Understanding Collision Avoidance Cues in Virtual Humans

Reiya Itatani
Universidad Politècnica de Catalunya
Barcelona, Spain
reiya.itatani@upc.edu

Tairan Yin
Universidad Politècnica de Catalunya
Barcelona, Spain
tairan.yin@upc.edu

Kexiang Huang
Beiging Institute of Technology
Beijing, China
hwangkx@gmail.com

Jose Luis Ponton
Universitat Politècnica de Catalunya
Barcelona, Spain
jose.luis.ponton@upc.edu

Oscar Argudo
Universitat Politècnica de Catalunya
Barcelona, Spain
oscar.argudo@upc.edu

Nuria Pelechano
Universitat Politècnica de Catalunya
Barcelona, Spain
npelechano@lsi.upc.edu

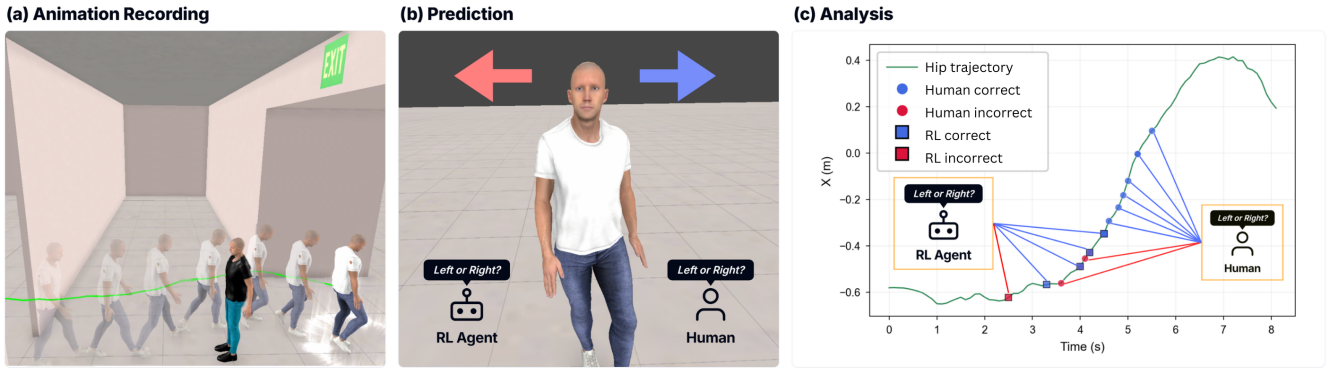


Figure 1: Approach overview. (a) We record locomotion animations in VR. (b) An RL agent and humans predict walking direction from them. (c) We analyze which body cues each relies on.

Abstract

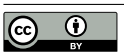
In the real world, humans can navigate through crowded spaces without collision because they are able to predict others' intentions and avoidance strategies. However, navigating among autonomous virtual humans in Virtual Reality (VR) can be challenging due to the absence of anticipatory cues that communicate turning intention. In this paper, we study which body-part kinematic cues are relevant using both a human perceptual experiment in Virtual Reality (VR) and a Reinforcement Learning (RL) analysis sharing the same task formulation. First, we examine participants' gaze behavior while predicting the avoidance direction of an approaching virtual pedestrian animated using full-body motion capture and gaze-tracking data from a real person. Second, we train RL agents on the same task and systematically ablate input features to quantify the contribution of each kinematic feature. Our gaze analysis shows that participants predominantly attend to the head and upper body, while lower-body fixations are associated with prediction errors. Our RL ablation identifies hip position as the dominant cue for accurate prediction, with head rotation providing a secondary but significant contribution to early decisions. Integrating both analyses, we identify a two-phase cue structure: an early orientation phase driven by head rotation, followed by a displacement phase driven by lateral hip movement. Our results show that realistic movement and gaze cues enable virtual humans to convey turning intentions, highlighting the need to incorporate these cues into simulation models. Moreover, the strong correlation between RL agents and perceptual results suggests that RL can help validate collision-avoidance models and support more natural human–agent interactions in virtual environments.

CCS Concepts

• **Human-centered computing** → **Virtual reality**; • **Computing methodologies** → *Reinforcement learning*.

Keywords

Virtual Reality, Virtual Humans, Collision Avoidance, Perceptual Evaluation



SAP '26, Rennes, France

ACM ISBN 979-8-4007-2799-3/26/08
https://doi.org/10.1145/3821409.3821476

ACM Reference Format:
Reiya Itatani, Tairan Yin, Kexiang Huang, Jose Luis Ponton, Oscar Argudo, and Nuria Pelechano. 2026. Understanding Collision Avoidance Cues in Virtual Humans. In *ACM Symposium on Applied Perception 2026 (SAP '26), August 26–27, 2026, Rennes, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3821409.3821476

1 Introduction

Realistic virtual characters are key for immersion in VR, but they often lack clear cues that let users understand their intentions [Yun et al. 2024], making path and avoidance prediction harder than observing real pedestrians. Prior VR studies have examined how virtual walkers' gaze behavior, proximity, and appearance affect avoidance perception [Berton et al. 2019; Bourgaize et al. 2024; Lynch et al. 2018b; Nelson et al. 2024]. However, it remains unclear which specific body-part cues virtual humans must reproduce to convey turning intentions.

This lack of anticipatory cues is rooted in current virtual-human pipelines, which combine path planning with collision-avoidance algorithms that move the character forward and rotate the root to initiate a turn [van Toll and Pettré 2021]. While recent motion matching [Büttner and Clavet 2015; Ponton et al. 2025] and machine-learning synthesis [Rempe et al. 2023] pair these trajectories with high-quality animations, the resulting motion remains reactive rather than proactive. The trajectory often changes before any bodily cues appear, leaving observers without the lead time to anticipate the character's next move. Understanding how humans anticipate others' avoidance behavior could inform how to integrate these cues into virtual human simulations.

VR has become a standard tool for studying collision avoidance in controlled, repeatable settings [Berton et al. 2019], but most VR analyses operate at the level of root motion, such as walking speed and time-to-collision [Lynch et al. 2018b; Varma et al. 2017]. In parallel, machine learning has advanced pedestrian trajectory prediction [Bae et al. 2025; Bremer et al. 2021; Rempe et al. 2023; Xu et al. 2024] and full-body motion generation [Tevet et al. 2023], but targets plausible motion rather than whether an observer can read avoidance intent from it. It thus remains open whether kinematic features alone suffice to reach the same avoidance-direction decisions as a human observer on the same task.

We address this by running human participants and a Reinforcement Learning (RL) agent on the same binary avoidance-direction task. Humans see a virtual pedestrian rendered in VR from prerecorded motion capture, stripped of conversational, contextual, and social cues, so that only body kinematics remain. The RL agent receives the corresponding kinematic features. Comparing their outputs tests whether kinematics alone suffice and lets us use RL for feature ablations that are infeasible with human observers.

We address the following research questions:

RQ1: To what extent does trajectory deviation serve as a primary cue for predicting turning intention?

RQ2: Which body-part movements carry directional information prior to a change in trajectory?

RQ3: Are kinematic features alone sufficient for an RL agent to reach the same binary avoidance-direction decisions as human observers on the same task?

The main contributions of this paper are:

(1) A VR user study showed that **correct predictions** are associated with **increased gaze** towards the **upper body**.
(2) An **RL-based feature ablation** showing that, among the kinematic features, hip position is the dominant cue, while head rotation contributes most to early decisions.
(3) Demonstrating high **alignment between human perception and RL**, with kinematic features alone sufficient to reach the same decisions, supporting RL as a complementary analysis tool for ablations infeasible with human observers.
(4) Demonstrating that virtual humans with realistic body motion and gaze can convey turning intentions similarly to real-world observations, supporting their inclusion in future simulation models.
(5) **Proposing a two-phase cue structure for animation design**, in which head rotation signals intention prior to trajectory displacement. Accordingly, virtual humans should initiate turns with head rotation preceding root motion.

2 Related Work

2.1 Locomotion Cues in Pedestrian Walking

Pedestrians routinely negotiate collision avoidance without speaking by mutually anticipating each other's path intentions [Murakami et al. 2019, 2021, 2022]. To communicate spatial information, humans rely on body-based locomotion cues [Varma et al. 2017], which are generally categorized into global motion (e.g., the displacement of the center of gravity) and local kinematics (e.g., the position and orientation of limbs or the trunk) [Lynch et al. 2018a]. While global motion alone suffices for successful collision avoidance, the addition of local kinematic information stabilizes perceptual-motor responses, enabling earlier, faster, and more accurate avoidance maneuvers [Lynch et al. 2018a; Meerhoff et al. 2014].

Building on this, recent efforts have focused on identifying the specific body parts that serve as predictive cues, primarily exploring body orientation [Murakami et al. 2022], head orientation [Prévost et al. 2003; Schulz and Stiefelhagen 2015; Varytimidis et al. 2018], and gaze behavior [Hessels et al. 2020; Holman et al. 2021; Raimbaud et al. 2023]. For instance, a systematic review by López et al. [López et al. 2019] concluded that gaze yaw significantly anticipates head yaw by 198 ms, which in turn anticipates trunk yaw by 271 ms. However, Cinelli and Warren [Cinelli and Warren 2012] cautioned that such sequences can be confounded by experimental context, reporting no significant temporal anticipation between gaze, head, and body rotation in their own experiments.

VR has become a widely adopted tool to isolate specific physical traits [Berton et al. 2020, 2019; Nelson et al. 2024; Patotskaya et al. 2024; Zibrek et al. 2020], and Berton et al. [Berton et al. 2019] showed that avoidance gaze behavior in VR is qualitatively similar to real-world environments. Yet, which body-part cues humans prioritize to predict walking direction, and when these cues become informative, remains largely unanswered [Lynch et al. 2018b; Varma et al. 2017]. Moreover, isolating the contribution of individual cues directly in human observers is difficult, since specific body-part signals cannot be selectively removed without introducing perceptual artifacts.

2.2 Walking Intention Prediction

Understanding pedestrian walking intention is crucial for human-crowd [Huber et al. 2014], human-robot [Conte and Furukawa 2021; Holman et al. 2021; Yu et al. 2024], and human-vehicle [Ling and Ma 2024; Ridel et al. 2018; Taha et al. 2026] interactions, where the

core task is forecasting a pedestrian's future position from current and past motion observations [Liu et al. 2025].

Classic crowd simulation methods, both macroscopic [Henderson 1971, 1974; Mirzabagheri et al. 2025] and microscopic [van Toll and Pettré 2021], abstract pedestrians as 2D points and rely solely on global acceleration. Local locomotion cues are therefore ignored, and turns are driven only by root rotation, with no early predictive body kinematics.

Recent deep-learning approaches instead use pedestrian-centric features such as joint kinematics, head orientation, and trajectories [Fang and López 2018; Gesnouin et al. 2020; Kothari et al. 2020; Varytimidis et al. 2018], with shoulder and leg joints reported as the most informative for activity estimation [Mirzabagheri et al. 2025].

However, these machine-side studies are typically conducted in isolation from human observers and benchmarked on geometric trajectory errors [Mirzabagheri et al. 2025; Taha et al. 2026], leaving open whether the same kinematic features are sufficient for a model to reach the same decisions humans reach on the same task.

2.3 Positioning

The two lines of work leave complementary gaps. On the perceptual side, body-part cues cannot be selectively removed from human observers without introducing uncanny motions. On the machine side, models are evaluated against trajectory error rather than human decisions. We address both within one framework. Humans and a Reinforcement Learning (RL) agent perform the same binary avoidance-direction task on the same kinematic motion, and we quantify their behavioral consistency in decision timing, accuracy, and lateral-displacement threshold. Strong consistency would imply both that kinematic features carry enough information to support human judgment and that RL enables controlled feature ablations infeasible with humans.

3 Experimental Data Preparation

The virtual character's walking animations were generated by capturing full-body motion of human users in a VR environment. We refer to them as **stimulus providers (SP)** in the following text.

Apparatus. Full-body motion was captured with an Xsens Awinda motion capture system; eye gaze and facial expressions were simultaneously recorded with the built-in sensors of the Meta Quest Pro, and the avatar was authored in Character Creator. The recorded motion was resampled to 10 Hz, one full-body pose every 0.1 s.

Virtual Environment. A minimalist hallway with no decoration, ambient agents, or auditory cues was used as the capture environment, so that the resulting animations carry body motion with little additional context. As visualized in Fig. 2A, the hallway had an entry and exit 7 m apart, with an idle avatar placed 2 m before the exit and a 90 cm-wide central opening 3 m before the avatar (±1 m across trials, $d \in [1, 3]$ m). The idle avatar showed no facial expression, gaze, or verbal signal. The exit was a double door in three configurations (Fig. 2B): **both-open** (free choice), **right-open**, and **left-open**.

Animation Capture. Three members of our research facility, naive to the study purpose, served as SPs. Wearing the VR HMD and MoCap suit, each SP was given a single instruction, to walk from entry to exit while avoiding the idle avatar, with no guidance on how or which side to take. The opening in front of the idle avatar ensured that SPs walked straight at first and could only initiate avoidance steering after passing it. The three door configurations were deliberately selected for each capture iteration to implicitly induce rightward or leftward avoidance. Each SP completed 20 capture iterations (10 both-open, 5 left-open, 5 right-open), yielding 60 walking animations.

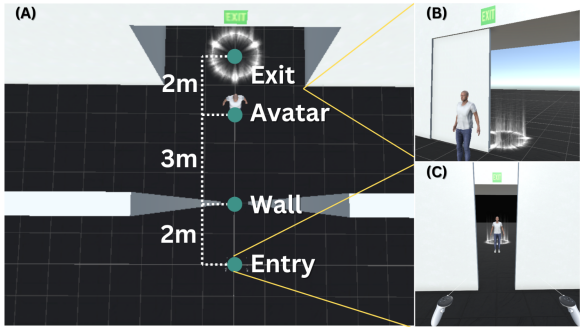


Figure 2: Capture environment. (A) Top–down hallway view with Entry, opening, Avatar, and Exit. (B) Right-open condition biasing the SP rightward. (C) First-person participant view.

4 Experiments

We evaluate the same task with both human and machine inference. The human experiment captures perceptual inference from motion rendered on a virtual avatar, using eye tracking to record which body parts participants look at before deciding. The RL agent predicts avoidance direction from the underlying kinematic data (trajectory and joint rotations), independent of perceptual and rendering constraints.

4.1 Human Prediction Experiment

4.1.1 Task. Participants were instructed to remain stationary while observing an avatar walking toward them (animation recorded from stimulus providers in Sec 3), and to predict whether the avatar would pass on their left or right to avoid a collision. Each trial was presented from a first-person perspective on an empty floor plane (Fig. 3A); the avatar walked from 5 m in front of the participant to 2 m behind them, following one pre-recorded sequence with an avoidance maneuver. The avatar was shown in isolation—without gestures, greetings, or other agents—so that predictions relied on locomotion cues alone, free from communicative biases.

Participants pressed a controller button as soon as they felt confident, which removed the avatar and showed a slider whose displacement encoded the predicted side and confidence (Fig. 3B). To balance speed and accuracy, correct predictions earned a score that decayed quadratically with response time, while incorrect predictions scored zero (Fig. 3C–D). Specifically, the score for a correct prediction was $s = 1 - p^2$, where $p = t/T \in [0, 1]$ is the playback progress at response time, with t the elapsed time and T the full animation duration. This score served only as an incentive

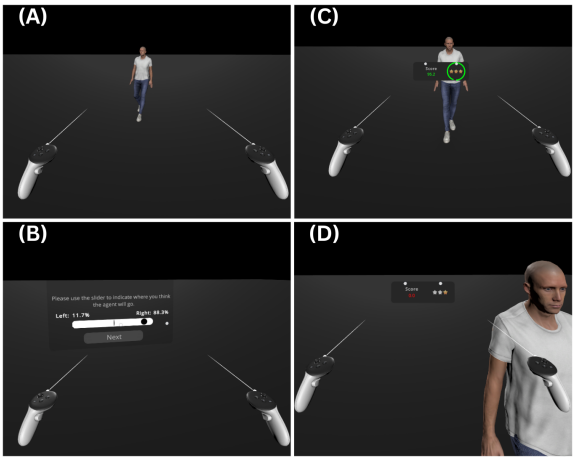


Figure 3: Prediction task. (A) First-person view of the approaching avatar. (B) Slider displayed after pressing the button. Feedback for a correct (C) and incorrect (D) prediction.

to keep participants from delaying their responses and was not used in any analysis. After responding, participants viewed the remainder of the animation.

4.1.2 Experiment. After providing informed consent and reviewing instructions in VR, participants completed a six-trial tutorial and then the main experiment. Of the 60 animations recorded from the SPs, 24 were presented in randomized order—balanced across conditions (12 *both-open*, 6 *left-open*, 6 *right-open*)—and reserved as the held-out test set for the RL experiment.

4.1.3 Participants. Thirteen participants (10 male, 3 female) took part in the experiment. Ages ranged from 18 to 55 years (18–25: $n = 3$, 26–35: $n = 9$, 36–45: $n = 0$, 46–55: $n = 1$). All reported normal or corrected-to-normal vision and provided informed consent prior to participation. A sensitivity power analysis indicated that with $N = 13$, $\alpha = .05$ (two-tailed), and statistical power $1 - \beta = .80$, the study could detect bivariate correlations of $|r| \geq .71$, corresponding to a large effect size [Cohen 1988].

4.1.4 Data Collection. For each trial, we recorded the prediction timestamp, reported confidence, and the participant's gaze. The avatar carried 15 body-part colliders (head, spine, pelvis, and the left/right upper/lower arms, hands, upper/lower legs, and feet); gaze–collider intersections were sampled every 0.1 s, yielding a per-trial time series of body-part fixations.

4.2 Reinforcement Learning Prediction

We formulate the same prediction task as a Markov Decision Process (MDP) and train RL agents in Unity ML-Agents with Proximal Policy Optimization (PPO) [Schulman et al. 2017]. Unlike static classifiers with a fixed data window, an RL formulation dynamically chooses the decision instant, balancing more information against the penalty of a delayed response. This sequential framework mirrors our human experiment, where participants had to decide under a time-decaying score.

4.2.1 Observation and Action Space. At each time step, the agent received motion features extracted from a sliding temporal window. The window advanced in 0.1 s steps, so successive windows overlapped. Four features were included, each defined relative to the first frame: hip position, hip rotation, head rotation, and upper-chest rotation. Models were trained with temporal windows ranging from 0.5 to 1.1 s to examine how the length of the observation history affected prediction performance.

The agent outputs a continuous action value $a \in [-1, 1]$ representing directional confidence from left to right. A continuous action space was chosen because a discrete formulation led the agent to commit to a direction immediately at the first time step, preventing meaningful analysis of prediction timing. A decision was triggered when $|a| \geq 0.9$, a threshold selected after evaluating candidates (e.g., 0.7, 0.8, 0.9); 0.9 yielded the highest accuracy while ensuring that decisions reflect strong directional commitment rather than transient fluctuations around zero.

4.2.2 Reward and Training. The reward decouples a terminal decision component ($R_{decision} = \pm|a|$ for correct/incorrect, -1 on timeout) from a per-step time penalty ($R_{time} = -0.01$) that encourages earlier predictions.

This minimalist reward avoids denser reward shaping that could bias the agent toward specific joints, so that accuracy drops during ablation reflect a feature's intrinsic predictive utility rather than reward engineering.

To stabilize early learning, observation noise is annealed and the time penalty is introduced via a curriculum schedule (full details in the supplemental material). Of the 60 walking animations (Sec 3), 36 unseen by participants are used for training and the 24 trials shown in the human experiment serve as the held-out test set. Each configuration is trained three times and results are averaged across runs.

5 Analysis and Results

5.1 Motion Analysis

5.1.1 Trajectory Turning Point. Trajectories from SPs' recorded data are obtained as the ground projection of the virtual avatar's hip over time. All trajectories follow the same pattern (Appendix A). We estimate the turning point as the intersection of two lines fitted to the initial and avoidance-displacement phases of each trajectory (Fig. 4). See Appendix A for the best-fit computation.

5.1.2 Body-part kinematic cues. To look for anticipatory cues, we computed whether the head, torso, and hip rotate toward the upcoming turn in the frames before the turning point. We define a *pre-turn* window ($[-3.0, -0.5]$ s) and a *late* window ($[-0.5, 0.0]$ s) relative to the turning point. The late window corresponds to the actual turning phase reported in the literature, which starts with head rotation, followed by torso and hips before the feet move [Akram et al. 2010]. We also observed this behavior in the aggregated proportion of body-part animations that rotate towards the turn direction (Fig. 5).

We then tested, for each body part, whether rotation was biased towards the goal in each window. For each animation and window, we computed the fraction of frames rotating towards the goal, and tested whether the mean fraction across the 60 animations differed

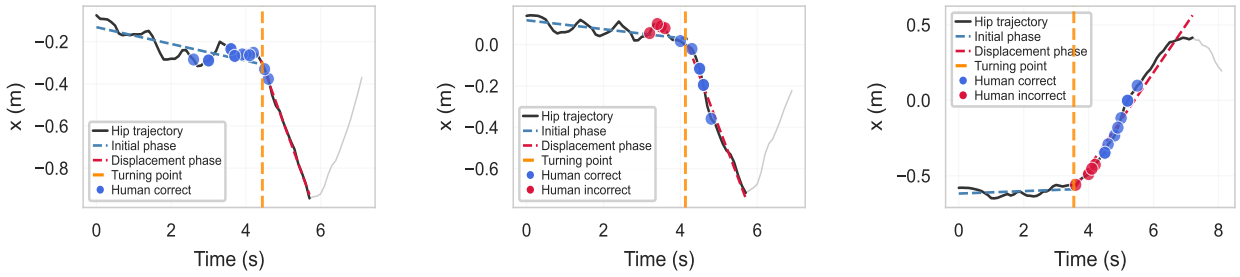


Figure 4: Human predictions relative to the detected turning point. The $x(t)$ trajectory (black) is fitted with a two-phase model (initial: blue, displacement: red); their intersection is the turning point (orange). Circles mark correct (blue) and incorrect (red) predictions. Left: mostly correct before the turning point. Middle: incorrect before, correct after. Right: mixed after the turning point.

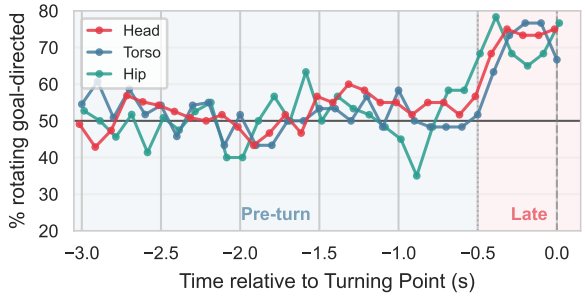


Figure 5: Proportion of animations in which each body part is rotating toward the upcoming turn at each frame relative to the turning point (50% indicates chance).

from chance (0.5, no directional bias) using a one-sided t-test. In the late window, all three body parts showed a strong goal-direction bias (head: 70.3%, torso: 68.1%, hip: 70.8%; all $p < .001$). In the pre-turn window, however, only the head was significant (52.3%, $t(59) = 2.13$, $p = .019$, $d = 0.28$); torso ($p = .084$) and hip ($p = .195$) did not reach significance. Although small, this effect suggests that head rotation carries a subtle anticipatory directional signal before the turn becomes apparent in the trajectory.

5.2 Human Prediction Experiment

5.2.1 Prediction Accuracy. We first examine participants' ability to predict the SPs' left-or-right walking intention. Prediction accuracy was computed for each participant as the number of correct predictions divided by the total number of displayed animations, yielding a high mean accuracy of 91.3%. Fig. 4 provides an overview of participants' predictions with respect to the SPs' lateral displacement, and shows three representative trajectories: (1) most participants predicted correctly even before the turning point, (2) most predicted correctly only after the turning point, and (3) some predicted incorrectly even after the turning point.

To better understand the perception–prediction mechanism, we analyze three factors associated with correct and incorrect predictions: (1) reported confidence, (2) prediction time, and (3) the SP's lateral displacement at the moment of prediction.

Reported Prediction Confidence. After verifying that the confidence scores significantly deviated from normality (Shapiro–Wilk test: $W = 0.850$, $p = .042$), we used a Wilcoxon signed-rank test to investigate how confidence related to prediction correctness. Participants exhibited significantly higher confidence when correct ($M = 0.75$, $SD = 0.16$) than when incorrect ($M = 0.43$, $SD = 0.14$; Wilcoxon $T^+ = 66$, $p < .001$, $r = 1.00$, $n = 11$), where r denotes the rank-biserial correlation: a large effect indicating that all participants showed higher confidence for correct predictions.

Decision Time. Since each prediction trial is a one-shot attempt over a walking motion sequence, participants might achieve good results by predicting late. To assess this, we computed the decision-time offset, $\Delta t = t_D - t_{\text{turn}}$, where negative values indicate decisions made before the turning point. As shown in Fig. 6, prediction accuracy was significantly positively correlated with mean decision-time offset (Pearson's $r = .79$, $p = .001$; Spearman's $\rho = .69$, $p = .009$; $n = 13$), a large effect: participants who committed later tended to be more accurate. This indicates a speed–accuracy trade-off in human inference, where delaying a decision allows participants to observe more of the movement and classify it more reliably.

Lateral Displacement of Stimulus Providers. As the SP approaches the turning point, lateral displacement gradually increases, providing a potential visual cue for direction prediction. Note that lateral displacement first exhibits a swaying phase before becoming large enough to be perceived as a trajectory deviation. We compared the lateral displacement at the time of prediction between correct ($n = 285$) and incorrect ($n = 27$) trials. The mean displacement at correct predictions ($M = 0.21$ m, $SD = 0.15$) was significantly greater than at incorrect predictions ($M = 0.10$ m, $SD = 0.05$; Mann-Whitney $U = 5519$, $p < .001$, $r = .43$), a medium-to-large effect. This indicates that participants might wait until lateral displacement is large enough to secure a correct prediction.

5.2.2 Participant Gaze.

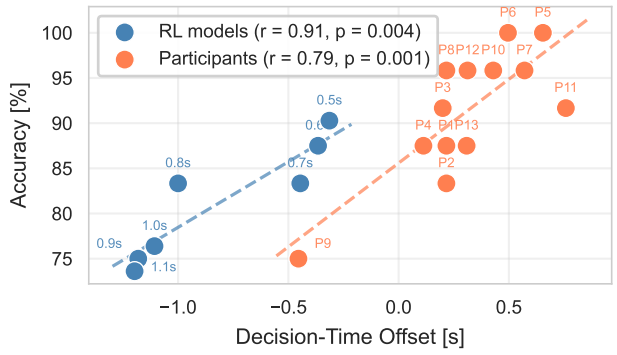


Figure 6: Decision-time offset vs. prediction accuracy for the human (Sec. 4.1) and RL (Sec. 4.2) prediction experiments. Dashed lines are linear regression fits, with Pearson correlations in the legend.

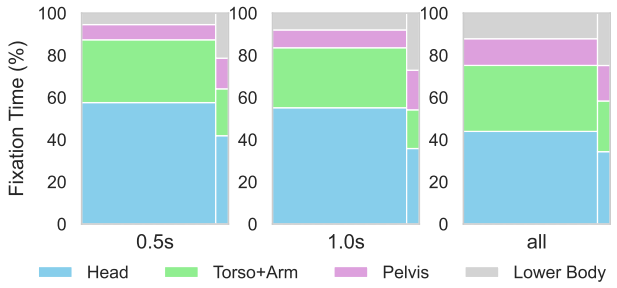


Figure 7: Marimekko charts of gaze distribution across body regions before prediction for the three time windows. Bar width is proportional to trial count (Correct: 285, Incorrect: 27).

Gaze Distribution. We aggregate the 15 colliders into 4 body-part groups to simplify gaze analysis: **Head** (head), **Torso+Arm** (spine, arms, hands), **Pelvis** (pelvis), and **Lower Body** (legs, feet). Because gaze recording for each trial ends at decision time t_D, the recordings have unequal lengths; we therefore analyzed three time windows preceding each decision—$[t_D - 0.5, t_D]$, $[t_D - 1, t_D]$, and $[0, t_D]$—and normalized each to express the gaze distribution as a percentage. As shown in Fig. 7, correctly predicted trials involved longer fixation on the head and torso+arms, whereas incorrectly predicted trials showed more fixation on the pelvis and lower body.

Gaze–Prediction Correlation. To examine whether gaze patterns predict trial outcomes while accounting for the repeated-measures structure of the data (multiple trials nested within each of the 13 participants), we fitted Generalized Linear Mixed Models (GLMMs) with trial outcome (correct/incorrect) as the dependent variable, standardized gaze percentage as the fixed effect, and participant as a random intercept. Holm correction was applied within each temporal window (four tests per window). Lower-body fixation was a consistent source of incorrect predictions across all temporal windows after correction (0.5 s: OR = 0.64, $p_{\text{adj}} = .011$; 1.0 s: OR = 0.59, $p_{\text{adj}} = .001$; all: OR = 0.60, $p_{\text{adj}} = .008$), where an OR below 1 indicates that higher fixation is associated with lower odds of a correct response. No other body region reached significance after correction (see Table S6 in the supplemental material).

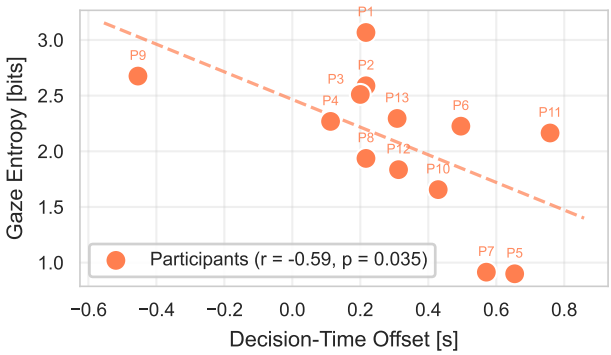


Figure 8: Gaze entropy vs. decision-time offset. Each point is one participant. The dashed line is a linear regression fit, with the Pearson correlation in the legend.

Gaze Entropy. To understand how participants distributed their visual attention across the avatar's body parts, we computed Shannon entropy $H(p) = -\sum_{i=1}^{n} p_i \log_2 p_i$ over the n body parts; higher values indicate a more broadly distributed gaze. Gaze entropy was significantly negatively correlated with decision-time offset (Pearson's $r = -.59$, $p = .035$; Spearman's $\rho = -.72$, $p = .006$; $n = 13$). Although the Pearson coefficient falls below the detectable threshold from our sensitivity analysis ($|r| \geq .71$), the non-parametric correlation exceeded it. This suggests that early predictors tended to distribute their gaze more broadly—probably looking for cues everywhere—whereas late predictors focused more on confirmation from specific parts (Fig. 8).

5.3 Reinforcement Learning

5.3.1 Prediction Accuracy. We analyzed RL prediction behavior using the same metrics applied to human participants.

RL Agent Prediction Confidence. RL agents maintained near-ceiling confidence (≈ 0.98) regardless of correctness, reflecting the fixed $|a| \geq 0.9$ decision threshold rather than a meaningful counterpart to human subjective confidence; full statistics are reported in the supplemental material.

Decision Time. Using the same decision-time offset (Sec. 5.1), models with longer temporal windows (0.8–1.1 s) produced the earliest predictions (most negative offsets) but lower accuracy, while shorter windows (0.5–0.7 s) decided closer to the turning point with higher accuracy (Fig. 6). Decision-time offset and prediction accuracy were strongly correlated across RL models (Pearson's $r = .91$, $p = .004$; Spearman's $\rho = .99$, $p < .001$; $n = 7$), paralleling the human speed–accuracy trade-off where late predictors classified more reliably.

Lateral Displacement. We compared the RL agents' lateral displacement at the time of prediction between correct ($n = 410$) and incorrect ($n = 94$) trials. The mean displacement at correct predictions ($M = 0.13$ m, $SD = 0.13$) was significantly greater than at incorrect predictions ($M = 0.08$ m, $SD = 0.07$; Mann-Whitney $U = 22{,}724.5$, $p = .007$, $r = .18$), a small effect. RL agents responding when the trajectory exhibited larger lateral displacement were more likely to predict the correct direction, mirroring the human pattern.

5.3.2 Feature Ablation. To quantify each feature's contribution, we performed a feature ablation in which individual features were set to zero at test time, measuring the resulting accuracy relative to the full-feature baseline. A two-way ANOVA with ablation condition (5 levels) and observation window size (7 levels) as factors revealed a significant main effect of condition ($F(4, 70) = 127.46$, $p < .0001$, $\eta^2 = .708$, large effect), window size ($F(6, 70) = 9.52$, $p < .0001$, $\eta^2 = .079$, medium effect), as well as a significant interaction ($F(24, 70) = 3.48$, $p < .0001$, $\eta^2 = .116$, medium effect). Tukey HSD post-hoc tests showed that removing hip position ($M = 53.57\%$, $p < .0001$) and head rotation ($M = 68.06\%$, $p < .0001$) significantly reduced accuracy relative to the full-feature baseline ($M = 81.35\%$), whereas removing hip orientation ($M = 80.95\%$, $p = .999$) and upper-chest orientation ($M = 80.95\%$, $p = .999$) did not. Although the main effect of window size was significant, no pairwise differences reached significance in Tukey HSD post-hoc comparisons. These results confirm that global hip translation is the dominant cue for locomotion prediction, with head rotation providing a substantial secondary contribution.

6 Discussion

The perception experiment let us investigate human attention via gaze and when humans commit to a decision, identifying the cues available before each prediction. Despite the design encouraging early responses, participants achieved a mean accuracy of 91.3%, well above chance. Decision-time analysis showed that 12 of 13 participants (92.3%) predominantly predicted after the turning point, suggesting that trajectory deviation is an important cue. However, all participants also made some predictions before the turning point, with a mean accuracy of 79.2% (SD=17.4%), indicating that additional cues such as body part orientation were used. We therefore studied the same task with RL, and given the behavioral consistency between humans and RL agents, used it to run an ablation study estimating the contribution of each cue.

6.1 Humans vs. RL Prediction

Our results show a strong correlation between decision time and accuracy for both humans and RL agents (Fig. 6). Similar to the speed-accuracy trade-off observed in humans, RL agents with different temporal windows exhibit a similar trade-off between earlier, less accurate predictions and later, more accurate ones. The two linear fits are nearly parallel, separated by an observed offset of approximately 0.5 s.

This offset is consistent with the fact that the RL agent decides as soon as cues become available, whereas humans require an observation-cognition-motor loop between cue availability and the button press. Accounting for visual perception and awareness (≈ 100 ms), cognitive processing for two-alternative decisions (≈ 300-400 ms), and motor execution (≈ 100 ms), the expected total delay is approximately 0.5-0.6 s between cue presentation and the recorded response time [Card et al. 1983; Thorpe et al. 1996]. Once this established human response delay is taken into account, the human and RL fits closely match in both decision time and accuracy.

Lateral displacement at the moment of prediction also matches once the response delay is taken into account. The human means shift from 0.21 m (correct) and 0.10 m (incorrect) to 0.14 m and 0.08 m, in line with the RL agent's values. Addressing RQ3, this behavioral consistency in timing, accuracy, and lateral-displacement threshold shows that kinematic features alone are computationally sufficient to reach the same binary decisions.

6.2 Prediction from Trajectory Cues

Aligned with previous research on human collision avoidance [Lynch et al. 2018a], global position was the most important feature for predicting avoidance direction, both for human observers and RL agents. Global position was provided to the RL agent as the hip position. However, humans can perceive this information from the hip, torso, or head, as these body parts move laterally together. Our gaze analysis showed that humans fixated mostly on the head and upper torso during correct predictions, consistent with lateral hip displacement being conveyed upward through the kinematic chain into upper-body motion.

The ablation study performed with the RL agents shows that the prediction accuracy dropped to nearly 50% when the hip position was removed, indicating that within the kinematic input space of the model the agent relies strongly on lateral displacement. Moreover, the RL agents mostly made their predictions before the estimated turning point, implying that the RL models have effectively learned to jointly use different locomotion cues.

To understand prediction based on trajectory, we also need to study the impact of the swaying hip displacement inherent in human locomotion, and analyze when this lateral displacement is interpreted as a possible trajectory turn. Addressing RQ1 and RQ3, we found that this interpretation happened when the stimulus provider was over 0.14 m from the central line of the virtual hallway, for both RL agents and humans (after accounting for the 0.5 s human response delay [Card et al. 1983; Thorpe et al. 1996]).

Thus, while accurate predictions frequently coincide with lateral displacement crossing the 0.14 m threshold, this is a temporal correlation rather than evidence that displacement is the sole causal driver. The final decision is likely triggered by an accumulation of coordinated kinematic cues.

6.2.1 Turning Point vs. Perceived Turning Point. Note that t_{turn} is a conservative estimate, computed from the full trajectory with lateral swaying averaged out by the line fit. We therefore distinguish the turning point, when the SP decided to change direction, from the perceived turning point. Because observers (human or RL) lack the full trajectory, the actual turn stays hidden until the motion clearly deviates beyond the swaying.

Consequently, the number of early predictions would likely be even higher (indeed, 79.2% of predictions made before the turning

point were already correct), highlighting the importance of considering other human movement cues that may anticipate turning intention for collision avoidance.

6.3 Prediction from Body Orientation Cues

Gaze results summarized in Fig. 7 show that participants who predicted correctly looked more often towards the stimulus provider's head, followed by Torso+Arm, whereas those who predicted incorrectly gazed more towards the pelvis and lower body. As noted in earlier work [Cinelli and Warren 2012; Lynch et al. 2018a], these gaze results should not be generalized to social-interaction settings, since they were obtained in a virtual environment without any social interaction. The body-part cues identified here are thus perceptual cues for locomotion prediction, not communicative signals.

These results affirm that humans do not rely on a single fixation point to predict the lateral movement of an incoming pedestrian. This is also supported by the linear regression in Fig. 8, which indicated that participants with higher gaze entropy generally tended to make earlier predictions. As noted in our limitations, the medium effect size of this specific relationship warrants cautious interpretation, but it aligns with the broader observation that excessive attention on lower body movement can be misleading. A possible explanation is that lower-limb kinematics during straight walking are largely periodic (e.g., leg swing) and do not carry directional information until the walker has already committed to a turn.

The RL ablation showed that head rotation was the second most important kinematic feature, while torso and pelvis orientation did not improve prediction accuracy on their own. Consistently, our analysis of all 60 animations (Fig. 5) found that all three body parts carried a strong goal-directed bias in the late window (−0.5 to 0 s), but only the head carried a small but significant directional signal in the pre-turn window (−3.0 to −0.5 s). Addressing RQ2 and RQ3, our findings highlight head rotation as the primary early anticipatory cue prior to trajectory deviation. In contrast, torso and pelvis orientations did not provide significant early predictive value on their own.

We expected agents with larger time windows to be more precise, since they observe a wider range of temporal data. Yet accuracy decreased as the window increased, and agents with larger windows tended to predict earlier. This is in line with the entropy vs. decision-time offset for humans (Fig. 8), where those who looked at more parts predicted earlier. With RL, larger input windows play the role of more gaze targets.

While the ablation shows that the RL models generally rely on hip position, only the variant without hip observation benefits from a larger time window, suggesting that in its absence the agent learns to exploit rotation information.

7 Limitations and Future Works

First, as mentioned in Sec 6, the observed offset of ~ 0.5s between human and RL prediction times is interpreted in light of previous work on HCI rather than measured directly in our task. Measuring this delay precisely for the collision-avoidance task would be valuable but requires additional interdisciplinary expertise and experimental tools.

Second, we did not provide the RL agent with information from every joint. Instead, we simplified the input to a subset of features based on previous empirical studies and on the analysis of human gaze behavior during our experiments. While providing the full body motion data could enable a broader exploration, the high dimensionality of this data (typically more than 22 joints) would complicate both the training and subsequent analysis.

Third, our human prediction experiment was conducted with a relatively small sample size ($N = 13$). As established in our sensitivity analysis, this restricted our statistical power to detecting only large effects. While this sample was sufficient to identify the primary macroscopic cues driving collision avoidance prediction, more subtle perceptual patterns, secondary postural cues, or individual differences in gaze strategy may have gone undetected.

Fourth, all walking animations originate from three stimulus providers in a single frontal hallway scenario, and the RL train/test split holds out animations but not actors. The conclusions about which kinematic features are sufficient should therefore be interpreted within this stimulus range, and broader generalization across walkers, layouts, and densities requires further data.

These limitations also point to future directions. Future work could consider collecting a continuous response for the participant's prediction of a virtual character's walking direction. In addition, our results indicated that lower-body movement can be misleading when humans predict turning intention. This finding contradicts the conclusion of previous work [Mirzabagheri et al. 2025] in machine-learning-based prediction tasks. Understanding this discrepancy is an interesting avenue for further investigation. Beyond visual cues, recent work shows that auditory cues such as footstep sounds can prime anticipation in pedestrian collision avoidance [Hufschmitt et al. 2026], making multi-modal integration a promising future direction.

8 Conclusion

We presented a combined VR perception and reinforcement learning (RL) framework to study the locomotion cues behind third-person avoidance direction prediction. Our RL agent matches human prediction accuracy and timing, showing that kinematic features alone are sufficient to reach the same decisions. Rather than a model of human perception, the RL side complements VR user studies by enabling controlled ablations that isolate each kinematic feature's contribution.

Across both humans and RL, hip position was the dominant cue and head rotation the key secondary cue for early decisions. In the gaze data, humans predominantly fixated on the head and upper body, with lower body fixation leading to incorrect predictions.

Modeling how humans infer directional intent from subtle locomotion cues can ultimately enable virtual humans that are not only visually plausible but also behaviorally predictable, supporting smoother and more natural interactions in immersive environments. Therefore, collision avoidance maneuvers should be led by animations following a head–torso–hip rotation sequence prior to trajectory deviation as opposed to the current sudden root rotation. Finally, equipping virtual humans to perceive such cues from users and fold them into local collision-avoidance algorithms could enable more natural and responsive crowd simulations.

Acknowledgments

This work has received funding from MCIN/AEI/10.13039/501100011033/FEDER, UE (Spain) in the framework of the project PID2021-122136OB-C21.

References

Sakineh B. Akram, James S. Frank, and Julia Fraser. 2010. Effect of walking velocity on segment coordination during pre-planned turns in healthy older adults. *Gait & Posture* 32, 2 (2010), 211–214. doi:10.1016/j.gaitpost.2010.04.017

Inhwan Bae, Junoh Lee, and Hae-Gon Jeon. 2025. Continuous Locomotive Crowd Behavior Generation. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2025), 22416–22431. https://api.semanticscholar.org/CorpusID:277621023

Florian Berton, Ludovic Hoyet, Anne-Hélène Olivier, Julien Bruneau, Olivier Le Meur, and Julien Pettré. 2020. Eye-Gaze Activity in Crowds: Impact of Virtual Reality and Density. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. doi:10.1109/VR46266.2020.00038

Florian Berton, Anne-Hélène Olivier, Julien Bruneau, Ludovic Hoyet, and Julien Pettré. 2019. Studying Gaze Behaviour during Collision Avoidance with a Virtual Walker: Influence of the Virtual Reality Setup. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 717–725. doi:10.1109/VR.2019.8798204

Sheryl Bourgaize, Michael Cinelli, Pierre Raimbaud, Ludovic Hoyet, and Anne-Hélène Olivier. 2024. Collision avoidance behaviours of young adult walkers: Influence of a virtual pedestrian's age-related appearance and gait profile *(SAP '24)*. Association for Computing Machinery, New York, NY, USA, Article 14, 10 pages. doi:10.1145/3675231.3675233

Gianni Bremer, Niklas Stein, and Markus Lappe. 2021. Predicting Future Position From Natural Walking and Eye Movements with Machine Learning. In *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. doi:10.1109/AIVR52153.2021.00026

Michael Büttner and Simon Clavet. 2015. Motion matching-the road to next gen animation. Nucl. ai. https://www.youtube.com/watch?v=z_wpgHFSWss

Stuart K Card, Thomas P. Moran, and Allen Newell. 1983. *The psychology of human-computer interaction*. Lawrence Erlbaum Associates.

Michael E. Cinelli and William H. Warren. 2012. Do walkers follow their heads? Investigating the role of head rotation in locomotor control. *Experimental Brain Research* 219 (2012), 175–190. https://api.semanticscholar.org/CorpusID:14635712

Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dean Conte and Tomonari Furukawa. 2021. Autonomous Robotic Escort Incorporating Motion Prediction and Human Intention. *2021 IEEE International Conference on Robotics and Automation (ICRA)* (2021), 3480–3486. https://api.semanticscholar.org/CorpusID:239040991

Zhijie Fang and Antonio M. López. 2018. Is the Pedestrian going to Cross? Answering by 2D Pose Estimation. In *2018 IEEE Intelligent Vehicles Symposium (IV)*. 1271–1276. doi:10.1109/IVS.2018.8500413

Joseph Gesnouin, Steve Pechberti, Guillaume Bresson, Bogdan Stanciulescu, and Fabien Moutarde. 2020. Predicting Intentions of Pedestrians from 2D Skeletal Pose Sequences with a Representation-Focused Multi-Branch Deep Learning Network. *Algorithms* 13, 12 (2020). doi:10.3390/a13120331

LF Henderson. 1971. The statistics of crowd fluids. *nature* 229, 5284 (1971), 381–383.

Leroy F Henderson. 1974. On the fluid mechanics of human crowd motion. *Transportation research* 8, 6 (1974), 509–515.

Roy S. Hessels, Jeroen S. Benjamins, Andrea J. van Doorn, Jan J. Koenderink, Gijs A. Holleman, and Ignace Th. C. Hooge. 2020. Looking behavior and potential human interactions during locomotion. *Journal of Vision* 20 (2020). https://api.semanticscholar.org/CorpusID:222168925

Blake Holman, Abrar Anwar, Akash Singh, Mauricio Tec, Justin Hart, and Peter Stone. 2021. Watch Where You're Going! Gaze and Head Orientation as Predictors for Social Robot Navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 3553–3559. doi:10.1109/ICRA48506.2021.9561286

Markus Huber, Yi-Huang Su, Melanie Krüger, Katrin Faschian, Stefan Glasauer, and Joachim Hermsdörfer. 2014. Adjustments of speed and path when avoiding collisions with another pedestrian. *PloS one* 9, 2 (2014), e89589.

Aline Hufschmitt, Agathe Bilhaut, Ludovic Hoyet, Julien Pettré, and Anne-Hélène Olivier. 2026. You Walkin' to Me? How Footstep Sound Primes Anticipation in Virtual Pedestrian Collision Avoidance. In *IEEE VR 2026 - IEEE International Conference on Virtual Reality and 3D User Interfaces*. Daegu, South Korea. https://inria.hal.science/hal-05514579

Parth Kothari, Sven Kreiss, and Alexandre Alahi. 2020. Human Trajectory Forecasting in Crowds: A Deep Learning Perspective. *IEEE Transactions on Intelligent Transportation Systems* 23 (2020), 7386–7400. https://api.semanticscholar.org/CorpusID:220381085

Yancheng Ling and Zhenliang Ma. 2024. Pedestrian Crossing Intention Prediction in the Wild: A Survey. *CHAIN* 1, 4 (2024), 263–279. doi:10.23919/CHAIN.2024.000008

Yu Liu, Zhijie Liu, Xiao Ren, You-Fu Li, and He Kong. 2025. Intention-Aware Diffusion Model for Pedestrian Trajectory Prediction. *ArXiv* abs/2508.07146 (2025). https://api.semanticscholar.org/CorpusID:280566304

Antonio M. López, Juan Carlos Alvarez, and Diego Álvarez. 2019. Walking Turn Prediction from Upper Body Kinematics: A Systematic Review with Implications for Human-Robot Interaction. *Applied Sciences* (2019). https://api.semanticscholar.org/CorpusID:115609618

Sean Lynch, Julien Pettré, Julien Bruneau, Richard Kulpa, Armel Crétual, and Anne-Hélène Olivier. 2018b. Effect of Virtual Human Gaze Behaviour during an Orthogonal Collision Avoidance Walking Task. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 136–144. doi:10.1109/VR.2018.8446180

Sean Dean Lynch, Richard Kulpa, Laurentius Antonius Meerhoff, Julien Pettré, Armel Crétual, and Anne-Hélène Olivier. 2018a. Collision Avoidance Behavior between Walkers: Global and Local Motion Cues. *IEEE Transactions on Visualization and Computer Graphics* 24, 7 (2018), 2078–2088. doi:10.1109/TVCG.2017.2718514

L. Rens A. Meerhoff, Harjo J. de Poel, and Chris Button. 2014. How visual information influences coordination dynamics when following the leader. *Neuroscience Letters* 582 (2014), 12–15. https://api.semanticscholar.org/CorpusID:33020977

Alireza Mirzabagheri, Majid Ahmadi, Ning Zhang, Reza Alirezaee, Saeed Mozaffari, and Shahpour Alirezaee. 2025. Navigating Uncertainty: Advanced Techniques in Pedestrian Intention Prediction for Autonomous Vehicles—A Comprehensive Review. *Vehicles* (2025). https://api.semanticscholar.org/CorpusID:279298971

Hisashi Murakami, Claudio Feliciani, and Katsuhiro Nishinari. 2019. Lévy walk process in self-organization of pedestrian crowds. *Journal of The Royal Society Interface* 16, 153 (04 2019), 20180939. arXiv:https://royalsocietypublishing.org/rsif/article-pdf/doi/10.1098/rsif.2018.0939/890454/rsif.2018.0939.pdf doi:10.1098/rsif.2018.0939

Hisashi Murakami, Claudio Feliciani, Yuta Nishiyama, and Katsuhiro Nishinari. 2021. Mutual anticipation can contribute to self-organization in human crowds. *Science Advances* 7, 12 (2021), eabe7758. arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.abe7758 doi:10.1126/sciadv.abe7758

Hisashi Murakami, Takenori Tomaru, Claudio Feliciani, and Yuta Nishiyama. 2022. Spontaneous behavioral coordination between avoiding pedestrians requires mutual anticipation rather than mutual gaze. *iScience* 25 (2022). https://api.semanticscholar.org/CorpusID:253464709

Michael G. Nelson, Fu-Chia Yang, Alexandros Koilias, Christos-Nikolaos Anagnostopoulos, and Christos Mousas. 2024. Avoiding Virtual Characters: The Effects of Proximity and Gesture. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 41–50. doi:10.1109/ISMAR62088.2024.00018

Yuliya Patotskaya, Ludovic Hoyet, Katja Zibrek, and Julien Pettré. 2024. Entropy and Speed: Effects of Obstacle Motion Properties on Avoidance Behavior in Virtual Environment. In *Proceedings of the ACM Symposium on Applied Perception (SAP)*. doi:10.1145/3675231.3675236

Jose Luis Ponton, Sheldon Andrews, Carlos Andujar, and Nuria Pelechano. 2025. Environment-aware Motion Matching. *ACM Trans. Graph.* (2025). doi:10.1145/3763334

Pascal Prévost, Yuri Ivanenko, Renato Grasso, and Alain Berthoz. 2003. Spatial Invariance in Anticipatory Orienting Behaviour during Human Navigation. *Neuroscience Letters* 339, 3 (2003), 243–247. doi:10.1016/S0304-3940(02)01390-3

Pierre Raimbaud, Alberto Jovane, Katja Zibrek, Claudio Pacchierotti, Marc Christie, Ludovic Hoyet, Julien Pettré, and Anne-Hélène Olivier. 2023. The stare-in-the-crowd effect when navigating a crowd in virtual reality. In *ACM Symposium on Applied Perception 2023*. 1–10.

Davis Rempe, Zhengyi Luo, Xue Bin Peng, Ye Yuan, Kris Kitani, Karsten Kreis, Sanja Fidler, and Or Litany. 2023. Trace and Pace: Controllable Pedestrian Animation via Guided Trajectory Diffusion. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), 13756–13766. https://api.semanticscholar.org/CorpusID:257921535

Daniela Ridel, Eike Rehder, Martin Lauer, Christoph Stiller, and Denis Wolf. 2018. A Literature Review on the Prediction of Pedestrian Behavior in Urban Scenarios. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. 3105–3112. doi:10.1109/ITSC.2018.8569415

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).

Andreas Th. Schulz and Rainer Stiefelhagen. 2015. Pedestrian intention recognition using Latent-dynamic Conditional Random Fields. In *2015 IEEE Intelligent Vehicles Symposium (IV)*. 622–627. doi:10.1109/IVS.2015.7225754

Mahmoud Taha, Ahmed Zaky, Hirozumi Yamaguchi, and Ahmed Fares. 2026. Pedestrian trajectory and intention prediction: a comprehensive review of models, datasets, and challenges. *Journal of Ambient Intelligence and Humanized Computing* (2026), 1–33.

Guy Tevet, Sigal Raab, Brian Gordon, Yoni Shafir, Daniel Cohen-or, and Amit Haim Bermano. 2023. Human Motion Diffusion Model. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=SJ1kSyO2jwu

Simon Thorpe, Denis Fize, and Catherine Marlot. 1996. Speed of processing in the human visual system. *Nature* 381 (1996), 520–522. doi:10.1038/381520a0

Wouter van Toll and Julien Pettré. 2021. Algorithms for Microscopic Crowd Simulation: Advancements in the 2010s. *Computer Graphics Forum* 40 (2021). https://api.semanticscholar.org/CorpusID:235337856

Kamala Varma, Stephen J Guy, and Victoria Interrante. 2017. Assessing the relevance of eye gaze patterns during collision avoidance in virtual reality. In *Proceedings of the 27th International Conference on Artificial Reality and Telexistence and 22nd Eurographics Symposium on Virtual Environments*. 149–152.

Dimitrios Varytimidis, Fernando Alonso-Fernandez, Boris Duran, and Cristofer Englund. 2018. Action and intention recognition of pedestrians in urban traffic. In *2018 14th International conference on signal-image technology & internet-based systems (SITIS)*. IEEE, 676–682.

Liang Xu, Xintao Lv, Yichao Yan, Xin Jin, Shuwen Wu, Congsheng Xu, Yifan Liu, Yizhou Zhou, Fengyun Rao, Xingdong Sheng, et al. 2024. Inter-x: Towards versatile human-human interaction analysis. In *CVPR*. 22260–22271.

Weiwei Yu, Siong Yuen Kok, and Gautam Srivastava. 2024. EmoiPlanner: Human emotion and intention aware socially acceptable robot navigation in human-centric environments. *Expert Systems* 42 (2024). https://api.semanticscholar.org/CorpusID:273011758

Haoran Yun, Jose Luis Ponton, Alejandro Beacco, Carlos Andujar, and Nuria Pelechano. 2024. Exploring the Role of Expected Collision Feedback in Crowded Virtual Environments. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 472–481. doi:10.1109/VR58804.2024.00068

Katja Zibrek, Benjamin Niay, Anne-Hélène Olivier, Ludovic Hoyet, Julien Pettre, and Rachel McDonnell. 2020. The Effect of Gender and Attractiveness of Motion on Proximity in Virtual Reality. *ACM Trans. Appl. Percept.* 17, 4, Article 14 (Nov. 2020), 15 pages. doi:10.1145/3419985

A Turning point computation

All recorded trajectories follow a similar pattern, with the SP initially walking straight ahead, passing through the wall opening, and subsequently initiating a turn to avoid the humanoid obstacle. Since the displacement along the forward direction is generally linear with time, our analysis mainly focuses on the trajectories' lateral displacement, denoted as $x(t)$. To investigate whether human avoidance prediction is triggered by observing a clear change in trajectory or by earlier perceptual cues, we identify a turning point in each trajectory as the moment at which it begins to deviate from the forward direction to execute collision avoidance. This point is computed as the intersection of two lines, corresponding to the fitted linear trajectories for the initial phase and the displacement or avoidance phase.

First, the most prominent peak of the lateral trajectory $x(t)$ is selected as the final displacement time T. The trajectory segment from $t = 0$ to $t = T$ is then divided into two phases at a candidate split time k. Each phase is approximated by a line, and the intersection of the two lines represents the turning point. For each candidate $k \in \{3, \ldots, T-2\}$, two least-squares linear models are fitted over the intervals $[0, k]$ and $[k, T]$, denoted by $\hat{x}_{0,k}(t)$ and $\hat{x}_{k,T}(t)$, respectively. The optimal split time is obtained by minimizing the total squared fitting error:

$$k^* = \arg\min_k \quad \sum_{t=0}^{k} (x(t) - \hat{x}_{0,k}(t))^2 + \sum_{t=k}^{T} (x(t) - \hat{x}_{k,T}(t))^2$$

The turning point t_{turn} is defined as the intersection of the two fitted lines $\hat{x}_{0,k^*}$ and $\hat{x}_{k^*,T}$.