants' decisions when interacting with an AV which performed subtle longitudinal nudging maneuvers, e.g. briefly decelerating and then accelerating back to its original speed. We found that participants' behavior was sensitive to deceleration nudges but not acceleration nudges. We compared the obtained data to predictions of several variants of a drift-diffusion model of human decision making. The most parsimonious model that captured the data hypothesized noisy integration of dynamic information on time-to-arrival and distance to a fixed decision boundary, with an initial accumulation bias towards the Go decision. Our model not only accounts for the observed behavior but can also flexibly generate predictions of human responses to arbitrary longitudinal AV maneuvers, and can be used for both informing future studies of human behavior and incorporating insights from such studies into computational frameworks for AV interaction planning.

1. Introduction

Future automated vehicles (AVs) can potentially facilitate mobility, boost economic growth, and improve public health (Clements and Kockelman, 2017; Pettigrew, 2017). However, despite unprecedented efforts into research and development of AVs, the existing technologies struggle to provide intuitive and safe interactions of these vehicles with human drivers, pedestrians, and cyclists (Milford et al., 2019). Addressing this issue requires both technological developments and understanding human behavior in interactions with AVs.

In the technical realm, the need for AVs to interact with human drivers around them stimulated development of new kinds of motion planning and control algorithms — *interaction-aware controllers*. Such controllers set out to generate an optimal motion plan that takes humans into account, typically with the help of a model that is used to predict human behavior (e.g. Sadigh et al., 2018; Schwarting et al., 2019; Jayaraman et al., 2020b; Fisac et al., 2019; Evestedt et al., 2016).

One recurring notion in the literature on interaction-aware controllers is that AVs could proactively influence the interaction instead of just predicting human behavior and then reacting to it. For instance, the controller proposed by Sadigh et al. (2018) generated emergent behaviors in this vein: their simulated AV learned to back down at an intersection to signal yielding and thereby influence the human driver at the other end of the intersection to go first, helping to avoid a traffic conflict. Sun et al. (2018) argued that AV designers should leverage the opportunity to influence the interactions with humans and incorporate the "courtesy" term in the AV optimization criteria. Such prosocial AV behaviors are especially relevant given the often-overlooked ethical challenge: The impact (both positive and negative) of even the most mundane AV decisions could be amplified in case the presence of AVs on the roads is scaled up (Himmelreich, 2018).

Understanding and quantifying impacts of proactive and courteous AV behaviors on the individual humans and society in general necessarily requires studying human behavior in interactions with such

* Corresponding author.
E-mail address: a.zgonnikov@tudelft.nl (A. Zgonnikov).

https://doi.org/10.1016/j.ijhcs.2024.103224
Received 31 January 2023; Received in revised form 22 December 2023; Accepted 8 January 2024
Available online 19 January 2024

AVs. However, engineering contributions proposing interaction-aware controllers have at best demonstrated the feasibility of these controllers in interactions with real humans (e.g. Sadigh et al., 2018; Schwarting et al., 2019), but so far have stopped short of rigorous behavioral experiments investigating human behavior when interacting with them.

Such detailed investigations of human-AV interactions exist in the current literature, but have largely been a part of a separate line of research within the fields of human factors and traffic psychology. In these studies, the behavior of AVs is typically scripted; this enables researchers to constrain the experiments and rigorously investigate reactions of human traffic participants to AV behaviors of interest. Existing human-AV interaction studies have explored a range of scenarios (AV interacting with human drivers (Imbsweiler et al., 2018; Miller et al., 2022; Rettenmaier et al., 2021), pedestrians (Terwilliger et al., 2019; Schmidt et al., 2019; Ackermann et al., 2019; Dey et al., 2019), or cyclists (Ackermann et al., 2021) around them), types of communication (explicit vs implicit Dey and Terken, 2017; de Winter and Dodou, 2022; Lee et al., 2020; Rettenmaier and Bengler, 2021), and specific means of communication (e.g. longitudinal Terwilliger et al., 2019; Miller et al., 2022; Schmidt et al., 2019; Ackermann et al., 2019; Tian et al., 2023, lateral Sripada et al., 2021; Miller et al., 2022, and pitch Bindschädel et al., 2022 maneuvers of AV). As a result, there are numerous insights that can guide AV design and implementation, for instance, in regard to the optimal strategy of handling interaction with a human (Schieben et al., 2019; Schmidt et al., 2019; Ackermann et al., 2019). Still, these studies have mostly focused on scenarios in which *the human has the right of way* and therefore the AV inevitably needs to take action. However, as argued above, AVs might need to interact with humans even when *the AV already has the right of way*. Yet it remains unclear how AVs could behave in such interactions and how humans would react to them.

In this paper, we aimed to pave way towards addressing this research gap. We investigated human behavior in a driving simulator experiment during interactions with a simulated AV. In the unprotected left turn scenario, the AV approached the intersection where the human driver was about to cross its path. The AV had the right of way yet was pre-programmed to execute a "nudging" maneuver, as if it aimed to help the human make a decision. We took inspiration in behavioral economics, where subtle changes in the design of an economic choice could help "nudge" humans' decisions in the right direction (Thaler and Sunstein, 2008). In this paper we use the term "nudge" when referring to a specific AV maneuver involving short longitudinal deceleration (or acceleration), followed by acceleration (or deceleration) of the same duration that brings the vehicle back to its original speed. We focused on this specific maneuver because it can hypothetically help AV proactively influence the behavior of the left-turning human (e.g. with the purpose of reducing uncertainty about that human's actions), while not deviating much from AV's existing motion plan. In addition, our experiment included longer acceleration and deceleration conditions, to understand whether more pronounced changes in AVs longitudinal dynamics lead to stronger effect on participants' decision making.

We hypothesized that, compared to AV moving with the constant speed, the participants' probability of taking the turn before the oncoming AV would increase if that AV performs a deceleration nudge, and decrease for the acceleration nudge maneuver. In addition to testing this hypothesis, we investigated the effect of the chosen AV maneuvers on the time it took the participants to make a decision. We then modeled human response to nudging movements of the AV, building up on a previously proposed cognitive model of left-turn gap acceptance (Zgonnikov et al., 2022). We assessed the model's consistency with the observed human behavior, as well as timing of human responses. Finally, we investigated the models' predictions of human behavior in response to a variety of AV nudge maneuvers that were not included in the experiment, in order to exemplify how cognitive models can help translate empirical findings into computational frameworks for AV interaction planning.

2. Methods

2.1. Summary

We performed a driving simulator experiment in which participants ($N = 19$) repeatedly made a left turn across path of an approaching automated vehicle (Fig. 1). The experiment followed a 2 × 5 within-subject design, manipulating two independent variables, initial time-to-arrival and acceleration profile of the AV. The dependent variables included the decision outcome (Go or Stay), response time, and occurrence of a subjective negative reaction to AV behavior by the participants. We quantified the statistical relationships between the independent and dependent variables using mixed-effects (multilevel) regression models. Finally, we designed and implemented multiple variants of a generalized drift-diffusion model and fitted it to the obtained data. All data and code required to replicate our findings, as well as online supplementary information are publicly available at https://osf.io/5hu7e/.

2.2. Experiment setup

This study was approved by the TU Delft research ethics committee. Nineteen participants (16 M, 3F, age range 22 to 28, all had a driving license) recruited at a university campus took part in the driving simulator experiment, compensated by a €25 gift voucher.

The experiment setup and the task were based on the paradigm reported by Zgonnikov et al. (2022); here we describe only the key aspects of that paradigm and all the deviations from it. The hardware included a 55-inch screen and a commercially available game controller (Logitech G29) comprising a steering wheel and a set of pedals. The software, running on a Windows-based desktop computer, was based on the open-source driving simulator CARLA (Dosovitskiy et al., 2017).

The participants were instructed to drive as they would normally drive in real traffic, following auditory instructions of a navigation system that guided them through a virtual urban environment (a square grid of roads). They had to complete 10 routes, each consisting of 24 left-turn and 6 right-turn trials. The right turns were filler trials and were not analyzed. In the left-turn trials, the ego vehicle driven by the participant approached the intersection where a truck was parked in their lane, blocking the participant's way. Once the participant stopped behind the truck, the truck started driving forward, revealing an oncoming automated vehicle that approached the intersection in the opposite lane. The moment when the oncoming AV appeared in the field of view of the participant was considered the start of the interaction ($t = 0$). During the interaction, the participant had to decide whether to turn left before or after the AV passes the intersection.

The participants were explicitly instructed that the oncoming vehicles they will interact with during the experiment is fully autonomous. The AV was a regular-sized sedan, and did not have any external human-machine interfaces. Both the driver and the passenger seats of the AV were empty. Besides the participant's vehicle, oncoming AV, and the truck initially blocking the participant's view, there were no other vehicles in the environment.

2.3. Conditions

The AV started its motion 80 meters away from the ego vehicle, and was randomly assigned an initial speed of either 17.8 m/s or 14.5 m/s, resulting in initial time-to-arrival (TTA_0) of 4.5 s or 5.5 s, respectively. We included two initial TTA values to investigate the participants' reactions to AV maneuvers in situations with lower and higher baseline probability of accepting the gap.

In all trials, AV moved with the initial speed for the first 0.25 s, but its acceleration profile for 2 s after that was manipulated, resulting in five acceleration conditions (Fig. 2, Table 1). We chose the timing of acceleration changes to be consistent with the typical human response

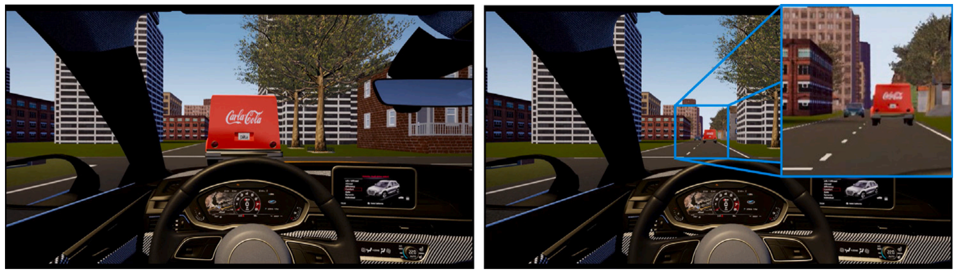

Fig. 1. Participants' view of the task in the driving simulator. As the vehicle driven by a participant approached an intersection, the truck in the participants' lane started moving away, revealing the oncoming car. The participants had to decide whether to go (accept the gap and turn left before the oncoming car), or stay (reject the gap and turn left after the oncoming car passes the intersection). Video recordings of example trials are available in online supplementary information at https://osf.io/5hu7e/.

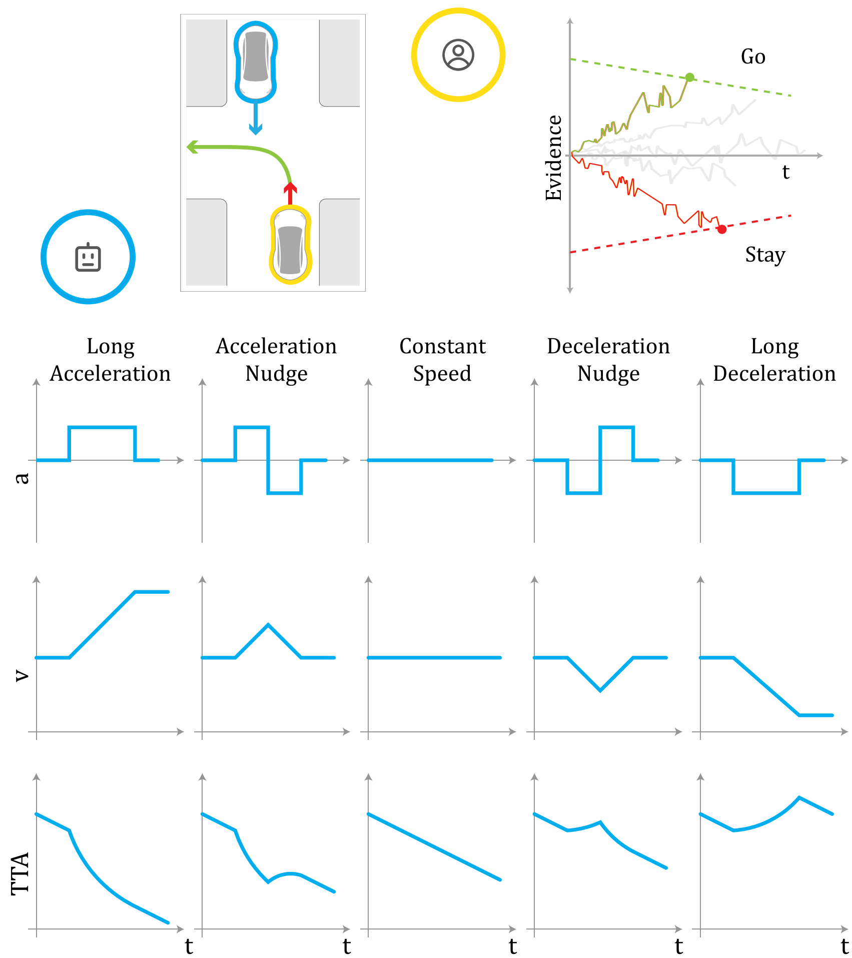

Fig. 2. Left-turn interaction setup, evidence accumulation modeling approach, and the five AV behaviors used in the experiment. The acceleration, velocity, and TTA profiles are schematic; the exact velocity and TTA profiles of the AV were dependent on the initial TTA value.

times in left-turn gap acceptance (on average 0.6 s Zgonnikov et al., 2022): the manipulations had to start early enough in order to have a chance of affecting the decisions. They also had to last long enough and have acceleration magnitudes that are large enough, in order to be noticeable by human participants: this was calibrated in a pilot experiment while keeping acceleration/deceleration rates physically plausible.

The Constant speed condition was a baseline, with AV approaching the intersection sticking to its initial speed. In the Deceleration nudge condition, after 0.25 s of driving with its initial speed the AV decelerated briefly between $t = 0.25$ s and $t = 1.25$ s, resulting in the speed drop of 4 m/s. Then the AV recovered back to its initial speed by accelerating for another second ($t = 1.25$ s to $t = 2.25$ s). The Acceleration nudge condition represented the reverse situation: the AV first accelerated

Table 1
Investigated maneuvers of the automated vehicle: acceleration values (m/s^2) in the five conditions. For the first 0.25 s of the interaction, as well as after 2.25 s, acceleration was zero in all conditions.

Condition	0.25 s to 1.25 s	1.25 s to 2.25 s
Long acceleration	4	4
Acceleration nudge	4	−4
Constant speed	0	0
Deceleration nudge	−4	4
Long deceleration	−4	−4

for 1 s and then decelerated back to its initial speed. In the Long acceleration/deceleration conditions, the AV accelerated/decelerated

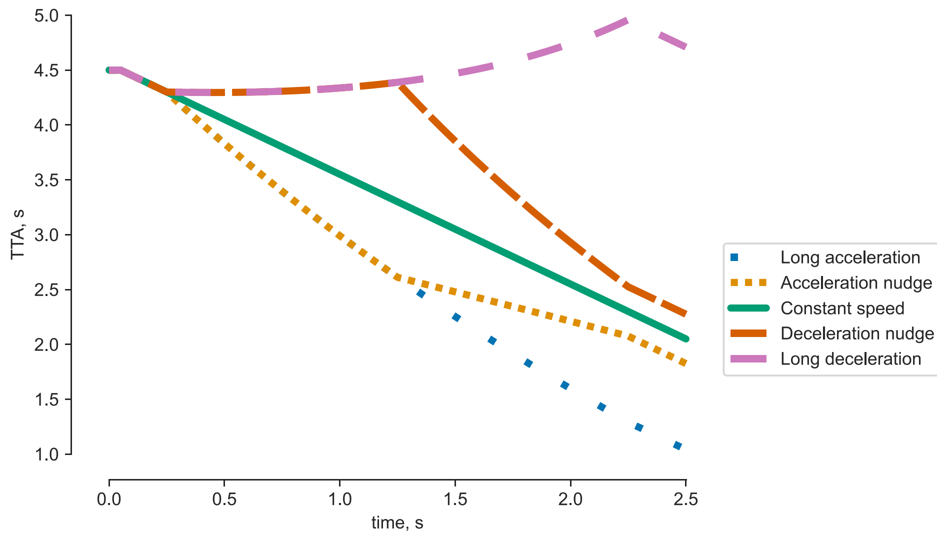

Fig. 3. Dynamics of time-to-arrival (TTA) of the oncoming vehicle over the course of the first 2.5 s of the trial for $TTA_0 = 4.5$ s.

for two seconds, which resulted in speed increase/decrease of 8 m/s by $t = 2.25$ s. In all conditions, the AV had zero acceleration after $t = 2.25$ s.

Note that different acceleration profiles which could potentially result in similar TTA values after the end of the maneuver still produce different TTA dynamics (e.g., acceleration nudge vs deceleration nudge, Fig. 3). We hypothesized this difference over the course of the trial would influence participants' behavior and hence manipulated the acceleration profile and not just the resulting TTA.

Each of the ten combinations of two initial TTA values and five acceleration conditions was repeated 20 times for each participant, randomly shuffled over the 10 routes the 19 participants completed during the experiment. This resulted in 3800 recorded left-turn decisions. In addition, 4 left turns per route included the AVs that either appeared very close to the intersection moving very fast or far from the intersection moving very slow. These extreme conditions were included to ensure that participants take into account the AV kinematics rather than defaulting to either Go or Stay decision; the data from these trials was not analyzed.

2.4. Measures

The first dependent variable of interest was the decision outcome in each trial (Go or Stay), determined based on whether the participant drove through the intersection before or after the AV.

Second, we measured response time for each decision, that is, the time the participant took to arrive to that decision, starting from the moment the AV appeared in their field of view. The end of the decision process was determined differently for Go and Stay decisions. For Go decisions, we assumed that the end of the decision process was marked by the onset of the first accelerator pedal press after the start of the interaction (Zgonnikov et al., 2022). However, when preprocessing the data, we found that only in 70% of the trials the onset of the pedal press could estimate the response time: participants often started pressing the accelerator pedal already while waiting for the truck to leave the intersection, even before they could see the oncoming AV. In these trials (representing 30% of all Go decisions) we therefore could not assess the response time.

For Stay decisions, in natural driving there are no candidate markers of the end of the decision like the accelerator pedal press for Go decisions. For this reason, we asked the participants to press a button on a steering wheel as soon as they decided to Stay. The time of the button press was considered to mark the end of the Stay decision. However, in 11% of Stay decisions participants failed to press the button, resulting in missing response time data.

Finally, we measured participants' subjective perception of the interaction by asking participants to press a "negative rating" button on the steering wheel "*at any time when the AV behavior is confusing, weird, or unintuitive*". We opted for only measuring negative reactions because we did not expect the investigated AV behaviors to induce any strong positive reactions. In addition, having to keep in mind yet another button to press could have further complicated the task for the participants.

2.5. Exclusion criteria

We excluded 24 outlier trials with response times longer than 4 s from all analyses. In addition, Go and Stay trials with missing response times were excluded from statistical analyses of response times and cognitive modeling (but included in statistical analyses of decision outcomes and negative ratings).

2.6. Data analysis

Statistical analyses were performed using mixed-effects regressions (logistic for decision and negative rating, linear for response time) implemented in pymer4 (Jolly, 2018). Decision and negative rating were coded as 0 (Stay/No rating) and 1 (Go/Negative rating). Dummy coding was used for AV maneuver (Table 1), with Constant speed as the reference category. To account for individual differences in baseline values of dependent variables, the participant ID was included as a random intercept in all regressions; regressions with random slopes per participant failed to converge. For the response time regression, we computed Type-III sum-of-squares ANOVA table using the Satterthwaite approximation of degrees of freedom, correcting for multiple comparisons with the Tukey method.

3. Results

We found that decision outcome was affected by both initial time gap (TTA_0) and the AV's approach dynamics (Fig. 4a, Table 2). Compared to the Constant speed condition, probability of a Go decision increased in response to AV's deceleration nudge ($b = 0.5$, $z = 3.4$, $p < 0.001$). We found no evidence of an effect of AV's acceleration nudge ($b = -0.13$, $z = -0.9$, $p = 0.35$) on the decision. Long acceleration and Long deceleration had a negative and positive effect on the probability of Go, respectively (Table 2). Across all acceleration/deceleration conditions, decision outcome depended on TTA_0: the larger the initial

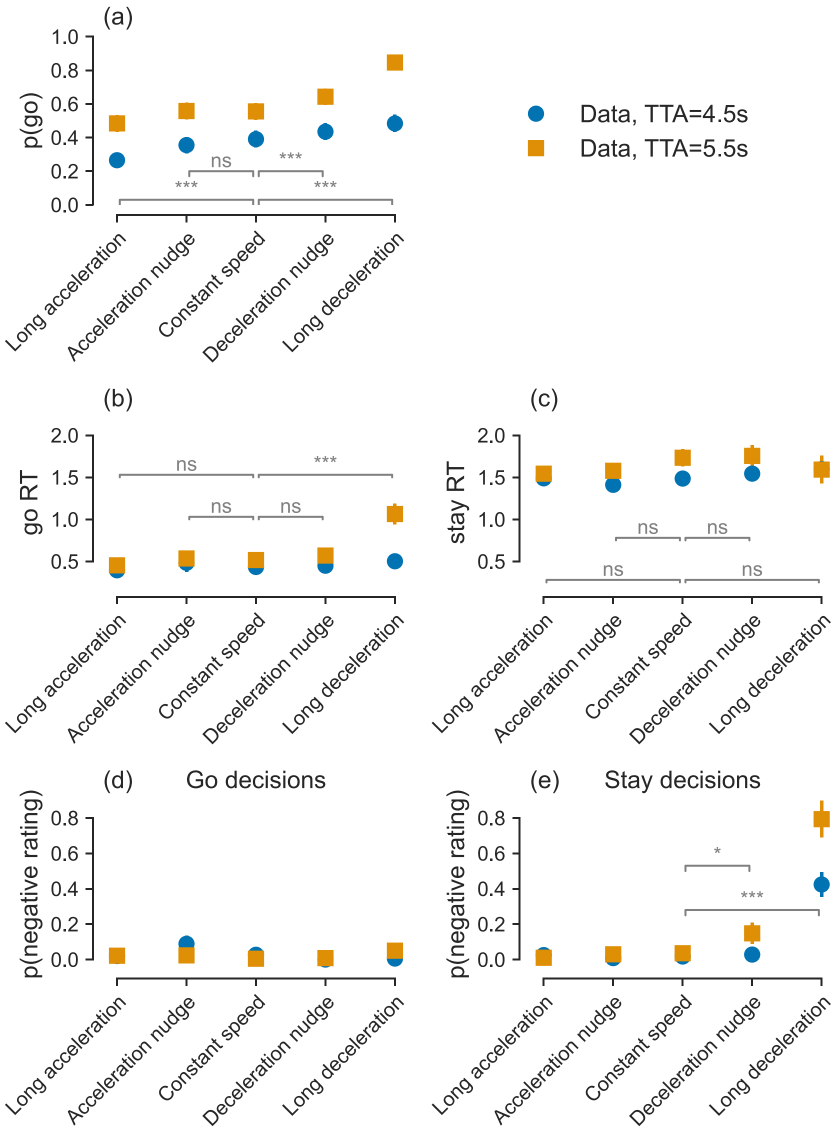

Fig. 4. Effect of condition and TTA_0 on (a) decision outcome, mean response times in (b) Go and (c) Stay decisions, and proportion of trials with a negative rating in (d) Go and (e) Stay decisions. The data was aggregated across all participants. Error bars represent 95% confidence intervals. Results of comparisons between conditions are indicated by brackets: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ns: $p > 0.05$; p-values were calculated based on mixed-effects models reported in the main text.

Table 2
Coefficients of the mixed-effects logistic regression describing the final decision as a function of TTA_0 and condition (with Constant speed as the reference category). Participant ID was included as a random intercept.

	b	SE	*z*	*p*
(Intercept)	−9.30	0.71	−13.10	<0.001
TTA_0	1.82	0.10	17.74	<0.001
Long acceleration	−0.78	0.15	−5.38	<0.001
Acceleration nudge	−0.13	0.14	−0.94	0.35
Deceleration nudge	0.50	0.14	3.44	<0.001
Long deceleration	1.50	0.15	9.89	<0.001

time-to-arrival, the more likely the participants were to accept the gap ($b = 1.82$, $z = 17.8$, $p < 0.001$)).

Analyzing response times, we found evidence for main effects of decision, TTA_0, and acceleration condition, as well as an intricate three-way interaction between them (Fig. 4b,c, Table 3). Post-hoc comparisons revealed that

- Go responses were substantially faster than Stay responses ($\Delta = 1.02$ s, $t = 39.6$, $p < 0.001$).
- There was no evidence that Deceleration nudges differed from the Constant speed condition in terms of response time ($\Delta = 0.001$ s, $t = 0.1$, $p \approx 1$ for Go decisions, $\Delta = 0.05$ s, $t = 1.2$, $p = 0.74$ for Stay decisions).
- Similarly, there was no evidence of response time differences between Acceleration nudge and Constant speed conditions ($\Delta = 0.04$ s, $t = 0.73$, $p = 0.95$ for Go decisions, $\Delta = -0.1$ s, $t = -2.6$, $p = 0.07$ for Stay decisions).
- Long deceleration induced significantly longer Go response times compared to Constant speed ($\Delta = 0.26$ s, $t = 5.8$, $p < 0.001$), although this was accounted mostly by trials with $TTA_0 = 5.5$ s. There was no evidence for a difference in Stay response times between these conditions ($\Delta = 0.04$, $t = 0.8$, $p = 0.93$), as well as any differences in response time between Long acceleration and Constant Speed ($\Delta = -0.04$, $t = 0.7$, $p = 0.96$ for Go trials, $\Delta = 0.08$, $t = 2$, $p = 0.28$ for Stay trials).
- TTA_0 positively affected Go response times in the Long deceleration condition ($b = 0.5$, $t = 9.2$, $p < 0.001$), but we found no evidence of this effect in other conditions, including the Constant speed condition.

Table 3
ANOVA table based on the mixed-effects linear regression describing response time as a function of decision, TTA_0, and condition.

	SS	MS	df	F	p
Decision	5.27	5.27	1	19.54	<0.001
TTA_0	11.79	11.79	1	43.73	<0.001
Condition	3.25	0.81	4	3.01	0.02
Decision: TTA_0	0.03	0.03	1	0.10	0.75
Decision: condition	7.71	1.93	4	7.15	<0.001
Condition: TTA_0	4.55	1.14	4	4.22	0.002
Decision:condition: TTA_0	8.41	2.10	4	7.80	<0.001

Table 4
Coefficients of the mixed-effects logistic regression describing occurrence of the negative rating as a function of TTA_0 and condition (with Constant speed as the reference category). Participant ID was included as a random intercept.

	b	SE	z	p
(Intercept)	−6.53	0.95	−6.90	<0.001
TTA_0	0.54	0.17	3.14	0.002
Decision	−2.67	0.25	−10.60	<0.001
Long acceleration	−0.15	0.37	−0.39	0.69
Acceleration nudge	0.50	0.34	1.46	0.15
Deceleration nudge	0.73	0.33	2.20	0.03
Long deceleration	3.42	0.30	11.26	<0.001

- Stay response times were positively affected by TTA_0 in the Acceleration nudge ($b = 0.18$, $t = 3.2$, $p < 0.001$), Constant speed ($b = 0.23$, $t = 4.1$, $p < 0.001$), and Deceleration nudge ($b = 0.21$, $t =$, $p < 0.001$) conditions.

Finally, we found that participants gave negative ratings overwhelmingly to AVs that performed Long deceleration, and mostly in Stay decisions (Fig. 4d,e, Table 4); more negative ratings were given in the larger TTA_0 condition ($b = 0.54$, $z = 3.1$, $p = 0.002$). Participants perceived Deceleration nudge slightly more negatively than Constant speed ($b = 0.73$, $z = 2.2$, $p = 0.03$).

Based on these findings, we conclude that

- Deceleration nudges increased the probability of Go decision. There was no evidence they required more or less time for the participants to make up their mind, compared to the Constant speed condition. In a small fraction of Stay trials, Deceleration nudges were negatively rated.
- Compared to Deceleration nudge, the Long deceleration maneuver further increased the probability of going, but it caused participants to spend significantly more time to arrive to a Go decision, and induced by far the strongest negative reaction from the participants.
- We found no evidence that the Acceleration nudge actually pushed the participants towards staying, but it did make their Stay decisions marginally faster. At the same time, Long acceleration did reduce the probability of going, compared to the Constant speed condition.

4. Cognitive process modeling

4.1. Basic drift-diffusion model

To gain a deeper insight into the experimental findings, we sought a process-level explanation of the observed behavior through cognitive modeling. We used the drift-diffusion modeling framework which is based on the notion of evidence accumulation (Fig. 2, top-right panel). Evidence accumulation models assume that humans integrate relevant perceptual information over time. This accumulation is noisy and happens until the amount of evidence in favor of either alternative reaches a certain boundary. Evidence accumulation models, despite their seeming simplicity, have been remarkably successful in explaining a variety of behavioral effects, including intricate relationships between the decision context, outcomes of the decision, and response times (Gold and Shadlen, 2007; Ratcliff and Smith, 2004; Ratcliff et al., 2016). For this reason, evidence accumulation has been one of the dominant paradigms of modeling human decision making in cognitive neuroscience and psychology over the past few decades.

The drift-diffusion model (Ratcliff, 1978; Ratcliff et al., 2016) is one of the simplest evidence accumulation models that represents the process of making a choice between two options as a random process

$$\frac{dx}{dt} = s(t) + \epsilon(t), \tag{1}$$

where $s(t)$ is the drift rate (the momentary evidence in favor of one option over the other, sampled from the environment at time t), $\epsilon(t)$ is the diffusion rate (random noise), and x is the evidence accumulated by time t as a result of the joint effect of drift and diffusion. The accumulation process starts at an initial position $x(0) = x_0$ and continues until x crosses either an upper decision boundary $x = b(t)$ or a lower boundary $x = -b(t)$ (each corresponding to one of the choice alternatives). Finally, drift-diffusion models typically account for cognitive processing not related to decision making (e.g. perceptual and motor delays, premature responses) by adding non-decision time to the response time produced by the evidence accumulation process. Here, we hypothesized that non-decision time is normally distributed: $t^{ND} \in \mathcal{N}(\mu_{ND}, \sigma_{ND})$.

Recently, drift-diffusion models have been applied to human decisions in traffic, e.g. gap acceptance in pedestrian crossing (Pekkanen et al., 2022; Markkula et al., 2023), unprotected left turns (Zgonnikov et al., 2022), and overtaking (Mohammad et al., 2023). The key assumption of these versions of the model is that the drift rate tracks the dynamically changing size of the gap, relative to a certain critical value. The decision boundary is usually assumed to be fixed (Ratcliff et al., 2016; Pekkanen et al., 2022) but can also be collapsing over time to reflect choice urgency (Zgonnikov et al., 2022; Drugowitsch et al., 2012).

4.2. Candidate variants of a drift-diffusion model

Our experimental setup was similar to that of Zgonnikov et al. (2022), hence we used the left-turn gap acceptance model they proposed as a basis for a cognitive process account of our findings. However, due to a number of changes introduced in our experiment compared to theirs, we investigated several variants of that model as potential candidate models. These variants differed based on their assumptions about three components of the model: (1) drift rate (whether or not it includes a separate acceleration term), (2) initial bias of the accumulator (starting at $x = 0$ or with a bias towards one of the options), and (3) decision boundary (whether it stays constant or collapses over time). We then tested eight configurations of the model resulting from all possible combinations of these three binary design choices.

4.2.1. Drift rate: Without or with acceleration term?

The original left-turn gap acceptance model (Zgonnikov et al., 2022) assumes that the drift rate in the drift-diffusion model (Eq. (1)) depends on the dynamically changing distance d and time-to-arrival TTA:

$$s(t) = \alpha\Big(\mathrm{TTA}(t) + \beta_d d(t) - \theta_{\mathrm{crit}}\Big), \tag{2}$$

where α, β_d, and θ_{crit} are free parameters. In Eq. (1), positive values of the accumulator state x are then associated with accumulated evidence for the Go decision (so negative $x(t)$ means the decision maker leans more to the Stay decision at time t). Then intuitively, the larger the gap to the oncoming vehicle (a weighted combination of TTA and d), compared to some critical value θ_{crit}, the more positive the drift rate becomes, and the more likely the decision maker is to arrive to the Go decision.

The study of Zgonnikov et al. (2022) only included constant-speed oncoming vehicles. In our data however, most trials included the oncoming vehicle that varied its speed by accelerating or decelerating. A model including a TTA-sensitive drift rate can potentially capture human behavior in response to such accelerating/decelerating AV simply through the associated changes in TTA. However, one might also argue that human decision makers can benefit from estimating acceleration in addition to TTA. Indeed, previous studies have shown that a separate term for acceleration in the drift rate can help a drift-diffusion model better capture human responses to AV's apparent yielding behavior in a pedestrian crossing scenario (Giles et al., 2019; Pekkanen et al., 2022). It is unclear though to what extent information on acceleration is accumulated by humans on top of TTA in our scenario, given that it involves a different kind of task, with larger distances, higher speeds, and lower accelerations of the oncoming vehicle compared to Giles et al. (2019), Pekkanen et al. (2022). Hence, we tested models with two drift rate variants: *Without acceleration term* (Eq. (2)) and *With acceleration term*:

$$s(t) = \alpha\Big((\mathrm{TTA}(t) + \beta_d d(t) - \theta_{\mathrm{crit}}) - \beta_a a(t)\Big). \tag{3}$$

Here the term $\beta_a a(t)$ negatively contributes to the drift rate, reflecting the assumption that positive acceleration of the AV would increase the probability of reaching the lower boundary (Stay decision), and negative acceleration would push the decision maker towards the upper boundary (Go).

4.2.2. Initial state of the accumulator: Without or with bias?

Zgonnikov et al. (2022) only measured response times in Go but not Stay decisions. Here we addressed this limitation, and measured Stay response times, which we found to be substantially longer compared to Go response times. We hypothesized this could have resulted from initial bias in the accumulation process which can reflect the presence of a default option. In our experiment, participants were instructed to turn left on every trial, so they might have been cognitively predisposed towards Go which would then explain shorter response times in Go decisions. The cognitive modeling approach enabled us to test this explanation by including an initial bias in the model and examining if this would allow the model to better describe the data.

In one version of this model component, the *initial bias* $x(0) = x_0$ is a free parameter expressed as a fraction of the decision boundary value, taking values between −1 (extreme bias towards Stay) and 1 (extreme bias towards Go). In the second version, there is *no initial bias*, that is, accumulation always starts at $x(0) = 0$ (which was also the case for the model reported by Zgonnikov et al. 2022).

4.2.3. Decision boundary: Constant or collapsing?

Zgonnikov et al. (2022) emphasized the positive effect of TTA_0 on Go response times as one of their main findings: the longer the time budget drivers have, the more time they are taking to make the Go decision, despite having more clear evidence in favor of Go. In terms of cognitive processes, Zgonnikov et al. (2022) explained this effect by having the decision boundary in the drift-diffusion model collapsing with $\mathrm{TTA}(t)$

$$b(t) = \pm b_0/(1 + e^{-k(\mathrm{TTA(t)}-\tau)}). \tag{4}$$

This collapsing boundary requires three free parameters to provide flexibility in the boundary baseline value at $t = 0$ (b_0), rate of collapsing (k), and characteristic scale of TTA at which collapsing occurs (τ). Despite its relative complexity, this boundary was shown to be essential for capturing the behavior observed by Zgonnikov et al. (2022), as the model with a constant boundary could not capture Go response times increasing with TTA_0. However, in our study we did not find evidence for a relationship between TTA_0 and Go response times in all conditions but one (Long deceleration). For this reason, we aimed to test two versions of the decision boundary in our models: the *constant boundary* ($b(t) \equiv \pm B$) and the *collapsing boundary* (Eq. (4)).

4.3. Model fitting and evaluation

We performed model selection between eight variants of the model based on qualitatively assessed fit to our data, as well as comparison of Bayes Information Criterion (BIC) values. The simplest model including no acceleration term, no initial bias, and constant boundary (Model 1) is characterized by 6 free parameters, while the most complex model with acceleration term, initial bias, and collapsing boundary (Model 8) has 10 free parameters. The model reported by Zgonnikov et al. (2022) has 8 free parameters, and here is represented as Model 3.

We aimed to investigate whether models can describe the behavior of the "average" participant, and therefore fitted each model to the data obtained from all 19 participants disregarding participant IDs. Fitting the model to individual participant's data is also possible and can provide additional insights into potential individual differences in cognitive processes, but deserves a separate investigation beyond the scope of this work.

To evaluate how well the models generalize to unseen conditions, we fitted the models to a subset of the data that included only Constant speed, Long acceleration, and Long deceleration conditions (thereby excluding the two nudge conditions). We performed posterior predictive checks on the same three conditions, as a way of confirming whether models did in fact fit the data used for parameter estimation. We then evaluated models' generalization to the two unseen conditions (Acceleration nudge and Deceleration nudge). Go and Stay trials with missing response times had to be excluded at this stage, because they cannot be handled by the existing model fitting frameworks.

After excluding trials with missing response time data and the hold-out conditions, the models were fitted to the remaining 1782 trials (out of 3776 trials with non-outlier RTs). The models were then evaluated on the dataset without the missing response time data (therefore, data estimates in Fig. 5 are slightly different from those in Fig. 4) but with the held-out nudge trials, which in total amounted to 2987 trials. Model fitting was performed via differential evolution optimization of Bayesian Information Criterion as implemented in pyddm, a framework for drift-diffusion model fitting in Python (Shinn et al., 2020).

4.4. Modeling results

We found that all candidate models captured the following qualitative patterns we observed in human behavior (Figs. 5 and S1 in online supplementary information):

- Probability of Go increased with TTA_0, and increased monotonically as the AV dynamics became more assertive (lowest p(Go) in the Long acceleration condition, highest p(Go) in the Long deceleration condition).
- Go response times were much shorter than Stay response times.
- Go response times were substantially longer in Long deceleration at $\mathrm{TTA}_0 = 5.5$ s, compared to other conditions.
- Stay response times increased with TTA_0 in the two nudge conditions and the Constant speed condition.

Despite the qualitative similarities, the models did differ in their ability to parsimoniously describe the data, as quantified by BIC (Table 5). The presence of initial bias had the strongest effect on BIC, with models including a free bias parameter consistently having substantially lower BIC than their counterparts without the bias parameter. Inclusion of the acceleration term in the drift rate or a more flexible boundary had almost no effect on descriptive capacity of the model, and even led to higher BIC values due to the added complexity in terms of the number of parameters. Upon inspection of fitted model parameters, the values of β_a for models with the acceleration term and k for models with collapsing boundary turned out to be very small (online supplementary information Table S1), further indicating that these extensions do not benefit the model. For this reason, in what follows we focus on three candidate models: the simplest one (Model

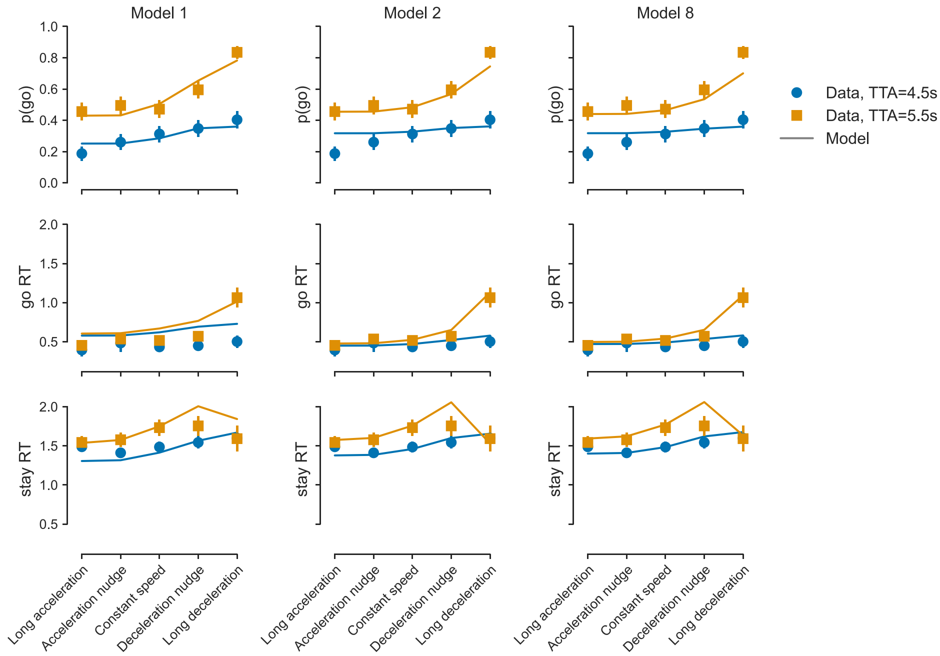

Fig. 5. Behavior of three candidate models compared to participants' behavior. The simplest model, Model 1, has no initial bias, no acceleration term, and constant boundaries. The model with the lowest BIC, Model 2, has initial bias, no acceleration term, and constant boundaries. The most complex model, Model 8, has initial bias, acceleration term, and collapsing boundaries. The models were fitted to the subset of data including only the Long acceleration, Constant speed, and Long deceleration trials, while the Acceleration nudge and Deceleration nudge conditions were held out to test generalization of the models to unseen scenarios. The data depicted here excludes the trials with missing response times (mostly Go trials), hence lower probabilities of Go compared to Fig. 4a. Other tested models are illustrated in online supplementary information (Figure S1).

Table 5
Eight tested model configurations with their number of free parameters (κ) and resulting Bayes Information Criterion (BIC) values (lower values correspond to models more consistent with the data, accounting for the number of free parameters). Model 2 is highlighted as the model with the lowest BIC value; the fitted parameters of Model 2 were $\alpha = 0.55$, $\beta_d = 0.005$, $\theta_{\mathrm{crit}} = 6.8$, $B = 1.3$, $\mu_{\mathrm{ND}} = 0.19$, $\sigma_{\mathrm{ND}} = 0.17$, $x_0 = 0.67$.

	Without acceleration term		With acceleration term	
	No initial bias	With initial bias	No initial bias	With initial bias
Constant boundary	Model 1 $\kappa = 6$ BIC = 5050	**Model 2** **$\kappa = 7$** **BIC=4610**	Model 5 $\kappa = 7$ BIC = 5063	Model 6 $\kappa = 8$ BIC = 4620
Collapsing boundary	Model 3 $\kappa = 8$ BIC = 5072	Model 4 $\kappa = 9$ BIC = 4622	Model 7 $\kappa = 9$ BIC = 5090	Model 8 $\kappa = 10$ BIC = 4638

1), the most complex one (Model 8), and the model providing the best trade-off between model performance and complexity, as measured by BIC (Model 2). The behavior of Models 3, 5, and 7 is similar to Model 1; Models 4 and 6 are similar to Models 2 and 8 (online supplementary information Figure S1).

Models without the initial accumulator bias (e.g., Model 1) captured decision probabilities better than models with the bias, but overestimated Go response times by approx. 0.3 s (a substantial difference given the mean Go response times about 0.6 s). These models however yielded implausible estimates of non-decision time (μ_{ND} close to 0, σ_{ND} approx. 0.25 s) which hints at model misspecification. In contrast, models with the initial bias towards the Go decision (e.g., Models 2 and 8) closely captured Go response times, but underestimated the effect of Long acceleration (Long deceleration) in the lowest (highest) TTA_0 on probability of Go. These models revealed strong initial bias towards the Go decision (65% to 69% of the initial boundary value). Interestingly, predictions of all models overestimated Stay response times in the Deceleration nudge condition (by approx. 0.3 s to 0.4 s depending on the model, with the average response time in this condition being 1.6 s). Overall, given much lower BIC for models with the initial bias, we conclude that Model 2 (drift-diffusion model with no acceleration term, initial bias, and constant boundary) provides the most parsimonious and plausible explanation of our data among the eight tested models. Crucially, it captured all qualitative relationships between AV motion dynamics and decision outcomes and response times without "seeing" any data from the nudge conditions, with the only quantitative discrepancy in these conditions being overly long Stay response times in the Deceleration nudge condition.

For a more detailed picture of the behavior of the model and its fit to the data, we analyzed full response time distributions generated by Model 2. We found that the model closely matched 19 of the 20 experimentally measured response time distributions (Fig. 6). Particularly notable is the model's ability to capture the bimodal response time distribution observed in Go response times in the Long deceleration condition at $\mathrm{TTA}_0 = 5.5$ s. Because in this condition the early TTA cue favors the Go decision, majority of Go decisions were made within the first second of the trial. But there were relatively rare cases in which, due to noise, the initially accumulated evidence favored the Stay decision. In such trials, the stronger late-coming evidence (acceleration cue in combination with increasing TTA) pushed evidence towards the Go decision after all, which altogether takes a much longer time than a typical decision, resulting in the second mode in the distribution.

Interestingly, Model 2 predicted that the response time would also be bimodally distributed in Stay decisions for Deceleration nudges at $\mathrm{TTA}_0 = 5.5$ s. The likely reason for this atypical distribution in the model is that it predicted the participants to be less likely to commit to

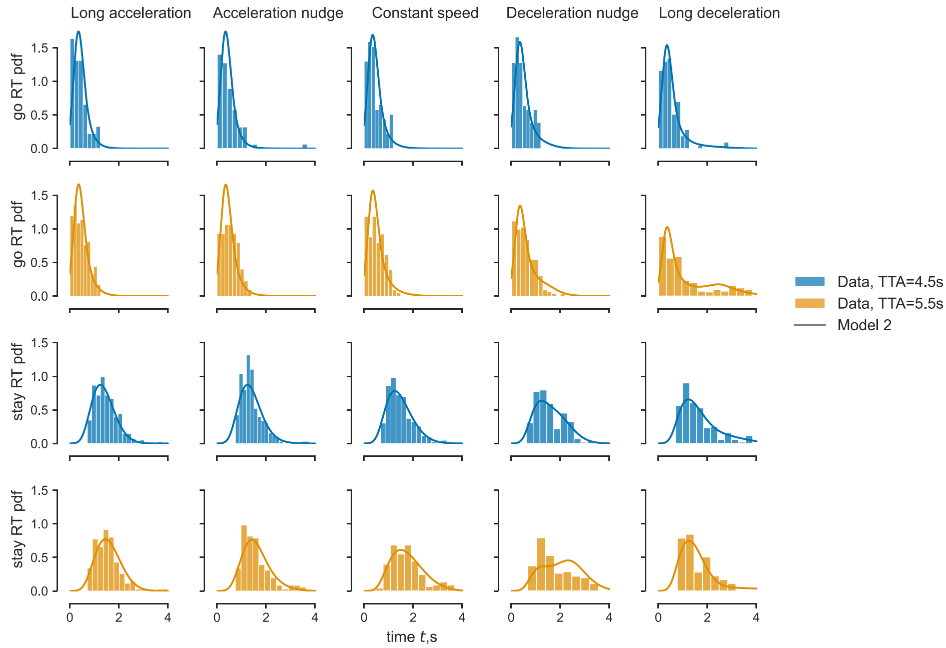

Fig. 6. Full response time distributions produced by Model 2 compared to the data. Similar to Fig. 5, Acceleration nudge and Deceleration nudge conditions were withheld during model parameter estimation.

their Stay decisions during the initial phase of the Deceleration nudge (as evidence in favor of Staying at $TTA_0 = 5.5$ s would be low, further lowered by AV's deceleration). Then, as the TTA started decreasing after the AV started accelerating in the latter phase of Deceleration nudge, the evidence in favor of Staying would increase again, leading to an increase in the likelihood of arriving to the Stay boundary around 2 s in the decision. This however did not match the unimodal distribution observed in the data, suggesting that participants often committed to the Stay decision early on, despite the evidence supporting Stay decision being comparatively low at that time.

Taken together, our modeling results strengthen the existing notion that noisy accumulation of evidence coming from relevant dynamic perceptual variables (distance and TTA) is a key mechanism underlying human tactical decisions in traffic (Pekkanen et al., 2022; Zgonnikov et al., 2022; Markkula et al., 2023; Mohammad et al., 2023). They also highlight that changes in acceleration of the oncoming vehicle could be processed via their effect on TTA rather than independently.

5. Using the drift-diffusion model to predict human responses to new AV maneuvers

Importantly, our model can take arbitrary longitudinal dynamics of the oncoming vehicle as an input ($TTA(t)$, $d(t)$, $a(t)$), and therefore does not rely on the exact configuration of the AV maneuvers used in the experiment. The model, even if fitted to a limited dataset like ours, can then generate testable predictions of human left-turn decision-making not only for different initial values of d and TTA (which is what most existing left-turn gap acceptance models can do), but also for arbitrary approach dynamics of the oncoming AV.

Here we illustrate the model's flexibility by simulating it under the conditions not included in our experiment. We explored the predictions of the model (using Model 2 with the parameters fitted to our data; see the caption of Table 5) for new initial conditions $d = 90$ m and $TTA = 6$ s and a multitude of variations of the deceleration nudge maneuver. We independently varied the duration (0.1 to 2.5 s) and the magnitude (0.5 to 5 m/s^2) of the initial deceleration, which was immediately compensated by acceleration of the same duration and magnitude. Similar to the experiment, the maneuver was initiated at 0.25 s after the start of the interaction.

The key prediction of the model is that increasing the duration and magnitude of AV's deceleration nudge increases the probability of the Go decision by the human (Fig. 7a) and simultaneously increases the response times in both Go and Stay decisions (except for extreme nudges, see below). This prediction is to be tested in future experiments.

Importantly, although somewhat unintuitively, the predicted probability of Go did not reach 100% even for extreme deceleration nudges, peaking at 98% in the scenario with the strongest nudge (the oncoming vehicle slowing down from 15 m/s to 2.5 m/s in 2.5 s). One would expect human drivers to choose Go 100% of the time if the oncoming vehicle almost comes to a full stop, even if they initially inclined to Stay; yet the model still predicts a fraction of Stay decisions. This can be explained by the scope of the model being limited to just the first decision resulting from evidence accumulation and not affording the possibility of a subsequent change-of-mind. In our simulations, even for strongest nudges the Stay boundary is occasionally reached; this typically happens early on in the accumulation process. In the current version of the model, this means that the Stay decision is finalized, and the evidence accumulation is terminated. However, in real interactions the drivers presumably continue accumulating evidence, and could change their mind after an initial Stay decision in case the new evidence favors the Go decision. Addressing this limitation would allow the model to account for finer details of human left-turn behavior, but requires substantial further research and is therefore not covered in this paper.

Overall, our simulations of the model estimate the expected consequences of deceleration nudge maneuvers with varying magnitude and duration. This exemplifies how the model can generate testable quantitative predictions of human behavior for arbitrary AV approach trajectories.

6. Discussion

Future automated vehicles will face a technically and ethically challenging problem: how to behave in seemingly non-safety-critical interactions with humans when the AV clearly has the right of way. A pedestrian who can unexpectedly start to jaywalk, or a driver trying to merge onto a highway who might suddenly decide to take their chances with a small gap in front of an AV — both of them can turn the situation

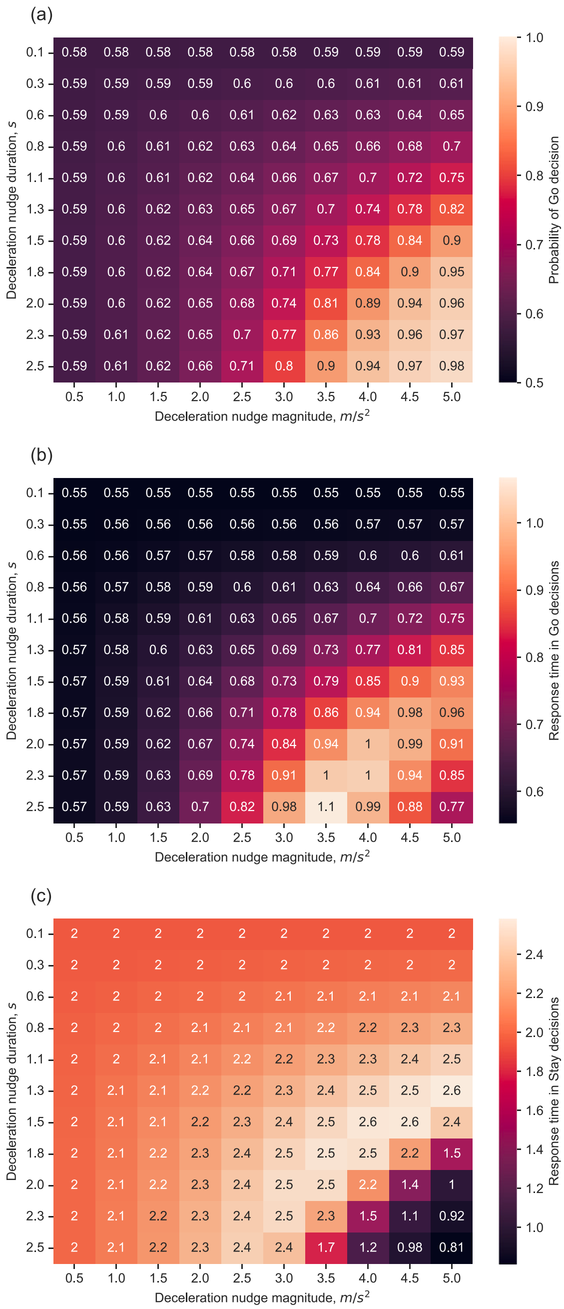

Fig. 7. Model predictions: Decision outcome (a) and response times in Go (b) and Stay (c) decisions as a function of duration and magnitude of a deceleration nudge. Predictions were generated by simulating the behavior of Model 2 for $d_0 = 90$ m and $\text{TTA}_0 = 6$ s.

from mundane to safety-critical in a fraction of second. Can AVs do something to reduce uncertainty about future human behavior, but without sacrificing much of its own plans? In this paper, we explored the idea of AV subtly influencing the left-turning human driver by adjusting its longitudinal motion. In a driving simulator experiment, we found that human drivers were more likely to go in front of the oncoming AV when that AV performed a deceleration nudge. At the same time, acceleration nudge did not lead to a marked change in human behavior. Our observations of participants' decisions, their timing, and the relationship between those and the AV dynamics were explained by a generalized drift-diffusion model. The modeling results suggest that human drivers integrate the noisy dynamic information on time-to-arrival and distance to a constant boundary, with an initial bias towards the Go decision. We believe that this model provides a promising framework for quantifying human responses not just to arbitrary nudges but a wide class of longitudinal AV behaviors in this kind of interactions. The ability of the model to generate predictions of human behavior in these interactions makes the model a potentially useful tool for training and testing of AVs in virtual simulation environments, as well as guiding the development of AV interaction planning frameworks.

6.1. Translating the concept of nudging to human-AV interactions in traffic

Our findings resonate with the recent literature on implicit human-AV communication. This literature emphasized longitudinal motion cues by AV as an important communication signal for humans that make a decision whether or not to cross paths with that AV (Ackermann et al., 2019; Rettenmaier et al., 2021; Terwilliger et al., 2019; Schmidt et al., 2019; Dey et al., 2019; Beggiato et al., 2018; Tian et al., 2023). Such studies however have so far focused almost exclusively on overt yielding maneuvers: either constant deceleration to a full stop, or prolonged deceleration to a low velocity (similar to our Long deceleration condition). Exceptions include investigations of adversarial AV movements that were deliberately designed to be confusing or malicious (Schmidt et al., 2019; Rettenmaier et al., 2021). The novel contribution of our paper to this body of literature lies in its focus on deliberately subtle AV maneuvers. Such maneuvers have potential to nudge other traffic participants without major changes in AV's own motion plan, potentially yielding benefits in interaction outcomes while minimizing impact on AV occupants' comfort.

Our work was loosely inspired by the idea of nudging from behavioral economics. Thaler and Sunstein (2008) define nudge as: "*any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid.*" In this exploratory paper, we aimed to test the effect of specific nudge-inspired AV behaviors on human decisions. Our results present preliminary evidence that such behaviors could be a useful strategy for proactively influencing behavior of human road users by AVs. However, we believe that further steps along these lines should necessarily involve generalizing our results through translating the concept of nudging to the AV realm in a more rigorous and systematic way. The reason for this is that many kinds of AV behaviors can be conceptualized as nudges, not only the specific longitudinal dynamics introduced here. For instance, in the context of the left-turn interaction, a brief lateral maneuver, or even a mere headlight blink by a vehicle that has the right of way can be considered a nudge. Future studies of human responses to other kinds of nudges need to be accompanied by relating them to theoretical frameworks of interactions in traffic (e.g. Markkula et al., 2020) to ensure generalizability and applicability of this concept in the real world. Computational cognitive modeling studies like ours, in turn, have potential to contribute towards such theoretical frameworks (Guest and Martin, 2021; Devezer and Buzbas, 2023).

6.2. Cognitive processes underlying responses to AV behaviors

This work not only investigated humans' responses to AV nudges, but also elucidated the cognitive mechanisms behind these responses. There is mounting evidence that human tactical behaviors in traffic interactions exploit the evidence accumulation mechanism (Giles et al., 2019; Pekkanen et al., 2022; Markkula et al., 2023; Zgonnikov et al., 2022; Mohammad et al., 2023), which has also been implicated in a wide range of other safety-critical decision-making tasks (Boag et al., 2023). Our results further reinforce evidence accumulation as the key mechanism underlying humans' judgments of gaps in traffic. We also compared multiple variants of the drift-diffusion model to investigate the relative importance of several aspects of evidence accumulation for explaining our results.

Based on a comparison of eight candidate models, we found that collapsing boundaries do not play a major role in explaining our data. This is in contrast to the study of Zgonnikov et al. (2022) who found collapsing boundaries to be critical for explaining the positive effect of TTA_0 on Go response times they observed in left-turn gap acceptance decisions. However, except for the Long deceleration condition, we did not find any evidence for such relationship between TTA_0 and Go response times, hence the lack of support for collapsing boundaries. There are two important differences between our experiment and modeling approach and those of Zgonnikov et al. (2022) which could explain this discrepancy. First, their experiment included longer initial distance conditions (90 m to 150 m, compared to 80 m in our study) which could have an impact on time pressure perceived by the participants, even though distance objectively does not influence the time budget for the decision. Second, Stay response times were not measured by Zgonnikov et al. (2022). We believe that in our study these response times more tightly constrained model fits, leading to more reliable estimation of the boundary parameters. Likely due to a combination of these two differences, our conclusion on collapsing boundary contradicts their findings. Therefore, further research is needed to understand the role of collapsing boundaries in evidence accumulation during traffic interactions.

We also found that including a separate acceleration term in the drift part of the model was not necessary for describing human responses to accelerating/decelerating maneuvers of the AV. The latter were described well already by simpler models integrating only dynamic TTA and distance cues. This suggests that our participants did not use acceleration motion cues over and above the variations in TTA. This is in contrast to previous studies on pedestrian crossing decisions (Giles et al., 2019; Pekkanen et al., 2022; Markkula et al., 2023; Tian et al., 2023) which emphasized the importance of independent acceleration-related cues (namely, time derivative of TTA). This mismatch could be explained by a) a relatively low visual fidelity of our driving simulator, and b) the longer initial distance we used (80 m is larger than the distances typically investigated in pedestrian crossing scenarios). Both these factors could have impaired visual perception of acceleration in our participants, leading to their low reliance on recognizing AV acceleration/deceleration. Another potential explanation is that acceleration cues might influence the evidence accumulation process in a non-linear way. For instance, subtle and short maneuvers might not be consciously perceived by the decision maker (yet still influence the decision via the perceived time gap), while stronger and longer accelerations can give a salient, consciously perceived yielding signal which strongly pushes the evidence accumulation process towards one of the boundaries. Future studies might be able to disambiguate between these potential explanations by systematically varying the strength of the acceleration cues. The latter however would require combining a large number of acceleration conditions with a large number of repetitions per participant, making it challenging to design a logistically feasible experiment.

Finally, we investigated the impact of including initial bias in the accumulation process, which has mostly been overlooked in previous studies on drift-diffusion modeling of driver behavior. The only previous study to model accumulation bias in traffic interactions is that of Mohammad et al. (2023). They aimed to explain the previously observed effect of initial speed of the ego vehicle on overtaking decisions and response times (Sevenster et al., 2023). Mohammad et al. (2023) compared the constant and speed-dependent versions of the accumulation bias, finding that the bias in favor of the Go decision that increased with the initial speed of the ego vehicle provided a better account of overtaking response times. Here, we aimed to answer a more basic question of whether initial bias is present in left-turn decisions, in which the ego vehicle has zero (or close to zero) momentum when making the gap acceptance decision. We found that models with the initial bias towards the Go decision described our data consistently better, being essential to capture the contrast between the fast Go responses and slow Stay responses. This has implications for future models of evidence accumulation in traffic. In the basic response time modeling literature, initial bias is routinely treated as one of the core parameters of the drift-diffusion model (Ratcliff and Smith, 2004; Ratcliff and McKoon, 2008; Ratcliff et al., 2016), the practice which we recommend to adopt when modeling human-AV interactions with these models.

6.3. Limitations

In our evaluation of the candidate models, we used held-out conditions (Acceleration nudge and Deceleration nudge) to test predictive capacity of the models. This enabled us to evaluate the models' predictive capacity on new observations in unseen *scenarios* (Yarkoni and Westfall, 2017), but did not allow assessment of how well they generalize to new observations in unseen *participants*. Both kinds of generalization are relevant for the purposes of building a coherent theory of human decision making in traffic. However, here we emphasized generalization to new scenarios (as opposed to new participants) as we foresee the model being useful in evaluating human responses to new dynamic maneuvers of the AV based on data including only a limited set of such maneuvers. Generalization to new participants, although perhaps equally important for such applications, is more difficult to assess due to substantial individual differences in decision outcomes and response times observed even in small samples of participants (e.g. Zgonnikov et al., 2022). In this paper, we opted to focus on the "average" participant to highlight behavioral patterns that are present across our whole sample, while controlling for aggregation bias by including random intercepts in our statistical models. This can substantiate claims about out-of-sample generalization of our empirical findings, but needs to be complemented with fitting per-participant (e.g. similar to Zgonnikov et al. 2022) or hierarchical drift-diffusion models to test whether the cognitive mechanisms discussed here apply to individual participants.

An important limitation of our study is that, despite quantifying the effect of AV nudges on drivers' decision making, we cannot yet provide informed recommendations on whether or not such nudges are actually useful for implicit communication by AVs. The first reason for this is that recent studies have highlighted detrimental effects of ambiguous communication patterns by AVs (Miller et al., 2022; Rettenmaier et al., 2021). Nudges in our study were ambiguous by design: for instance, an AV performing a deceleration nudge in essence just sends a brief yielding signal quickly succeeded by the signal of asserting priority. Related to that, the second reason for caution in translating our results into AV design is that the specific deceleration nudge we tested was effective in terms of influencing the decision but was accompanied by occasional negative subjective ratings by the participants (Table 4, Fig. 4). Finally, besides ambiguity for the human road users interacting with the AV, communicative maneuvers of the AV could be considered undesirable by the occupants of the AV. Much research is needed to understand to what extent the beneficial effects of AV maneuvers-as-communication (including nudges) could be traded off with potentially detrimental effects of such maneuvers on the comfort and trust of the AV occupants as well as their willingness to use the AV.

A more fundamental potential limitation is the relative complexity of our model variants compared to traditional evidence accumulation models. The traditional strength of these models is that they capture not just decision probabilities and mean response times, but also often intricate relationships between decision outcomes and response times, as well as full response time distributions (Ratcliff et al., 2016; Evans and Wagenmakers, 2020). These models typically have just 3 to 5 free parameters, hence one could argue that our model (the most successful variant we identified has 7 free parameters) is overly complex and might lack the ability to generate emergent behavior (e.g. due to having an explicit representation of characteristic gap values in its parameters). However, most existing work on evidence accumulation considered very specific, distilled tasks with stimuli characterized by one (typically static) perceptual quantity (e.g., coherence in the random dot kinematogram task, or angle in orientation discrimination tasks). Our driving task, in comparison, is characterized by (a) complexity of the perceptual information, (b) the interplay between different cues available to the driver, and (c) the dynamic nature of the task (where Go decisions must be made fast in order for the driver to still have enough time to execute the decision). Previous work has shown that even simple models are able to explain the decision probabilities in such tasks, yet intricacies of response timing might require making non-trivial assumptions that do not map directly to experimental manipulations (Zgonnikov et al., 2022). Here, we extended this work by measuring response times in Stay decisions; as a result, in order to describe our data, the model needs to provide a good match not just to ten mean decision outcomes and response times (Fig. 5) but also to 20 continuous response time distributions (Fig. 6), at least one of which turned out to be bimodal, and 8 of which are unseen during model fitting. The fact that our 7-parameter model does so successfully with only isolated discrepancies is striking, and, we believe, signifies great potential of our approach for modeling human decisions in complex dynamic scenarios.

6.4. Wider implications

We hope this work contributes to bridging empirical studies of human behavior in interactions with AVs and development of AV interaction planning algorithms. Existing literature on interaction-aware AV controllers shares some common limitations. First, the models used by these controllers to predict human behavior are virtually never directly compared to the very behavior they are supposed to predict (Siebinga et al., 2022) (see Jayaraman et al. 2020a,b for a notable exception). Second, human behavior when interacting with these controllers is usually only demonstrated in a few hand-picked scenarios, rather than systematically evaluated across variations of scenarios. These two limitations can have implications for generalization of these controllers outside of their testing environments, and exemplify the lack of link between the existing literature on interaction-aware AV controllers and the actual human behavior. On the other hand, despite providing valuable qualitative insights, most existing empirical research on human behavior when interacting with AVs falls short of providing computational accounts of this behavior in the form of generative (as opposed to statistical) human models (Haines et al., 2020). This obscures the path to incorporating the obtained understanding of human behavior into computational frameworks for AV interaction planning. Our work exemplifies how insights from empirical research on human-AV interactions can be translated into the computational realm, contributing to the recent efforts in this direction (Rettenmaier and Bengler, 2020; Pekkanen et al., 2022; Zgonnikov et al., 2022; Markkula et al., 2023; Mohammad et al., 2023).

Finally, our work can have conceptual implications for a wider class of human-agent interactions beyond traffic. Any artificial agent that regularly encounters space-sharing conflicts with humans can potentially benefit from strategies that avoid interacting "the hard way" by an early preventive action. We hope our speculative example of nudging the hesitating human in traffic will inspire interaction design in other domains. The usefulness of the particular modeling approach used here is likely specific to binary decision-making scenarios that typically occur in highly structured environments like roads. Alternative approaches of cognitive modeling of human behavior in human-agent interaction exist that are less situation-specific (Thomaz et al., 2016; Hiatt et al., 2017), although integration of cognitive models into computational frameworks for interaction planning remains an open problem (Ho and Griffiths, 2022; Schürmann and Beckerle, 2020). Yet, we believe the research of this kind will be instrumental in enabling agents to have appropriate representations of humans around them, which is critical for responsible development and deployment of artificial agents in the real world (Cavalcante Siebert et al., 2023).

CRediT authorship contribution statement

Arkady Zgonnikov: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Niek Beckers:** Conceptualization, Investigation, Methodology, Project administration, Software, Writing – review & editing. **Ashwin George:** Data curation, Investigation, Software, Visualization, Writing – review & editing. **David Abbink:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. **Catholijn Jonker:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Online supplementary information, data, and code required to replicate our analyses are publicly available at https://osf.io/5hu7e/.

Acknowledgments

This material is based upon work supported by the Dutch Research Council (NWO) funded project "Hybrid Intelligence Center" under grant no. 1136993, and by the European Commission funded project "Humane AI: Toward AI Systems That Augment and Empower Humans by Understanding Us, our Society and the World Around Us" (grant no. 820437). The support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these institutions.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijhcs.2024.103224.

References

Ackermann, C., Beggiato, M., Bluhm, L.-F., Löw, A., Krems, J.F., 2019. Deceleration parameters and their applicability as informal communication signal between pedestrians and automated vehicles. Transp. Res. F Traffic Psychol. Behav. 62, 757–768. http://dx.doi.org/10.1016/j.trf.2019.03.006, URL: https://www.sciencedirect.com/science/article/pii/S1369847818306600.

Ackermann, C., Trommler, D., Krems, J., 2021. Exploring cyclist-vehicle interaction – results from a naturalistic cycling study. In: Black, N.L., Neumann, W.P., Noy, I. (Eds.), Proceedings of the 21st Congress of the International Ergonomics Association. IEA 2021, In: Lecture Notes in Networks and Systems, Springer International Publishing, Cham, pp. 533–540. http://dx.doi.org/10.1007/978-3-030-74608-7_65.

Beggiato, M., Witzlack, C., Springer, S., Krems, J., 2018. The right moment for braking as informal communication signal between automated vehicles and pedestrians in crossing situations. In: Stanton, N.A. (Ed.), Advances in Human Aspects of Transportation. In: Advances in Intelligent Systems and Computing, Springer International Publishing, Cham, pp. 1072–1081. http://dx.doi.org/10.1007/978-3-319-60441-1_101.

Bindschädel, J., Krems, I., Kiesel, A., 2022. Active vehicle pitch motion for communication in automated driving. Transp. Res. F Traffic Psychol. Behav. 87, 279–294. http://dx.doi.org/10.1016/j.trf.2022.04.011, URL: https://www.sciencedirect.com/science/article/pii/S1369847822000791.

Boag, R.J., Strickland, L., Heathcote, A., Neal, A., Palada, H., Loft, S., 2023. Evidence accumulation modelling in the wild: understanding safety-critical decisions. Trends in Cognitive Sciences 27 (2), 175–188. http://dx.doi.org/10.1016/j.tics.2022.11.009, URL: https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(22)00294-7. Publisher: Elsevier.

Cavalcante Siebert, L., Lupetti, M.L., Aizenberg, E., Beckers, N., Zgonnikov, A., Veluwenkamp, H., Abbink, D., Giaccardi, E., Houben, G.-J., Jonker, C.M., van den Hoven, J., Forster, D., Lagendijk, R.L., 2023. Meaningful human control: actionable properties for AI system development. AI Ethics 3, 241–255. http://dx.doi.org/10.1007/s43681-022-00167-3, URL: https://link.springer.com/10.1007/s43681-022-00167-3.

Clements, L.M., Kockelman, K.M., 2017. Economic effects of automated vehicles. Transp. Res. Rec. 2606 (1), 106–114. http://dx.doi.org/10.3141/2606-14, Publisher: SAGE Publications Inc.

de Winter, J., Dodou, D., 2022. External human–machine interfaces: Gimmick or necessity? Transp. Res. Interdiscip. Perspect. 15, 100643. http://dx.doi.org/10.1016/j.trip.2022.100643, URL: https://www.sciencedirect.com/science/article/pii/S2590198222001051.

Devezer, B., Buzbas, E.O., 2023. Rigorous exploration in a model-centric science via epistemic iteration. J. Appl. Res. Memory Cogn. 12 (2), 189–194. http://dx.doi.org/10.1037/mac0000121, URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/mac0000121.

Dey, D., Martens, M., Eggen, B., Terken, J., 2019. Pedestrian road-crossing willingness as a function of vehicle automation, external appearance, and driving behaviour. Transp. Res. F Traffic Psychol. Behav. 65, 191–205. http://dx.doi.org/10.1016/j.trf.2019.07.027, URL: https://www.sciencedirect.com/science/article/pii/S1369847818306326.

Dey, D., Terken, J., 2017. Pedestrian interaction with vehicles: Roles of explicit and implicit communication. In: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, Oldenburg Germany, pp. 109–113. http://dx.doi.org/10.1145/3122986.3123009, URL: https://dl.acm.org/doi/10.1145/3122986.3123009.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V., 2017. CARLA: An open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning. PMLR, pp. 1–16, URL: https://proceedings.mlr.press/v78/dosovitskiy17a.html. ISSN: 2640-3498.

Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N., Pouget, A., 2012. The cost of accumulating evidence in perceptual decision making. J. Neurosci. 32 (11), 3612–3628. http://dx.doi.org/10.1523/JNEUROSCI.4010-11.2012, URL: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4010-11.2012.

Evans, N.J., Wagenmakers, E.-J., 2020. Evidence accumulation models: Current limitations and future directions. Quant. Methods Psychol. 16 (2), 73–90. http://dx.doi.org/10.20982/tqmp.16.2.p073, URL: http://www.tqmp.org/RegularArticles/vol16-2/p073.

Evestedt, N., Ward, E., Folkesson, J., Axehill, D., 2016. Interaction aware trajectory planning for merge scenarios in congested traffic situations. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems. ITSC, pp. 465–472. http://dx.doi.org/10.1109/ITSC.2016.7795596, ISSN: 2153-0017.

Fisac, J.F., Bronstein, E., Stefansson, E., Sadigh, D., Sastry, S.S., Dragan, A.D., 2019. Hierarchical game-theoretic planning for autonomous vehicles. In: 2019 International Conference on Robotics and Automation. ICRA, pp. 9590–9596. http://dx.doi.org/10.1109/ICRA.2019.8794007, ISSN: 2577-087X.

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. Cognitive Science Society, pp. 366–372.

Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30 (1), 535–574. http://dx.doi.org/10.1146/annurev.neuro.29.051605.113038, URL: http://www.annualreviews.org/doi/10.1146/annurev.neuro.29.051605.113038.

Guest, O., Martin, A.E., 2021. How computational modeling can force theory building in psychological science. Perspect. Psychol. Sci. 16, 789–802. http://dx.doi.org/10.1177/1745691620970585, Publisher: Sage Publications.

Haines, N., Kvam, P.D., Irving, L.H., Smith, C., Beauchaine, T.P., Pitt, M.A., Ahn, W.-Y., Turner, B., 2020. Theoretically Informed Generative Models Can Advance the Psychological and Brain Sciences: Lessons from the Reliability Paradox. Technical Report, http://dx.doi.org/10.31234/osf.io/xr7y3, PsyArXiv. URL: https://psyarxiv.com/xr7y3/. type: article.

Hiatt, L.M., Narber, C., Bekele, E., Khemlani, S.S., Trafton, J.G., 2017. Human modeling for human–robot collaboration. Int. J. Robot. Res. 36 (5–7), 580–596. http://dx.doi.org/10.1177/0278364917690592, URL: http://journals.sagepub.com/doi/10.1177/0278364917690592.

Himmelreich, J., 2018. Never mind the trolley: The ethics of autonomous vehicles in mundane situations. Ethical Theory Moral Pract. 21 (3), 669–684. http://dx.doi.org/10.1007/s10677-018-9896-4, URL: http://link.springer.com/10.1007/s10677-018-9896-4.

Ho, M.K., Griffiths, T.L., 2022. Cognitive science as a source of forward and inverse models of human decisions for robotics and control. Ann. Rev. Control Robot. Autonomous Syst. 5 (1), 33–53. http://dx.doi.org/10.1146/annurev-control-042920-015547.

Imbsweiler, J., Ruesch, M., Weinreuter, H., Puente León, F., Deml, B., 2018. Cooperation behaviour of road users in t-intersections during deadlock situations. Transp. Res. F Traffic Psychol. Behav. 58, 665–677. http://dx.doi.org/10.1016/j.trf.2018.07.006, URL: https://www.sciencedirect.com/science/article/pii/S1369847818300068.

Jayaraman, S.K., Robert, L.P., Yang, X.J., Pradhan, A.K., Tilbury, D.M., 2020a. Efficient behavior-aware control of automated vehicles at crosswalks using minimal information pedestrian prediction model. In: 2020 American Control Conference. ACC, pp. 4362–4368. http://dx.doi.org/10.23919/ACC45564.2020.9147248, ISSN: 2378-5861.

Jayaraman, S.K., Tilbury, D.M., Jessie Yang, X., Pradhan, A.K., Robert, L.P., 2020b. Analysis and prediction of pedestrian crosswalk behavior during automated vehicle interactions. In: 2020 IEEE International Conference on Robotics and Automation. ICRA, pp. 6426–6432. http://dx.doi.org/10.1109/ICRA40945.2020.9197347, ISSN: 2577-087X.

Jolly, E., 2018. Pymer4: Connecting R and python for linear mixed modeling. J. Open Source Softw. 3 (31), 862. http://dx.doi.org/10.21105/joss.00862, URL: http://joss.theoj.org/papers/10.21105/joss.00862.

Lee, Y.M., Madigan, R., Giles, O., Garach-Morcillo, L., Markkula, G., Fox, C., Camara, F., Rothmueller, M., Vendelbo-Larsen, S.A., Rasmussen, P.H., Dietrich, A., Nathanael, D., Portouli, V., Schieben, A., Merat, N., 2020. Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. Cogn., Technol. Work http://dx.doi.org/10.1007/s10111-020-00635-y, URL: http://link.springer.com/10.1007/s10111-020-00635-y.

Markkula, G., Lin, Y.-S., Srinivasan, A.R., Billington, J., Leonetti, M., Kalantari, A.H., Yang, Y., Lee, Y.M., Madigan, R., Merat, N., 2023. Explaining human interactions on the road by large-scale integration of computational psychological theory. PNAS Nexus 2 (6), pgad163. http://dx.doi.org/10.1093/pnasnexus/pgad163.

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. Theoret. Issues Ergon. Sci. 1–24. http://dx.doi.org/10.1080/1463922X.2020.1736686, URL: https://www.tandfonline.com/doi/full/10.1080/1463922X.2020.1736686.

Milford, M., Anthony, S., Scheirer, W., 2019. Self-driving vehicles: Key technical challenges and progress off the road. IEEE Potent. 39 (1), 37–45, Publisher: IEEE.

Miller, L., Leitner, J., Kraus, J., Baumann, M., 2022. Implicit intention communication as a design opportunity for automated vehicles: Understanding drivers' interpretation of vehicle trajectory at narrow passages. Accid. Anal. Prev. 173, 106691. http://dx.doi.org/10.1016/j.aap.2022.106691, Publisher: Elsevier.

Mohammad, S.H.A., Farah, H., Zgonnikov, A., 2023. A cognitive process approach to modeling gap acceptance in overtaking. http://dx.doi.org/10.48550/arXiv.2306.05203, arXiv. URL: http://arxiv.org/abs/2306.05203. arXiv:2306.05203 [cs, q-bio].

Pekkanen, J., Giles, O.T., Lee, Y.M., Madigan, R., Daimon, T., Merat, N., Markkula, G., 2022. Variable-drift diffusion models of pedestrian road-crossing decisions. Comput. Brain Behav. 5 (1), 60–80. http://dx.doi.org/10.1007/s42113-021-00116-z, URL: https://link.springer.com/10.1007/s42113-021-00116-z.

Pettigrew, S., 2017. Why public health should embrace the autonomous car. Australian New Zealand J. Public Health 41 (1), 5–7. http://dx.doi.org/10.1111/1753-6405.12588, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1753-6405.12588. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1753-6405.12588.

Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85 (2), 59–108.

Ratcliff, R., McKoon, G., 2008. The diffusion decision model: Theory and data for two-choice decision tasks. Neural Comput. 20 (4), 873–922. http://dx.doi.org/10.1162/neco.2008.12-06-420, URL: http://www.mitpressjournals.org/doi/10.1162/neco.2008.12-06-420.

Ratcliff, R., Smith, P.L., 2004. A comparison of sequential sampling models for two-choice reaction time. Psychol. Rev. 111 (2), 333–367. http://dx.doi.org/10.1037/0033-295X.111.2.333, URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.111.2.333.

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. http://dx.doi.org/10.1016/j.tics.2016.01.007, URL: https://linkinghub.elsevier.com/retrieve/pii/S1364661316000255.

Rettenmaier, M., Bengler, K., 2020. Modeling the interaction with automated vehicles in road bottleneck scenarios. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 64. No. 1. pp. 1615–1619. http://dx.doi.org/10.1177/1071181320641391, URL: https://journals.sagepub.com/doi/abs/10.1177/1071181320641391. Publisher: SAGE Publications Inc.

Rettenmaier, M., Bengler, K., 2021. The matter of how and when: Comparing explicit and implicit communication strategies of automated vehicles in bottleneck scenarios. IEEE Open J. Intell. Transp. Syst. 2, 282–293. http://dx.doi.org/10.1109/OJITS.2021.3107678, Conference Name: IEEE Open Journal of Intelligent Transportation Systems.

Rettenmaier, M., Dinkel, S., Bengler, K., 2021. Communication via motion – Suitability of automated vehicle movements to negotiate the right of way in road bottleneck scenarios. Applied Ergon. 95, 103438. http://dx.doi.org/10.1016/j.apergo.2021.103438, URL: https://www.sciencedirect.com/science/article/pii/S0003687021000855.

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. Auton. Robots 42 (7), 1405–1426. http://dx.doi.org/10.1007/s10514-018-9746-1.

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. Cogn., Technol. Work 21 (1), 69–85. http://dx.doi.org/10.1007/s10111-018-0521-z, URL: http://link.springer.com/10.1007/s10111-018-0521-z.

Schmidt, H., Terwilliger, J., AlAdawy, D., Fridman, L., 2019. Hacking nonverbal communication between pedestrians and vehicles in virtual reality. arXiv preprint arXiv:1904.01931.

Schürmann, T., Beckerle, P., 2020. Personalizing human-agent interaction through cognitive models. Front. Psychol. 11, http://dx.doi.org/10.3389/fpsyg.2020.561510, URL: https://www.frontiersin.org/article/10.3389/fpsyg.2020.561510.

Schwarting, W., Pierson, A., Alonso-Mora, J., Karaman, S., Rus, D., 2019. Social behavior for autonomous vehicles. Proc. Natl. Acad. Sci. 201820676. http://dx.doi.org/10.1073/pnas.1820676116, URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1820676116.

Sevenster, A., Farah, H., Abbink, D., Zgonnikov, A., 2023. Response times in drivers' gap acceptance decisions during overtaking. Transp. Res. F Traffic Psychol. Behav. 94, 329–340. http://dx.doi.org/10.1016/j.trf.2023.03.003, URL: https://www.sciencedirect.com/science/article/pii/S1369847823000517.

Shinn, M., Lam, N.H., Murray, J.D., 2020. A flexible framework for simulating and fitting generalized drift-diffusion models. eLife 9, e56938. http://dx.doi.org/10.7554/eLife.56938, URL: https://elifesciences.org/articles/56938.

Siebinga, O., Zgonnikov, A., Abbink, D., 2022. A human factors approach to validating driver models for interaction-aware automated vehicles. ACM Trans. Hum.-Robot Interact. 11 (4), 47:1–47:21. http://dx.doi.org/10.1145/3538705.

Sripada, A., Bazilinskyy, P., Winter, J.d., 2021. Automated vehicles that communicate implicitly: Examining the use of lateral position within the lane. Ergonomics 64, 1416–1428. http://dx.doi.org/10.1080/00140139.2021.1925353, URL: https://www.tandfonline.com/doi/full/10.1080/00140139.2021.1925353.

Sun, L., Zhan, W., Tomizuka, M., Dragan, A.D., 2018. Courteous autonomous cars. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, pp. 663–670. http://dx.doi.org/10.1109/IROS.2018.8593969, ISSN: 2153-0858.

Terwilliger, J., Glazer, M., Schmidt, H., Domeyer, J., Toyoda, H., Mehler, B., Reimer, B., Fridman, L., 2019. Dynamics of Pedestrian Crossing Decisions Based on Vehicle Trajectories in Large-Scale Simulated and Real-World Data. Technical Report, arXiv. URL: http://arxiv.org/abs/1904.04202. arXiv:1904.04202 [cs] type: article.

Thaler, R.H., Sunstein, C.R., 2008. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press.

Thomaz, A., Hoffman, G., Cakmak, M., 2016. Computational human-robot interaction. Found. Trends Robot. 4 (2–3), 104–223. http://dx.doi.org/10.1561/2300000049, URL: http://www.nowpublishers.com/article/Details/ROB-049.

Tian, K., Tzigieras, A., Wei, C., Lee, Y.M., Holmes, C., Leonetti, M., Merat, N., Romano, R., Markkula, G., 2023. Deceleration parameters as implicit communication signals for pedestrians' crossing decisions and estimations of automated vehicle behaviour. Accid. Anal. Prev. 190, 107173. http://dx.doi.org/10.1016/j.aap.2023.107173, URL: https://www.sciencedirect.com/science/article/pii/S0001457523002208.

Yarkoni, T., Westfall, J., 2017. Choosing prediction over explanation in psychology: Lessons from machine learning. Perspect. Psychol. Sci. 12 (6), 1100–1122. http://dx.doi.org/10.1177/1745691617693393, URL: http://journals.sagepub.com/doi/10.1177/1745691617693393.

Zgonnikov, A., Abbink, D., Markkula, G., 2022. Should I stay or should I go? Cognitive modeling of left-turn gap acceptance decisions in human drivers. Hum. Fact. The J. Hum. Fact. Ergon. Soc. 001872082211445. http://dx.doi.org/10.1177/00187208221144561, URL: http://journals.sagepub.com/doi/10.1177/00187208221144561.