

# Foothold selection during locomotion in uneven terrain: Results from the integration of eye tracking, motion capture, and photogrammetry

Reviewed Preprint

v2 • September 24, 2024


Revised by authors

Reviewed Preprint

v1 • October 23, 2023

Karl S Muller, Dan Panfili, Stephanie Shields, Jonathan S Matthis, Kathryn Bonnen, Mary M Hayhoe 

Center for Perceptual Systems, The University of Texas at Austin, Austin, United States • School of Optometry, Indiana University, Bloomington, United States • Department of Biology, Northeastern University, Boston, United States

 [https://en.wikipedia.org/wiki/Open\\_access](https://en.wikipedia.org/wiki/Open_access)
 Copyright information

## Abstract

Relatively little is known about the way vision is used to guide locomotion in the natural world. What visual features are used to choose paths in natural complex terrain? To answer this question, we measured eye and body movements while participants walked in natural outdoor environments. We incorporated measurements of the 3D terrain structure into our analyses and reconstructed the terrain along the walker's path, applying photogrammetry techniques to the eyetracker's scene camera videos. Combining these reconstructions with the walker's body movements, we demonstrate that walkers take terrain structure into account when selecting paths through an environment. We find that they change direction to avoid taking steeper steps that involve large height changes, instead of choosing more circuitous, relatively flat paths. Our data suggest walkers plan the location of individual footholds and plan ahead to select flatter paths. These results provide evidence that locomotor behavior in natural environments is controlled by decision mechanisms that account for multiple factors, including sensory and motor information, costs, and path planning.

### eLife assessment

This **fundamental** study has the potential to substantially advance our understanding of human locomotion in complex real-world settings and opens up new approaches to studying (visually guided) behavior in natural settings outside the lab. The evidence supporting the conclusions is overall **compelling**. Whereas detailed analyses represent multiple ways to visualize and quantify the rich and complex natural behavior, some of the specific conclusions remain more suggestive at this point. The work will be of interest to neuroscientists, kinesiologists, computer scientists, and engineers working on human locomotion.

<https://doi.org/10.7554/eLife.91243.2.sa4>

## Introduction

Sensory input guides actions, and in turn, those actions shape the sensory input. Consequently, to develop a sophisticated scientific understanding of even simple actions in the natural world, we must monitor both the sensory input and the actions. While technology for monitoring gaze and body position during natural behavior is both readily available and widely used in vision science and in movement science, the use of technology for the measurement of the visual input in natural environments has been limited. In this paper, we aim to bridge that gap by using photogrammetry techniques from computer vision to reconstruct the environment and subsequently approximate the visual input. The combination of the reconstructions with body pose information and gaze data allows a full specification of how walkers interact with complex real-world environments. These data can help expand our understanding of how visual information about the structure of the environment drives locomotion via sensorimotor decision-making.

Natural visually-guided behaviors, like visually-guided walking, can be characterized as a sequence of complex sensorimotor decisions (Hayhoe, 2017 [↗](#); Gallivan et al., 2018 [↗](#); Domínguez-Zamora and Marigold, 2021 [↗](#)). However, much of our current understanding of locomotion comes from work characterizing steady state walking in laboratory settings — most commonly with participants walking on treadmills. That work has shown that humans converge towards energetic optima. For example, walkers adopt a preferred gait that constitutes an energetic minimum given their own biomechanics (Warren, 1984 [↗](#); Warren et al., 1986 [↗](#); Kuo et al., 2005 [↗](#); Selinger et al., 2015 [↗](#); Finley et al., 2013 [↗](#); Lee and Harris, 2018 [↗](#); Rock et al., 2018 [↗](#); Yokoyama et al., 2018 [↗](#); O'Connor et al., 2012 [↗](#)).

There are a number of problems in generalizing these findings to walking in natural environments. In particular, locomotion over rough terrain depends on both the biomechanics of the walker and visual information about the structure of the environment. When the terrain is more complex, walkers use visual information to find stable footholds (Matthis et al., 2018 [↗](#)). There are also other factors to consider, such as the need to reach a goal or attend to the navigational context (Warren et al., 2001 [↗](#); Rio et al., 2014 [↗](#); Logan et al., 2010 [↗](#); Patla and Vickers, 1997 [↗](#)). Thus, the sensorimotor decisions in natural locomotion will be shaped by more complex cost functions than in treadmill walking. Furthermore, in the face of this complexity, individuals may be adopting heuristics rather than converging upon optimal solutions.

Previous studies tracking the eyes during outdoor walking have found that gaze patterns change with the demands of the terrain (Pelz and Rothkopf, 2007 [↗](#); Foulsham et al., 2011 [↗](#); 't Hart and Einhäuser, 2012 [↗](#)). However, in those studies, foot placement was not measured, making it impossible to analyze the relationship between gaze and foot placement. Recent work by Matthis et al. (2018 [↗](#)) and Bonnen et al. (2021 [↗](#)) integrated gaze and body measurements of walking in outdoor environments. Those papers demonstrated that walkers modulate gait speed in order to gather visual information necessary for the selection of stable footholds as the terrain became more irregular. Walkers spent more time looking at the ground close to their body (2–3 steps ahead) with increasing terrain complexity. While gaze and gait were tightly linked, the absence of terrain measurements made it impossible to ask what visual terrain features walkers use to choose footholds and navigate toward the goal.

In this paper, we ask how vision is used to identify viable footholds and choose paths in natural environments. In particular, what are the visual features of the terrain that underlie path choice? How do walkers use visual information to alter the preferred gait cycle appropriately for the upcoming path? We accomplish this by reconstructing the terrain and aligning the gaze and gait data to that reconstruction. Then we perform a series of analyses of walkers' body movements and the terrain, demonstrating that: (1) depth information available to walkers is predictive of

upcoming footholds; (2) walkers prefer flatter paths, and (3) walkers choose indirect routes to avoid height changes. These findings shed light on how walkers use visual information to find stable footholds and choose paths, a crucial everyday function of the visual system.

## Results

We analyzed data recorded while participants walked over rough terrain ( $n=9$ ). The data were collected by the authors for two separate, previously-published studies of visually guided walking (Bonnen et al. 2021 [↗](#),  $n=7$ ; Matthis et al. 2022 [↗](#),  $n=2$ ). Walkers' eye and body movements were recorded using a Pupil Labs Core mobile binocular eye tracker and a Motion Shadow full-body motion capture suit. Additionally, the walker's view of the scene was recorded by the eye tracker's outward-facing scene camera. As the scene camera moves with the head, the camera's view of the terrain changes along with the walker's. Due to those changes in viewpoint, the scene videos contain information about the terrain's depth structure, which we aimed to recover via photogrammetry.

### Terrain reconstruction

To reconstruct the 3D environment from the scene videos recorded by the eye tracker's scene camera, we used the photogrammetry software package Meshroom (Griwodz et al., 2021 [↗](#)), which combines multiple image processing and computer vision algorithms. The terrain reconstruction procedure uses the viewpoint changes across each scene video's frames to recover the environment's depth structure. The outputs, generated per scene video, are (1) a 3D textured mesh of the terrain and (2) an estimate of the 6D camera pose (both 3D location and 3D orientation) within the terrain's coordinate system. To give a sense of the quality of these reconstructions, **Figure 1** [↗](#) shows an example comparison of (A) the original scene camera image and (B) a corresponding view of the reconstructed terrain.

### Aligned gaze, body, and terrain data

We aimed to analyze the body and gaze data in the context of the reconstructed environment, which meant that we needed to align our data on the walker's movements to the terrain's coordinate system. We determined how to position the eye and body movement data relative to the terrain by aligning the head pose measured by the motion capture system to the estimated scene camera pose (**Figure 2** [↗](#)). To visualize the fully aligned data, we created videos showing the walker's skeleton moving through the associated textured terrain mesh (for an example, see Supplementary Video 1).

We also visualized the different paths that walkers took through the terrain. **Figure 3** [↗](#) shows an overhead view of the reconstructed terrains from the Austin dataset, with the paths chosen by the two Austin subjects overlaid onto the terrain. (For examples from the Berkeley dataset, see **Supplementary Figure 13** [↗](#).) The recorded paths were certainly not identical, indicating that foothold locations were not highly constrained. However, the two subjects' paths show considerable regularities. Visual inspection of the paths, particularly in 3D, gives the impression that the terrain's structure impacts the regularity of paths. In other words, features of the 3D environment might impact the degree of variability between paths, suggesting that there may be some identifiable visual features that underlie path choices.

To further illustrate the information present in our dataset, **Figure 4** [↗](#) shows an excerpt of the terrain from **Figure 3** [↗](#) with the following aligned data: gaze locations (green and blue dots), foothold locations (pink dots), and head locations (orange dots).

Figure 1.

### Example comparison of original and rendered video frames

We used the scene videos recorded by the eye tracker's outward-facing camera to estimate the structure of the environment and the scene camera's pose in each frame of the video. By moving a virtual camera to those poses and rendering the camera's view of the textured mesh, we can generate comparison images to help assess the reconstruction's accuracy. A. Frame from original scene video. B. Corresponding rendered image.

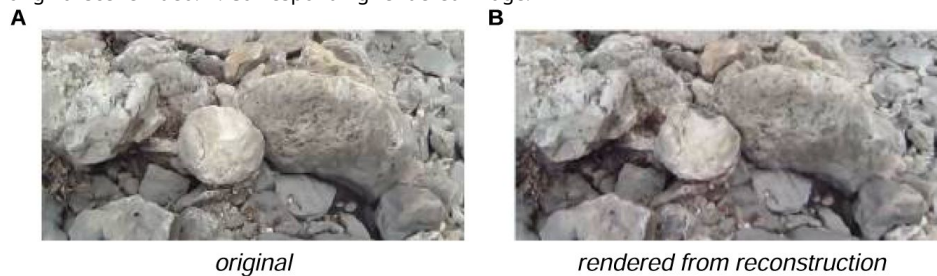
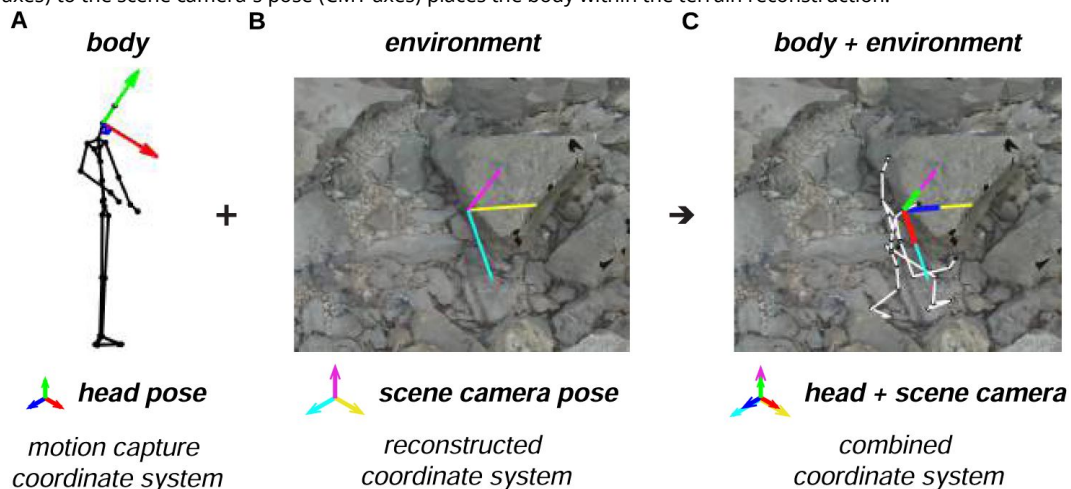
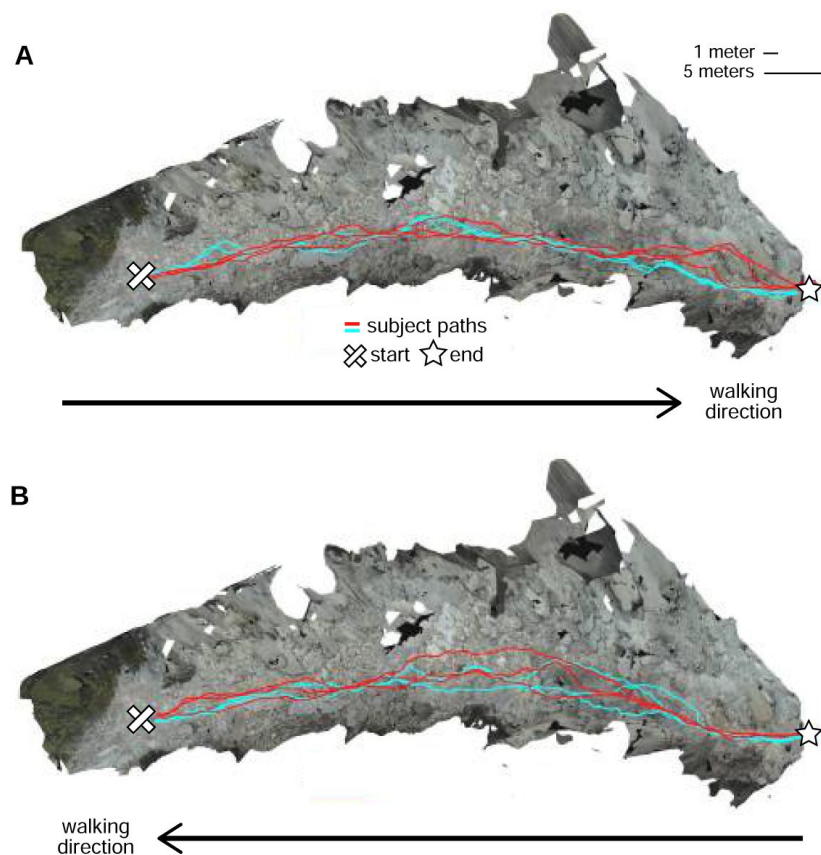


Figure 2.

### Alignment of motion capture data and terrain reconstruction

We combine the motion capture data with the reconstructed environment (photogrammetry data) by aligning the head's pose (RGB axes) to the scene camera's pose (CMY axes). A. Motion capture data provides body pose (i.e., position and orientation) information, including the head's pose (RGB axes). B. The process of reconstructing the environment via photogrammetry produces a 3D terrain mesh (image) and scene camera's poses (CMY axes). C. Aligning the head's pose (RGB axes) to the scene camera's pose (CMY axes) places the body within the terrain reconstruction.

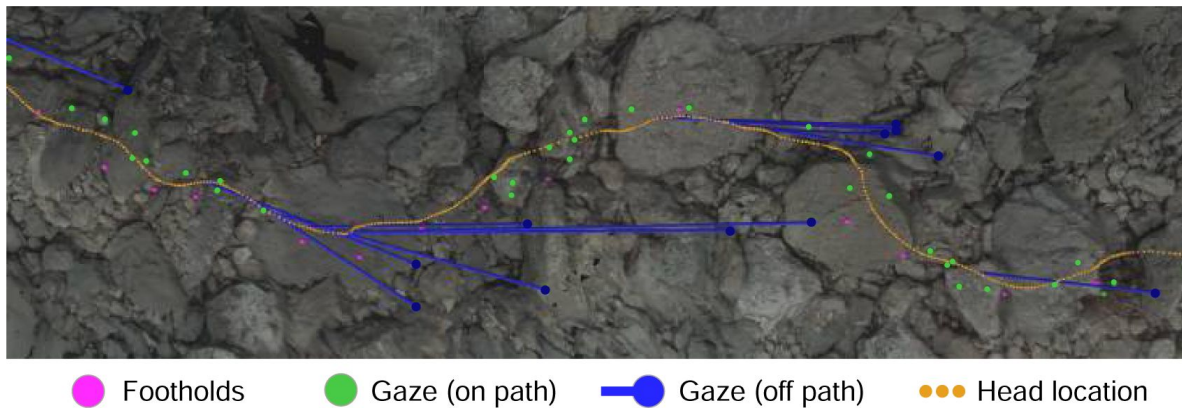




**Figure 3.**

### Repeated walks across rough terrain in the Austin data

Each of the two Austin subjects walked out and back over a rocky trail 3 times. These overhead views show the textured 3D terrain mesh along with the paths the subjects took through that terrain. Each color (red and cyan) corresponds to a different subject. Note that in some sections of the terrain, paths were highly similar across repetitions and across subjects, while in other sections, paths differed notably. A. Both subjects' 3 walks from the start of the path to the end. B. Both subjects' 3 walks returning to the start location.



**Figure 4.**

#### **Gaze and body data embedded in the corresponding reconstructed terrain**

This overhead view shows a representative excerpt of 20 steps from one of the Austin traversals. The walker was, in this overhead view, moving left to right. Dots mark the footholds locations (pink), gaze locations near the path (green), gaze locations off the path (blue), and head locations (orange). To illustrate the relationship between “off-the-path” gaze and head location, blue lines connect each blue gaze point to the simultaneous location of the head.



The close relationship between the gaze locations near the path (green dots) and the foothold locations was the concern of the investigations that generated our dataset: [Matthis et al. \(2018\)](#) and [Bonnen et al. \(2021\)](#). Those studies showed that gaze was clustered most densely in the region 2-3 steps ahead of the walker's current foothold and ranged between 1 and 5 steps ahead. In other words, we previously found that walkers look close to the locations where the feet will be placed, up to 5 steps ahead of their current location.

Relevant to the work in this article are the gaze locations “off-the-path” (blue dots) and the concurrent head locations (connected by blue lines). Those gaze points are off of the walker's chosen path but are still on the ground. Further, they seem to precede turns — a pattern which we observed throughout the data. In later sections, we provide evidence that walkers make a trade-off between maintaining a straight path versus maintaining a flat path; this gaze pattern points to how the visual system might collect the information used to make that trade-off.

## Benefits of reconstructing terrain

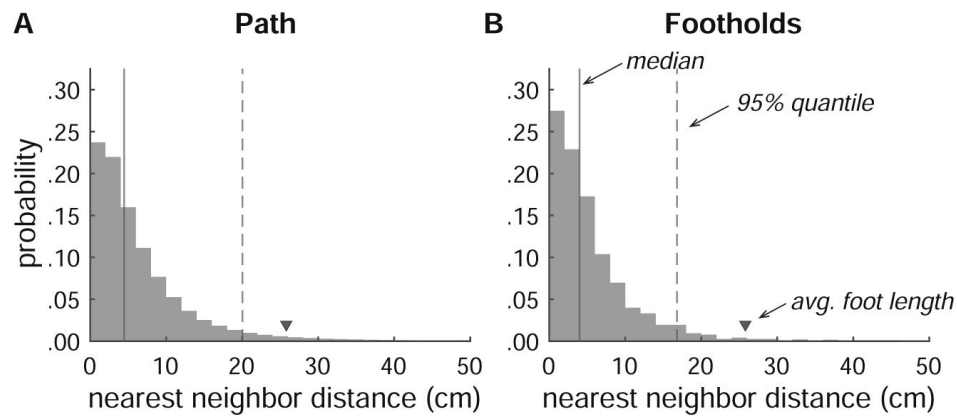
Incorporating the reconstructed terrain into our analyses has several advantages. Perhaps the most obvious is that having information about the terrain's depth structure allows us to analyze the relationship between that depth structure and the walkers' chosen foothold locations. Another major advantage of incorporating the terrain reconstruction is that it enables more accurate gaze and body localization. In previous work, we assumed a flat ground plane, which led to parallax error in gaze location estimates ([Matthis et al., 2018](#)). Here, we used the reconstructed 3D terrain surface, eliminating the need to assume a flat ground plane and thus eliminating that source of error. Additionally, body position estimates were previously negatively impacted by inertial measurement unit (IMU) drift, which results from the accumulation of small errors in the accelerometer and gyroscope measurements over time. This drift causes global, not local error — i.e., error in the overall localization of the motion capture skeleton, not in the localization of different body parts relative to one another. We were able to address this source of error by fixing the body's reference frame to that of the environment (**Figure 2**). Thus, by eliminating both of these sources of error, utilizing photogrammetry allowed us to more precisely estimate gaze and foothold locations.

## Evaluating reliability of terrain reconstruction

To evaluate the reliability of the 3D reconstruction procedure, we compared the terrain meshes calculated from multiple traversals of the same terrain. We used the Austin dataset for this reliability analysis because the terrain is contiguous, the walking paths have clearer start/end points, and there are 12 traversals of the terrain (2 subjects, 6 traversals each; see **Figure 3**).

Since we performed the reconstruction procedure on each traversal separately, we generated 12 Austin meshes that represented the same physical terrain. Each mesh contained a cloud of points, which we aligned and compared using CloudCompare (<https://www.danielgm.net/cc/>). For each pair of meshes, one mesh served as the baseline mesh, and one mesh served as the comparison mesh. For each point in the baseline mesh that was within 2 meters of the walking path, we found its nearest neighbor in the comparison mesh and calculated the distance, resulting in a distribution of distances (errors) between the two meshes.

The aggregate distribution across all pairwise mesh comparisons is shown in **Figure 5A**. The median error was 4.5cm, and the 95% quantile was 20.0cm. To evaluate reliability specifically at footholds, we also isolated the points in each baseline mesh associated with foothold locations (**Figure 5B**). For the foothold locations, the median error was 4.0cm, and the 95% quantile was 16.8cm. To put these numbers into context, the average foot length of a person from North America is 25.8cm ([Jurca et al., 2019](#)), so in both cases, the majority of mesh errors fall below 20% of the average foot length. Thus, our terrain reconstruction procedure produces reasonably reliable reconstructions of a walker's 3D environment.



**Figure 5.**

### Accuracy of terrain reconstructions

**A.** Nearest neighbor error distribution for the whole terrain (median=4.5cm, 95% quantile=20.0cm). **B.** Nearest neighbor error distribution for individual footholds (median=4.0cm, 95% quantile=16.8cm).



## Retinocentric depth information affects foothold selection

The results of [Bonnen et al. \(2021\)](#) suggest depth judgments are important in foothold finding, as the removal of depth information shifts gaze to foothold locations that are closer to the walker. Taking advantage of the output of the terrain reconstruction procedure, we sought to confirm that depth information from the walker's point of view could be used to predict the upcoming foothold locations.

We first used the reconstructed terrain — along with the aligned foothold and gaze information — to generate retinocentric depth images that approximate the visual information subjects have access to during walking (for an example, see [Figure 11](#)). Note that for each frame in the training dataset, the camera field of view includes multiple future footholds (up to 5; depicted as green circles in [Figure 6A-B](#)). We used the location of the footholds in these retinocentric depth images to create a training dataset. If a CNN can predict foothold locations above chance based on these retinocentric depth images, that would suggest that terrain depth structure plays a role in foothold selection.

Per subject, we trained the network on half of the terrain and tested on the remaining half (ensuring that the network was tested on terrain that it had not previously seen). For each depth image, we calculated the AUC, which quantifies the discriminability between image locations that show footholds and image locations that do not show footholds. [Figure 6C](#) shows that, per subject, the median AUC value for depth images from the test set was above chance. We can thus conclude that the network was able to find the potential footholds in the depth images, suggesting that retinocentric depth information contributes to foothold finding.

## Walkers prefer flatter paths

Our CNN analysis suggested that depth features in the upcoming terrain have some predictive value in the selection of footholds. We next decided to narrow our focus to examine whether terrain height, specifically, might play a role in foothold selection. Stepping up and down is energetically costly, and previous work in simpler environments has demonstrated that humans attempt to minimize energy expenditure during locomotion ([Selinger et al., 2015](#); [Finley et al., 2013](#); [Lee and Harris, 2018](#)). Furthermore, a walker avoiding large steps up and down would be choosing to take a flatter path, and walking on a flatter path would result in less deviation from their preferred gait cycle and thus more stable locomotion. To test the hypothesis that walkers seek out a flatter path to avoid large changes in terrain height, we measured the slope of steps chosen by our walkers and compared them to the slope of steps in paths we simulated along the same terrain.

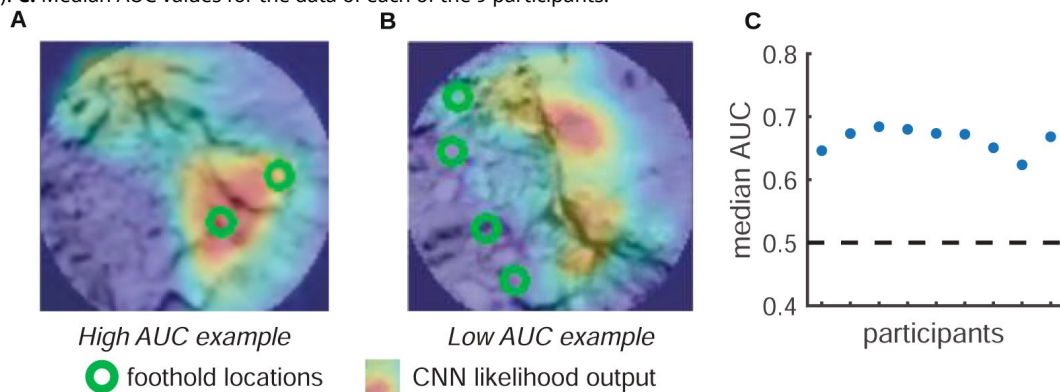
We simulated plausible paths for each walker that were comparable to their chosen paths. To ensure plausibility, we constrained potential foothold locations and potential steps based on human walking behavior. We identified potential foothold locations based on the maximum walk-on-able surface slant reported in [Kinsella-Shaw et al. \(1992\)](#), excluding areas of the terrain with an average local surface slant greater than 33°. We identified potential steps between foothold locations based on each walker's data, excluding steps with a step slope ([Equation 7](#)), step ground distance ([Equation 2](#)), and/or step deviation from goal ([Equation 5](#)) greater than the corresponding maximum absolute value for that walker's chosen steps ([Figure 7](#)). Only foothold locations and steps that met these conditions were considered viable.

To ensure comparability, we split each walker's path into 5-step (i.e., 6-foothold) segments, and we simulated corresponding 5-step sequences by starting at the walker's chosen foothold and randomly choosing each subsequent step from the available viable options ([Figure 8A](#)). (Note that, for this simulation, we did not predefine the endpoints of path segments.)

**Figure 6.**

### Predicting foothold locations from depth information

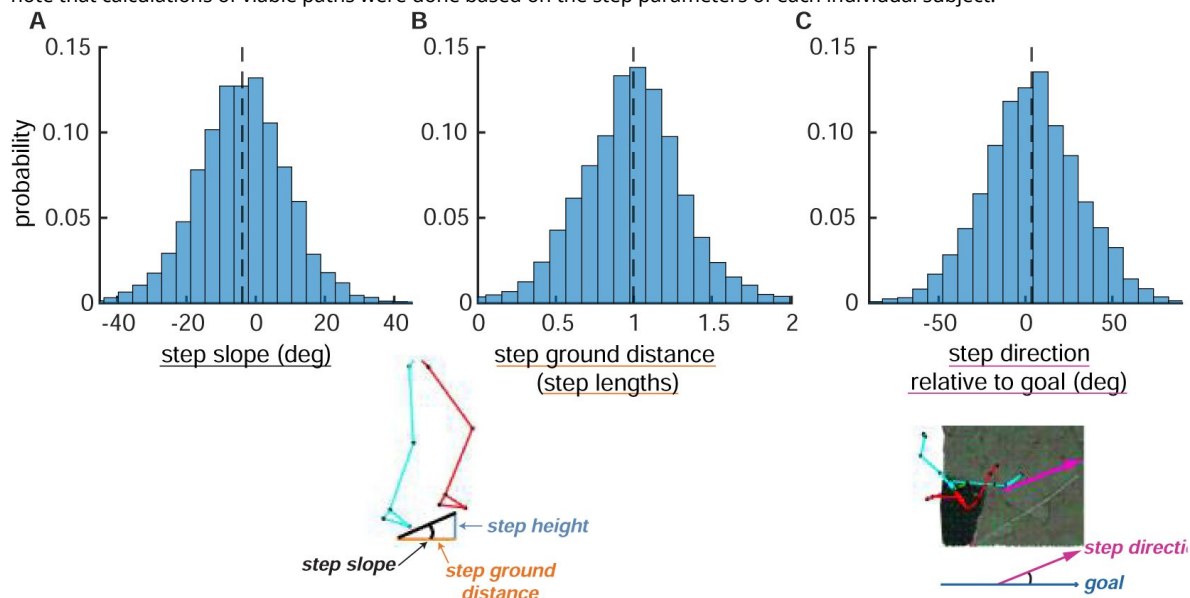
A CNN was trained to predict foothold locations in retinocentric depth images. **A.** Example retinocentric depth image associated with relatively good CNN performance (i.e., a high AUC value). The image is overlaid with the foothold locations (green) and a heatmap showing the CNN's likelihood output, which indicates the likelihood of finding a foothold in a particular location. **B.** Example retinocentric depth image associated with relatively poor CNN performance (i.e., a low AUC value). **C.** Median AUC values for the data of each of the 9 participants.

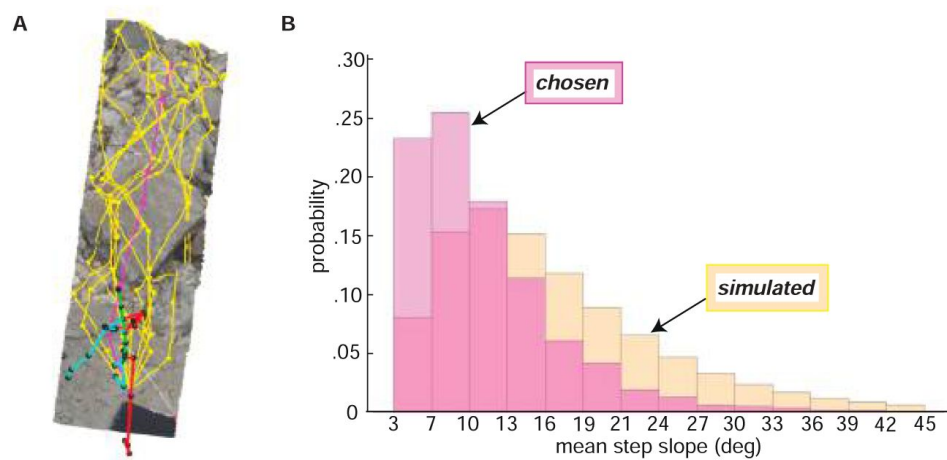


**Figure 7.**

### Step parameter distributions help define feasible alternative paths

The histograms show the distributions of (A) step slopes, (B) step lengths, and (C) step direction relative to goal direction. These distributions define the set of feasible next steps for a given foothold, allowing the calculation of feasible alternative paths to the one actually chosen by the subject. This figure shows histograms of these quantities pooled over subjects, but note that calculations of viable paths were done based on the step parameters of each individual subject.





**Figure 8.**

### Paths chosen by walkers have a lower step slope

We simulated path segments composed of viable steps and compared them to subjects' chosen step sequences. **A.** Overhead view of an example chosen step sequence (magenta), along with a subset of the corresponding simulated viable step sequences (yellow). The cyan and red lines show the walker's skeleton. **B.** Histograms of mean step slope for chosen and simulated step sequences for one participant.

After simulating walkable path segments that we could directly compare with walkers' chosen path segments, we calculated the overall slope of each path segment by averaging the absolute slope of the steps in the sequence. The resulting path segment slope values quantify the net flatness of each path. The per-subject median chosen path segment slope ranged from 7.7° to 11.8°, with a mean of 9.5° and a standard deviation of 1.7°. This corresponds to quite a small change in height; for a step of average length, a 9.5° slope corresponds to a height change of just a few inches. The per-subject median simulated path segment slope ranged from 9.1° to 19.2°, with a mean of 14.9° and a standard deviation of 3.3°.

As is evident from the per-subject medians, the slopes of chosen path segments tended to be lower than the slopes of simulated path segments, consistent with the idea that walkers seek to minimize energetic costs by taking flatter routes. The bias in the chosen path segment slope distribution toward lower slopes (vs. the simulated distribution) can be seen in **Figure 8**, where both path segment slope distributions for one example subject are plotted. For every subject, their median chosen path segment slope was lower than their median simulated path slope, with the simulated slopes being 5.56° larger on average ( $SD = 2.18^\circ$ ). A paired sample t-test confirmed that the differences between the two medians were statistically significant,  $t(8) = 7.64, p < .001$ .

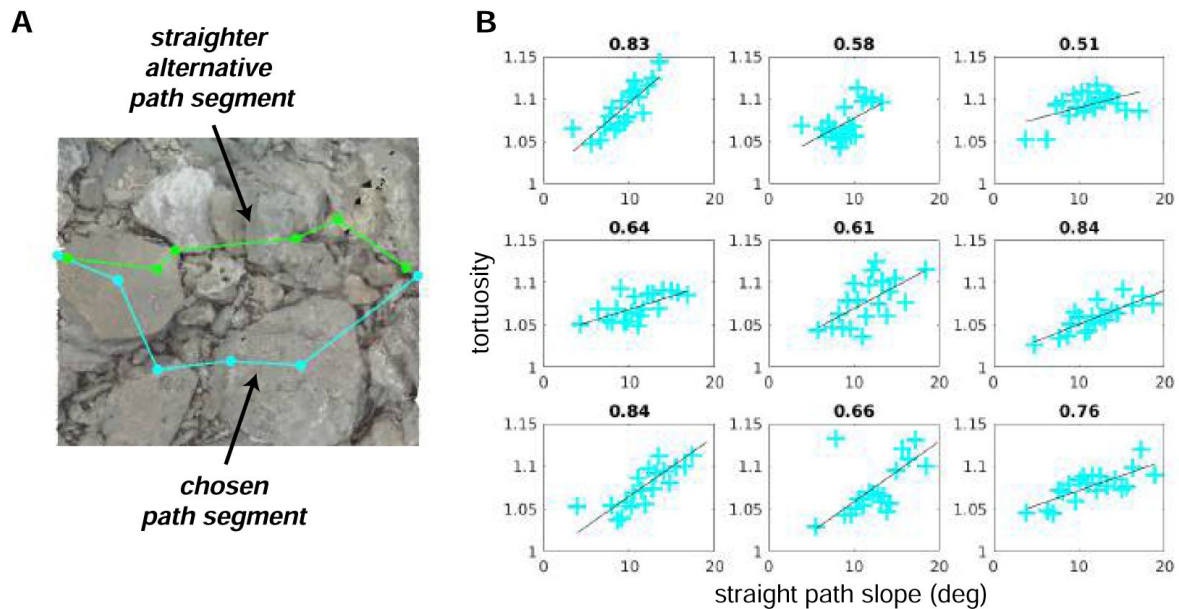
Our results suggest that walkers prefer taking flatter paths. This does not mean that they categorically avoid large steps up and down. Clearly they do sometimes choose paths with greater slopes, as is indicated by the tail of the chosen distribution. However, on average they tend toward flatter paths than would be predicted by the terrain alone.

## Walkers choose indirect routes to avoid height changes

Another factor that influences the energetic cost of taking a particular path is how straight the path is. Changing direction requires more energy than walking on an equivalent straight path (McNarry et al., 2017), as one might expect since curvier (i.e., more tortuous) paths are longer and require walkers to deviate from their preferred gait cycle. However, walkers frequently alter their direction of travel in rocky terrain (see, e.g., **Figure 4**). If there is a large height change along the straight path, turning might require less energy than stepping up or down while following the straight path. Therefore, building on our finding that walkers prefer flatter paths, we hypothesized that walkers choose to turn when turning allows them to avoid notable changes in terrain height. We evaluated that possibility by examining the relationship between the tortuosity of their chosen path segments and the slope of corresponding straight alternative path segments.

As in the prior section, we simulated path segments to compare walkers' chosen steps to the viable steps along that specific terrain. We followed a similar procedure with one notable difference: Here, for each chosen path segment, we simulated steps between the first and sixth footholds in the chosen path segment. In other words, we predefined both the starting and ending locations of simulated path segments (**Figure 9A**), whereas above, we predefined only the starting locations (**Figure 8A**).

We used the simulated path segments to quantify, for each chosen path segment, the average step slope a subject would encounter if they tried to take a straighter path between the segment's endpoints. To accomplish that, we quantified the straightness of all path segments via a tortuosity metric (Batschelet, 1981; Benhamou, 2004), and per terrain mesh, we used the tortuosity of chosen path segments to compute a conservative "straightness" threshold. We then selected the simulated path segments with a tortuosity below the computed threshold, computed their slope, and averaged across those path segment slopes. Those calculations resulted in, for each path segment, the mean slope of relatively straight alternative paths, which we refer to here as "straight path slope". **Figure 9A** shows one example straight path segment, together with the path that the subject actually chose.



**Figure 9.**

### Average tortuosity of chosen path increases with increased straight path slope

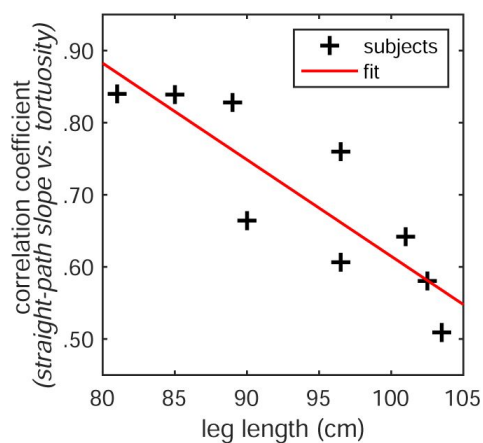
**A.** An example chosen path segment (cyan; 5-step sequence), along with one straighter alternative path segment (green). **B.** An illustration of the relationship between chosen path segment tortuosity and the slope of “straight” path segments that were simulated across the same terrain. Each subpanel depicts one subject’s data. To summarize the large amount of data per subject (317–497 path segments), we binned the data into 20 quantiles of straight path slope and averaged tortuosity per bin, generating one summary tortuosity value per slope level. These scatterplots show the average tortuosity as a function of the straight path slope quantile (cyan crosses), along with best fit lines (black). (For scatterplots showing data per chosen path segment, see [Supplementary Figure 18A](#).) Associated correlation values (Pearson’s  $r$ ) are shown at the top of each panel.

To determine whether subjects chose longer paths as the slope of relatively straight options increased, we compared path segment tortuosity to the corresponding straight path slope. The tortuosity of chosen path segments ranged from 1 (which denotes a straight path) to 14.73 (which denotes a quite curvy path), with a mean of 1.09 and a standard deviation of 0.29. For all walkers in our dataset, the distribution of tortuosity values was concentrated near 1, with median values per subject of 1.04–1.07 ( $M=1.06$ ,  $SD=0.01$ ; **Supplementary Figure 17** [↗](#), left). The simulated straight path slopes ranged from  $1.89^\circ$  to  $21.72^\circ$ , with a mean of  $10.31^\circ$  and a standard deviation of  $3.44^\circ$ . There was some variability in the per-subject distributions, with median values of  $7.83^\circ$ – $11.94^\circ$  ( $M=10.12^\circ$ ,  $SD=1.20^\circ$ ; **Supplementary Figure 17** [↗](#), right), but the primary difference was between Austin and Berkeley participants, suggesting that variability was at least somewhat a function of the terrain.

We expected that walkers would choose curvier (i.e., more tortuous) paths when the relatively straight alternatives were also relatively steep (i.e., have a relatively high slope). Thus, we expected that, when directly comparing the tortuosity and straight path slope values for each chosen path segment, (a) relatively low straight path slope values would be associated with tortuosity values fairly close to 1 and (b) tortuosity values would tend to increase as straight path slope increased. For each subject, we calculated the average tortuosity of the chosen paths for different levels of “straight path slope” (**Figure 9B** [↗](#)), and we found that, indeed, (a) average tortuosity values were near 1 at the lowest slope level and (b) average tortuosity increased across increasing quantiles of straight path slope. This relationship suggests that walkers made a trade-off, choosing paths that were flatter but more tortuous over paths that were steeper but more direct. Such choices may reflect decisions that the cost of taking the longer, flatter path is ultimately less than cost of taking the straighter, steeper path. All subjects show this relationship, though its strength does vary between subjects.

The energetic cost of taking steeper steps likely varies with factors affecting a walker’s biomechanics. The question of whether a flatter, more tortuous path is a more energetically efficient path than a steeper, straighter path likely has a walker-specific answer. Therefore, one might expect that the some of the between-subject variability in the strength of the slope-tortuosity correlation is due to biomechanically relevant factors, such as the walker’s leg length. We thus asked whether the strength of the trade-off between tortuosity and straight path slope (i.e., the correlation coefficient) varied with walkers’ leg lengths. Subject leg lengths ranged from 81 cm–103.5 cm, and slope-tortuosity correlation values ranged from 0.51 (–0.84 (**Figure 9B** [↗](#))). Longer leg lengths were associated with lower correlation coefficients ( $r = -.86$ ,  $p = .003$ ; **Figure 10** [↗](#)), suggesting that subjects with the shortest legs are more likely to choose longer paths when the straight path becomes less flat (that is, with increasing values of straight path slope).

Note that **Figure 9** [↗](#) and **Figure 10** [↗](#) both use the aforementioned binned data, with the average tortuosity of chosen paths calculated for each of 20 “straight path slope” quantiles to summarize the per-path-segment data. Parallel plots made using the per-path-segment data are shown in **Supplementary Figure 18** [↗](#). Importantly, the analysis of per-path-segment data reveals similar relationships as those described above. For eight of the nine subjects, the correlations are significant. The correlation coefficients between straight path slope and chosen path tortuosity are substantially smaller for the per-path-segment data. The lower correlation values result from the amount of variability across path segments. The variability can be seen in the spread of the points shown in **Supplementary Figure 18A** [↗](#) (vs. **Figure 9** [↗](#)), but note that even those subplots do not show the full extent of the variability. Those scatterplots include the full range of straight path slope values ( $1.89^\circ$ – $21.72^\circ$ ) but, in effect, omit the tail of the tortuosity distribution ( $\max = 14.73$ ) to ensure that the majority of the data is readily visible. This cross-path variability suggests that subjects are not using strict criteria to make decisions about the trade-off between path slope and path curvature. Rather, the trends we observe likely reflect learned heuristics about which paths are more or less preferable, which walkers can use to flexibly select paths.



**Figure 10.**

### Relationship between leg length and the correlation between straight path step slope and path tortuosity

Subjects' leg lengths (in centimeters) are plotted on the horizontal axis. The correlation coefficients drawn from the analyses depicted in [Figure 9B](#) are plotted on the vertical axis. The scatterplot shows one point per subject (black crosses). The linear trendline is also shown (red line). We found that there was a statistically significant negative correlation between subjects' leg lengths and the straight path slope vs. average path tortuosity correlations in their data ( $r = -.86$ ,  $p = .003$ ). (For a comparable plot showing the correlations derived from data per chosen path segment, see [Supplementary Figure 18B](#).)



## Discussion

In this work, we present novel analyses of natural terrain navigation that take advantage of the 3D terrain reconstructions we generated using photogrammetry. The terrain reconstructions allowed for greater precision than was possible in previous studies of walking in natural outdoor environments (Matthis et al., 2018 [DOI](#); Bonnen et al., 2021 [DOI](#)). We were able to more accurately calculate both gaze and foothold locations. Most importantly, the quantification of terrain geometry allowed us to examine how the structure of the visual environment influences foothold selection. An analysis of this relationship — between the structure of the visual environment and selected footholds/paths — has been missing in much previous work on visually-guided action in the natural world, where the depth structure is typically not measured.

After developing the reconstruction and data alignment procedure, our next challenge was to develop a strategy for identifying visual features that influenced subjects' foothold and path selection. We noted regularities in the paths chosen by walkers, both across individuals and across repeats of the same walk, suggesting that there are some terrain features that serve as a basis for path choice. Previous work suggested a role for depth features in visually-guided walking Bonnen et al. (2021) [DOI](#). Using a CNN to predict foothold locations on the basis of retinocentric depth images, we confirmed a role for depth information in foothold selection. This result justified the further exploration of depth variation in the terrain (e.g., changes in terrain height) as a potential feature used by walkers in foothold selection.

To ask whether changes in terrain height (i.e., depth structure) influenced path selection, we simulated viable paths that could be compared with the chosen paths. Comparing the sets of chosen and viable paths, we found that walkers prefer flatter paths and avoid regions with large height irregularities. While in some ways this might not be a surprising result, the data reveal that this is a strong constraint on path choice. The median slope of 5-step paths was less than 10 degrees, which corresponds to a quite small height change of about 14 cm–17 cm.

This work did not investigate which depth cues walkers used to make these path choices, but we highlight that gap in knowledge here as an avenue ripe for future study. A variety of depth cues might be relevant to such sensorimotor decisions, including motion parallax generated by the movement of the head, binocular disparity, local surface slant, and the size of the step-able area. Determining how depth cues are used to make these sensorimotor decisions will require more controlled experiments.

Finally, we observed that walkers chose longer paths when the straightest viable paths involved greater height changes (Figure 9 [DOI](#)), and further, our data suggest that walkers were more likely to choose longer paths if their legs were shorter (Figure 10 [DOI](#)). This suggests that the sensorimotor decision-making that supports walking complex terrain is highly body-specific, taking into account the details of a walker's body, like leg length. This suggests that any cost function or model describing the sensorimotor decision making processes that support walking in complex natural terrains will also need to be body-specific.

### Cost functions in visually-guided walking

While we do not know what contributes to the internal cost functions that determine walkers' choices, the preference for flatter paths is likely driven in part by the energetic cost of stepping up or down. On flat ground, humans converge to an energetic optimum consistent with their passive dynamics (Kuo et al., 2005 [DOI](#); Selinger et al., 2015 [DOI](#); Finley et al., 2013 [DOI](#); Lee and Harris, 2018 [DOI](#)). Deviations from this gait pattern, including turns and changes in speed, are energetically costly (Voloshina et al., 2013 [DOI](#); Soule and Goldman, 1972 [DOI](#)). Recent work by Darici and Kuo (2023) [DOI](#)

also showed that subjects are able to minimize energetic cost on uneven ground planes by modulating speed. Our findings suggest that walkers may be adjusting their behavior to minimize energetic costs in natural outdoor terrains as well. Future work should examine more directly how particular walking decisions impact energetic costs in natural outdoor terrains.

## Path planning

Our analyses show that vision is used to locate flatter paths in upcoming steps. We found that the average step slope of the chosen path was significantly lower than simulated paths, suggesting that walkers were intentionally maintaining a flatter path. Furthermore, our findings suggest that walkers turn to avoid paths with large changes in terrain height. To accomplish this, walkers must plan ahead, demonstrating that planning is an important component of path selection in rugged terrain. Though this study has not explicitly examined the role of gaze in walking, future studies of gaze during walking will be critical to understanding how individuals perform path planning.

Laboratory studies suggest that walkers need to look two steps ahead to preserve inverted pendulum dynamics [Matthis et al. \(2017\)](#). Biomechanical models indicate that walkers can adjust their gait to accommodate upcoming obstacles and may plan up to 8 or 9 steps ahead ([Darici and Kuo, 2023](#)). Our previous work studying gaze suggests that, in rocky terrain, walkers distribute most of their gaze on the ground to footholds up to 5 steps ahead ([Matthis et al., 2018](#); [Bonnen et al., 2021](#)). Because of the differences between these studies, it is difficult to say exactly what causes the discrepancy (5 steps vs. 9 steps) in the planning horizons reported in these two studies. However, there are notable differences between the laboratory obstacle paths they used and our natural environments. Their walking paths involved height changes of no more than 7.5 cm, the surfaces themselves were flat, and the path required no changes in direction. Our terrains involved greater height changes, irregular and sloping surfaces, large boulders, and frequent direction changes based on visual information. More complex terrains may also impose a greater load on visual working memory ([Lin and Lin, 2016](#)). Thus a shorter planning horizon in our data might be expected as individuals adjust their planning horizon depending on the nature of the terrain. On the other hand, because there is no eyetracking in [Darici and Kuo \(2023\)](#), we cannot rule out the possibility that these two planning horizons are in fact the same — individuals may be able to get information about 8-9 steps ahead from their peripheral vision. More study is needed on the details of planning horizons in walking and how individuals adjust them depending on the task and terrain.

## Conclusion

In conclusion, we have integrated eye tracking, motion capture, and photogrammetry to create a visuomotor dataset that includes gaze information, body position data, and accurate 3D terrain representations. The reconstructed 3D terrains were a valuable addition to our methodology because they allowed a much more direct, more precise investigation of the visual terrain features that are used to guide path choice. Previous investigations of walking in natural outdoor environments have been limited to video recordings. The reconstruction and integration procedures outlined in this paper should be generally useful for the analysis of visually guided behavior in natural environments. In our analyses, we observed that visual information about depth appeared to play a role in path choice. Despite the complexity of the sensory-motor decisions in natural, complex terrain, we observed that there were consistencies in the paths walkers chose. In particular, walkers chose to take more indirect routes to achieve flatter paths, which required them to plan ahead. Taken together, these findings suggest that walkers' locomotor behavior in complex terrain reflects sensorimotor decision mechanisms that involve different costs, sensory and motor information, and path planning.

## Methods

### Experimental data

The data used here was collected by the authors in two separate studies, conducted in two separate locations: Austin, Texas (Matthis et al., 2022 [↗](#)) and Berkeley, California (Bonnen et al., 2021 [↗](#)). The studies were approved by the Institutional Review Boards at the University of Texas at Austin and the University of California, Berkeley, respectively. All participants gave informed consent and were treated according to the principles set forth in the Declaration of Helsinki of the World Medical Association. Both studies used the same eye and body tracking equipment as well as the same data collection methods. Additionally, both included multiple terrain conditions. One terrain condition common to both was rough terrain, which consisted of large rock obstacles with significant height deviations.

### Data selection

In our combined dataset, we included data from only (a) rough terrain, (b) participants walking with normal or corrected-to-normal vision, and (c) participants with scene videos of sufficiently high quality for terrain reconstruction. We therefore did not include any of the Berkeley data used to study the impact of binocular visual impairments. Further, we excluded one Austin participant and one Berkeley participant because the quality of their scene videos caused issues with the terrain reconstruction process, which was essential for the analyses we describe here. (More specifically, one Austin participant was excluded because the scene camera was angled too far upward, limiting the view of the ground, and one Berkeley participant was excluded because their scene videos were too low contrast due to the dim outdoor lighting conditions at the time of the recording.)

### Participants

We used data from 9 participants: 2 from the Austin study and 7 from the Berkeley study (Table 1 [↗](#)). All had normal or corrected-to-normal vision. There were 5 male and 4 female subjects. They were 23–54 years old at the time of data collection, with an average age of 31 years (median: 27). Their legs were 81–103.5 cm long, with an average of 93.9 cm (median: 96.5 cm).

The amount of data recorded per participant varied since they were tasked with walking along loosely defined paths, rather than walking for a fixed duration, number of steps, etc. In Table 1 [↗](#), we represent the amount of data recorded via the number of steps in rough terrain in each participant's processed data. In total, the dataset included 4230 steps. Per participant, there were 347–603 steps, with an average of 470 steps (median: 468). Overall, participants with longer legs took fewer steps ( $r = -0.57$ ), and the Berkeley participants took approximately 125 steps more than Austin participants with similar leg lengths.

### Equipment

Eye movements were recorded using a Pupil Labs Core mobile eye tracker and the Pupil Capture app (Pupil Labs, Berlin, Germany). The eye tracker had two infrared, eye-facing cameras, which recorded videos of the eyes at 120 Hz with 640×480 pixel resolution. Additionally, there was an outward-facing camera mounted 3 cm above the right eye, which recorded the scene in front of the subject at 30 Hz with 1920×1080 pixel resolution and a 100° diagonal field of view. A pair of dilation glasses was fitted over the eyes and eye-facing cameras to protect the infrared eye cameras from interference due to the sun. For participants, this felt like wearing a pair of sunglasses.

<b>Location</b>	TX	TX	CA	CA	CA	CA	CA	CA	CA
<b>Age</b>	23	25	27	39	34	29	24	24	54
<b>Gender</b>	F	M	M	M	M	F	F	F	M
<b>Leg length (cm)</b>	89	102.5	103.5	101	96.5	81	85	90	96.5
<b>Step count</b>	468	347	462	489	385	537	486	603	453

**Table 1.**

Information about participants included in dataset. Table includes the location of data collection, key demographics, and the amount of data recorded per participant (quantified as the number of steps in rough terrain in participants' processed data).

Body movements were recorded using Motion Shadow's full-body motion capture suit and the Shadow app (Motion Shadow, Seattle, WA, USA). The suit included 17 inertial measurement units (IMUs), which each contained three 3-axis sensors: an accelerometer, a gyroscope, and a magnetometer. The Shadow app recorded data from the suit at 100 Hz and simultaneously estimated the joint poses (i.e., positions and orientations) for the full 30-node 3D skeleton. IMUs were placed on the head, chest, and hips as well as on both the left and right shoulders, upper arms, forearms, hands, thighs, calves, and feet. The 3D skeleton then included nodes for the head, head top, neck, chest, body, hips, mid spine, and low spine as well as the left and right shoulders, arms, forearms, hands, fingers, thighs, legs, feet, toes, heels, and foot pads.

In addition to the eye tracker and motion capture suit, subjects wore a backpack-mounted laptop, which was used to record all raw data. Importantly, using the same computer to record both data streams meant both were recording timestamps queried from the same internal clock. Their timestamps were therefore already synchronized.

## Task

At the beginning of each recording, participants performed a 9-point vestibuloocular reflex (VOR) calibration task. They were instructed to stand on a calibration mat 1.5 m from a calibration point marked on the mat in front of them. This distance was chosen based on the most frequent gaze distance in front of the body during natural walking in these terrains (Matthis et al., 2018). They were instructed to fixate on the calibration point while rotating their head along each of the 8 principal winds, i.e., the 4 cardinal and 4 ordinal directions. Their resulting VOR eye movements were later used to calibrate the eye tracking data and to spatially align the eye and body data.

Participants' primary task was to walk along a trail. The Austin participants walked along a rocky, dried out creek bed in between two specific points that the experimenters had marked (**Figure 3**). They traversed the trail 3 times in each direction, for a total of 6 traversals per subject. The Berkeley participants walked along a hiking trail in between two distinctive landmarks. They traversed the trail once in each direction, for a total of 2 traversals per subject. The trail included pavement, flat terrain, medium terrain, and rough terrain, so as with the walk's start and end points, the experimenters used existing landmarks in the environment to mark the transitions between terrain types. We found the sections of recordings marked as rough terrain and included only that subset of the data in this study.

## Data processing

Following data collection, we performed a post-hoc eye tracking calibration using the 9-point VOR calibration task data. Per subject, we placed a reference marker on the calibration fixation target at 10 timepoints in the recording (corresponding to the 9 points of the VOR calibration task, plus an additional repeat of the center marker at the end). With the Pupil Player app in natural features mode (Pupil Labs, Berlin, Germany), we used those markers to perform gaze mapping, generating 3D gaze vectors for both eyes.

We then had three sets of tracking data recorded at two different timescales and expressed in three different coordinate systems: the left eye's gaze data, the right eye's gaze data, and the 3D skeleton's pose data. Thus, our next step was to temporally and then spatially align the recordings via a procedure detailed in Matthis et al. (2022).

The timestamps from both systems were already synchronized to the same clock since they were recorded by the same computer, but the sampling rates (motion capture, 100 Hz; eye tracking 120 Hz) and specific timestamps were different. Using MATLAB's "resample" function (Signal Processing Toolbox; MathWorks, Natick, MA, USA), we performed interpolation so that the motion capture and eye tracking data streams had the same sampling rate and time stamps (120 Hz). The result was that the left eye, right eye, and kinematic data streams were temporally aligned.

Once the three sets of data were temporally aligned, we used the VOR calibration data to spatially align them. During the VOR task, participants were fixating on a single point while moving their head, so we aligned each eye's coordinate system to that of the 3D skeleton by shifting and rotating them, such that the eyes were in an appropriate location relative to the head and the gaze vectors remained directed at the calibration fixation target as the head and eyes moved. To determine the shift per eye coordinate system, we estimated the position of each eye's center in 3D skeleton coordinates. We based our estimate on (a) the position and orientation of the skeleton's head node and (b) average measurements of where the eyes are located within the human head. To determine the rotation per eye coordinate system, we found the single optimal rotation that minimized the distance between the calibration fixation target and the gaze vector's intersection with the mat. Note that because the head's position and orientation changed throughout the recording, we applied transformations relative to the position and orientation of the skeleton's head node in each frame.

Once the eye and body tracking data were fully aligned, we used the data from the body pose foot nodes to find the time and location of each step (both heel strike and toe off), following the velocity-based step-finding procedure outlined in [Zeni et al. \(2008\)](#).

Additionally, we identified periods of time when subjects were potentially collecting visual information by differentiating between when they were fixating and when they were making saccadic eye movements. We only used periods of fixation when considering the visual information available to participants, as mid-saccade visual input is unlikely to be used for locomotor guidance due to saccadic suppression and image blur.

We identified fixations by applying an eye-in-orbit velocity threshold of 65 deg/s and an acceleration threshold of 5 deg/s<sup>2</sup>. (Note that the velocity threshold is quite high to avoid including the smooth counter-rotations that occur during eye stabilization.) If both values were below threshold for a given frame, we classified that frame as containing a fixation; if not, the frame was classified as containing a saccade.

## Terrain reconstruction

To factor terrain height into our analyses, we needed information about the 3D structure of the terrain that participants walked over. Our dataset did not include data on terrain depth, but the scene videos recorded by the eye tracker's outward-facing camera did provide us with 2D images of the terrain. In principle, photogrammetry should allow us to extract accurate 3D information about the terrain's structure from those 2D video frames, so our first step beyond the original analyses of these data ([Bonnen et al., 2021](#); [Matthis et al., 2022](#)) was to use photogrammetry to generate 3D terrain reconstructions.

## Photogrammetry pipeline

To reconstruct the 3D environment from our 2D videos, we used an open-source software package called Meshroom, which is based on the AliceVision Photogrammetric Computer Vision framework ([Griwodz et al., 2021](#)). Meshroom combines multiple image processing and computer vision algorithms, ultimately allowing the user to estimate both environmental structure and relative camera position from a series of images.

The steps in the AliceVision photogrammetry pipeline are (1) natural feature extraction, (2) image matching, (3) features matching, (4) structure from motion, (5) depth maps estimation, (6) meshing, and (7) texturing ([Griwodz et al., 2021](#)).

To summarize in greater detail: (1) Features that are minimally variant with respect to viewpoint are extracted from each image. (2) To find images that show the same areas of the scene, images are grouped and matched on the basis of those features. (3) The features themselves are then

matched between the two images in each candidate pair. (4) Those feature matches are then used to infer rigid scene structure (3D points) and image pose (position and orientation) via an incremental pipeline that operates on each image pair, uses the best pair to compute an initial two-view reconstruction, and then iteratively extends that reconstruction by adding new views. (5) The inferred 3D points are used to compute a depth value for each pixel in the original images. (6) The depth maps are then merged into a global octree, which is refined through a series of operations that ultimately produce a dense geometric surface representation of the scene. (7) Texture is added to each triangle in the resulting mesh via an approach that factors in each vertex's visibility and blends candidate pixel values with a bias toward low frequency texture.

The most relevant outputs of this pipeline for our analyses are the 3D triangle mesh and the 6D camera trajectory (i.e., the estimated position and orientation of the camera, per input frame). We also used the textured triangle mesh but solely for visualization (e.g., **Figure 3** [↗](#)).

## Reconstruction procedure

Prior to terrain reconstruction, we processed the raw scene videos recorded by the eye tracker's outward-facing camera. We first used the software package “ffmpeg” to extract the individual frames from the videos. We then undistorted each frame using a camera intrinsic matrix, which we estimated via checkerboard calibration ([Zhang and Pless, 2004](#) [↗](#)).

Then we used Meshroom to process the scene video frames, one traversal at a time, specifying the camera intrinsics (focal length in pixels and viewing angle in degrees) and using Meshroom's default parameters. Meshroom processed the scene video frames according to the pipeline described above, producing both a terrain mesh and a 6D camera trajectory (3D position and 3D orientation), with one 6D vector for each frame of the original video. To give a sense of the mesh output, we have provided a rendered image of a small section of the textured Meshroom output in **Figure 1** [↗](#).

## Data alignment

**Figure 2** [↗](#) illustrates the data alignment process which positions the body and eye movement data within the reconstructed terrain. Data alignment was performed on a per-traversal basis. First the body/eye tracking data was translated, pinning the location of the head node to Meshroom's estimated camera position. Next, the orientation of the head node was matched to Meshroom's estimated camera orientation by finding a single 3-element Euler angle rotation that minimized the L2 error (i.e., the sum of squared errors) across frames using MATLAB's “fminsearch” function. After applying that rotation, the body/eye tracking data was scaled so that, across all heel strikes in a given recording, the distance between the skeleton's heel and the closest point on the mesh at the time of that heel strike was minimized.

A visualization of the aligned motion capture, eye tracking, and terrain data for one traversal of the Austin trail can be seen in Supplementary Video 1.

## Evaluating terrain reconstruction

To evaluate the accuracy of the 3D reconstruction, we used the terrain meshes estimated from different traversals of the same terrain, both by an individual subject and also by the different subjects. We used only the Austin data here, as that dataset included 6 traversals per subject (vs. 2) and was collected over a much shorter time span (5 days) and thus was less likely to physically change.

The meshes were aligned using the open-source software package CloudCompare. To align two meshes in CloudCompare, one mesh needs to be designated as the fixed “reference” mesh and the other as the moving “aligned” mesh (i.e., the mesh that will be moved to align with the reference).



We first coarsely aligned the meshes via a similarity transform. That step requires an initial set of keypoints, so we chose 5 easily discernible features in the environment (e.g., permanent marks on rocks) that were visible in each terrain mesh and manually marked their locations. We then completed the point cloud registration on a finer scale using the iterative closest point (ICP) method. That procedure involves locating, for each point in the moving mesh, the closest point in the fixed point cloud (i.e., the moving point's nearest neighbor). The distance between nearest neighbors is then iteratively minimized via best-fitting similarity transforms.

We can then evaluate the reliability of terrain reconstruction by measuring the distance between nearest neighbors on the two meshes. We make two key comparisons: (1) measuring nearest neighbor distances for all points on the terrain mesh within 2 meters of the path (see [Figure 5A](#)) and (2) measuring nearest neighbor distances for all foothold locations by taking the nearest neighbor distance for the mesh point closest to that foothold (see [Figure 5B](#)). For the path comparison, the median error was 4.5cm, and the 95% quantile was 20.0cm ([Figure 5A](#)). For the foothold comparison, the median error was 4.0cm, and the 95% quantile was 16.8cm ([Figure 5B](#)). To put this into context, the average foot of a person from North America is 25.8cm ([Jurca et al., 2019](#)). In both cases, the majority of mesh errors fall below 20 % of the average foot size.

## Retinocentric analysis

To assess whether depth features can be used to explain some variation in foothold selection, we trained a convolutional neural network (CNN) to predict future foothold locations given the walker's view of the terrain's depth. This analysis involved computing both the retinocentric depth images that served as the input data ([Figure 11A](#)) and the foothold likelihood maps that served as target output for training ([Figure 11B](#)). The retinocentric depth images approximate the visual information subjects have when deciding on future foothold locations, and the foothold likelihood maps represent the subjects' subsequent decisions. After training and testing the CNN, we quantified its prediction accuracy by calculating the AUC. With that metric, scores above chance (50%) would indicate that the depth information plays some role in determining where individuals will put their feet.

### Retinocentric depth images

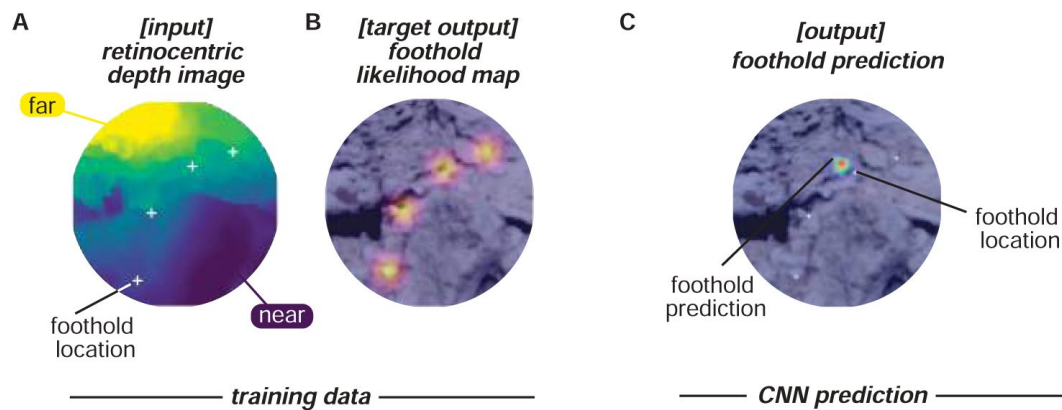
We computed subject-perspective depth images (e.g., [Figure 11A](#)) using the aligned eye tracking, motion capture, and photogrammetry data in the open-source 3D computer graphics software Blender. Per subject, we moved a virtual camera through the reconstructed terrain, updating its position and orientation in each frame based on Meshroom's estimate of the scene camera's location in the corresponding frame of the scene video. Blender's "Z-buffer" method then captured an image showing the depth of the 3D triangle mesh representation relative to the camera.

The resulting egocentric depth images were then transformed to polar coordinates (polar angle  $\theta$  and eccentricity  $\rho$ ) to approximate a retinocentric perspective. We defined the diameter of images as 45 degrees of visual angle, and we used 2D interpolation to compute the pixel values at each polar coordinate from the pixel values in Cartesian coordinates.

The depth values in the images were then converted into relative depth values. To make that shift, we subtracted the gaze point's depth (i.e., the value at the center pixel) from the entire depth image. The value at center thus became 0, and the rest of the depth values were relative to the depth of the gaze point.

### Foothold likelihood maps

To calculate the ground-truth of the future foothold locations in each depth image, we found the point at which the line between the current camera position and the foothold intersected the camera's image plane.



**Figure 11.**

### CNN inputs and outputs

Schematic shows the inputs and outputs for one example frame. **A.** Input: retinocentric depth image. **B.** Target output: foothold likelihood map. **C.** Output: predicted foothold locations.

The ground truth foothold locations in the world video frame were converted to likelihood maps (e.g., [Figure 11B](#)) by smoothing foothold locations with a Gaussian kernel:  $\sigma = 5$  pixels. (In degrees of visual angle, the kernel size was roughly 1 degree. That value is not exact because the conversion between pixels and degrees is not constant throughout the visual field.) This smoothing mitigated the impact of any noise in our estimation of foothold location to allow more robustness in the CNN learned features.

## Convolutional neural network (CNN)

The retinocentric depth images and foothold likelihood maps were then used to train a custom convolutional neural network (CNN) to predict the probability that each location in the retinocentric depth images was a foothold location. The network input was a depth image ([Figure 11A](#)), and the target output was a foothold likelihood map ([Figure 11B](#)).

The CNN had a convolutional-deconvolutional architecture with three convolutional layers followed by three transposed convolutional layers ([Table 2](#)). Training was performed using KL-divergence between the CNN output ([Figure 11C](#)) and the foothold likelihood maps ([Figure 11B](#)). Data was split so that half of the pairs of depth images and likelihood maps were used to train the network and the other half was reserved for testing. This split ensured that the network was tested on terrain that it had not previously “seen”.

To evaluate performance, we calculated the area under the ROC curve (AUC) per depth image. The true foothold locations per image were known, and the CNN generated a probability per pixel per image. To generate the ROC curve, we treated the CNN task as a binary classification of pixels, and we calculated the rate of false positives and true positives at different probability criterion values, increasing from 0 to 1. Calculating AUC was then just a matter of computing the area under the resulting ROC curve.

## Step analysis

We sought to better understand how subjects chose their footholds by analyzing the properties of their chosen steps and step sequences. Throughout this work, a foothold location is defined as the 3D position of the left or right foot marker at the time of heel strike, and a step is defined as the transition between two footholds. To analyze sets of steps, we segmented participants’ paths into 5-step sequences, consisting of 6 consecutive footholds.

## Step properties

For each step in the dataset, we computed seven properties: distance, ground distance, direction, goal direction, deviation from goal, height, and slope.

To illustrate how we compute a step’s properties, consider a step vector  $\vec{s}$  that starts at a foothold with coordinates  $(x_1, y_1, z_1)$  and ends at a foothold with coordinates  $(x_2, y_2, z_2)$ , where the y-axis corresponds to gravity.

## In 3D

We define step **distance**  $D$  as the magnitude of step vector  $\vec{s}$ , i.e., the three-dimensional Euclidean distance between the start and end footholds:

$$D = |\vec{s}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (1)$$

Layer	Output Shape	# Params
Conv2D	(100, 100, 4)	404
BatchNormalization	(100, 100, 4)	16
MaxPooling2D	(50, 50, 4)	0
Conv2D	(50, 50, 8)	3208
BatchNormalization	(50, 50, 8)	32
MaxPooling2D	(25, 25, 8)	0
Conv2D	(25, 25, 16)	12816
BatchNormalization	(25, 25, 16)	64
Conv2DTranspose	(25, 25, 16)	25616
BatchNormalization	(25, 25, 16)	64
UpSampling2D	(50, 50, 16)	0
Conv2DTranspose	(50, 50, 8)	12808
BatchNormalization	(50, 50, 8)	32
UpSampling2D	(100, 100, 8)	0
Conv2DTranspose	(100, 100, 4)	3204
Conv2DTranspose	(100, 100, 1)	401
BatchNormalization	(100, 100, 1)	4
Flatten	(10000)	0
Softmax	(10000)	0
Reshape	(100, 100)	0

**Table 2.**

**Layers of custom convolutional neural network (CNN)**

## In 2D, from overhead

To focus on the progression of the step along the subject's route and ignore the step's vertical component, we project step vector  $\vec{s}$  onto the ground plane ( $xz$ -space), producing ground vector  $\vec{g}$ . We define step **ground distance**  $G$  as the magnitude of ground vector  $\vec{g}$ , i.e., the two-dimensional Euclidean distance in  $xz$ -space between the start and end footholds:

$$G = |\vec{g}| = \sqrt{(x_B - x_A)^2 + (z_B - z_A)^2} \quad (2)$$

We define step **direction**  $\gamma$  as the direction of ground vector  $\vec{g}$ :

$$\gamma = \arctan \frac{z_B - z_A}{x_B - x_A} \quad (3)$$

We then consider the endpoint of the current terrain traversal, foothold  $E$ . Along the ground plane, the vector  $\vec{e}$  connects the step's starting foothold ( $x_A, z_A$ ) to that traversal's endpoint ( $x_E, z_E$ ). That vector represents the most direct path the participant could take to reach their current goal. We refer to the direction of vector  $\vec{e}$  as the **goal direction**  $\omega$ :

$$\omega = \arctan \frac{z_E - z_A}{x_E - x_A} \quad (4)$$

We use that angle to calculate step **deviation from goal**  $\delta$ , which we define as the angle between the step direction  $\gamma$  and goal direction  $\omega$ :

$$\delta = \gamma - \omega \quad (5)$$

## In 2D, from the side

To analyze the step's vertical component, we calculate step **height**  $\Delta h$  by finding the change in vertical position between footholds  $A$  and  $B$ :

$$\Delta h = y_B - y_A \quad (6)$$

We then compute step **slope** by dividing step height  $\Delta h$  by step distance  $D$ :

## Properties of step sequences

$$\theta = \arcsin \frac{\Delta h}{D} \quad (7)$$

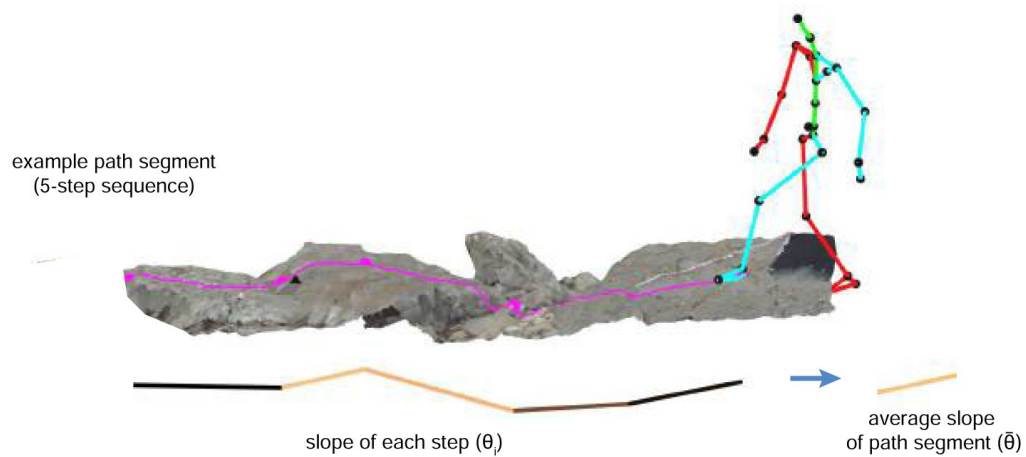
Step sequences (sometimes also called paths or path segments) are composed of a series of steps. In this paper, we primarily focused on 5-step sequences. We calculated two key properties: mean slope and tortuosity.

## Mean slope

Each step-sequence has a **mean slope**,  $\bar{\theta}$ , which is defined as the average step slope across steps within the sequence:

## Tortuosity

$$\bar{\theta} = \frac{\sum_i^n \theta_i}{n} \quad (8)$$



**Figure 12.**

### **Schematic depicting the calculation of the mean slope of a step sequence**

To calculate the slope of a step sequence — chosen or simulated — we first calculated the step slope for each step in the path, and we then averaged the absolute values of those slopes.

We quantified the curvature of each step sequence by calculating **tortuosity**,  $T$ :

$$T = \frac{\sum_i^n D_i}{D_s} \quad (9)$$

where  $n$  is the number of steps in the sequence,  $D_i$  is the magnitude of a given step vector, and  $D_s$  is the magnitude of the vector connecting the start and end foothold locations.

Thus, we quantify tortuosity as the ratio of the cumulative step distance to the distance between the first and final footholds. This metric is the inverse of the straightness index formula proposed in Batschelet (1981) [\[1\]](#), which has been shown to be a reliable estimate of the tortuosity of oriented paths (Benhamou, 2004 [\[2\]](#)). A tortuosity of 1 indicates a straight path, while a tortuosity greater than 1 indicates a curved path. A perfect semi-circle would have a tortuosity of  $\pi/2$  (approximately 1.57), and a circle would be infinitely tortuous.

## Path simulation

We sought to evaluate differences between the paths subjects chose and the alternative paths they could have chosen. To do so, we simulated 5-step sequences and compared the properties of chosen paths to those of simulated paths.

## Identifying viable footholds

Previous work has found that subjects are able to walk on surfaces slanted up to approximately 33 degrees (Kinsella-Shaw et al., 1992 [\[3\]](#)). We thus constrained possible foothold locations to those with a local surface slant below that value.

To calculate the slant of possible footholds, we computed the surface normal vector for each triangle in the 3D terrain mesh. We then calculated the mean local surface slant for each point in the mesh's point cloud representation by averaging the surface slants of all triangles within a radius of one foot length (25.8 cm).

## Identifying viable steps

After identifying viable foothold locations, viable steps between viable foothold locations were determined based on 3 constraints (**Figure 7** [\[4\]](#)). Per subject, we found the distributions of (a) step slope (**Equation 7** [\[5\]](#)), (b) step ground distance (**Equation 2** [\[6\]](#)) relative to that participant's leg length, and (c) step deviation from goal (**Equation 5** [\[7\]](#)). We computed the maximum observed absolute values, and we then deemed a step viable only if its properties fell within those maxima.

## Simulating possible paths

### Paths with a fixed start point and a random end point

For each foothold that a subject chose, we simulated possible alternative paths consisting of 5 steps (i.e., 6 footholds). These simulated path segments started at the chosen foothold, and subsequent footholds were iteratively selected, in accordance with the foothold constraint and step constraints defined above. The resulting set of path segments could then be directly compared to the path segment that the walker actually chose.

### Paths with fixed start and end points

We also simulated paths that started and ended at chosen foothold locations. To accomplish that, we treated the set of possible footholds and viable steps between them as a directed graph. We then used MATLAB's 'maxflow' function to find the subset of footholds that have non-zero flow



values in a directed graph between the two selected footholds (starting point and ending point). The ‘maxflow’ function then returns a set of footholds that can be visited from the starting foothold and still have available paths to the final foothold (i.e., 6<sup>th</sup> foothold in path).

Possible paths connecting the two end points of the actual path are then generated from this subset of possible foothold locations following the procedure in the previous section, iteratively selecting footholds in accordance with the step constraints defined above.

### Estimating straight path slope

When walking from one point to another in flat terrain, straight paths are almost certainly the preferable option. In rough terrain, however, there may be obstacles that make walking straight impossible — or at least less preferable — than taking a slightly longer curved path. To analyze this potential trade-off, for each step sequence, we estimated the slope a walker would encounter if they tried to take a relatively straight path.

To compute those values, we first found the median tortuosity of all chosen 5-step sequences in a particular terrain traversal (i.e., across a particular mesh). That gave us a conservative, terrain-specific tortuosity threshold that we could use to determine which of the possible paths were relatively straight. For each step sequence in that traversal, we then identified simulated paths with a tortuosity below that threshold and calculated the average of their mean step slopes ( $\bar{\theta}$ ; Equation 8). The resulting values are treated as the average step slope the subject would encounter if they tried to take a straighter path for that segment of terrain.

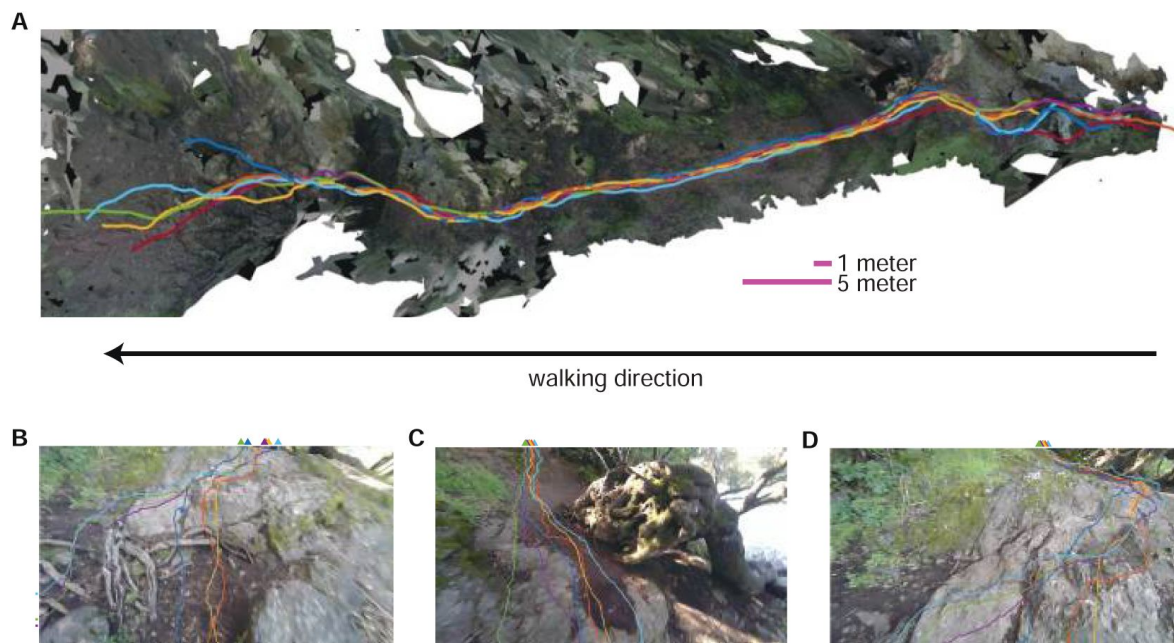
## Acknowledgements

This work was supported by NIH Grants EY05729 and K99 EY 028229.

## Supplementary Information

Video 1. Visualization of the aligned motion capture, eye tracking, and terrain data for one traversal of the Austin trail: [https://youtu.be/TzrA\\_iEtj1s](https://youtu.be/TzrA_iEtj1s). The video shows the 3D motion capture skeleton walking over the textured mesh. Gaze vectors are illustrated as blue lines. On the terrain surface, the heatmap shows gaze density, and the magenta dots represent foothold locations.

Video 2. Visualization of foothold locations in the scene camera’s view for one traversal of the Austin trail: <https://youtu.be/llulrzhIAVg>. Computed foothold locations are marked with cyan dots.

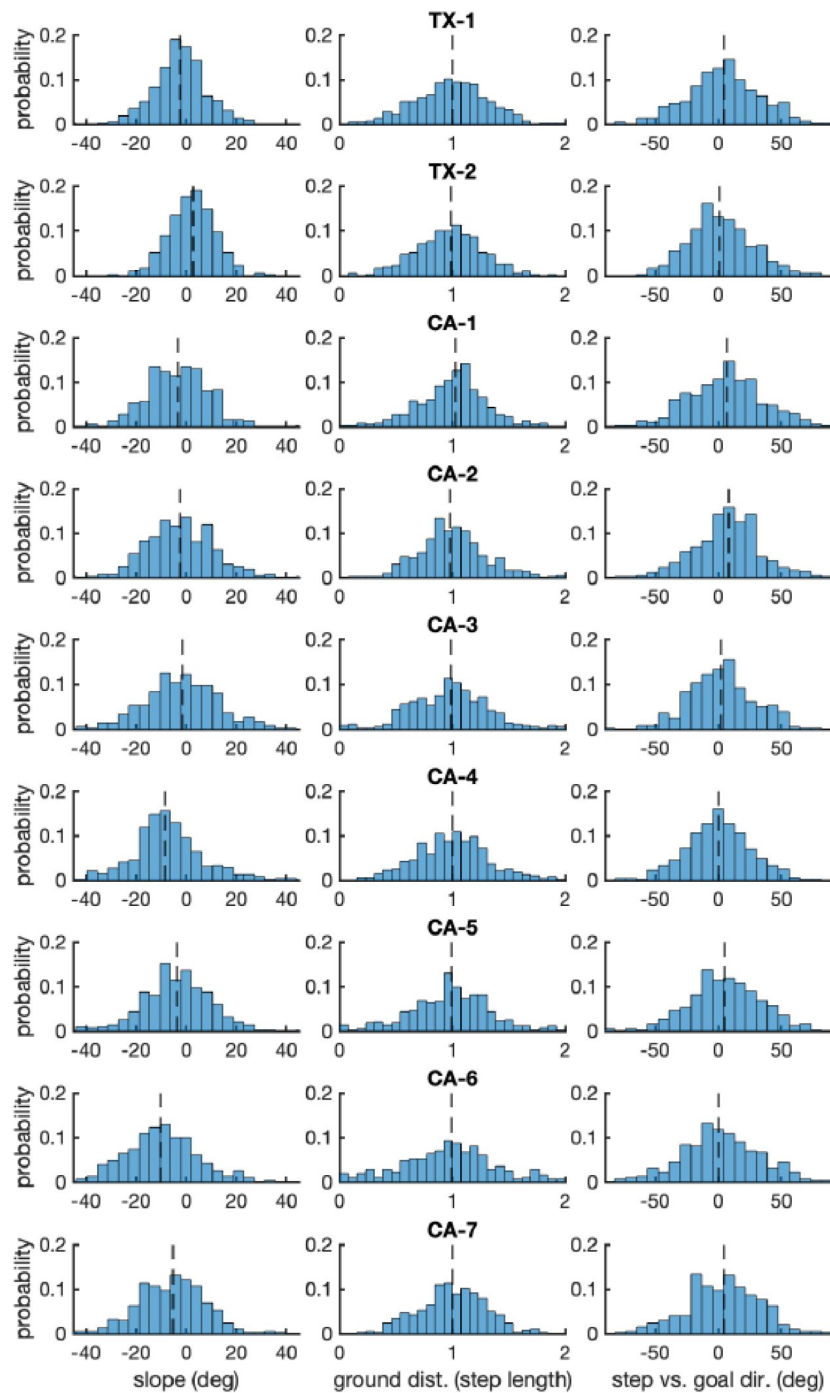


**Figure 13.**

### **Berkeley path consistency, convergence and divergence**

**A.** Overhead view of paths taken (colored lines correspond to individual subjects) through a portion of the Berkeley terrain.

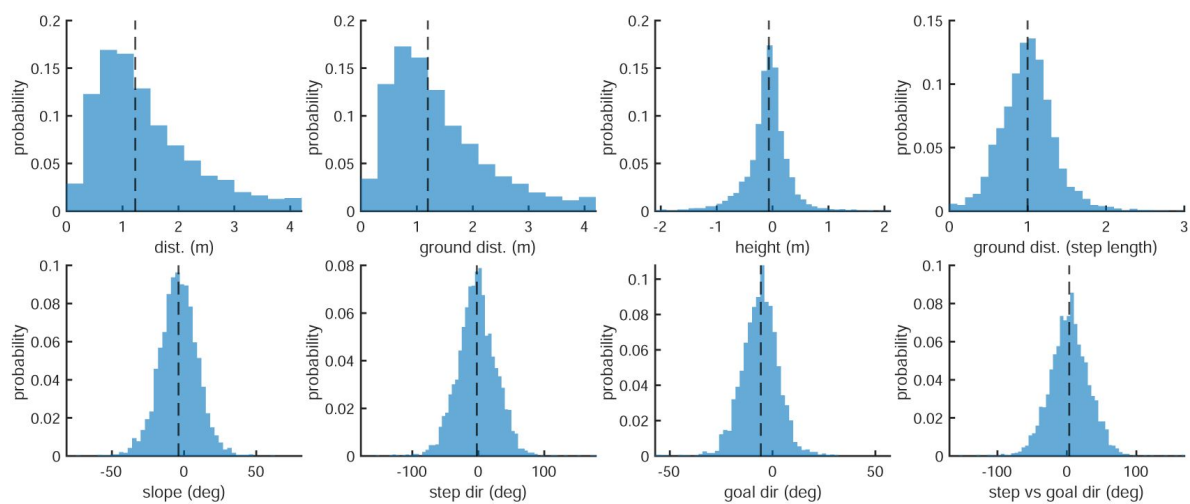
**B–D.** These panels show examples of path convergence and divergence. The colored lines indicate the paths that subjects traveled in this section of the terrain, with each color representing a different subject. For each path, the walk in this section of terrain begins at the tick mark near the bottom of the image and ends at the colored ▲. In (B), subjects diverge by choosing two different routes around a root, but then converge again. In (C) subjects paths converge to avoid the large outcrop. In (D) subject paths converge around a mossy section of a large rock.



**Figure 14.**

**Per-subject histograms of the step parameters shown in Figure 7**

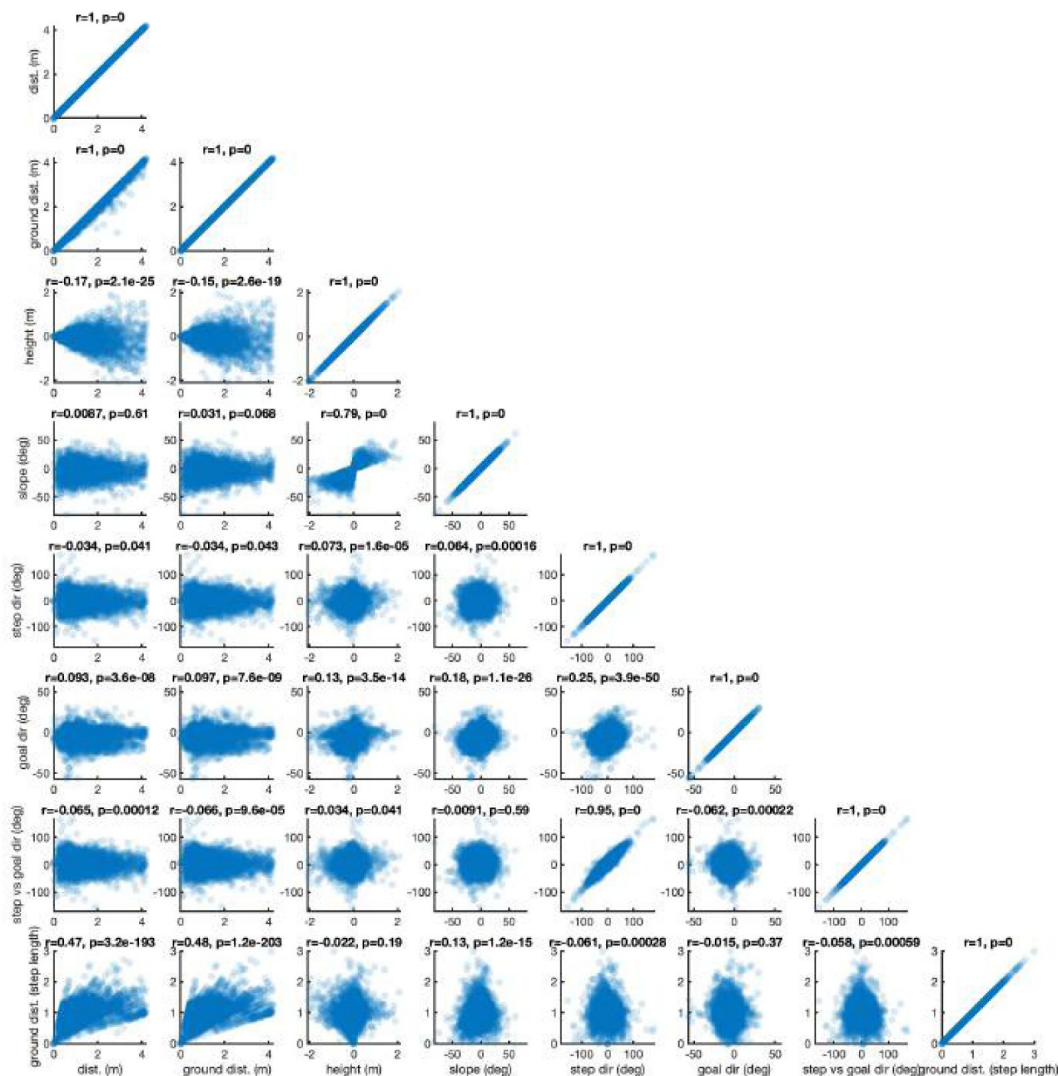
Each row shows one subject's data. Each column shows data for one of the step parameters used to constrain simulated steps: (1) step slope, (2) step length normalized by the average step length, and (3) step direction relative to goal direction.



**Figure 15.**

**Aggregate histograms of the step parameters defined in the methods section titled “Step analysis”**

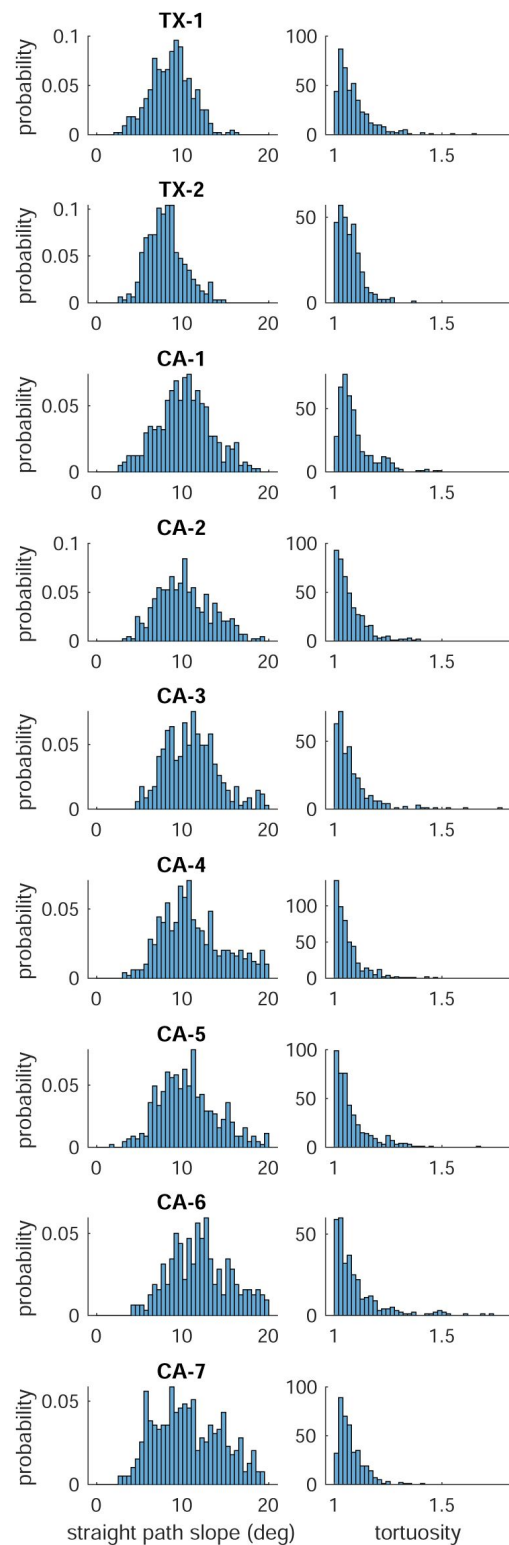
Top row, from left to right: (1) step distance (meters), (2) step ground distance (meters), (3) step height (meters), (4) step ground distance (step length; see also [Figure 7B](#)). Bottom row: (1) step slope (deg; see also [Figure 7A](#)), (2) step direction (deg), (3) goal direction (deg), (4) step direction relative to goal direction (deg; see also [Figure 7C](#)).



**Figure 16.**

### Scatter plots between all parameters defined in the methods section titled "Step analysis"

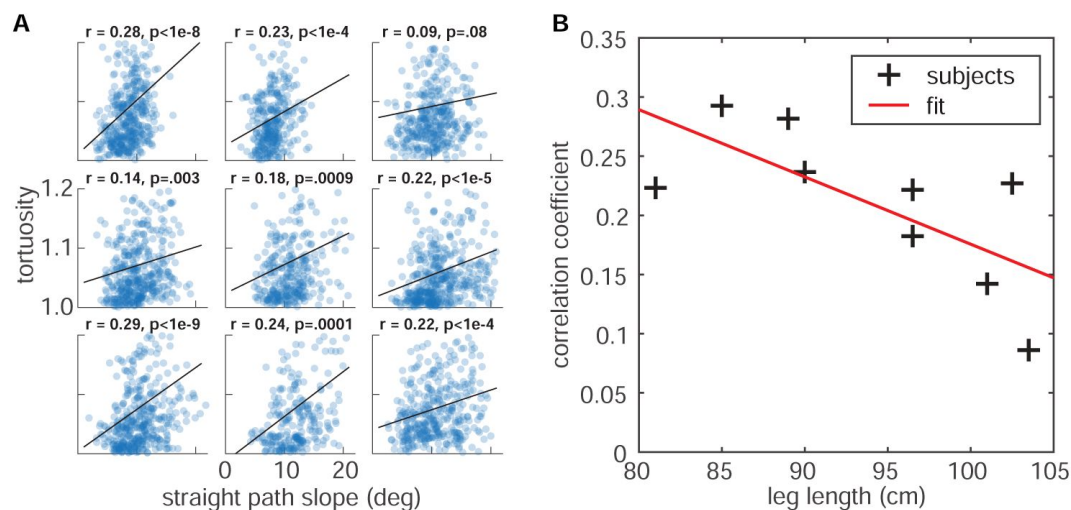
The order of the step parameters along the x- and y-subplot axes are: (1) step distance, (2) step ground distance, (3) step height, (4) step slope, (5) step direction, (6) goal direction, (7) step direction relative to goal direction, and (8) step ground distance in step lengths. The title of each subplot contains the correlation value  $r$  and its associated  $p$ -value.



**Figure 17.**

### Distributions of straight path slope and chosen path segment tortuosity

Each row presents one subject's data. The left column contains histograms of the average slope  $\bar{\theta}$  for simulated "straight" path segments, and the right column contains histograms of the tortuosity of that subject's chosen path segments (5-step sequences).



**Figure 18.**

### Relationship between average straight path segment slope and the tortuosity of each chosen path segment

**A.** Tortuosity of the chosen path segments (5-step sequences) vs. slope  $\bar{\theta}$  for the simulated “straight” path segments for each of the 9 subjects. Correlation values and corresponding  $p$ -values are indicated at the top of each panel. **B.** Correlation between subject leg length and the strength of the straight path slope vs. tortuosity relationship in their foothold selection data. Leg lengths (in centimeters) are plotted on the horizontal axis, and the correlation coefficients for each of the plots in panel A are plotted on the vertical axis. We found a statistically significant negative correlation between those two values ( $r = -.70$ ,  $p = .04$ ).



## References

- Batschelet E. (1981) **Circular Statistics in Biology** *Mathematics in Biology*
- Benhamou S (2004) **How to Reliably Estimate the Tortuosity of an Animal's Path: Straightness, Sinuosity, or Fractal Dimension?** *Journal of Theoretical Biology* **229**:209–220 <https://doi.org/10.1016/j.jtbi.2004.03.016>
- Bonnen K, Matthis JS, Gibaldi A, Banks MS, Levi DM, Hayhoe M (2021) **Binocular Vision and the Control of Foot Placement during Walking in Natural Terrain** *Scientific Reports* **11** <https://doi.org/10.1038/s41598-021-99846-0>
- Darici O, Kuo AD (2023) **Humans Plan for the near Future to Walk Economically on Uneven Terrain** *Proceedings of the National Academy of Sciences of the United States of America* **120** <https://doi.org/10.1073/pnas.2211405120>
- Domínguez-Zamora FJ, Marigold DS. (2021) **Motives Driving Gaze and Walking Decisions** *Current biology: CB* **31**:1632–1642 <https://doi.org/10.1016/j.cub.2021.01.069>
- Finley JM, Bastian AJ, Gottschall JS (2013) **Learning to Be Economical: The Energy Cost of Walking Tracks Motor Adaptation** *The Journal of Physiology* **591**:1081–1095 <https://doi.org/10.1113/jphysiol.2012.245506>
- Foulsham T, Walker E, Kingstone A. (2011) **The Where, What and When of Gaze Allocation in the Lab and the Natural Environment** *Vision Research* **51**:1920–1931 <https://doi.org/10.1016/j.visres.2011.07.002>
- Gallivan JP, Chapman CS, Wolpert DM, Flanagan JR (2018) **Decision-Making in Sensorimotor Control** *Nature Reviews Neuroscience* **19**:519–534 <https://doi.org/10.1038/s41583-018-0045-9>
- Griwodz C, Gasparini S, Calvet L, Gurdjos P, Castan F, Maujean B, De Lillo G, Lanthony Y. (2021) **AliceVision Meshroom: An Open-Source 3D Reconstruction Pipeline** *Proceedings of the 12th ACM Multimedia Systems Conference* :241–247 <https://doi.org/10.1145/3458305.3478443>
- 't Hart BM, 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**:233–249
- Hayhoe MM (2017) **Vision and Action** *Annual Review of Vision Science* **3**:389–413 <https://doi.org/10.1146/annurev-vision-102016-061437>
- Jurca A, Žabkar J, Džeroski S. (2019) **Analysis of 1.2 Million Foot Scans from North America, Europe and Asia** *Scientific Reports* **9** <https://doi.org/10.1038/s41598-019-55432-z>
- Kinsella-Shaw JM, Shaw B, Turvey MT. (1992) **Perceiving 'walk-on-Able' Slopes** *Ecological Psychology* :223–239 [https://doi.org/10.1207/s15326969eco0404\\_2](https://doi.org/10.1207/s15326969eco0404_2)
- Kuo AD, Donelan JM, Ruina A (2005) **Energetic Consequences of Walking like an Inverted Pendulum: Step-to-Step Transitions** *Exercise and Sport Sciences Reviews* **33**:88–97 <https://doi.org/10.1097/00003677-200504000-00006>

Lee DV, Harris SL (2018) **Linking Gait Dynamics to Mechanical Cost of Legged Locomotion** *Frontiers in robotics and AI* **5** <https://doi.org/10.3389/frobt.2018.00111>

Lin MIB, Lin KH (2016) **Walking while performing working memory tasks changes the prefrontal cortex hemodynamic activations and gait kinematics** *Frontiers in Behavioral Neuroscience* **10**

Logan D, Kiemel T, Dominici N, Cappellini G, Ivanenko Y, Lacquaniti F, Jeka JJ (2010) **The many roles of vision during walking** *Experimental brain research* **206**:337–350

Matthis JS, Barton SL, Fajen BR (2017) **The Critical Phase for Visual Control of Human Walking over Complex Terrain** *Proceedings of the National Academy of Sciences of the United States of America* **114**:E6720–E6729 <https://doi.org/10.1073/pnas.1611699114>

Matthis JS, Muller KS, Bonnen KL, Hayhoe MM (2022) **Retinal Optic Flow during Natural Locomotion** *PLoS computational biology* **18** <https://doi.org/10.1371/journal.pcbi.1009575>

Matthis JS, Yates JL, Hayhoe MM. (2018) **Gaze and the Control of Foot Placement When Walking in Natural Terrain** *Current Biology* **28**:1224–1233 <https://doi.org/10.1016/j.cub.2018.03.008>

McNarry M, Wilson R, Holton M, Griffiths I, Mackintosh K. (2017) **Investigating the relationship between energy expenditure, walking speed and angle of turning in humans** *PLoS One* **12**

O'Connor SM, Xu HZ, Kuo AD (2012) **Energetic Cost of Walking with Increased Step Variability** *Gait & Posture* **36**:102–107 <https://doi.org/10.1016/j.gaitpost.2012.01.014>

Patla AE, Vickers JN (1997) **Where and when do we look as we approach and step over an obstacle in the travel path?** *Neuroreport* **8**:3661–3665

Pelz JB, Rothkopf C (2007) **Oculomotor Behavior in Natural and Man-Made Environments** *In: Eye Movements Elsevier* :661–676 <https://doi.org/10.1016/B978-008044980-7/50033-1>

Rio KW, Rhea CK, Warren WH (2014) **Follow the leader: Visual control of speed in pedestrian following** *Journal of vision* **14**:4–4

Rock CG, Marmelat V, Yentes JM, Siu KC, Takahashi KZ (2018) **Interaction between Step-to-Step Variability and Metabolic Cost of Transport during Human Walking** *The Journal of Experimental Biology* **221** <https://doi.org/10.1242/jeb.181834>

Selinger JC, O'Connor SM, Wong JD, Donelan JM (2015) **Humans Can Continuously Optimize Energetic Cost during Walking** *Current Biology* **25**:2452–2456 <https://doi.org/10.1016/j.cub.2015.08.016>

Soule RG, Goldman RF (1972) **Terrain Coefficients for Energy Cost Prediction** *Journal of Applied Physiology* **32**:706–708 <https://doi.org/10.1152/jappl.1972.32.5.706>

Voloshina AS, Kuo AD, Daley MA, Ferris DP (2013) **Biomechanics and Energetics of Walking on Uneven Terrain** *The Journal of Experimental Biology* **216**:3963–3970 <https://doi.org/10.1242/jeb.081711>

Warren WH. (1984) **Perceiving Affordances: Visual Guidance of Stair Climbing** *Journal of Experimental Psychology: Human Perception and Performance* **10**:683–703 <https://doi.org/10.1037/0096-1523.10.5.683>

Warren WH, Kay BA, Zosh WD, Duchon AP, Sahuc S (2001) **Optic flow is used to control human walking** *Nature neuroscience* **4**:213–216

Warren WHJ, Young DS, Lee DN (1986) **Visual Control of Step Length during Running over Irregular Terrain** *Journal of Experimental Psychology: Human Perception and Performance* **12**:259–266 <https://doi.org/10.1037/0096-1523.12.3.259>

Yokoyama H, Sato K, Ogawa T, Yamamoto SI, Nakazawa K, Kawashima N (2018) **Characteristics of the Gait Adaptation Process Due to Split-Belt Treadmill Walking under a Wide Range of Right-Left Speed Ratios in Humans** *PloS One* **13** <https://doi.org/10.1371/journal.pone.0194875>

Zeni JA, Richards JG, Higginson JS (2008) **Two Simple Methods for Determining Gait Events during Tread-mill and Overground Walking Using Kinematic Data** *Gait & Posture* **27**:710–714 <https://doi.org/10.1016/j.gaitpost.2007.07.007>

Zhang Q, Pless R (2004) **Extrinsic Calibration of a Camera and Laser Range Finder (Improves Camera Calibration)** *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)* :2301–2306 <https://doi.org/10.1109/IROS.2004.1389752>

## Editors

Reviewing Editor

**Miriam Spering**

The University of British Columbia, Vancouver, Canada

Senior Editor

**Tirin Moore**

Stanford University, Howard Hughes Medical Institute, Stanford, United States of America

## Reviewer #1 (Public review):

Summary:

The work of Muller and colleagues concerns the question where we place our feet when passing uneven terrain, in particular how we trade-off path length against the steepness of each single step. The authors find that paths are chosen that are consistently less steep and deviate from the straight line more than an average random path, suggesting that participants indeed trade off steepness for path length. They show that this might be related to biomechanical properties, specifically the leg length of the walkers. In addition, they show using a neural network model that participants could choose the footholds based on their sensory (visual) information about depth.

Strengths:

The work is a natural continuation of some of the researchers' earlier work that related the immediately following steps to gaze. Methodologically, the work is very impressive and presents a further step forward towards understanding real-world locomotion and its interaction with sampling visual information. While some of the results may seem somewhat

trivial in hindsight (as always in this kind of studies), I still think this is a very important approach to understand locomotion in the wild better.

Weaknesses:

The concerns I had regarding the initial version of the manuscript have all been fixed in the current one.

<https://doi.org/10.7554/eLife.91243.2.sa3>

#### **Reviewer #2 (Public review):**

This manuscript examines how humans walk over uneven terrain and use vision to decide where to step. There is a huge lack of evidence about this because the vast majority of locomotion studies have focused on steady, well-controlled conditions, and not on decisions made in the real world. The author team has already made great advances in this topic by pioneering gaze recordings during locomotion, but there has been no practical way to map the gaze targets, specifically the 3D terrain features in naturalistic environments. The team has now developed a way to integrate such measurements along with gaze and step tracking. This allows quantitative evaluation of the proposed trade-offs between stepping vertically onto vs. stepping around obstacles, along with how far people look to decide where to step. The team also introduces several new analysis techniques to accompany these measurements. They use machine learning techniques to examine whether retinocentric depth helps predict footholds and develop simulations to assess possible alternative footholds and walking paths. The technical achievement is impressive.

This study addresses several real-world questions not normally examined in the laboratory. First, do humans elect to walk around steeper footholds rather than over them? Second, is there a quantifiable benefit to walking around, such as allowing for a flatter path? Third, does visual depth of terrain contribute to selection of footholds? Fourth, are there scale effects, where for example a tall adult can easily walk over an obstacle that a toddler must walk around. One might superficially answer yes to all of these questions, but it is highly nontrivial to answer them quantitatively. As for the conclusions, my feelings are mixed. I find strengths in answers to two of the questions, and weaknesses in the other two.

Strengths:

I consider the evidence strongest for the first of the main questions. The results show subjects walking with more laterally deviating paths, measured by a quantity called "tortuosity," when the direct straight-ahead paths appear to have steeper ups and downs (Fig. 9). The measure of straight-ahead steepness is fairly complicated (discussed below), but is shown to be well correlated with tortuosity, effectively predicting when subjects will not walk straight ahead.

There is also good evidence for the third question, showing that retinocentric depth is predictive of chosen footholds. Retinocentric depth was computed by a series of steps, starting with scene capture to determine a 3D terrain mesh, projecting that mesh into the eye's perspective, and then discarding all but the depth information. This highly involved process is only the beginning, because the depth was then used to train a neural network classifier with chosen footholds. That network was found to predict footholds better than chance, using a test set independent from the training set, each using half the recorded data. The results are strong and are best interpreted along with a previous study (Bonnen et al. 2021) showing that subjects gaze nearer ahead on rougher terrain, and slightly more so when binocular vision was disrupted. Depth information seems important for foothold selection.

As an aside, humans presumably also select footholds and estimate depth from a number of monocular visual cues, such as shading, shadows, color, and self-motion information.

Interestingly, the terrain mesh and depth data here were computed from monocular images, suggesting that monocular vision can in principle be predictive of both depth and footholds. Binocular human vision presumably improves on monocular depth estimation, and so it would be interesting to see whether binocular scene cameras would predict footholds better. In an earlier review, I had suggested other avenues for exploration, but these are not weaknesses so much as opportunities not yet taken. I believe much could be learned from deeper analysis of the neural network, and future experiments using variations of this technique.

There is much to be appreciated about this study. I was impressed by the overarching outlook and ambitiousness of the team. They seek to understand human decision-making in real-world locomotion tasks, a topic of obvious relevance to the human condition but not often examined in research. The field has been biased toward well-controlled, laboratory studies, which have undeniable scientific advantages but are also quite different from the real world. The present study discards all of the usual advantages of the laboratory, yet still finds a way to explore real-world behaviors in a quantitative manner. It is an exciting and forward-thinking approach, used to tackle an ecologically relevant question.

I also appreciate the numerous technical challenges of this study. The state of the art in real-world locomotion studies has largely been limited to kinematic motion capture. This team managed to collect and analyze an unprecedented, one-of-a-kind dataset. They applied a number of non-trivial methods to assess retinocentric depth, simulate would-be walking paths and steepness, and predict footholds from neural network. Any of these could and probably will merit individual papers, and to assemble them all at once is quite beyond other studies I am aware of. I hope this study will spur more inquiries of this type, leveraging mobile electronics and modern machine learning techniques to answer questions that were previously only addressable qualitatively.

#### Weaknesses:

Although I am highly enthusiastic about this study, I was not entirely convinced by the evidence for the second and fourth questions. Some of this is because I was confused by aspects of the analysis, limiting my understanding of the evidence. But I also question some of the basic conclusions, whether the authors indeed proved that (from Abstract, emphasis mine) "[walkers] change direction TO AVOID taking steeper steps that involve large height changes, instead of [sic] choosing more circuitous, RELATIVELY FLAT paths." (I interpret the "of" as a typo that should have been omitted.) I think it is more objective to say, "walkers changed direction more when straight-ahead paths seemed to have steeper height changes."

I say "seemed" because it is unknown whether humans would have experienced greater height changes if they walked straight ahead (the second main question). The comparison shown is between human tortuous paths taken and simulated straight-ahead paths never experienced by human. Ignoring questions about the simulations for now (discussed below), it is not an apples-to-apples comparison, say between the tortuous paths humans preferred and straight-ahead paths they didn't. The authors determined a measure of steepness, "straight path slope" (Fig. 9), that predicts when humans circuitously, but that is the same as the steepness that humans would actually experience if they had walked straight ahead. That could have been measured with an appropriate control condition, for example asking subjects to walk as straight ahead as they can manage. That also would have eliminated the need for simulations, because the slope of each step actually taken could simply have been measured and compared between conditions. Instead, two different kinds of simulations are compared, where steeper paths are fully simulated, and the circuitous paths are partially simulated but partially based on data. It seems that every fifth circuitous step coincides with a human foothold, but the intervening ones are somewhat random. I don't find this especially strong evidence that the chosen paths were indeed relatively flatter. I would prefer to be convinced by hard data than by unequal simulations.

I also have trouble accepting "TO AVOID" because it implies a degree of intent not evident in the data. I suppose conscious intent could be assessed subjectively by questionnaire, but I don't know how unconscious intent could be tested objectively. I believe my suggested interpretation above is better supported by evidence.

My limited acceptance is due in part to confusion about the simulations. I was especially confused about the connection between feasible steps drawn from the distribution in Figure 7, and the histograms of Figure 8. The feasible steps have clear peaks near zero slope, unity step length, and zero step direction (let's call them Flat). If 5-step simulations of Figure 8 draw from that distribution, why is there zero probability for the 0-3 deg bin (which is within {plus minus}3 deg due to absolute values)? It seems to me that Flat steps were eminently available, so why were they completely avoided? It seems that the simulations were probabilistic (and not just figurative) random walks, which implies they should have had about the same mean as Figure 7 but a wider variance, and then passed through absolute value. They look like something else that I cannot understand. This is important because the RELATIVELY FLAT conclusion is based on the chosen walks apparently being skewed flatter than random simulated walks. I have trouble accepting those distributions because Flat steps were unaccountably never taken by either simulation or human. (This issue is less concerning for Figure 9, because one can accept that some simulation measure is predictive of tortuosity even if the measure is hard to understand.)

I was also confused why Figure 7 distances and directions are nearly normally distributed and not more uniform. The methods only mention constraints to eliminate steps, which to me suggests a truncated uniform distribution. It is not clear to me why the terrain should have a high peak at unity step length, which implies that the only feasible footholds were almost exclusively straight ahead and one step length away. It is possible that the "feasible" footholds are themselves drawn from a "likely" normal distribution, perhaps based on level walking data. It could be argued that simulated steps should be performed by drawing from typical step distributions for level ground, eliminating non-viable footholds, and then repeating that across multiple steps. That would explain the normality, but it is not stated in the Methods, and even if they were "feasible and likely" it would not explain the distributions of Figure 8.

I had some misgivings about the fourth question, where Figure 10 suggests that shorter subjects had greater correlation between straight-path slope and tortuosity than taller ones, who tended to walk straighter ahead. I agree with the authors' rebuttal to my previous review that "the data are the data" but I still have doubts. Now supplied as suggested by another reviewer, Figure 18 provides more detail of the underlying data, with considerably lower correlations. I now suspect that Figure 10 benefits from some statistical artifacts due to binning and other operations, and the weaker correlations of Fig. 18A are closer to reality. I am rather suspicious of correlations of correlations (Figure 18B), which lose some statistical grounding because the second correlation treats all data on equal footing, effectively whitewashing the first correlations of their varying significance (p-values 0.008 to 1e-9).

Furthermore, I am also unsure about Figure 10's comparison of tortuosity vs. straight path slope against leg length. Both tortuosity and straight path slope are already effectively dimensionless and therefore already seem to eliminate scale. It is my understanding that the simulated paths were recomputed for each subject's parameters, and the horizontal axis, slope, is already an angular measure that should affect short and tall people similarly. Shouldn't all subjects equally avoid steep angles, regardless of their dimensional height? If there is indeed a scale effect, then I would expect it to be demonstrated with a dimensional measure (vertical axis) that depends on leg length.

I certainly agree with the hypothetical prior that tall adults walk straight over obstacles that shorter adults (or children) walk around. But I feel that simpler tests would better evidence, perhaps in future work. Did shorter subjects walk with greater tortuosity than taller ones on



the same terrain? Did shorter subjects take relatively more steps even after normalizing for leg length? A possible comparison would be  $(\text{number of steps}) \times (\text{leg length}) / (\text{start to end distance})$ . I feel that the evidence from this study is not that strong.

Although it is a strength of this study that so much can be learned from pure observation, that does not mean controlled conditions are not scientifically helpful. As mentioned earlier, a helpful control could have been to ask subjects to walk straighter but less preferred paths on the same terrain, treating human paths as an independent variable. Another would be to treat terrain as an independent variable, by using level ground and intermediate terrain conditions. This would make it easier to test whether taller subjects walk straighter ahead on more uneven terrain than shorter subjects. Indeed, the data set already includes some patches of flatter terrain, not included here. Additional and simpler tests might be possible based on existing data.

### Conclusion

This is an ambitious undertaking, presenting a wealth of unprecedented data to quantitatively test basic ecological questions that have long been unanswered. There are a number of considerable strengths that merit appreciation, especially the ability to quantitatively predict when humans will walk more circuitously. The weaknesses are about limitations in the conclusions that can be drawn thus far rather than the correctness of the study. I consider this to be a first step that will hopefully enable and inspire a long line of future work that will address these questions more in depth.

<https://doi.org/10.7554/eLife.91243.2.sa2>

### Reviewer #3 (Public review):

#### Summary:

The systematic way in which path selection is parametrically investigated is the main contribution.

#### Strengths:

The authors have developed an impressive workflow to study gait and gaze in natural terrain. They are able to determine footholds and gaze points in the 3D world, and explore different path selections in the terrain.

#### Weaknesses:

The finding that walkers prefer less tortuous, demanding paths is hardly surprising, and from the data it is still not clear what actual visual features are used to choose among alternative routes or what the nature of the decision process is. The authors discuss energetic cost and other "factors" that might influence path selection, but as yet there is no way to express these ideas rigorously in such complex natural settings.

<https://doi.org/10.7554/eLife.91243.2.sa1>

### Author response:

The following is the authors' response to the original reviews.

We thank the reviewers for their constructive reviews. Taken together, the comments and suggestions from reviewers made it clear that we needed to focus on improving the clarity of the methods and results. We have revised the manuscript with that in mind. In particular, we



have restructured the results to make the logic of the manuscript clearer and we have added details to the methods section.

### **Public Reviews:**

#### **Reviewer #1 (Public Review):**

##### *Summary:*

*The work of Muller and colleagues concerns the question of where we place our feet when passing uneven terrain, in particular how we trade-off path length against the steepness of each single step. The authors find that paths are chosen that are consistently less steep and deviate from the straight line more than an average random path, suggesting that participants indeed trade-off steepness for path length. They show that this might be related to biomechanical properties, specifically the leg length of the walkers. In addition, they show using a neural network model that participants could choose the footholds based on their sensory (visual) information about depth.*

##### *Strengths:*

*The work is a natural continuation of some of the researchers' earlier work that related the immediately following steps to gaze [17]. Methodologically, the work is very impressive and presents a further step forward towards understanding real-world locomotion and its interaction with sampling visual information. While some of the results may seem somewhat trivial in hindsight (as always in this kind of study), I still think this is a very important approach to understanding locomotion in the wild better.*

##### *Weaknesses:*

*The manuscript as it stands has several issues with the reporting of the results and the statistics. In particular, it is hard to assess the inter-individual variability, as some of the data are aggregated across individuals, while in other cases only central tendencies (means or medians) are reported without providing measures of variability; this is critical, in particular as  $N=9$  is a rather small sample size. It would also be helpful to see the actual data for some of the information merely described in the text (e.g., the dependence of  $\Delta H$  on path length). When reporting statistical analyses, test statistics and degrees of freedom should be given (or other variants that unambiguously describe the analysis).*

There is only one figure (Figure 6) that shows data pooled over subjects and this is simply to illustrate how the random paths were calculated. The actual paths generated used individual subject data. We don't draw our conclusions from these histograms – they are instead used to generate bounds for the simulated paths. We have made clear both in the text and in the figure legends when we have plotted an example subject. Other plots show the individual subject data. We have given the range of subject medians as well as the standard deviation for data illustrated in Figure (random vs chosen), we have also given the details of the statistical test comparing the flatness of the chosen paths versus the randomly generated paths. We have added two supplemental figures to show individual walker data more directly: (Fig. 14) the per subject histograms of step parameters, (Fig. 18) the individual subject distributions for straight path slopes and tortuosity.

*The CNN analysis chosen to link the step data to visual sampling (gaze and depth features) should be motivated more clearly, and it should describe how training and test sets were generated and separated for this analysis.*

We have motivated the CNN analysis and moved it earlier in the manuscript to help clarify the logic the manuscript. Details of the training and test are now provided, and the data have been replotted. The values are a little different from the original plot after making a correction in the code, but the conclusions drawn from this analysis are unchanged. This analysis simply shows that there is information in the depth images from the subject's perspective that a network can use to learn likely footholds. This motivates the subsequent analysis of path flatness.

*There are also some parts of figures, where it is unclear what is shown or where units are missing. The details are listed in the private review section, as I believe that all of these issues can be fixed in principle without additional experiments.*

Several of the Figures have been replotted to fix these issues.

#### **Reviewer #2 (Public Review):**

##### *Summary:*

*This manuscript examines how humans walk over uneven terrain using vision to decide where to step. There is a huge lack of evidence about this because the vast majority of locomotion studies have focused on steady, well-controlled conditions, and not on decisions made in the real world. The author team has already made great advances in this topic, but there has been no practical way to map 3D terrain features in naturalistic environments. They have now developed a way to integrate such measurements along with gaze and step tracking, which allows quantitative evaluation of the proposed trade-offs between stepping vertically onto vs. stepping around obstacles, along with how far people look to decide where to step.*

##### *Strengths:*

*(1) I am impressed by the overarching outlook of the researchers. They seek to understand human decision-making in real-world locomotion tasks, a topic of obvious relevance to the human condition but not often examined in research. The field has been biased toward well-controlled studies, which have scientific advantages but also serious limitations. A well-controlled study may eliminate human decisions and favor steady or periodic motions in laboratory conditions that facilitate reliable and repeatable data collection. The present study discards all of these usually-favorable factors for rather uncontrolled conditions, yet still finds a way to explore real-world behaviors in a quantitative manner. It is an ambitious and forward-thinking approach, used to tackle an ecologically relevant question.*

*(2) There are serious technical challenges to a study of this kind. It is true that there are existing solutions for motion tracking, eye tracking, and most recently, 3D terrain mapping. However most of the solutions do not have turn-key simplicity and require significant technical expertise. To integrate multiple such solutions together is even more challenging. The authors are to be commended on the technical integration here.*

*(3) In the absence of prior studies on this issue, it was necessary to invent new analysis methods to go with the new experimental measures. This is non-trivial and places an added burden on the authors to communicate the new methods. It's harder to be at the forefront in the choice of topic, technical experimental techniques, and analysis methods all at once.*

##### *Weaknesses:*

*(1) I am predisposed to agree with all of the major conclusions, which seem reasonable and likely to be correct. Ignoring that bias, I was confused by much of the analysis. There is an argument that the chosen paths were not random, based on a comparison of probability distributions that I could not understand. There are plots described as "turn probability vs. X" where the axes are unlabeled and the data range above 1. I hope the authors can provide a clearer description to support the findings. This manuscript stands to be cited well as THE evidence for looking ahead to plan steps, but that is only meaningful if others can understand (and ultimately replicate) the evidence.*

We have rewritten the manuscript with the goal of clarifying the analyses, and we have re-labelled the offending figure.

*(2) I wish a bit more and simpler data could be provided. It is great that step parameter distributions are shown, but I am left wondering how this compares to level walking. The distributions also seem to use absolute values for slope and direction, for understandable reasons, but that also probably skews the actual distribution. Presumably, there should be (and is) a peak at zero slope and zero direction, but absolute values mean that non-zero steps may appear approximately doubled in frequency, compared to separate positive and negative. I would hope to see actual distributions, which moreover are likely not independent and probably have a covariance structure. The covariance might help with the argument that steps are not random, and might even be an easy way to suggest the trade-off between turning and stepping vertically. This is not to disregard the present use of absolute values but to suggest some basic summary of the data before taking that step.*

We have replotted the step parameter distributions without absolute values. Unfortunately, the covariation of step parameters (step direction and step slope) is unlikely to help establish this tradeoff. Note that the primary conclusion of the manuscript is that works make turns to keep step slope low (when possible). Thus, any correlation that might exist between goal direction and step slope would be difficult to interpret without a direct comparison to possible alternative paths (as we have done in this paper). As such we do not draw our conclusions from them. We use them primarily to generate plausible random paths for comparison with the chosen paths. We have added two supplementary figures including distributions (Fig 15) and covariation of all the step parameters discussed in the methods (Fig 16).

*(3) Along these same lines, the manuscript could do more to enable others to digest and go further with the approach, and to facilitate interpretability of results. I like the use of a neural network to demonstrate the predictiveness of stepping, but aside from above-chance probability, what else can inform us about what visual data drives that?*

The CNN analysis simply shows that the information is there in the image from the subject's viewpoint and is used to motivate the subsequent analysis. As noted above, we have generally tried to improve the clarity of the methods.

*Similarly, the step distributions and height-turn trade-off curves are somewhat opaque and do not make it easy to envision further efforts by others, for example, people who want to model locomotion. For that, clearer (and perhaps) simpler measures would be helpful.*

We have clarified the description of these plots in the main text and in the methods. We have also tried to clarify why we made the choices that we did in measuring the height-turn trade-off and why it is necessary in order to make a fair comparison.

*I am absolutely in support of this manuscript and expect it to have a high impact. I do feel that it could benefit from clarification of the analysis and how it supports the conclusions.*

**Reviewer #3 (Public Review):**

**Summary:**

*The systematic way in which path selection is parametrically investigated is the main contribution.*

**Strengths:**

*The authors have developed an impressive workflow to study gait and gaze in natural terrain.*

**Weaknesses:**

*(1) The training and validation data of the CNN are not explained fully making it unclear if the data tells us anything about the visual features used to guide steering. It is not clear how or on what data the network was trained (training vs. validation vs. un-peaked test data), and justification of the choices made. There is no discussion of possible overfitting. The network could be learning just e.g. specific rock arrangements. If the network is overfitting the "features" it uses could be very artefactual, pixel-level patterns and not the kinds of "features" the human reader immediately has in mind.*

The CNN analysis has now been moved earlier in the manuscript to help clarify its significance and we have expanded the description of the methods. Briefly, it simply indicates that there is information in the depth structure of the terrain that can be learned by a network. This helps justify the subsequent analyses. Importantly, the network training and testing sets were separated by terrain to ensure that the model was being tested on “unseen” terrain and avoid the model learning specific arrangements. This is now clarified in the text.

*(2) The use of descriptive terminology should be made systematic.*

*Specifically, the following terms are used without giving a single, clear definition for them: path, step, step location, foot plant, foothold, future foothold, foot location, future foot location, foot position. I think some terms are being used interchangeably. I would really highly recommend a diagrammatic cartoon sketch, showing the definitions of all these terms in a single figure, and then sticking to them in the main text.*

We have made the language more systematic and clarified the definition of each term (see Methods). Path refers to the sequence of 5 steps. Foothold is where the foot was placed in the environment. A step is the transition from one foothold to the next.

*(3) More coverage of different interpretations / less interpretation in the abstract/introduction would be prudent. The authors discuss the path selection very much on the basis of energetic costs and gait stability. At least mention should be given to other plausible parameters the participants might be optimizing (or that indeed they may be just satisficing). That is, it is taken as "given" that energetic cost is the major driver of path selection in your task, and that the relevant perception relies on internal models. Neither of these is a priori obvious nor is it as far as I can tell shown by the data (optimizing other variables, satisficing behavior, or online "direct perception" cannot be ruled out).*

The abstract has been substantially rewritten. We have adjusted our language in the introduction/discussion to try to address this concern.

### **Recommendations for the authors:**

#### **Reviewing Editor comments**

*You will find a full summary of all 3 reviews below. In addition to these reviews, I'd like to highlight a few points from the discussion among reviewers.*

*All reviewers are in agreement that this study has the potential to be a fundamental study with far-reaching empirical and practical implications. The reviewers also appreciate the technical achievements of this study.*

*At the same time, all reviewers are concerned with the overall lack of clarity in how the results are presented. There are a considerable number of figures that need better labeling, text parts that require clearer definitions, and the description of data collection and analysis (esp. with regard to the CNN) requires more care. Please pay close attention to all comments related to this, as this was the main concern that all reviewers shared.*

*At a more specific level, the reviewers discussed the finding around leg length, and admittedly, found it hard to believe, in short: "extraordinary claims need strong evidence". It would be important to strengthen this analysis by considering possible confounds, and by including a discussion of the degree of conviction.*

We have weakened the discussion of this finding and provided some an additional analyses in a supplemental figure (Figure 17) to help clarify the finding.

#### **Reviewer #1 (Recommendations For The Authors):**

*First, let me apologize for the long delay with this review. Despite my generally positive evaluation (see public review), I have some concerns about the way the data are presented and questions about methodological details.*

*(1) Representation of results: I find it hard to decipher how much variability arises within an individual and how much across individuals. For example, Figure 7b seems to aggregate across all individuals, while the analysis is (correctly) based on the subject medians.*

Figure 7b That figure was just one subject. This is now clarified.

*It would be good to see the distribution of all individuals (maybe use violin plots for each observer with the true data on one side and the baseline data on the other, or simple histograms for each). To get a feeling for inter-individual and intra-individual variability is crucial, as obviously (see the leg-length analysis) there are larger inter-individual differences and representations like these would be important to appreciate whether there is just a scaling of more or less the same effect or whether there are qualitative differences (especially in the light of N=9 being not a terribly huge sample size).*

The medians for the individual subjects are now provided with the standard deviations between subjects to indicate the extent of individual differences. Note that the random paths were chosen from the distribution of actual step slopes for that subject as one of the constraints. This makes the random paths statistically similar to the chosen paths with the differences only being generated by the particular visual context. Thus the test for a difference between chosen and random is quite conservative

*Similarly, seeing  $\Delta H$  plotted as a function of steps in the path as a figure rather than just having the verbal description would also help.*

To simplify the discussion of our methods/results we have removed the analyses that examine mean slope as a function of steps. Because of the central limit theorem the slopes of the chosen paths remain largely unchanged regardless of the choice path length. The slopes of the simulated paths are always larger irrespective of the choice of path length.

*(2) Reporting the statistical analyses: This is related to my previous issue: I would appreciate it if the test statistics and degrees-of-freedom of the statistical tests were given along with the p-values, instead of only the p-values. This at some points would also clarify how the statistics were computed exactly (e.g., "All subjects showed comparable difference and the difference in medians evaluated across subjects was highly significant ( $p < 0.0001$ ).", p. 10, is ambiguous to me).*

Details have been added as requested.

*(3) Why is the lower half ("tortuosity less than the median tortuosity") of paths used as "straight" rather than simply the minimum of all viable paths)?*

The benchmark for a straight path is somewhat arbitrary. Using the lower half rather than the minimum length path is more conservative.

*(4) For the CNN analysis, I failed to understand what was training and what was test set. I understand that the goal is to predict for all pixels whether they are a potential foothold or not, and the AUC is a measure of how well they can be discriminated based on depth information and then this is done for each image and the median over all images taken. But on which data is the CNN trained, and on which is it tested? Is this leave-n-out within the same participant? If so, how do you deal with dependencies between subsequent images? Or is it leave-1-out across participants? If so, this would be more convincing, but again, the same image might appear in training and test. If the authors just want to ask how well depth features can discriminate footholds from non-footholds, I do not see the benefit of a supervised method, which leaves the details of the feature combinations inside a black box. Rather than defining the "negative set" (i.e., the non-foothold pixels) randomly, the simulated paths could also be used, instead. If performance (AUC) gets lower than for random pixels, this would confirm that the choice of parameters to define a "viable path" is well-chosen.*

This has been clarified as described above.

*Minor issues:*

*(5) A higher tortuosity would also lead a participant to require more steps in total than a lower tortuosity. Could this partly explain the correlation between the leg length and the slope/tortuosity correlation? (Longer legs need fewer steps in total, thus there might be less tradeoff between  $\Delta H$  and keeping the path straight (i.e., saving steps)). To assess this, you could give the total number of steps per (straight) distance covered for leg length and compare this to a flat surface.*

The calculations are done on an individual subject basis and the first and last step locations are chosen from the actual foot placements, then the random paths are generated between those endpoints. The consequence of this is that the number of steps is held constant for the analysis. We have clarified the methods for this analysis to try to make this more clear.



(6) *As far as I understand, steps happen alternatingly with the two feet. That is, even on a flat surface, one would not reach 0 tortuosity. In other words, does the lateral displacement of the feet play a role (in particular, if paths with even and paths with odd number of steps were to be compared), and if so, is it negligible for the leg-length correlation?*

All the comparisons here are done for 5 step sequences so this potential issue should not affect the slope of the regression lines or the leg length correlation.

(7) *Is there any way to quantify the quality of the depth estimates? Maybe by taking an actual depth image (e.g., by LIDAR or similar) for a small portion of the terrain and comparing the results to the estimate? If this has been done for similar terrain, can a quantification be given? If errors would be similar to human errors, this would also be interesting for the interpretation of the visual sampling data.*

Unfortunately, we do not have the ground truth depth image from LIDAR. When these data were originally collected, we had not imagined being able to reconstruct the terrain. However, we agree with the reviewers that this would be a good analysis to do. We plan to collect LIDAR in future experiments.

To provide an assessment of quality for these data in the absence of a ground truth depth image, we have performed an evaluation of the reliability of the terrain reconstruction across repeats of the same terrain both between and within participants. We have expanded the discussion of these reliability analyses in the results section entitled “Evaluating Terrain Reconstruction”, as well as in the corresponding methods section (see Figure 10).

(8) *The figures are sometimes confusing and a bit sloppy. For example, in Figure 7a, the red, cyan, and green paths are not mentioned in the caption, in Figure 8 units on the axes would be helpful, in Figure 9 it should probably be "tortuosity" where it now states "curviness".*

These details have been fixed.

(9) *I think the statement "The maximum median AUC of 0.79 indicates that the 0.79 is the median proportion of pixels in the circular..." is not an appropriate characterization of the AUC, as the number of correctly classified pixels will not only depend on the ROC (and thus the AUC), but also on the operating point chosen on the ROC (which is not specified by the AUC alone). I would avoid any complications at this point and just characterize the AUC as a measure of discriminability between footholds and non-footholds based on depth features.*

This has been fixed.

(10) *Ref. [16] is probably the wrong Hart paper (I assume their 2012 Exp. Brain Res. [<https://doi.org/10.1007/s00221-012-3254-x>] paper is meant at this point)*

Fixed

*Typos (not checked systematically, just incidental discoveries):*

(11) *"While there substantial overlap" (p.10)*

(12) *"field.." (p.25)*



(13) "Introduction", "General Discussion" and "Methods" as well as some subheadings are numbered, while the other headings (e.g., Results) are not.

Fixed

**Reviewer #2 (Recommendations For The Authors):**

*The major suggestions have been made in the Public Review. The following are either minor comments or go into more detail about the major suggestions. All of these comments are meant to be constructive, not obstructive.*

*Abstract. This is well written, but the main conclusions "Walkers avoid...This trade off is related...5 steps ahead" sound quite qualitative. They could be strengthened by more specificity (NOT p-values), e.g. "positive correlation between the unevenness of the path straight ahead and the probability that people turned off that path."*

The abstract has been substantially rewritten.

*P. 5 "pinning the head position estimated from the IMU to the Meshroom estimates" sounds like there are two estimates. But it does not sound like both were used. Clarify, e.g. the Meshroom estimate of head position was used in place of IMU?*

Yes that's correct. We have clarified this in the text.

*Figure 5. I was confused by this. First, is a person walking left to right? When the gaze position is shown, where was the eye at the time of that gaze? There are straight lines attached to the blue dots, what do they represent? The caption says gaze is directed further along the path, which made me guess the person is walking right to left, and the line originates at the eye. Except the origins do not lie on or close to the head locations. There's also no scale shown, so maybe I am completely misinterpreting. If the eye locations were connected to gaze locations, it would help to support the finding that people look five steps ahead of where they step.*

We have updated the figure and clarified the caption to remove these confusions. There was a mistake in the original figure (where the yellow indicated head locations, we had plotted the center of mass and the choice of projection gave the incorrect impression that the fixations off the path, in blue, were separated from the head).

The view of the data is now presented so the person is walking left to right and with a projection of the head location (orange), gaze locations (blue or green) and feet (pink).

*Figure 6. As stated in the major comments, the step distributions would be expected to have a covariance structure (in terms of raw data before taking absolute values). It would be helpful to report the covariances (6 numbers). As an example of a simple statistical analysis, a PCA (also based on a data covariance) would show how certain combinations of slope/distance/direction are favored over others. Such information would be a simple way to argue that the data are not completely random, and may even show a height-turn trade-off immediately. (By the way, I am assuming absolute values are used because the slopes and directions are only positive, but it wasn't clear if this was the definition.) A reason why covariances and PCA are helpful is that such data would be helpful to compute a better random walk, generated from dynamics. I believe the argument that steps are not random is not served by showing the different histograms in Figure 7, because I feel the random paths are not fairly produced. A better argument might draw randomly from the same distribution as the data (or drive a*

*dynamical random walk), and compare with actual data. There may be correlations present in the actual data that differ from random. I could be mistaken, because it is difficult or impossible to draw conclusions from distributions of absolute values, or maybe I am only confused. In any case, I suspect other readers will also have difficulty with this section.*

This has been addressed above in the major comments.

*p. 9, "average step slope" I think I understand the definition, but I suggest a diagram might be helpful to illustrate this.*

There is a diagram of a single step slope in Figure 6 and a diagram of the average step slope for a path segment in Figure 12.

*Incidentally, the "straight path slope" is not clearly defined. I suspect "straight" is the view from above, i.e. ignoring height changes.*

Clarified

*p. 11 The tortuosity metric could use a clearer definition. Should I interpret "length of the chosen path relative to a straight path" as the numerator and denominator? Here does "length" also refer to the view from above? Why is tortuosity defined differently from step slope? Couldn't there be an analogue to step slope, except summing absolute values of direction changes? Or an analogue to tortuosity, meaning the length as viewed from the side, divided by the length of the straight path?*

We followed the literature in the definition of tortuosity. We have clarified the definition of tortuosity in the methods, but yes, you can interpret the length of the chosen path relative to a straight path, as the numerator and denominator, and length refers to 3D length. We agree that there are many interesting ways to look at the data but for clarity we have limited the discussion to a single definition of tortuosity in this paper.

*Figure 8 could use better labeling. On the left, there is a straight path and a more tortuous path, why not report the metrics for these? On the right, there are nine unlabeled plots. The caption says "turn probability vs. straight path slope" but the vertical axis is clearly not a probability. Perhaps the axis is tortuosity? I presume the horizontal axis is a straight path slope in degrees, but this is not explained. Why are there nine plots, is each one a subject? I would prefer to be informed directly instead of guessing. (As a side note, I like the correlations as a function of leg length, it is interesting, even if slightly unbelievable. I go hiking with people quite a bit shorter and quite a lot taller than me, and anecdotally I don't think they differ so much from each other.)*

We have fixed Figure 8 which shows the average "mean slope" as a function of tortuosity. We have added a supplemental figure which shows a scatter plot of the raw data (mean slope vs. tortuosity for each path segment).

Note that when walking with friends other factors (e.g. social) will contribute to the cost function. As a very short person my experience is that it is a problem. In any case, the data are the data, whatever the underlying reasons. It does not seem so surprising that people of different heights make different tradeoffs. We know that the preferred gait depends on individual's passive dynamics as described in the paper, and the terrain will change what is energetically optimal as described in the Darici and Kuo paper.

Figure 9 presumably shows one data point per subject, but this isn't clear.

The correlations are reported per subject, and this has been clarified.

*p. 13 CNN. I like this analysis, but only sort of. It is convincing that there is SOME sort of systematic decision-making about footholds, better than chance. What it lacks is insight. I wonder what drives peoples' decisions. As an idle suggestion, the AlexNet (arXiv: Krizhevsky et al.; see also A. Karpathy's ConvNETJS demo with CIFAR-10) showed some convolutional kernels to give an idea of what the layers learned.*

Further exploration of CNN's would definitely be interesting, but it is outside the scope of the paper. We use it simply to make a modest point, as described above.

*p. 15 What is the definition of stability cost? I understand energy cost, but it is unclear how circuitous paths have a higher stability cost. One possible definition is an energetic cost having to do with going around and turning. But if not an energy cost, what is it?*

We meant to say that the longer and flatter paths are presumably more stable because of the smaller height changes. You are correct that we can't say what the stability cost is and we have clarified this in the discussion.

*p. 16 "in other data" is not explained or referenced.*

Deleted

*p. 10 5 step paths and p. 17 "over the next 5 steps". I feel there is very little information to really support the 5 steps. A p-value only states the significance, not the amount of difference. This could be strengthened by plotting some measures vs. the number of steps ahead. For example, does a CNN looking 1-5 steps ahead predict better than one looking  $N < 5$  steps ahead? I am of course inclined to believe the 5 steps, but I do not see/understand strong quantitative evidence here.*

We have weakened the statements about evidence for planning 5 steps ahead.

*p. 25 CNN. I did not understand the CNN. The list of layers seems incomplete, it only shows four layers. The convolutional-deconvolutional architecture is mentioned as if that is a common term, which I am unfamiliar with but choose to interpret as akin to encoder-decoder. However, the architecture does not seem to have much of a bottleneck ( $25 \times 25 \times 8$  is not greatly smaller than  $100 \times 100 \times 4$ ), so what is the driving principle? It's also unclear how the decoder culminates, does it produce some  $m \times m$  array of probabilities of stepping, where  $m$  is some lower dimension than the images? It might be helpful also to illustrate the predictions, for example, show a photo of the terrain view, along with a probability map for that view. I would expect that the reader can immediately say yes, I would likely step THERE but not there.*

We have clarified the description of the CNN. An illustration is shown in Figure 11.

### **Reviewer #3 (Recommendations For The Authors):**

*(This section expands on the points already contained in the Public Review).*

*Major issues*

*(1) The training and validation data of the CNN are not explained fully making it unclear if the data tells us anything about the visual features used to guide steering. A CNN was used on the depth scenes to identify foothold locations in the images. This is the bit of the methods and the results that remains ambiguous, and the authors may need to revisit the methods/results. It is not clear how or on what data the network was trained (training vs. validation vs. un-peeked test data), and justification of the choices made. There is no discussion of possible overfitting. The network could be learning just for example specific rock arrangements in the particular place you experimented. Training the network on data from one location and then making it generalize to another location would of course be ideal. Your network probably cannot do this (as far as I can tell this was not tried), and so the meaning of the CNN results cannot really be interpreted.*

*I really like the idea, of getting actual retinotopic depth field approximations. But then the question would be: what features in this information are relevant and useful for visual guidance (of foot placement)? But this question is not answered by your method.*

*"If a CNN can predict these locations above chance using depth information, this would indicate that depth features can be used to explain some variation in foothold selection." But there is no analysis of what features they are. If the network is overfitting they could be very artefactual, pixel-level patterns and not the kinds of "features" the human reader immediately has in mind. As you say "CNN analysis shows that subject perspective depth features are predictive of foothold locations", well, yes, with 50,000 odd parameters the foothold coordinates can be associated with the 3D pixel maps, but what does this tell us?*

See previous discussion of these issues.

It is true that we do not know the precise depth features used. We established that information about height changes was being used, but further work is needed to specify how the visual system does this. This is mentioned in the Discussion.

*You open the introduction with a motivation to understand the visual features guiding path selection, but what features the CNN finds/uses or indeed what features are there is not much discussed. You would need to bolster this, or down-emphasize this aspect in the Introduction if you cannot address it.*

*"These depth image features may or may not overlap with the step slope features shown to be predictive in the previous analysis, although this analysis better approximates how subjects might use such information." I do not think you can say this. It may be better to approximate the kind of (egocentric) environment the subjects have available, but as it is I do not see how you can say anything about how the subject uses it. (The results on the path selection with respect to the terrain features, viewpoint viewpoint-independent allocentric properties of the previous analyses, are enough in themselves!)*

We have rewritten the section on the CNN to make clearer what it can and cannot do and its role in the manuscript. See previous discussion.

*(2) The use of descriptive terminology should be made systematic. Overall the rest of the methodology is well explained, and the workflow is impressive. However, to interpret the results the introduction and discussion seem to use terminology somewhat inconsistently. You need to dig into the methods to figure out the exact operationalizations, and even then you cannot be quite sure what a particular term refers to. Specifically, you use the following terms without giving a single, clear definition for them (my interpretation in parentheses):*

*foothold (a possible foot plant location where there is an "affordance"? or a foot plant location you actually observe for this individual? or in the sample?)*

*step (foot trajectory between successive step locations)*

*step location (the location where the feet are placed)*

*path (are they lines projected on the ground, or are they sequences of foot plants? The figure suggests lines but you define a path in terms of five steps.*

*foot plant (occurs when the foot comes in contact with step location?)*

*future foothold (?)*

*foot location (?)*

*future foot location (?)*

*foot position (?)*

*I think some terms are being used interchangeably here? I would really highly recommend a diagrammatic cartoon sketch, showing the definitions of all these terms in a single figure, and then sticking to them in the main text. Also, are "gaze location" and "fixation" the same? I.e. is every gaze-ground intersection a "gaze location" (I take it it is not a "fixation", which you define by event identification by speed and acceleration thresholds in the methods)?*

We have cleaned up the language. A foothold is the location in the terrain representation (mesh) where the foot was placed. A step is the transition from one foothold to the next. A path is the sequences of 5 steps. The lines simply illustrate the path in the Figures. A gaze location is the location in the terrain representation where the walker is holding gaze still (the act of fixating). See Muller et al (2023) for further explanation.

*(3) More coverage of different interpretations / less interpretation in the abstract/introduction would be prudent. You discuss the path selection very much on the basis of energetic costs and gait stability. At least mention should be given to other plausible parameters the participants might be optimizing (or that indeed they may be just satisficing). Temporal cost (more circuitous route takes longer) and uncertainty (the more step locations you sample the more chance that some of them will not be stable) seem equally reasonable, given the task ecology / the type of environment you are considering. I do not know if there is literature on these in the gait-scene, but even if not then saying you are focusing on just one explanation because that's where there is literature to fall back on would be the thing to do.*

*Also in the abstract and introduction you seem to take some of this "for granted". E.g. you end the abstract saying "are planning routes as well as particular footplants. Such planning ahead allows the minimization of energetic costs. Thus locomotor behavior in natural environments is controlled by decision mechanisms that optimize for multiple factors in the context of well-calibrated sensory and motor internal models". This is too speculative to be in the abstract, in my opinion. That is, you take as "given" that energetic cost is the major driver of path selection in your task, and that the relevant perception relies on internal models. Neither of these is a priori obvious nor is it as far as I can tell shown by your data (optimizing other variables, satisficing behavior, or online "direct perception" cannot be ruled out).*

We have rewritten the abstract and Discussion with these concerns in mind.

*You should probably also reference:*

Warren, W. H. (1984). *Perceiving affordances: Visual guidance of stair climbing*. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 683-703. <https://doi.org/10.1037/0096-1523.10.5.683>

Warren WH Jr, Young DS, Lee DN. *Visual control of step length during running over irregular terrain*. *J Exp Psychol Hum Percept Perform*. 1986 Aug;12(3):259-66. doi: 10.1037//0096-1523.12.3.259. PMID: 2943854.

We have added these references to the introduction.

*Minor point*

*Related to (2) above, the path selection results are sometimes expressed a bit convolutedly, and the gist can get lost in the technical vocabulary. The generation of alternative "paths" and comparison of their slope and tortuousness parameters show that the participants preferred smaller slope/shorter paths. So, as far as I can tell, what this says is that in rugged terrain people like paths that are as "flat" as possible. This is common sense so hardly surprising. Do not be afraid to say so, and to express the result in plain non-technical terms. That an apple falls from a tree is common sense and hardly surprising. Yet quantifying the phenomenon, and carefully assessing the parameters of the path that the apple takes, turned out to be scientifically valuable - even if the observation itself lacked "novelty".*

Thanks. We have tried to clarify the methods/results with this in mind.

<https://doi.org/10.7554/eLife.91243.2.sa0>