Specific Aims

Vision provides crucial information for successfully moving through the environments of daily life. There is a rich and growing body of literature that describes vision and visual perception in response to dynamic and realistic environments [CITE], as well as the details of the biomechanics of walking in natural environments [CITE]. However, the vast majority of experiments investigating the sensorimotor processes underpinning the visually-guided walking are conducted in isolation, focusing primarily on visual perception or motor function [CITE]. As a result, there is a lack of data to support the development of a normative description of the sensorimotor processes involved in walking. This significantly hinders the development of models of the cognitive planning and visual information gathering processes that integrate the details of visual processing AND the bimoechanics of human movement. Understanding these basic sensorimotor processes is critical to human health as we age, as there is considerable evidence that visual impairment and other changes associated with aging put individuals at a high risk for falls.

The overarching goal of this proposal is to **develop an integrated model of the visuomotor processes that support movement through real-world environments**. It will provide a detailed and integrated account of the visual information gathering and cognitive/motor planning processes that support walking. We will take into consideration the role of divided attention and the way that it shapes the coordination of gaze and gait by limiting visual information gathering and cognitive processing. This work will be informed by the collection of an **integrated visuomotor dataset (eye tracking and full-body movement through real-world environments)** and a series of controlled-laboratory experiments with protocols designed to mimic the visually-guided walking observed in the natural environment. These complementary approaches will enable the observation of real-world behavior, while still providing precise laboratory measurements to test specific hypotheses related to the **dynamics of visual information-gathering and motor planning**.

We are uniquely positioned tackle this set of scientific questions, having developed both environment- and laboratory-based data collection techniques that produce integrated visuomotor datasets for full-body movement. The fields of vision science, neuroscience, and biomechanics are at a critical junction as advances in machine learning increase the capacity for processing and analyzing big multi-modal data. However, the success of such efforts is dependent on the content and quality of the data that exist. Our proposed work will result in a high-quality, open-source, visuomotor dataset. Furthermore, each aim includes planned technical deliverables. These open-source solutions will lower the barrier for creating integrated visuomotor datasets. Because of the rarity of such datasets for full-body movement and the current difficulty of producing them, these technical deliverables are a central contribution of the proposal, but one that is deeply intertwined with the scientific goals.

SA1: Information gathering and motor planning during full-body movement through real-world environments. There is a lack of normative baseline data on how individuals use their vision to actively select the information that guides walking through complex environments. Our approach will be to collect an integrated visuomotor dataset (including body movements, eye movements and the environment) in 50 adults with typical vision and motor function. We will analyze motor planning strategies, assess how the moment-to-moment instability of gait impacts gaze behavior, and identify adaptive gaze patterns. More generally, we will model the coordination of gaze/gait during visually-guided walking, establishing typical gaze patterns in the context of the motor behavior. *SA1* will provide a comprehensive description of the sensorimotor processes that underlie visually-guided walking. *Technical Deliverables (SA1):* comprehensive visuomotor dataset; open-source documentation of the hardware infrastructure for collecting low-cost, high-quality integrated visuomotor data; open-source software processing pipeline for integrating multi-modal dataset into the same spatial reference frame.

SA2: Testing the spatial and temporal dynamics of visual information gathering and motor planning. We have developed a body- and gaze-contingent augmented reality ground plane (3m x 10m) for the presentation of arbitrary 2D walking paths. Preliminary results show that the manipulation of foothold sparsity results in modulation of gaze/gait behavior that mimics the changes due to different terrain complexity observed in natural environments. We will manipulate the availability of visual information based on current body position and gaze location to identify the role of temporal dynamics and peripheral processing in the visual information gathering that supports walking. *SA2* will identify when and where in the visual field the critical information for foothold selection during walking occurs. *Technical Deliverables (SA2):* open source hardware specifications for "augmented reality ground-plane"; open-source software processing pipeline for integrating laboratory-based multi-modal data into the same spatial reference frame

SA3: Impact of Divided on Attention on the visuomotor control of walking. The data collection effort described in SA1 will include a divided attention condition. Participants will be asked to walk while talking to an experimenter and/or completing tasks on their phone. Prior work across gait and postural control studies demonstrates that there is a cost to divided attention [CITE]. We will measure the impact of divided attention on the coordination of gaze/gait during full-body movement through real-world environments. *SA3* will provide insight into how visual information gathering and motor planning resources are allocated when individuals are simultaneously engaged in two tasks. *Technical Deliverables (SA3): divided attention extension to comprehensive visuomotor dataset (see SA1)*.

——

1 FROM GOOGLE DOC

When humans walk through the world, their eye movements and body movements are precisely coordinated enabling impressive performance in complicated environments. Past studies have generally been small, and either focused on vision or on the biomechanics of walking, with relatively little crosstalk. A major impediment for the combination is a lack of good datasets, we would want to know body movements, eye movements, and the environment, all at the same time. Such datasets also hold back the development of strong predictive models, as neither eye movements nor body movements can meaningfully be separated during real-world walking. It also holds back research to ask to which level eye-movements and body movements can be seen as jointly optimized. What is missing is a big meaningful dataset, along with the relevant descriptive and normative modeling.

Multiple developments in movement science and neural networks make addressing these problems now realistic. Eye and body movements can now be efficiently traced using pose tracking technologies that I learned to use during my Postdoc. Overall movement of eyes and body can now be modeled with neural networks, that I have used since my PhD thesis. And asking questions about optimality of movements has been the core of my postdoctoral work. The combination of these skills now promises to allow us to move the joint study of walking and eye movement from the laboratory into the real world.

A. Significance

Walking through complex natural environments requires the robust integration of our visual and motor systems. The visual and motor components of visually-guided walking, visual search and the biomechanics of bipedal locomotion, have been studied extensively, but largely in isolation.

- Recent work (that we have been heavily involved in) takes important first steps to integrated understanding of the visuomotor control of walking.

Visual search

Gap 1. how visual processing supports human movement in the real-world

Testing in real life. the sequence of events that happen in the real world.

The large datasets that are being generated only have eye movements and world content: ego 4d and other long list of stuff. An old idea in cognitive science and visual perception that has gained traction in the ML community via reinforcement learning approaches is that the ability to take action and adjust your own input influences learning. We need datasets with stored actions to make this happen. . .

Gap 2. The tech to actually study this doesn't exist. We're uniquely positioned to study it.

New technology is a gap – it's still a lot of work. We're going

- Established the general patterns of visuomotor control, but now we need to:
  - 1. Leverage improved processes for collecting integrated visuomotor datasets to collect more precise, expanded datasets.
  - 2. Develop laboratory protocols that allow for the

that allow for the

B. Innovation

- Integrated visuomotor datasets from humans walking in real-world environments.
  - Extended to include new technologies: photogrammetry methods which extracts the scene structure and location of the camera (i.e. the head) in space, improving the accuracy of body/gaze calibration.
  - A dataset that is one order of magnitude larger than the existing datasets
  - Re-usable data collection pipelines
- augmented reality ground plane
  - Body + eye movement contingent
  - Complementary to outdoor data collection
  - Generation of complex scenes that are more precisely controlled.
  - Approach will allow us to isolate/test real-world observations

Text from Trent's F32:

- Previous laboratory investigations into the visual control of foot placement have been greatly restricted by the length of the walkable space17–19, where only a few steps are recordable in a given trial and (in some cases) no feedback is provided to the participant to tell them if their foot placement was successful. Additionally, some of the best work investigating the visual control of foot placement took place at a time where there was a technological inability to measure precise eye movements in relation to the gait cycle in real time. Here, we develop our new experimental paradigm using an indoor, 14m long (3m wide) projector-based Augmented Reality (AR) groundplane (*Fig. 2*, *Fig. 3*), which participants will actively walk across, enabling us to measure ~25 footsteps made in a given trial. We will measure participant eye movements with a mobile, binocular eye tracker (Pupil Labs), as well as their full body kinematics and position tracking using maker based motion capture (Qualisys). By combining new technologies with a much larger laboratory space (as well as an improved capacity to measure visuo-locomotor in natural, outdoor terrains), we provide the first experimental designs which will be sufficient to understand the influence of biomechanical constraint in visual search.

•

C. Approach

A Note on technical deliverables

1.1 Aim 1: Visually-guided walking in complex terrains

Rationale:

?? because it's awesome.

Furthermore the size of the dataset (n=75) will enable us to establish estimates of natural individual variability for our measurements. (All previous studies are too small for this and too narrow in their data collection purposes to effectively be combined).

General Methods:

Participants will be recruited and screened to include healthy young adults: ages 18-35 who have normal, or corrected-to-normal vision and typically functioning motor systems (details JON!?!). Screenings will include measurement of visual acuities, stereo-vision, Y balance test (what else). We will recruit a diverse set of participants, advertising our experiment in the IUSO optometry clinic, at Ivy Tech Community College (Bloomington), the public library, at 5 local community/recreation centers and other community bulletin boards.

We will conduct data collection from participants (n=75) over the first 2-3 years of the grant period. The methods we will use in Aim 1 are protocols that the co-PIs developed during their postdoctoral work. These methods are described in detail in [CITE]. Data collection relies on three primary pieces of equipment (see Figure XXx): a mobile eye tracker (Pupil Labs), a body motion suit (Rokoko), and a lightweight computer (e.g., Macbook Air). A 9-point VOR eye tracking calibration procedure [CITE] is used to calibrate the eyetracker and place the body motion data in the same reference frame as the gaze data. Finally, a photogrammetry pipeline, described in detail in [CITE], provides a reconstruction of the environment that individuals walked through and the location of the head-mounted camera moving through the environment. The camera location is used to localize the walker in the reconstructed environment.

As in the previous data collection locations [CITE], we have identified locations near Indiana University with easy access to multiple types of terrain: pavement (smooth, flat), medium (pebbles, tree roots), rough (similar to a rocky hiking trail or a dry creekbed), see Figure (XXb). Participants will perform out-and-back walks twice (double-pass) in each of the three conditions (single task, divided attention - talking, divided attention - phone). We expect each walking bout to take 10-15 minutes and for all walking bouts to be completed across two sessions.

Participants will complete walking bouts in all terrains for three different conditions: single task, divided attention - talking, divided attention - phone. In the single task condition, participants will be instructed to walk along the path. In the two divided attention conditions there will be a second task. For the divided attention - talking condition, an experimenter will walk along behind the participant talking to them about their day and their plans for the weekend. For the divided attention - phone condition we will ask participants to load a web experiment created with Pavlovia/Psychopy [CITE] on their phone to play a puzzle game (like Candy-crush) while they walk along the path.

Data analysis will be performed using a set of custom open-source processing pipelines, written in python. These will integrate the three data streams: eye movements, body movements, world content (see technical deliverables below for more details).

Technical Deliverables: A major contribution of this grant will be to lower the barrier to the collection and analysis of integrated visuomotor datasets collected in real-world environments.

[Hardware / Data Collection Documentation] We will provide detailed open source documentation for the hardware integration of a motion capture suit (Rokoko), eye tracker (Pupil Labs), and data collection device (tablet/computer). We will also create detailed documentation and video tutorials for how to collect these data.

[Data Processing Pipelines] There are several steps required to convert the raw data from these devices into a calibrated and integrated visuomotor dataset. In our previous work together, we created an initial version of these pipelines. However they are not easily re-usable by other groups or updated for new technologies. This is necessary due to how quickly the eye tracking and motion capture technologies are changing. Since the creation of those initial pipelines, Matthis (NU) has had significant training in open source software development. Using our internal pipelines as a starting point we will develop a suite of packages that: 1. Perform photogrammetry on the world video content, relying on Meshroom (https://alicevision.org/). 2. Spatially calibrate the gaze data, body motion and environment into the same reference frame (see methods CITE). 3. Generate visualizations of walkers moving through the terrain (e.g., youtube link). All three packages will be written in python and shared as open-source packages on Github.

[Data] The integrated visuomotor dataset (n=75) collected in Aim 1 will be made available as a data repository on Zenodo. This dataset will include the original raw data as well as the fully processed data that includes the photogrammetry reconstruction and spatially calibrated data.

Research Design:

The resulting data will provide the opportunity to answer several questions about visual perception, gaze, gait, the environment, and the coordination of these during walking.

i. *How do people allocate gaze (spatially and temporally) during foothold finding? What is the planning horizon?* Previous work has established that people allocate their gaze to locations 2-5 footholds ahead during foothold finding [CITE], shown also in Figure XXx. These findings did not have the precision afforded by the photogrammetry pipeline described above. Preliminary analyses confirm previous conclusions that walkers are allocating gaze to upcoming footholds, as the increased precision leads to the emergence of "hills of gaze", Figure XXx... For each individual we will quantify the allocation of gaze to upcoming footholds, establishing the variability in gaze allocation across a larger population of individuals. The increased precision also allows the classification of gaze locations by their foothold, or as a non-foothold search location, further distinguishing between non-foothold search locations near the path and those that are part of search sequences before turns (i.e., paths not taken; see Figure XXx [sequence of gaze locations, classified]). For each individual we will quantify the % time they spend on search locations that become footholds, the % time they spend on search locations that don't become footholds and how often those are part of 'paths-not-taken'. As participants walk each path twice, we will measure the variability in gaze allocation both with-in and across participants to establish measurements of variability in gaze allocation for individuals with typical visual and motor systems. Establishing the individual variability for those with typical vision is important as previous work demonstrates that an impairment to visual processing leads to the allocation of gaze closer to the body [CITE binocular walking paper]. Thus, changes in gaze allocation may serve as a behavioral biomarker of issues in visual processing.

Relying on the specified gaze locations (e.g., foothold 1, foothold 2, . . . , non-foothold location, path-not-taken location), we will quantify/model the gaze sequences present during walking in complex terrains. Such sequences can easily be summarized as a markov chain, quantifying the transition probability from one gaze location to the next. We expect that this analysis will capture the broad structure of eye movements in this task. However, we also recognize that the choice of the next gaze location is likely dependent on more than just the current gaze location, a known limitation of this type of Markovian analysis and there are likely common sequences (gaze strategies) that are not captured. Recent work in task-based eye movement analysis has had success in identifying eye movement strategies using Hidden Markov Models (HMMs) and predicting intent using Recurrent Neural Networks (RNNs, specifically LSTM). We will use these modeling approaches to capture the additional complexity in the gaze sequences of walkers.

The larger size of the dataset will offer a unique opportunity to study the sequences of searches on 'paths-not-taken' and examine sensorimotor decision-making in the context of path planning. As these events are less common on any individual walk, access to a large dataset is necessary. Preliminary analyses suggest that [FIGURE OUT HOW TO SAY THIS PROPERLY] humans explore their path options in stereotyped ways... blah blah blah

ii. *How is gaze allocation modulated by step-to-step gait efficiency?* From previous work, we know that terrain difficulty impacts gaze [CITE]. However, many factors change across terrains. In the more difficult terrains, one of central changes to gait is a greater deviation from the preferred gait (i.e., decreased gait efficiency). This accommodates the fewer available footholds due to increased terrain complexity. Here we look to isolate within a terrain the impact of the current level of stability on gaze allocation.

All people have a preferred gait that is dependent on their weight, height, and leg length. Using body measurements and gait measurements during walking on flat terrain (pavement), we can measure preferred gait parameters for each participant. Then we can measure the step-to-step instability as a deviation from that preferred gait (see [CITE] for example) and (using GLM approaches, with history kernels) relate that deviation to look-ahead distance, look-ahead time and gaze location. We expect that individuals will modulate their gaze to do longer term path planning and foothold finding during periods of instability, i.e. their gaze will be farther down the path during periods of stability.

iii. *What makes a "good" foothold?* Foothold locations are determined by a variety of factors: getting to a location, minimizing the energetic costs of walking, avoiding paths that change in height or direction, stability and/or "flatness" of the ground. Basically, you want to get from point A to B, but without getting too tired or stepping on wobbly rocks. How do walkers trade off these different costs? Our preliminary work on this question demonstrates that walkers select paths that are flatter, electing to take more circuitous paths to keep the change in height across steps lower [CITE]. The photogrammetry pipeline was used to build depth maps of the walking terrain and generate viable alternative routes to compare with those chosen by walkers. Additional analyses (using a CNN trained on depth maps) demonstrated that subject-perspective depth maps contain sufficient information to classify foothold locations.

In order to address this question more broadly, we will: 1. Repeat the path height analyses, examining the impact of path height and depth content on foothold selection in our larger dataset, establishing the level of individual variability in the tradeoff between changing path height and more circuitous routes 2. Analyze the role of texture cues (i.e., monocular features) available in the scene on foothold selection. We will train classifiers for foothold and non-foothold locations, using both standard visual processing models (e.g., Portilla-Simoncelli,... 1-2 others) and pretrained ML models (e.g., VGG, ResNet, ViT) generate an embedding of the visual input. The variety of approaches, combined with the filtering out of different types of information (spatial frequency, orientation, regions of interest) will allow us to determine what types of information can be used by such a classifier to distinguish a good foothold location. 3. Develop a binocular RNN model that identifies good foothold locations in a visual scene, learning from the human data. We will examine how generalizable models are across individuals. We will look at the role of monocular vs binocular cues by building comparable monocular and perspective depth map versions of the RNN model. Taken together these analyses will build a better understanding of what makes a good foothold and how visual processing supports the selection of good footholds.

iv. *What is the individual variability in ground clearance?* In studies of walking, mobility, obstacle clearance and stair climbing, foot ground clearance (or more simply step height) is often associated with visual impairment, decreased mobility, or aging [CITE SOME THINGS]. It is thought to be a biomarker of increased caution, and an inability to be as precise either because of visual or motor restrictions. With the photogrammetry reconstruction of the terrain we can make a rough estimate of step height when walking in natural terrains. Initial analyses will examine how step height is modulated by terrain difficulty (likely increasing in both mean and variability) in more difficult terrains.

We will also examine how much of the fluctuations in step-to-step efficiency are due to changing step height and step length and how that is related to the change in path height. Previous work looking at treadmill-walking over uneven terrains examined the energetic tradeoff between the cost of changing step height/length vs. "scuffing" your foot against the ground. Here we can examine SOMETHING

v. *How does divided attention impact the visuomotor control of walking?* When humans walk through the world they are frequently doing more than just finding safe footholds. They might also talk to a friend, or check directions on their phone, or be planning out the rest of their day. Divided attention and/or task switching is required in these contexts. This puts limits on the resources available for the visuomotor control of walking. Previous work on .. has found this to ... Because these contexts are common in everyday life, it is critical that we measure and compare performance to multi-task conditions. In this project we will use two divided attention tasks: 1. Having a conversation with another person and 2. Using a cell phone to play a game or text. Below we briefly describe some of the major hypotheses we have about how divide attention will impact the measured outcomes in our previous research questions (i-iv).

[i. Gaze Allocation] We expect that people will spend less time attending to footholds, allocating some fixations to the task. We will measure how the cost of the second task changes fixation location, fixation duration and fixation sequences. We can further explore this attentional tradeoff by comparing how gaze allocation changes across our two divided attention conditions, which have very different visual information needs. *[ii. Step-to-step gait efficiency]* Is there a decrease in step-to-step efficiency with attention engaged elsewhere? We will measure the step-to-step efficiency observed in the single task and divided attention conditions. The cost of the second task may decrease step-to-step efficiency, and we are curious how variable this decrease is across individuals and whether it is predicted by the measurement of their dynamic balance (NAME OF TEST). *[iii Foothold properties]* Does the choice of footholds become less precise in quantifiable ways? In the divided attention task the number of fixations allocated to footholds will be decreased. Using classification analyses like those described in (iii) we can ask whether there is discernible difference between the visual properties available during the ground fixations performed in the single task condition and the divided attention conditions. *[iv. Ground Clearance]* Finally we are interested in understanding whether walkers adjust their ground clearance in order to avoid "scuffs" or tripping that might be caused by the lowered attention to foothold finding.

vi. [COMMENTS ON MODELING? That point to the end of Aim II]

Experimental concerns.

- This seems hard → Validated in previous published studies that we were authors in.
- What if there's learning? → We will complete the walking bouts in a randomized order twice. Participants will walk the path for the experiment once before. We don't expect measurable effects of learning on this time scale but we will be able to check. If the first 10 participants show evidence of learning effects across sessions we can consider collecting data at two sites.

1.2 Aim 2 - ARGP Stuff

AUGMENTED REALITY GROUND PLANE - Laboratory-precision measurements of visually-guided walking

Rationale:

Why do we need Laboratory-precision measurements?

General Methods: The Augmented Reality Ground-Plane (ARGP, developed at Northeastern) is a projector-based, 14m-long indoor walking path. The content of the groundplane is displayed using a series of X projectors and is body-contingent so it can be updated as a function of the current position of the participant.

For each of the experiments described below, we will recruit groups of participants (n>=10) from the Northeastern Community.

Preliminary data demonstrates that the manipulation of the foothold density (# of available footholds) broadly mimics the changes observed across terrains of different difficulty. As in the outdoor data, walkers fixate the footholds in the upcoming terrain. Like rough terrain, more difficult ARGP walking trials with few available footholds result in gaze allocation closer to the body.

Manipulations: foothold density, path tortuosity, paths with variable foothold density

(by decreasing density in one part of the path we can create path situations.)

Double-pass experiments

Research Design:

i.

A. Are speed fluctuations and deviations from preferred gait repeatable with-in and across participants dependent on the path content? Are gaze paths repeatable within and across participants?

B. What is the planning horizon? Measured as in Darici & Kuo 2023. Measured by gaze paths as well, keeping in mind that we might use central and peripheral vision. Look at where footholds land in the visual field.

C. Do people exhibit optimal control when we take into account the changes to the path characteristics (path shape and foothold density)? [Is there previous work to suggest that lateral deviations in path should cause the kinds of changes that the vertical blocks cause in Darici & Kuo 2023? Or do we have to establish this? Jon says yes... if you have to take a shorter step or move more laterally you should adjust (read his PNAS paper: https://www.pnas.org/doi/epdf/10.1073/pnas.1611699114https://www.pnas.org/doi/epdf/10.1073/pnas.1611

D. Varying "terrain" step density over the course of a single path to force deviations in efficiency

E. Divided Attention – it's all gonna change friend

END OF GRANT NOW (we're dumping the separate models section)

MODELS

These should account for laboratory and world data

A. Inverted optimal control model of walking. Apply Darici and Kuo 2023 to our outdoor data. We know the environment, height changes, overall terrain, possible foothold locations.... Can we predict the deviations for the idealized pavement walker with an inverse optimal control model that takes into account the environment (height changes, possible foothold locations)?

   a. Combining optimal control model developed in Aim II and the one in Darici & Kuo.

   b.

B. GPT

C. RNNs to predict next x visual frames, next footsteps, next gaze locations, next head movements

   a.

Predict saccades (or footholds) from world-reference frame information

Predict saccades (or footholds) from retino-centric information

Use to synthesize artificial terrains to understand how these different sources of information work and what tradeoffs there are.