Multisensory cues for walking in virtual reality: Humans combine conflicting visual and self-motion information to reproduce distances

Karl Kopiske [*], Elisa-Maria Heinrich [†], Georg Jahn [‡], Alexandra Bendixen [§], Wolfgang Einhäuser [¶]

Abstract

When humans walk, it is important for them to have some measure of the distance they have travelled. Typically, many cues from different modalities are available, as humans perceive both the environment around them (for example, through vision and haptics) and their own walking. Here, we investigate the contributions of visual cues and non-visual self-motion cues to distance reproduction when walking on a treadmill through a virtual environment by separately manipulating the speed of a treadmill belt and of the virtual environment. Using mobile eye tracking, we also investigate how our participants sampled the visual information through gaze. We show that – as predicted – both modalities affected how participants (N = 28) reproduced a distance. Participants weighed non-visual self-motion cues more strongly than visual cues, corresponding also to their respective reliabilities, but with some inter-individual variability. Those who looked more towards those parts of the visual scene that contained cues to speed and distance tended also to weigh visual information more strongly, although this correlation was non-significant, and participants generally directed their gaze towards visually informative areas of the scene less than expected. As measured by motion capture, participants adjusted their gait patterns to the treadmill speed but not to walked distance. In sum, we show in a naturalistic virtual environment how humans use different sensory modalities when reproducing distances, and how the use of these cues differs between participants and depends on information sampling.

Keywords: virtual reality, multisensory perception, treadmill walking, distance perception, eye movements, walking

Running head: Visual and self-motion cues to distance

Corresponding author: Karl Kopiske (karl.kopiske@physik.tu-chemnitz.de)

---

[*]Cognitive Systems Lab, Institute of Physics, Chemnitz University of Technology, Reichenhainer Str. 70, D-09126 Chemnitz, Germany. Corresponding author, correspondence to be addressed to karl.kopiske@physik.tu-chemnitz.de

[†]Physics of Cognition Group, Institute of Physics, Chemnitz University of Technology, and Cognitive Systems Lab, Institute of Physics, Chemnitz University of Technology, Reichenhainer Str. 70, D-09126 Chemnitz, Germany; now at Professorship for Predictive Analytics, Faculty of Behavioural and Social Sciences, Chemnitz University of Technology, Thüringer Weg 11, D-09126 Chemnitz, Germany

[‡]Professorship for Applied Geropsychology and Cognition, Faculty of Behavioural and Social Sciences, Wilhelm-Raabe-Straße 43, D-09120 Chemnitz, Germany

[§]Cognitive Systems Lab, Institute of Physics, Chemnitz University of Technology, Reichenhainer Str. 70, D-09126 Chemnitz, Germany

[¶]Physics of Cognition Group, Institute of Physics, Chemnitz University of Technology, Reichenhainer Str. 70, D-09126 Chemnitz, Germany

New and noteworthy

Combining virtual reality with treadmill walking, we measured the relative importance of visual cues and non-visual self-motion cues for distance reproduction. Participants used both cues, but put more weight on self-motion; weight on visual cues had a trend to correlate with looking at visually informative areas. Participants overshot distances, especially when self-motion was slow; they adjusted steps to self-motion cues, but not to visual cues. Our work thus quantifies the multimodal contributions to distance reproduction.

1 Introduction

Finding back to a place previously visited or knowing how much distance is left before reaching a target are fundamental to successful locomotion in many species (Mittelstaedt & Glasauer, 1991). One key aspect of these operations is the ability to reliably estimate the distance traveled. In humans, many sources of information about the walked distance can be used: Humans can estimate the distances to landmarks, for example, but even in the absence of these, other visual cues are available (Lappe et al., 1999; Thomson, 1980), given for example by optic flow of the environment over time, as well as self-motion cues such as the number of steps taken, and otolithic or vestibular signals (Israël et al., 1997).

The question then is how humans sample, use, and combine these cues to enable them to better navigate the environment. In real-world situations, one may remove one source of information to investigate its contribution to participants' responses (Klatzky et al., 1990; Mittelstaedt & Glasauer, 1991), while in the lab, we can separately manipulate cues. This way, influences of both visual cues and non-visual self-motion cues on estimates of distances and speed have been shown (Campos et al., 2010; Durgin et al., 2005; Frenz & Lappe, 2005; Sun et al., 2004; Varraine et al., 2002). Their interactions can reduce the biases humans show when just one modality is available (Klatzky et al., 1990) and are well described by Bayesian cue combination (Chen et al., 2017; Nardini et al., 2008). Specifically, the modalities are combined with different respective weights depending on the task demands and the perceptual input. In short, manipulating what people see (Prokop et al., 1997) and what they feel (Sun et al., 2004) will affect how far they estimate to have walked.

Of course, humans are not mere passive observers of the world they walk through and do not merely make estimates based on static information. As their position in the world changes, so does the viewpoint and the available information, while eye movements and attending to different aspects of the environment will further enrich information as walking continues. Again, both those factors have been shown to impact distance and speed estimates: It matters whether humans walk or passively cover a distance (Sun et al., 2004) and how they walk, for example whether they walk at a familiar speed or not (Mittelstaedt & Mittelstaedt, 2001). Similarly, neural activity in the hippocampus and parahippocampal areas during wayfinding and navigation tasks has also been reported to differ depending on whether humans or animals cover distances themselves or are moved around passively (Lappe & Frenz, 2009; Terrazas et al., 2005; Winter et al., 2015). Unsurprisingly then, humans are most accurate during active, 'normal' walking with few constraints and a high-fidelity environment (Durgin, 2009). It also matters where they look, as an accuracy advantage has been found when optic flow is not seen in central vision but peripherally (Banton et al., 2005; Durgin et al., 2005). Further, distance estimates serve behaviours that may themselves give humans different information to compare their estimate to. In particular, just as it makes a difference whether humans cover a distance passively or walk it actively, it makes a difference whether they then compare this distance to something they estimate once (Mittelstaedt & Mittelstaedt, 2001), view and match (Caramenti et al., 2018), actively

35 walk again (Sun et al., 2004), reproduce observed distances at self-chosen speeds (Bremmer & Lappe, 1999),
36 or perform a homing task in which one returns to a starting location (Harrison et al., 2022; Nardini et al.,
37 2008). Indeed, such task differences affect not only how humans perform on the task at hand, but also
38 to what degree they learn transfers to another task (Bruggeman et al., 2007). That said, task-dependent
39 response patterns may still be consistent with shared mechanisms underlying the tasks (Lappe et al., 2007).
40   To summarise, there is plenty of evidence that humans can gather speed and travelled distances from
41 visual cues or from non-visual self-motion information, but perform substantially better when both sources
42 of information are available. This is especially true when the task is to reproduce rather than to estimate a
43 distance; that is, when the same kind of information is available during learning and reproduction. However,
44 many open questions remain as to how the two sources of information are sampled, used, and weighed
45 relative to one another. Visual sampling can be controlled by directing one's gaze, but how relevant is
46 the gaze direction for gathering visual information about speed and distance? One would assume that gaze
47 matters, but perhaps not in the intuitive way that bringing optic-flow patterns into central vision is best. And
48 if gaze not towards but past optic-flow patterns may be more adaptive, do participants still focus their gaze
49 where most information is, as they typically do in walking (Hollands et al., 1995; Matthis et al., 2018)? Does
50 the sampling of visual information modulate how it is weighed relative to non-visual self-motion information?
51 In turn, how does gait influence the sampling of non-visual self-motion information and how it is weighed
52 to judge distances?. Does gaze and do gait kinematics differ between learning and reproduction? And what
53 are their differential roles in either phase? How are gaze and gait affected when visual and self-motion
54 provide conflicting information about speed and distance? How are these cues then weighed and integrated
55 to reproduce distance? We aimed to answer these questions in an experiment that had participants walk
56 and reproduce distances in a virtual environment on a treadmill as we manipulated the treadmill's speed as
57 well as the speed of optic flow, and tracked participants' head, body, and eye movements.
58   In the present study, we investigated the roles of visual cues and of non-visual self-motion cues in distance
59 reproduction. We abbreviate the latter cues henceforth as "self-motion cues" for simplicity; this term will en-
60 compass all non-visual information, whereas the term "visual cues" will refer to all visual information, which
61 includes visual cues modulated by participants' self-motion (e.g., optic flow). We manipulated visual cues
62 and self-motion cues independently in a high-fidelity setup, but without removing either modality and while
63 having participants view a screen rather than a head-mounted display. This allowed us to address these
64 questions in an environment in which we could expect participants to show relatively natural behaviour.
65 Specifically, we could (i) isolate and model the respective contributions of visual information and self-motion
66 information, (ii) test whether participants specifically attended to visual information about speed and dis-
67 tance, and (iii) whether there is a relationship between sampling and weighing of information. We expected,
68 as has been shown for speed perception and in navigation tasks, that both visual cues and self-motion cues
69 affect reproduced distances, that the more reliable modality would be more heavily weighed, and that there
70 would be a relationship between time spent sampling and weight given to visual motion cues.

Methods

Participants

73 A total of N = 28 participants (mean age 25.7 years, 16 women and 12 men; mean height 177 cm ± a
74 standard deviation of 10 cm, mean body mass 73 kg ± 15 kg, mean leg length 95 cm ± 5 cm) took part in
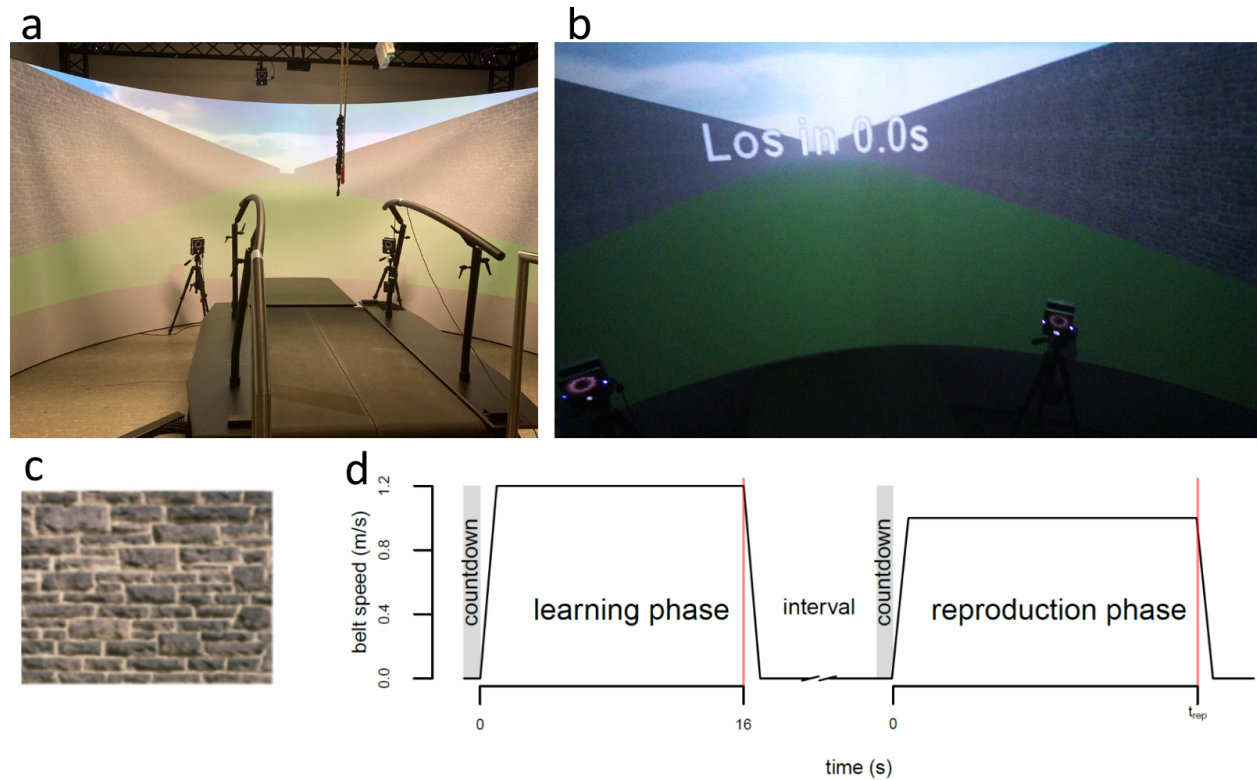
Figure 1: The GRAIL system and the participant view. **a**: Dual-belt treadmill, 240° projection screen, and infrared Vicon cameras for visual presentation, motion capture, and force measurement throughout the experiment. **b**: Participant view of the endless virtual road, with a count-down indicating a trial about to start. The floor was a uniform green to prevent participants from counting or comparing the visual texture with that of the treadmill belts, virtual walls on either side of the road moved to induce optic flow. **c**: Example of a wall texture. **d**: Schematic depiction of a trial. The y-axis shows belt speed, x-axis shows time. Learning phases were always 16 s long (including 1-s onset ramp), but differed in terms of belt speed and visual gain. Reproduction phases were always at 1 m/s belt speed with a visual gain of 1, but differed in length as they ended when the participant pressed a button. Recording started after the countdown and with the start of the onset ramp and ended with the red screen and the start of the offset ramp.

our experiment. All participants had normal or corrected-to-normal vision and no walking impairments and were naive to the hypotheses of the experiment. They received course credit or 8€/h as reimbursement. We planned to include at least N = 24 participants to achieve 80 % power at a medium effect size of Cohen's $f = 0.25$ and $\alpha = .05$ (Cohen, 1988). We invited a small number of extra participants in case some data would need to be excluded due to missing eye-tracking or motion-capture data - this turned out not to be the case, so we analysed the full data set of N=28 participants. All procedures including data management were approved by the Chemnitz University of Technology, Faculty of Behavioural and Social Sciences ethics committee (V-421-PHKP-WET-GRAIL Distanz-15012021).

Setup and procedure

Participants walked on a dual-belt treadmill while viewing a virtual scene on a 240° screen placed 2.5 m ahead of the treadmill's centre, in a Gait Real-Time Analysis Interactive Lab (GRAIL, Motek Medical, Amsterdam, Netherlands) while secured by a harness that was attached to the ceiling. Ten Vicon motion-

capture cameras (Vicon Motion Systems, Yarnton, UK) were placed around the treadmill to enable full-body motion tracking. The virtual scene consisted of a green virtual road with brown brick walls on each side of it (4.5 m from the centre of the road, 3 m high; Figure 1). These walls were the only areas of the visual scene that contained information about motion and distance because of their moving irregular texture, while the ground was chosen to be homogenous providing no motion information. The wall consisted of bricks of irregular sizes (Figure 1c) between 0.6 and 1.3 degrees of visual angle (dva) wide and between 0.3 dva and 0.6 dva high at the closest simulated distance of 4.5 m.

Before each participant's arrival, we calibrated the Vicon cameras. Participants then were instructed about the experiment, gave written, informed consent to take part and filled in a brief questionnaire, providing information on their age and gender (optional) as well as confirming that they had normal or corrected-to-normal vision, were healthy and alert enough to take part in the experiment. Then, participants changed into tight-fitting athletic apparel and we took anthropometric measurements such as body mass, height, and leg length, and applied 35 retro-reflective markers for a Vicon Full-Body model, always applied by the same experimenter for consistent measurements (McGinley et al., 2009). Participants then put on our mobile eye-tracking device, a pair of Tobii Pro Glasses 2 (Tobii AB, Stockholm, Sweden), which had four additional markers attached to track its position and orientation. Next, we calibrated first the body model using approximately 10 s of motion-capture consisting of a T-pose and light walking and then the Tobii glasses using a circle on a standard card provided by the manufacturer. This calibration was validated by having participants fixate 20 red dots presented in sequence on the screen. After eye-tracker validation, the experiment started, consisting of 55 trials of walking and then reproducing a distance. At the end of the experiment, another eye-tracker validation was performed, and the experimenter administered a structured interview consisting of nine questions regarding the experiment (see Table 3 in the appendix), the virtual environment, and any strategies participants might have used.

Stimuli and manipulations

Our virtual environment allowed us to manipulate not just the speed of the treadmill (which on each trial was running at either 0.8 m/s, 1.0 m/s, or 1.2 m/s), but also the gain at which the virtual scene was moving (0.667, 1.0, and 1.5) relative to the speed of the belt. This leads to a 3 × 3 factorial design with factors *belt speed* and *visual gain* and two control conditions in which only one source of information was provided, see Table 1.

Each of the resulting 11 combinations was presented five times for 55 trials total per participant. The trial order was pseudo-randomised such that five blocks of 11 trials contained each combination exactly once, in randomised order, allowing participants to take breaks between blocks when necessary. A trial always consisted of a learning phase and a reproduction phase. The learning phase started with a 1-s countdown when the participant indicated that they were ready, followed by 16 s of walking (including an onset ramp of 1 s to accelerate the belt linearly to its target speed, see Figure 1d). The end was signalled by the visual scene being overlaid by a transparent red, at which point a 1-s linearly decelerating off-ramp started. Then followed a brief interval until the participant indicated they were ready for the reproduction phase. The reproduction phase again started with a countdown and a 1-s ramp to accelerate the belt to 1 m/s and lasted until participants pressed a button on a hand-held controller when they were satisfied that the distance matched the one during learning phase, at which point the screen again turned red and the belt was decelerated in a 1-s off-ramp (Figure 1d). Participants were not told in advance that visual and non-visual information could be incongruent, and thus not instructed to follow one source of distance information over

Table 1: Conditions in our experiments and their basic characteristics. Rows show gain, columns show belt speed, so that each cell shows the resulting speed of the visual scene.

speed - gain	0 m/s	0.8 m/s	1 m/s	1.2 m/s
-	1 m/s		0 m/s	
0.667		0.53 m/s	0.667 m/s	0.8 m/s
1		0.8 m/s	1.0 m/s	1.2 m/s
1.5		1.2 m/s	1.5 m/s	1.8 m/s

the other in the reproduction phase. All participants took at least one extended break after the third block of five and were allowed to take more. In total, an experiment lasted around 30 to 40 minutes, excluding preparation.

Data processing and analysis

We recorded the timing of participants' button presses, as well as their eye and body movements. A total of 18 trials (1.2 % of all trials) had to be discarded due to technical difficulties. Motion-tracking data were recorded at 250 Hz throughout the experiment using the Vicon cameras, as were data from force plates under each treadmill belt. Eye-tracking data were recorded at 100 Hz using the Tobii glasses. To process and analyse motion-tracking and eye-tracking data, we largely followed the analysis pipeline from Kopiske et al. (2021), which used the same setup. Specifically, eye- and motion-tracking data were synchronised by identifying the time of trial onset in the eye tracker's scene-camera video. We then applied a cubic-spline interpolation to the motion-tracking data to deal with missing data (on average, 0.3 % of frames contained a missing head marker; 1.5 % a missing pelvis marker; and 0.1 % a missing foot marker) and to down-sample it to 100 Hz so that it could be merged with the eye-tracking data. On average, a relatively low proportion of 4.5 % of eye-tracking data including blinks (Kopiske et al., 2021) were missing. These were interpolated in the same way as the motion-tracking data. Merged data were then smoothed using a third-order Savitzky-Golay filter (Savitzky & Golay, 1964) with a filter window of 110 ms. This allowed us to calculate gaze positions continuously in real-world coordinates.

For robust step detection, force data on both belts were added and a Savitzky-Golay filter with 524 ms width – wide enough to contain roughly one entire step, but never two – was applied. The peaks of the resulting force profiles were counted as step onsets. We verified this against an automated online step-detection algorithm using force threshold for foot-off and foot-down. We found that the methods matched perfectly (except for a somewhat displaced onset timing, which we did not analyse as it made no difference for any of our analyses) when participants walked on both belts, but that the offline algorithm using filtering was more robust when cross-stepping occurred (i.e., when participants placed both their feet on one belt). To calculate the length of each step, we computed the distance between the toe marker of the front foot and the heel marker of the hind foot at the time point of each step.

Two measures were taken to ensure the validity of our gaze data: One, since the glasses typically did not sit perfectly horizontally on the participant's head, the slope between head markers at the back and the front of the head would have given a biased estimate of the actual head inclination. We corrected for this by comparing the height of the horizon in the scene-camera videos as detected using a Hough transform (Duda & Hart, 1972) to the height inferred from motion-capturing data. From this, we computed a correction for

each participant. On average, the slope between the markers was -15.1 °. Two, we compared the position of the validation dots in the scene-camera videos to participants' gaze position while fixating them and applied a corresponding correction. The mean absolute error was 0.9 ° vertically and 0.5 ° horizontally.

This way, we obtained three main types of data for each trial: (i) Data of the reproduced distance, (ii) continuous gaze data in real-world coordinates showing us where participants looked while completing the task, and (iii) step lengths and step counts to determine if and how participants adjusted their gait. The reproduced distances were submitted to a 3×3 repeated-measures analysis of variance (rmANOVA) with factors *belt speed* (0.8 m/s, 1.0 m/s, 1.2 m/s) and *visual gain* (0.667, 1, 1.5) to assess whether each source of information had any impact on participants' performance. To then get an estimate of how strongly each cue was weighed by each participant, we modelled the reproduced distance d_r as a linear combination of the treadmill distance during learning (d_m, for self-motion distance), optic-flow distance during learning (d_v, for visual distance), and a static prior to account for each participant's bias (Petzschner et al., 2015), which we modelled as consisting of a distance d_p towards which the participant was biased, as well as a weight w_p with which they weighed the bias towards this distance. Equation 1 specifies this model of what contributed to reproduced distance d_r, while equation 2 and equation 3 show how to then calculate the weight of self-motion information (w_m) and of visual information (w_v), respectively. Each participant's distance prior (d_p) and prior weight (w_p) were estimated following equation 4. Using the usual normalization that the three weights add up to 1, we can then calculate w_m and w_v for each trial where the gain was smaller or larger than one, that is, for each trial where $d_m \neq d_w$. Specifically, the formula for d_r

$$d_r = w_m d_m + w_v d_v + w_p d_p \tag{1}$$
$$= w_m d_m + (1 - w_m - w_p)d_v + w_p d_p$$

can be rewritten to give the parameters of interest, w_m and w_v, as a function of the participant's prior d_p, its weight w_p, trial conditions (visual distance d_v, self-motion distance d_m), as well as the reproduced distance d_r:

$$w_m = \frac{(d_r - (1 - w_p)d_v - w_p d_p)}{d_m - d_v} \tag{2}$$

and

$$w_v = \frac{(d_r - (1 - w_p)d_m - w_p d_p)}{d_v - d_m} \tag{3}$$

To estimate each participant's prior (d_p) and the prior's weight (w_p), we used the data from trials with a gain of 1, taking advantage of the fact that $d_m = d_v$ allows us to simplify equation 1 to

$$d_r = d_m(1 - w_p) + w_p d_p \tag{4}$$

Filling in the three unique values of d_m and the corresponding averages for d_r, we then obtained three linear equations with two unknown variables, w_p and d_p, which we fitted using the *nls* function in R (R Core Team, 2022). We bounded w_p between -0.2 and 1.2 (rather than 0 and 1), since a hard bound at 0 could result in artificially observing mean weights larger than 0 due to statistical noise. For w_v and w_m (which

we did not need to fit, but obtained from equations 2 and 3), participant means also fell within this range when w_p was bounded.

Gaze allocation was analysed descriptively (Figure 4), as well as by looking at the proportion of gaze towards the walls in each trial, as this was where optic-flow information was shown. These values were then correlated with the visual weights to test the prediction that looking at the visually informative areas more would be associated with weighing visual information more strongly, as only the wall patterns provided visual information about distance. Gait parameters (step count, step length) were also assessed descriptively. All data and analyses are available at: https://osf.io/cbvpa/?view_only=c7b33fba59564ed68d0b740074448618
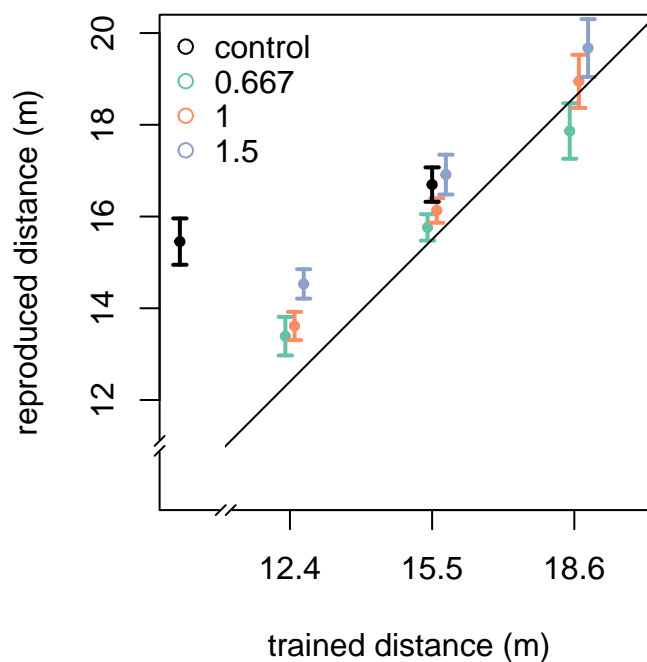
Results

Figure 2: Distances reproduced, by self-motion distances and visual gain, averaged across all N=28 participants by computing the arithmetic mean first by condition and then across conditions. Colours indicate the different visual gains, see figure legend. Black circle on the left shows reproduced distance in our "optic flow only" condition, where the visual scene moved at 1 m/s, but the belt did not move at all, while the other black circle (at *trained distance* = 15.5 m) shows the mean distance for the condition in which the belt speed was 1 m/s but the visual scene did not move. Error bars indicate between-participant SEMs. Data points for different visual gains are jittered on the x-axis to facilitate viewing, actual belt distances were identical at 12.4 m, 15.5 m and 18.6 m for 0.8 m/s, 1 m/s and 1.2 m/s, respectively. Solid line indicates unity.

To investigate whether participants reproduced distances according to self-motion information or visual information, we analysed reproduced distances, participants' gaze, and participants' steps; where they looked while doing so, and if this affected the way they reproduced distances; and how they adjusted their gait to reach different speeds and distances.
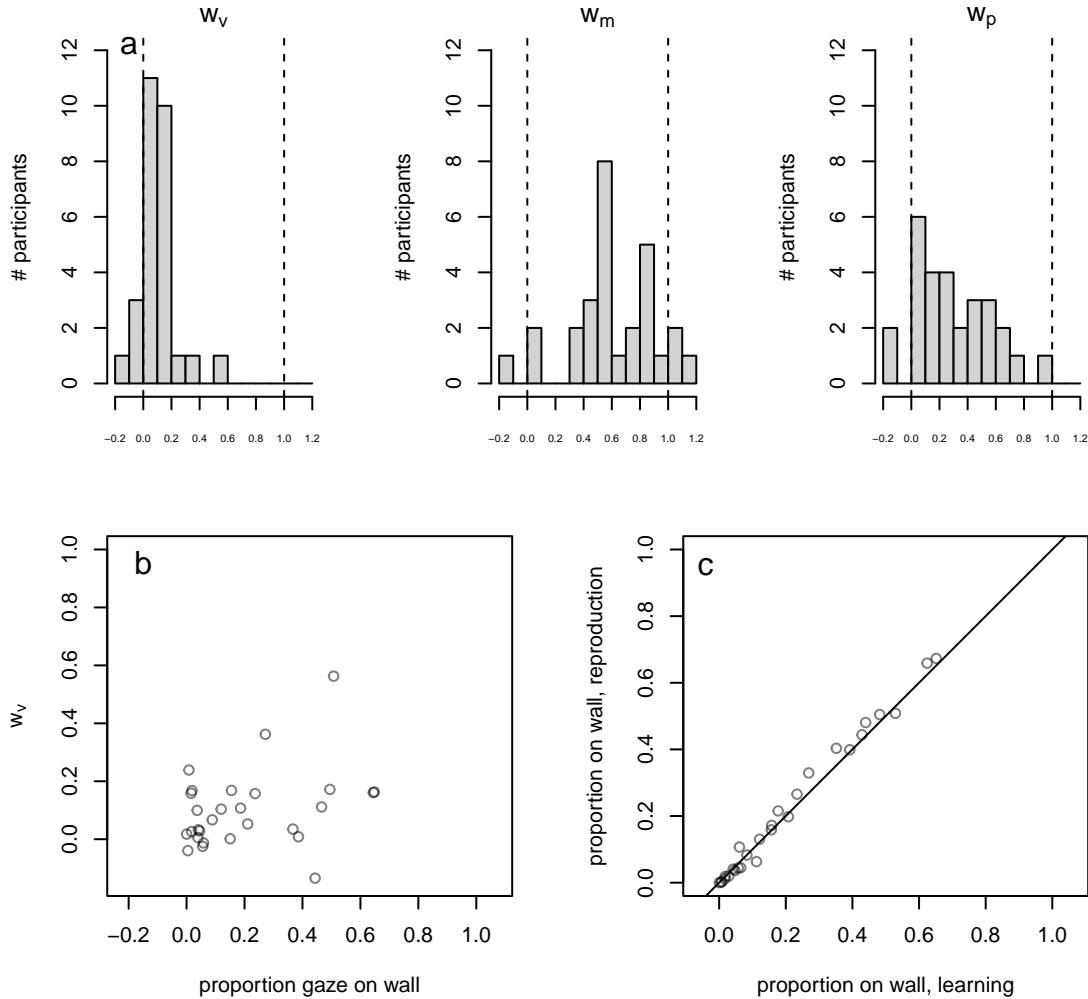
Figure 3: a: Histograms of modelled weights for visual cues, self-motion cues, and the distance prior, by participant (N=28). We see that on average but not for all participants, self-motion cues and to a lesser degree the prior are weighed more strongly than visual cues. Weights calculated according to formulas 2, 3, and 4. **b**: Weights for visual information, by participant, plotted against proportion of gaze directed towards the virtual walls. Each dot represents one participant. We see a large amount of inter-individual variability in gaze allocation, with some participants looking predominantly towards the walls, while others never looked towards them. Data from trials in which belt speed or visual gain was 0 are excluded, since no weights could be calculated for these. We show averages across all trials (excluding those with gain = 1), each circle represents one participant. **c**: Proportion of gaze towards the walls during learning phases and during reproduction phases, by participant. Shown are averages across all trials, each circle represents one participant. Solid line indicates unity. We see that virtually all participants were very consistent in how much they gazed towards the walls of the virtual scene with very similar proportions during learning phases and during reproduction phases.

Combining self-motion and visual information

To investigate the impact of self-motion (treadmill distance) and visual information (optic-flow gain) during the learning phase on how far participants walked during reproduction, we submitted the time until participants pressed the button during reproduction to a 3×3 rmANOVA with factors *belt speed* and *visual gain*. As expected, we found main effects of both *belt speed*, $F(2, 54) = 64.07$, $p < .001$, and of *visual gain*,

Table 2: Within-participant means and standard deviations for visual weight. Third column shows the means and standard deviations across participants for the within-participant mean of visual weight for each combination of belt speed and visual gain.

visual gain	belt speed	visual weight, mean (SD)
0.667	0.8	0.22 (0.43)
1.5	0.8	0.03 (0.35)
0.667	1	0.17 (0.41)
1.5	1	0.04 (0.27)
0.667	1.2	0.15 (0.41)
1.5	1.2	0.09 (0.26)

208 $F(2, 54) = 11.37$, $p = .001$. There was no interaction, $F(4, 108) = 0.86$, $p = .435$. Specifically, participants
209 reproduced a longer distance both when the walked distance was longer and when optic flow was faster
210 (Figure 2). On average, participants tended to overshoot the distance (Figure 2), but less so the faster the
211 belt speed was, such that they were almost perfectly accurate at the fastest speed used in our experiment,
212 1.2 m/s. The mean slope of the response function $d_r \sim d_m$ was 0.8 ± 0.47 (standard deviation), significantly
213 different from a slope of 1 ($t(27) = 2.23$, $p = .035$).

214   To quantify the influence of each factor, we modelled the reproduced distances following equation 1,
215 which allowed us to get estimates of participants' weights of visual cues (w_v) and of self-motion cues (w_m),
216 as described in equation 2. Again as expected, self-motion information was weighed more strongly, with a
217 mean w_m of $.60 \pm .32$, compared to a mean w_v of $.10 \pm .13$. In fact, only two participants weighed visual
218 information more strongly, and all but three participants weighed self-motion information at least twice as
219 strongly, see also Figure 3a. That said, the mean w_v was still statistically significantly different from 0, $t(27)$
220 $= 3.92$, $p < .001$. Consistent with the notion that more reliable cues are also weighed more strongly (Landy
221 et al., 1995), we found that the mean standard deviation in trials in which participants had only visual cues
222 was higher numerically than in those in which only self-motion cues were available (2.57 m to 2.07 m, $t(27)$
223 $= 1.67$, $p = .107$). Note that since these unimodal cases would also contain variability from the prior, which
224 can only be estimated very noisily from relatively few trials, we refrained from attempting to isolate these
225 variances and standard deviations and compute Bayesian model predictions. We show within-participant
226 means and standard deviations of w_v in Table 2 as an estimate for how variable the use of vision was
227 depending on the condition. Weights and their respective standard deviations were numerically higher when
228 the visual gain was low, indicating that when the belt was moving faster than the visual scene, participants
229 tended to use the visual information more, but also in a more variable manner.

Sampling visual information for walking

231 Combining eye-tracking and motion-tracking data allowed us to analyse participants' gaze in real-world
232 coordinates. Gaze maps for different types of trials can be seen in Figure 4, showing that participants
233 tended to fixate close to the horizon, but also directed their gaze to each side a substantial amount. We
234 also specifically determined what proportion of the time participants looked at the walls on each side of
235 the endless road (with a tolerance of 2 degrees to account for measurement inaccuracies and the size of the
236 fovea), as in our scene only these regions contained a texture that moved in accordance to the movement of
237 the virtual camera and thus information about speed and distance (the road was uniformly green, see Figure
238 1). On average, participants looked at the road and towards the horizon much more than the walls, but
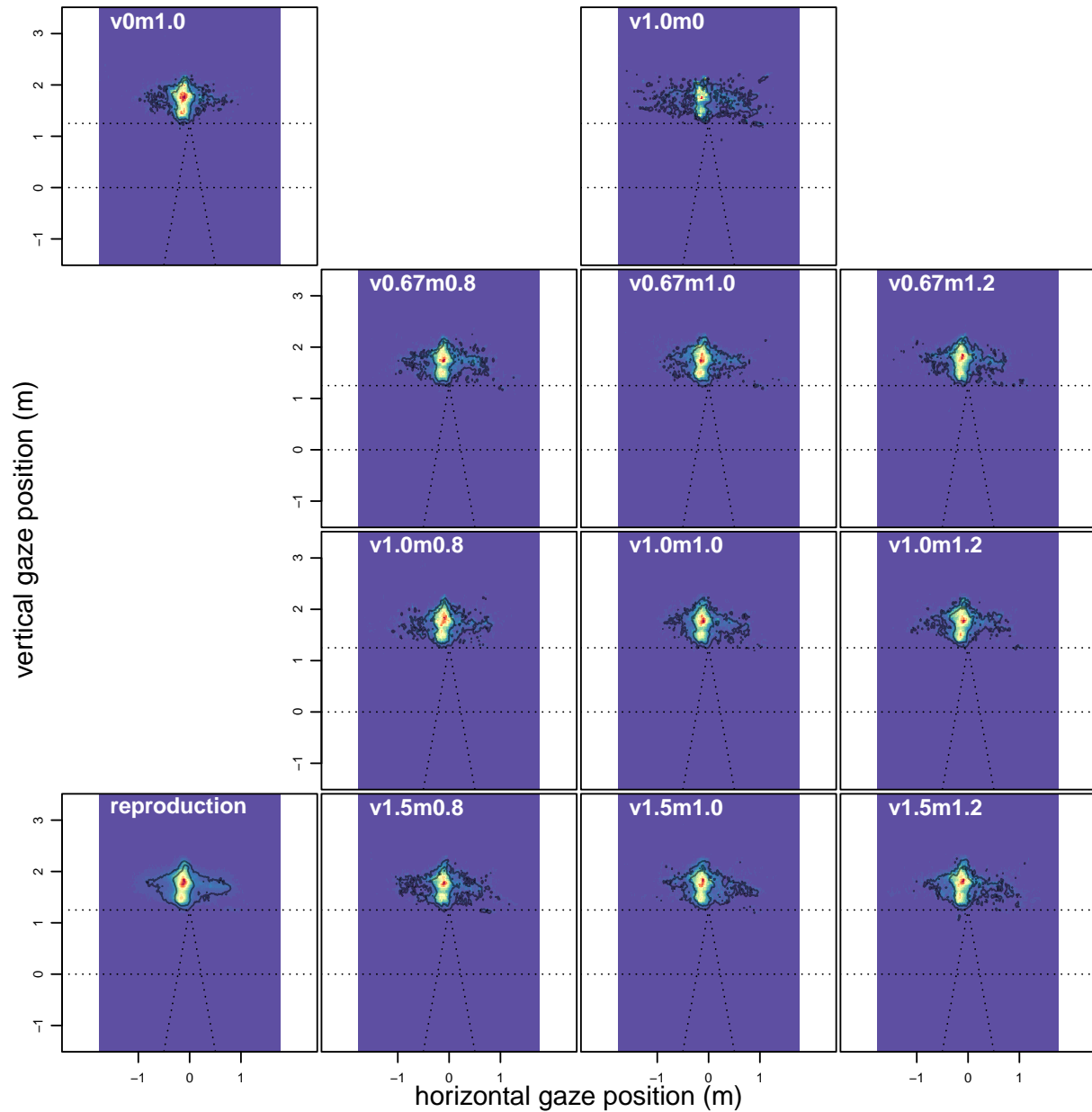
Figure 4: Gaze maps of eye-in-world positions, across all participants (N=28), by condition. Data from learning phases are split up by visual speed (labelled v) and the gain factor for self-motion, that is, for the belt speed (m). Data from reproduction phases are collapsed since here, both speeds were kept constant at 1 m/s. Dashed lines indicate the outlines of the virtual road and horizon, solid black outlines indicate areas where gaze was allocated 20 % (inner outline) and 80 % (outer outline) of the time. We see most gaze directed towards the virtual horizon, as well as the upper halves of the virtual walls.

with considerable inter-individual variability (mean proportion gaze on wall: 21.1 %, $\pm$ 20.7 %; see Figure 3b). Compared to this, there was relatively less intra-individual variability, with a mean within-participant standard deviation of 11.5 %. Gaze-on-wall proportions barely differed between learning phase (20.6 %, $\pm$ 20.1 %) and reproduction phase (21.5 %, $\pm$ 21.3 %), with the two being very highly correlated on a trial-wise basis (r (1520) = .90) and almost perfectly per participant at r (26) = .99 (see Figure 3c).

²⁴⁴ We also expected that spending a longer time looking where visual information was given would cause
²⁴⁵ participants to weigh this information more strongly. We found a weak correlation between the proportion
²⁴⁶ of gaze towards the walls and w_v of $\rho = .32$, which was not statistically significant, $p = .097$ (Spearman's
²⁴⁷ rank correlation), see Figure 3. That said, looking at the walls more did not necessarily make participants
²⁴⁸ better at estimating visual distance, as there was virtually no relation between a participant's variance of
²⁴⁹ reproduced distances in vision-only trials and the participant's proportion of gaze towards the walls in those
²⁵⁰ same trials, $\rho = .02$, $p = .91$. When considering the correlation between w_v and gaze towards the walls in
²⁵¹ each participant individually, we do not see any consistent relation that trials where the walls were fixated
²⁵² more had higher visual weight, either: this correlation ranges between -.48 and .31 across individuals with a
²⁵³ median of $\rho = 0.02$ and mean $(\rho = .01)$ that is indistinguishable from zero, $t(27) = 0.26$, $p = .796$.
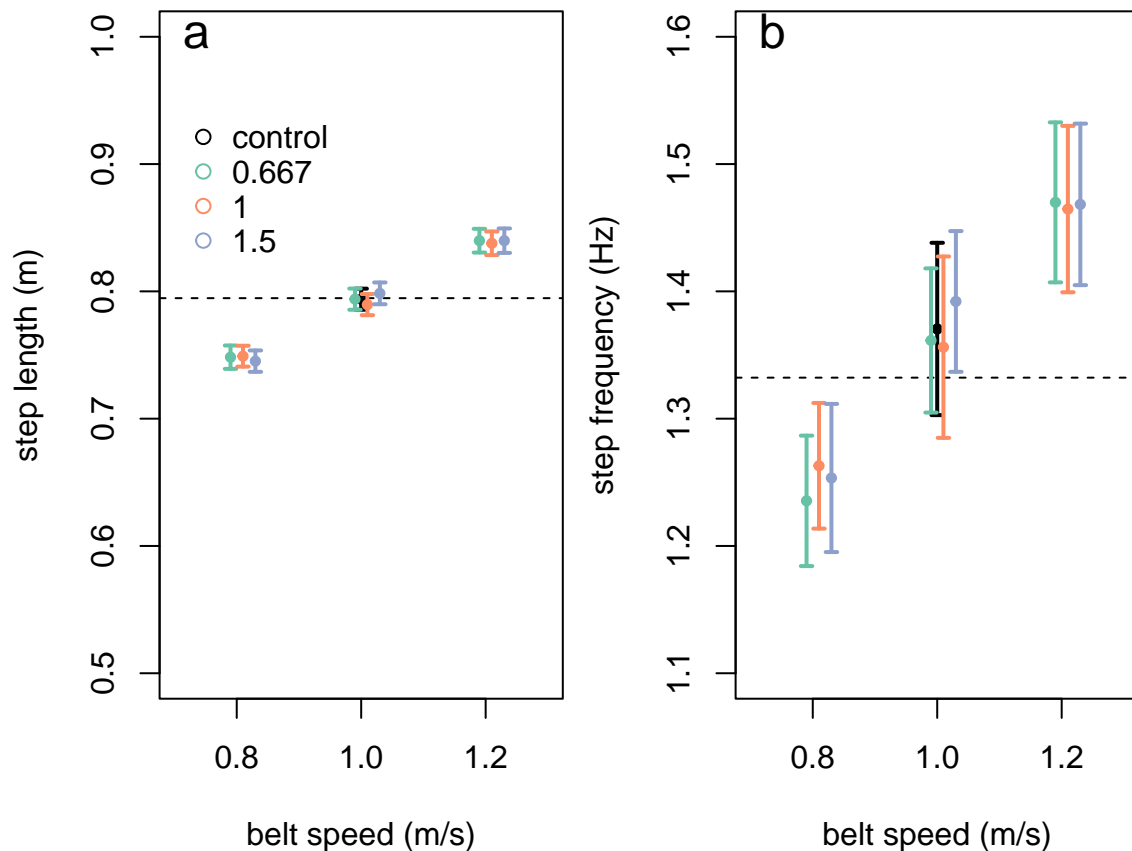
Gait adjustments

Figure 5: Step length and step frequency during learning phases, by condition, averaged across all N=28 participants by computing the arithmetic mean first by condition and then across conditions. Colours indicate the different visual gains, see figure legend, error bars indicate between-participant SEMs. Data points for different visual gains are jittered on the x-axis to facilitate viewing, actual belt speeds were identical. Dashed horizontal line shows parameter mean during reproduction phase. **a**: Average step length, showing mild scaling with belt speed. **b**: Average step frequency, also scaling with a slope below unity.

²⁵⁵ Finally, we assessed how participants adjusted their gait when walking at different speeds during learning,
²⁵⁶ as well as when controlling the distance walked during reproduction, where different patterns have been found

(Durgin et al., 2007). We found that both step length and step frequency were adjusted when the belt was running at different speed (see also Figure 5): Participants' mean step length increased from 74.8 cm to 79.4 cm to 83.9 cm as belt speed increased from 0.8 m/s to 1.2 m/s. At the same time, step frequency also increased from 1.25/s to 1.37/s to 1.47/s. Meanwhile, optic flow had virtually no impact on either measure, with mean step length varying only between 79.2 cm and 79.5 cm between different visual gains, and corresponding step frequencies ranging from 1.35/s to 1.37/s. Two 3×3 rmANOVAs with dependent variables *step length* and *step frequency* corroborated these findings, as the factor *belt speed* had a main effect on both variables (on step length, $F(2, 54) = 242.30$, $p < .001$, and on step frequency, $F(2, 54) = 46.78$, $p < .001$), which was not the case for *visual gain* (neither on step length, $F(2, 54) = 0.89$, $p = .397$, nor on step frequency, $F(2, 54) = 0.35$, $p = .710$).

During reproduction, we saw that neither step length nor step frequency correlated strongly with the reproduced distance. Within each participant, the mean correlation between step frequency and reproduced distance was $r_{d,SF}$ (26) = -.13, $p = .51$, and $r_{d,SL}$ (26) = .02, $p = .91$ for step length. Rather, the deciding variable was number of steps, as a longer reproduced distance was highly correlated with more steps taken (mean correlation within each participant: $r_{d,SC}$ (26) = .71, $p < .001$).

Post-experiment interview and reported counting strategy

In our structured interview administered after the experiment, 20 out of 28 participants reported when prompted that they had noticed that visual and belt speeds were not always congruent. Six participants did not notice this, while two were not sure. Other mentions of things that participants noticed included different belt speeds in learning and reproduction, respectively, and pixel aliasing at the walls' edges and the virtual horizon.

When asked about how they attempted to solve the task, the most frequent response was counting steps (16 out of 28), followed by comparing belt speeds between learning and reproduction (9 mentions). Other strategies included estimating the time (5), paying special attention to the moving walls (5), and walking in a certain rhythm (3). Of these, only counting steps appeared to be descriptively an adaptive strategy, with the mean absolute error reduced from 2.9 m to 2.2 m, although a Welsh's two-sample t-test revealed that this difference was not statistically significant, $t(17.34) = 1.58$, $p = .133$. Similarly, the mean standard deviation for each condition was also descriptively lower among participants who counted their steps compared to those who did not (2.3 m to 2.9 m), but this difference was not statistically significant, either, $t(25.93) = -1.26$, $p = .219$.

To analyze whether the reported strategy of step counting was indeed effective, we compared participants who reported to have counted steps ("counters") to participants who did not count steps ("non-counters"). At a constant step length, counting steps would yield a correct distance reproduction only if the number of steps was adjusted for belt speed. Alternatively, reproducing the step count exactly would yield a correct distance reproduction only if step length scaled with belt speed. To assess whether participants used either of these strategies, we first compared the number of steps in learning to reproduction. At 0.8 m/s belt speed (Figure 6a), on average participants had a count ratio of 0.95 (0.95 for counters, and 0.94 for non-counters) between learning and reproduction, with no difference between counters and non-counters ($t(19.81) = 0.05$, $p = .960$). This is below unity, but above the 0.8 ratio (dotted line in Figure 6a) expected for constant step length. For 1 m/s (Figure 6b) the deviation between learning and training on average was 1.03, with no difference between counters and non-counters (1.00 and 1.07, respectively, $t(16.94) = 1.38$, $p = .185$). For 1.2 m/s (Figure 6c), there was a shift above unity (mean ratio: 1.11, counters: 1.08, non-counters: 1.17, no
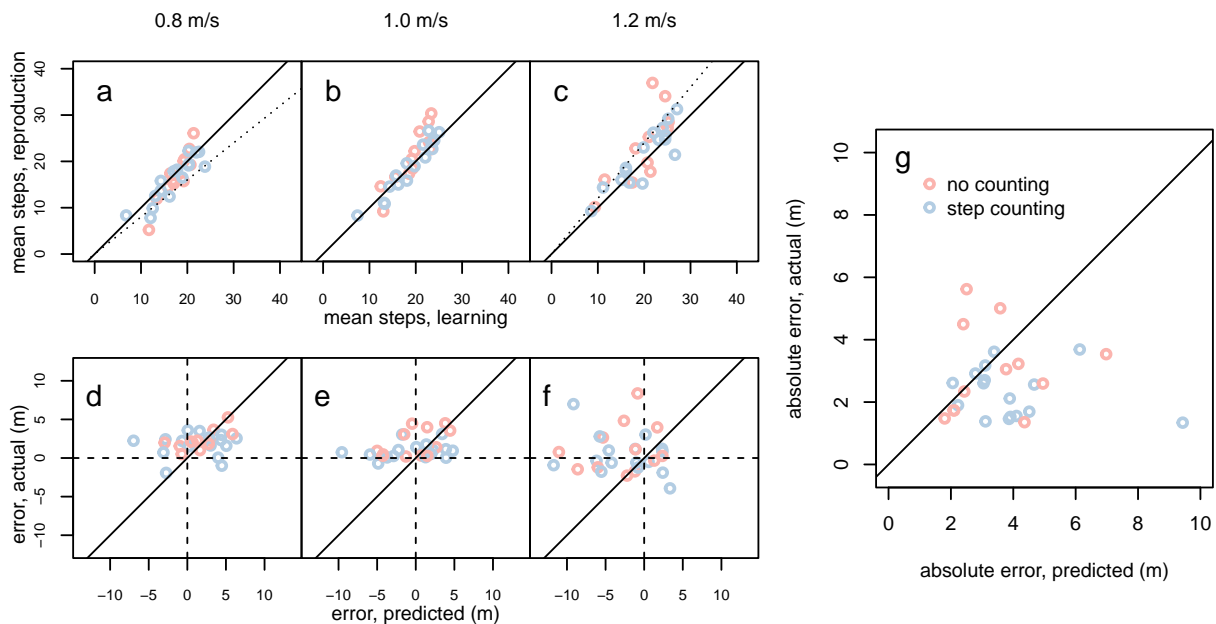
Figure 6: Steps and reproduction errors, by reported step-counting strategy. Top left, **a-c:** Mean steps taken in the learning phase (x-axis) and the reproduction phase (y-axis), split up by participant (N=28), belt speed, and by whether participants reported counting their steps (light blue) or not (light red). Solid line indicates unity, dotted line shows a slope corresponding to the ratio of learning-phase speed to reproduction-phase speed, so a participant with no error and constant step length would be on this line. Bottom left, **d-f:** Mean error (y-axis) plotted against the error predicted if each participant had in each reproduction phase merely reproduced the number of steps during learning phase with their average step length for a self-motion speed of 1 m/s (x-axis). Diagonal line indicates unity, colours as in panels a-c. **g:** Mean absolute error, by participant. Each circle represents one participant, colours as in other panels. X-axis shows the expected absolute error if each participant had in each reproduction phase merely reproduced the number of steps during learning phase with their average step length for a self-motion speed of 1 m/s. Diagonal line indicates unity.

statistically significant difference, $t(16.31) = 1.29$, $p = .215$), which remained below the ratio (1.2) needed at constant step length. In sum, the reproduction of steps was adjusted to belt speed, similarly in counters and non-counters, but not to an extent that step length would not need adjustment (see also Figure 5). In turn, the distance reproduction error could not be explained by participants reproducing the number of step counts using an average step length, as there was no systematic relation between the thus predicted error and the actual reproduction error (Figure 6d-f), consistent with the fact that step length increased slightly with higher belt speeds (Figure 5a). We neither observe a systematic over- or underestimation nor a marked difference between counters and non-counters. However, it is noteworthy that the most extreme errors arose for non-counters. This is even more evident when performing the same analysis but considering the absolute error pooled across all belt speeds (Figure 6g), which is possible as there is no systematic over- or underestimation tendency. The three participants with the highest error were non-counters, but beyond this, there are few patterns to be seen. This illustrates that counting steps was only a small aspect of our task, even among those participants who reported using it as a strategy.

Finally, we also asked participants whether they paid special attention to any part of the visual scene. Only two participants answered no, among those who responded yes, by far the most frequent responses were

314 the walls and the virtual horizon (12 each). Other participants mentioned looking at the pixel aliasing (3),
315 the clouds (2), edges of the walls (2), and one participant reported looking at the motion-tracking cameras.
316 Note, however, that responses of having paid attention to the walls were only weakly and non-significantly
317 associated with the proportion of gaze actually directed at walls (24.7 % compared to 17.2 % for other
318 participants, $t(22.33) = 0.92$, $p = .367$).

Discussion

320 When walking over a flat surface, humans typically do not need vision. They can walk fine with their
321 eyes closed, and will even have some idea of how far they have walked based on non-visual self-motion
322 cues alone, a critical bit of information for finding one's way. However, vision helps by providing additional
323 information about one's position and movement. Here, we investigated to what degree vision and self-motion
324 cues, respectively, affect performance in a distance-reproduction task, how this was related to gaze patterns,
325 and how participants walked at different speeds and to reproduce different distances. Our setup allowed
326 us to investigate these question by manipulating vision and walking speed independently in a virtual but
327 naturalistic environment in which participants could move their eyes freely through the real world while
328 walking, yet received only experimentally controlled visual cues to distance. As expected and in line with
329 previous results, we found that both visual cues and self-motion cues matter for reproducing distances.
330 This was true across a range of walking speeds, with participants being most accurate (i.e., overshooting
331 the least) at the highest tested walking speed (1.2 m/s). Self-motion cues were more reliable and weighed
332 more strongly than optic flow, with participants who directed their gaze towards visual information more
333 descriptively showing some tendency to also weigh vision more. Gaze was directed mainly towards the
334 horizon, with some participants focussing mainly on the virtual walls where a repeating pattern provided
335 cues to speed and distance, both in learning and reproduction phases. Gait patterns differed during learning
336 phases depending on belt speed, with both step length and step frequency increasing for higher speeds.
337 During reproduction, neither of these two parameters differed by trial type, as participants reproduced
338 longer distances purely by walking more steps. Indeed, our structured interview revealed step counting to
339 be a common explicit strategy.

340 We found that across participants, self-motion cues (belt speed) were weighed much more strongly than
341 visual cues (optic flow given by the virtual walls), which is consistent with prior work (Campos et al.,
342 2010, 2012) and partly explainable in a Bayesian framework: Here, one would expect different sources of
343 sensory information to be combined by weighing each according to their reliability (Landy et al., 1995),
344 and indeed, trials in which participants relied only on visual motion were descriptively more variable than
345 those where participants had only self-motion information available. The pattern could also be driven
346 partly by participants deliberately focussing primarily on the motion of the treadmill rather than the visual
347 environment, as we also note that many participants reported using some motion-cue based strategy to
348 reproduce distances, such as counting steps, but only very few reported incorporating visual cues into any
349 explicit strategy.

350 That said, both the rmANOVAs and the modelling of weights clearly show that visual cues had a small,
351 but persistent effect on reproduced distances and most participants completed the task not solely based on
352 self-motion cues – despite the fact that visual cues were presented only in a pre-defined area of the visual
353 scene. This finding that both self-motion and the motion of the visual scene matter is in line not just
354 with the distance-perception and navigation literature, but also object perception during walking. Here, it

has also been reported that both vision and non-visual self-motion cues contribute to estimation of object motion (Fajen & Matthis, 2013), and that participants' estimates are inaccurate once either is removed (Xie et al., 2020). Along these lines, we see not only that participants responded to visual information, they also reproduced distances less accurately when vision was uninformative – i.e., overshot the distance more in the control condition with visual speed set to 0 (Figure 2).

Our visual scene (Figure 1) offered three major advantages: One, we could easily manipulate its moving speed independently of the treadmill's belt, two, since only the virtual walls contained a pattern, only that part of the scene provided any optic flow and thus visual cues to speed and distance, and three, since we used a screen and not a head-mounted display, participants experienced real, non-simulated changes to visual input following head and eye movements, enhancing the scene's naturalness. The environment thus allowed us to analyse directly when participants looked towards those areas, and whether this affected their behaviour. Most participants mainly looked ahead, as they would during normal walking (Hart & Einhäuser, 2012; Matthis et al., 2018), as opposed to towards the more informative walls to sample relevant visual information (Domínguez-Zamora et al., 2018; Marigold & Patla, 2007), although walking was sufficiently easy that there was no need to look in front of one's feet. This pattern was the same in learning phases and in reproduction phases – despite considerable inter-individual variability in how much the walls were gazed towards (Figure 3), participants were remarkably consistent in their individual tendencies. We know that humans may still detect motion well in the periphery (McKee & Nakayama, 1984) and in previous works (Banton et al., 2005) it has even been suggested that optic flow may even be more useful in the periphery as opposed to the fovea - perhaps indicative of participants using lamellar flow. Fixating the centre of the screen also has the benefit that optic flow is present in both sides of the visual field rather than just one, albeit peripherally. Note, however, that while humans can detect self-motion from both, perceiving optic flow centrally and perceiving lamellar flow in the periphery are considered separate systems that have different properties with respect to both detecting direction (Crowell & Banks, 1993; Warren & Kurtz, 1992) and detecting speed or distance (Harris et al., 2012; McManus et al., 2017). We also would have expected the relatively fine-grained textures of our virtual walls to make fixation necessary to extract movement, so another possible interpretation is that looking towards them briefly was sufficient. Critically, there was only a non-significant correlation between the proportion of time a participant looked towards the walls and how strongly they weighed visual information. In contrast, if participants relied heavily on the information gained from fixating the walls to perform the task, a much clearer relationship would be predicted. This may, however, deserve further investigation. We also found no evidence of an advantage of optic flow in the periphery or the fovea, as the proportion of gaze towards the walls was virtually uncorrelated with performance. We should note that participants were not instructed to look towards any part of the visual scene in particular, and in fact not even about the fact that the visual scene could provide information beyond the belt. There is also the practical consideration that walking on a treadmill necessitates very straight walking, which can be difficult when looking far to one side, as participants typically tend to veer to the side they gaze towards (Cutting et al., 2002). Certain participants nevertheless reported using the virtual walls as a cue for the task, but some others did not even notice the discrepancy between visual and belt speed, even with gain values of 1.5 and 0.667. Note that it has been reported that participants tend to perceive a slightly faster visual scene (i.e., gain > 1) as equal to a given belt speed (Caramenti et al., 2018; Durgin et al., 2005; Pelah & Barlow, 1996), but this would only explain difficulties detecting the mismatch for gain $= 1.5$, not for gain $= 0.667$.

We also found that participants tended to overshoot distances, consistent with previous results (Banton et al., 2005; Mittelstaedt & Mittelstaedt, 2001) – though this tendency is task-dependent (Lappe et al.,

2007) – while also replicating the finding that accuracy differs by walking speed, as average overshoot was highest in trials with 0.8 m/s belt speed and almost zero in 1.2 m/s trials. Previously, it has been suggested that participants may perform best at speeds they are most accustomed to (Mittelstaedt & Mittelstaedt, 2001), which would be plausible for the speeds tested here. Gait patterns reflected the different speeds in the learning phases, as both average step length and average step frequency increased at higher speeds. The same was not true for reproduction phases, as participants adjusted the distance walked purely through the number of steps, which is unsurprising given that walking speed – the main variable affecting step frequency (Grieve & Gear, 1966) – was constant in reproduction phases. In fact, counting steps was by far the most frequently reported strategy to solve our task. We note that for this strategy, adjusting step length during learning phases is actually not adaptive, as a constant step length would allow this strategy to work across different speeds, and such behaviour had previously been reported (Durgin et al., 2007).

On the topic of strategies, another frequent strategy other than counting steps was looking at the virtual walls. A number of other strategies were mentioned, such as paying attention to the cadence of walking, but, interestingly, most showed at best a very moderate relation to participants' behaviour. Even reporting that one paid attention to the walls was not very strongly correlated with time directing one's gaze towards the wall, which is in line with previous findings that participants often misjudge where they looked (Võ et al., 2016), and only step counting showed any measurable performance benefit to the participants. Thus, while the results of our structured interview are a valuable window into how participants went about the reproduction task, they do not tell us much about which ways would have been more beneficial – in line with the fact that walking is, in large part, not cognitively controlled (Varraine et al., 2002). In fact, gait patterns were remarkably unresponsive to any manipulations other than belt speed, as even though our visual gain had clear effects on the reproduced distances, none were visible on gait.

There are, of course, a number of factors that limit the generalisability of our study. For one, both the use of cues (Harrison et al., 2022) and sampling of visual information in walking (Tong et al., 2017) are task-dependent. We used a reproduction task here, as we see it as the most ecologically valid task, but of course, an estimation task may have produced different patterns. There is also our use of a treadmill and a very specific virtual scene. Treadmill walking has been shown to differ from regular walking (Dingwell et al., 2001), and specifically also with respect to distance and speed perception (Durgin et al., 2007; Solini et al., 2021). We would argue that our environment is rather high-fidelity for a virtual environment as it contains not just optic-flow patterns but a real environment in which participants can move their heads and eyes, and we have previously found patterns very similar to real-world situations in this very setting (Kopiske et al., 2021). That said, another reason to vary the complexity of the visual scene might be to investigate this factor's impact on eye movements – while our design had the advantage of allowing us to see clearly when participants fixated the informative areas of the scene, i.e., the walls, this could have made eye movements less natural. Another obvious variation of the present experiment that could allow broader generalisation would be to have participants choose their own walking pace, both in learning phases in certain trials and in reproduction phases. This would allow them to employ more strategies, eliminate the difficulty of choosing speeds that work for all participants (nearly) equally well and enhance performance (Mittelstaedt & Mittelstaedt, 2001), and provide an even more natural setting, thereby promising many further insights into how humans are able to walk a certain predefined distance. In such a design, a strategy such as step counting might in fact be even more adaptive as the belt would not force participants towards a certain gait – this would remain to be investigated, perhaps in conjunction with a secondary task to prevent step counting.

441   We conclude that humans use both visual cues and non-visual self-motion cues to reproduce distances.
442 Quantifying their relative contributions showed the greater importance of non-visual self-motion cues in the
443 task we employed. Gaze measures were correlated with responses and cue weights, but only weakly and non-
444 significantly, as were strategies. These findings shed an important light on sensory integration, information
445 sampling, and the use of strategies in distance reproduction.

446 Acknowledgements

Appendix

Following the main experiment, we conducted a structured interview with each participant. The purpose of this interview was to assess whether participants were aware of the manipulations in the main experiment, and whether they had used deliberate strategies to complete the main task. The original questions (in German) are provided, in the order they were administered, along with their English translations in Table 3.

Table 3: Questions of the structured interview administered directly after the experiment.

German (original)	English translation
Ist Ihnen irgendetwas aufgefallen?	Did you notice anything?
Welche Strategien haben Sie angewandt, um die Aufgabe zu lösen?	Which strategies did you use to complete the task?
Haben Sie auf einen speziellen Teil der Landschaft geachtet?	Did you pay attention to a particular part of the scene?
Haben Sie irgendwo besonders hingeschaut?	Did you look anywhere in particular?
Haben Sie das während des Versuchs geändert?	Did you change this over the course of the experiment?
Haben Sie versucht, Ihr Gehen der Aufgabe anzupassen?	Did you try to adapt your gait to the task?
Haben Sie bemerkt, dass die Landschaft sich anders bewegt hat als das Laufband?	Did you notice that the environment moved differently from the treadmill?
Wenn ja: langsamer oder schneller?	If so: faster or slower?
Haben Sie versucht, Ihre Schritte zu zählen?	Did you try to count your steps?

References

Banton, T., Stefanucci, J., Durgin, F., Fass, A., & Proffitt, D. (2005). The perception of walking speed in a virtual environment. *Presence: Teleoperators and Virtual Environments*, *14*(4), 394–406. https://doi.org/10.1162/105474605774785262

Bremmer, F., & Lappe, M. (1999). The use of optical velocities for distance discrimination and reproduction during visually simulated self motion. *Experimental Brain Research*, *127*(1), 33–42. https://doi.org/10.1007/s002210050771

Bruggeman, H., Zosh, W., & Warren, W. H. (2007). Optic flow drives human visuo-locomotor adaptation. *Current Biology*, *17*, 2035–2040. https://doi.org/10.1016/j.cub.2007.10.059

Campos, J. L., Butler, J. S., & Bülthoff, H. H. (2012). Multisensory integration in the estimation of walked distances. *Experimental Brain Research*, *218*(4), 551–565. https://doi.org/10.1007/s00221-012-3048-1

Campos, J. L., Byrne, P., & Sun, H.-J. (2010). The brain weights body-based cues higher than vision when estimating walked distances. *European Journal of Neuroscience*, *31*(10), 1889–1898. https://doi.org/10.1111/j.1460-9568.2010.07212.x

Caramenti, M., Lafortuna, C. L., Mugellini, E., Abou Khaled, O., Bresciani, J. P., & Dubois, A. (2018). Matching optical flow to motor speed in virtual reality while running on a treadmill. *PLOS ONE*, *13*(4), e0195781. https://doi.org/10.1371/journal.pone.0195781

Chen, X., McNamara, T. P., Kelly, J. W., & Wolbers, T. (2017). Cue combination in human spatial navigation. *Cognitive Psychology*, *95*, 105–144. https://doi.org/10.1016/j.cogpsych.2017.04.003

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Psychology Press.

Crowell, J. A., & Banks, M. S. (1993). Perceiving heading with different retinal regions and types of optic flow. *Perception & Psychophysics*, *53*(3), 325–337. https://doi.org/10.3758/BF03205187

Cutting, J. E., Readinger, W. Q., & Wang, R. F. (2002). Walking, looking to the side, and taking curved paths. *Perception & Psychophysics*, *64*, 415–425.

Dingwell, J. B., Cusumano, J. P., Cavanagh, P. R., & Sternad, D. (2001). Local dynamic stability versus kinematic variability of continuous overground and treadmill walking. *Journal of Biomechanical Engineering*, *123*, 27–32. https://doi.org/10.1115/1.1336798

Domínguez-Zamora, F. J., Gunn, S. M., & Marigold, D. S. (2018). Adaptive gaze strategies to reduce environmental uncertainty during a sequential visuomotor behaviour. *Scientific Reports*, *8*(1), 14112. https://doi.org/10.1038/s41598-018-32504-0

Duda, R. O., & Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, *15*(1), 11–15. https://doi.org/10.1145/361237.361242

Durgin, F. H. (2009). When walking makes perception better. *Current Directions in Psychological Science*, *18*(1), 43–47.

Durgin, F. H., Gigone, K., & Scott, R. (2005). The perception of visual speed while moving. *Journal of Experimental Psychology: Human Perception & Performance*, *31*(2), 339–353. https://doi.org/10.1037/0096-1523.31.2.339

Durgin, F. H., Reed, C., & Tigue, C. (2007). Step frequency and perceived self-motion. *ACM Transactions on Applied Perception*, *4*(1), 5. https://doi.org/10.1145/1227134.1227139

Fajen, B. R., & Matthis, J. S. (2013). Visual and non-visual contributions to the perception of object motion during self-motion. *PLoS ONE*, *8*(2), e55446. https://doi.org/10.1371/journal.pone.0055446

Frenz, H., & Lappe, M. (2005). Absolute travel distance from optic flow. *Vision Research*, *45*(13), 1679–1692. https://doi.org/10.1016/j.visres.2004.12.019

Grieve, D. W., & Gear, R. J. (1966). The relationships between length of stride, step frequency, time of swing and speed of walking for children and adults. *Ergonomics*, *9*(5), 379–399. https://doi.org/10.1080/00140136608964399

Harris, L. R., Herpers, R., Jenkin, M., Allison, R. S., Jenkin, H., Kapralos, B., Scherfgen, D., & Felsner, S. (2012). The relative contributions of radial and laminar optic flow to the perception of linear self-motion. *Journal of Vision*, *12*(10), 7–7. https://doi.org/10.1167/12.10.7

Harrison, S. J., Reynolds, N., Bishoff, B., Stergiou, N., & White, E. (2022). Homing tasks and distance matching tasks reveal different types of perceptual variables associated with perceiving self-motion

during over-ground locomotion. *Experimental Brain Research*, *240*(4), 1257–1266. https://doi.org/10.1007/s00221-022-06337-3

Hart, B. M. 't, & Einhäuser, W. (2012). Mind the step: Complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental Brain Research*, *223*(2), 233–249. https://doi.org/10.1007/s00221-012-3254-x

Hollands, M. A., Marple-Horvat, D. E., Henkes, S., & Rowan, A. K. (1995). Human eye movements during visually guided stepping. *Journal of Motor Behavior*, *27*(2), 155–163.

Israël, I., Grasso, R., Georges-François, P., Tsuzuku, T., & Berthoz, A. (1997). Spatial memory and path integration studied by self-driven passive linear displacement. I. Basic properties. *Journal of Neurophysiology*, *77*(6), 3180–3192. https://doi.org/10.1152/jn.1997.77.6.3180

Klatzky, R. L., Loomis, J. M., Golledge, R. G., Cicinelli, J. G., Doherty, S., & Pellegrino, J. W. (1990). Acquisition of route and survey knowledge in the absence of vision. *Journal of Motor Behavior*, *22*(1), 19–43. https://doi.org/10.1080/00222895.1990.10735500

Kopiske, K., Koska, D., Baumann, T., Maiwald, C., & Einhäuser, W. (2021). Icy road ahead—rapid adjustments of gaze–gait interactions during perturbed naturalistic walking. *Journal of Vision*, *21*(8), 11. https://doi.org/10.1167/jov.21.8.11

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*(3), 389–412.

Lappe, M., Bremmer, F., & van den Berg, A. V. (1999). Perception of self-motion from visual flow. *Trends in Cognitive Sciences*, *3*(9), 329–336. https://doi.org/10.1016/S1364-6613(99)01364-9

Lappe, M., & Frenz, H. (2009). Visual estimation of travel distance during walking. *Experimental Brain Research*, *199*(3-4), 369–375. https://doi.org/10.1007/s00221-009-1890-6

Lappe, M., Jenkin, M., & Harris, L. R. (2007). Travel distance estimation from visual motion by leaky path integration. *Experimental Brain Research*, *180*(1), 35–48. https://doi.org/10.1007/s00221-006-0835-6

Marigold, D. S., & Patla, A. E. (2007). Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience*, *144*, 302–313. https://doi.org/10.1016/j.neuroscience.2006.09.006

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, *28*, 1224–1233.e5. https://doi.org/10.1016/j.cub.2018.03.008

McGinley, J. L., Baker, R., Wolfe, R., & Morris, M. E. (2009). The reliability of three-dimensional kinematic gait measurements: A systematic review. *Gait and Posture*, *29*, 360–369. https://doi.org/10.1016/j.gaitpost.2008.09.003

McKee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision Research*, *24*, 25–32. https://doi.org/10.1016/0042-6989(84)90140-8

McManus, M., D'Amour, S., & Harris, L. R. (2017). Using optic flow in the far peripheral field. *Journal of Vision*, *17*(8), 3. https://doi.org/10.1167/17.8.3

Mittelstaedt, M. L., & Glasauer, S. (1991). Idiothetic navigation in gerbils and humans. *Zoologische Jahrbücher: Abteilung Für Allgemeine Zoologie Und Physiologie Der Tiere*, *95*, 427–435.

Mittelstaedt, M. L., & Mittelstaedt, H. (2001). Idiothetic navigation in humans: Estimation of path length. *Experimental Brain Research*, *139*(3), 318–332. https://doi.org/10.1007/s002210100735

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*(9), 689–693. https://doi.org/10.1016/j.cub.2008.04.021

Pelah, A., & Barlow, H. B. (1996). Visual illusion from running. *Nature*, *381*, 283.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285–293. https://doi.org/10.1016/j.tics.2015.03.002

Prokop, T., Schubert, M., & Berger, W. (1997). Visual influence on human locomotion: Modulation to changes in optic flow. *Experimental Brain Research*, *114*(1), 63–70. https://doi.org/10.1007/PL00005624

R Core Team. (2022). *R: A language and environment for statistical computing.* https://www.R-project.org/

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Solini, H. M., Bhargava, A., & Pagano, C. C. (2021). The effects of testing environment, experimental design, and ankle loading on calibration to perturbed optic flow during locomotion. *Attention,*

Perception, & Psychophysics, *83*(1), 497–511. https://doi.org/10.3758/s13414-020-02200-1

Sun, H. J., Campos, J. L., Young, M., Chan, G. S. W., & Ellard, C. G. (2004). The contributions of static visual cues, nonvisual cues, and optic flow in distance estimation. *Perception*, *33*(1), 49–65. https://doi.org/10.1068/p5145

Terrazas, A., Krause, M., Lipa, P., Gothard, K. M., Carnes, C. A., & McNaughton, B. L. (2005). Self-motion and the hippocampal spatial metric. *Journal of Neuroscience*, *25*(35), 8085–8096. https://doi.org/10.1523/JNEUROSCI.0693-05.2005

Thomson, J. A. (1980). How do we use visual information to control locomotion? *Trends in Neurosciences*, *3*, 247–250.

Tong, M. H., Zohar, O., & Hayhoe, M. M. (2017). Control of gaze while walking: Task structure, reward, and uncertainty. *Journal of Vision*, *17(1)*(28), 1–19. https://doi.org/10.1167/17.1.28

Varraine, E., Bonnard, M., & Pailhous, J. (2002). Interaction between different sensory cues in the control of human gait. *Experimental Brain Research*, *142*(3), 374–384. https://doi.org/10.1007/s00221-001-0934-3

Võ, M. L. H., Aizenman, A. M., & Wolfe, J. M. (2016). You think you know where you looked? You better look again. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1477–1481. https://doi.org/10.1037/xhp0000264

Warren, W. H., & Kurtz, K. J. (1992). The role of central and peripheral vision in perceiving the direction of self-motion. *Perception & Psychophysics*, *51*(5), 443–454. https://doi.org/10.3758/BF03211640

Winter, S. S., Mehlman, M. L., Clark, B. J., & Taube, J. S. (2015). Passive transport disrupts grid signals in the parahippocampal cortex. *Current Biology*, *25*(19), 2493–2502. https://doi.org/10.1016/j.cub.2015.08.034

Xie, M., Niehorster, D. C., Lappe, M., & Li, L. (2020). Roles of visual and non-visual information in the perception of scene-relative object motion during walking. *Journal of Vision*, *20*(10), 15. https://doi.org/10.1167/jov.20.10.15