Contents

Intention to submit document for the Work with US researchers BBSRC-NSF/BIO lead agency 2024 funding opportunity

Enabling Naturalistic, Long-Duration and Continual Neuroscience Experimentation with Advanced Machine Learning

October 26, 2024

1 Summary

Word limit: 2 A summary is not required for this section, please write 'N/A' in the textbox. Please still include a title for your project.

N/A

2 Core team

List the key members of your team and assign them roles from the following:

- project lead (PL)
- project co-lead (UK) (PcL)
- specialist
- professional enabling staff
- research and innovation associate
- technician
- researcher co-lead (RcL)

Only list one individual as project lead.

The core team section must only contain details of the UK applicants. The US applicant information should be listed in the 'US applicants' section.

Find out more about UKRI's core team roles in funding applications.

project lead (PL) Prof. Maneesh Sahani

project co-lead (UK) (PcL) Prof. Tiago Branco, Prof. Thomas Mrsic-Flogel

researcher co-lead (UK) (RcL) Dr. Joaquin Rapela, Dr. Dario Campagner

professional enabling staff Dr. Adam Tyson

3 Application questions

3.1 Research theme

Word limit: 5 Please state the research theme you are applying under. Choose one of the following research themes:

1. biological informatics
2. understanding host-microbe interactions
3. synthetic cells and cellular systems
4. synthetic microbial communities

biological informatics

3.2 Vision

Word limit: 500

What are you hoping to achieve with your proposed work?

What the assessors are looking for in your response

Your vision should clearly address:

- one of the opportunity research themes (biological informatics, understanding host-microbe interactions, synthetic cells and cellular systems or synthetic microbial communities)
- the remit of the BBSRC and the NSF/BIO division associated with your chosen research theme

References may be included within this section, but this will count towards your word count.

Images are not required for this section.

3.2.1 Context

Conventional systems neuroscience experiments are typically short in duration and often place significant constraints on subject behaviour to simplify data analysis. However, these restrictions may limit our ability to observe critical aspects of brain function and behaviour that only manifest in more naturalistic and extended conditions.

At the Sainsbury Wellcome Centre (SWC) and Gatsby Computational Neuroscience Unit (GCNU) we are pioneering Naturalistic, Long-Duration, and Continual (NaLoDuCo) foraging experiments in mice that span weeks to months. During these experiments, we collect high-resolution behavioural and neural recordings in naturalistic settings.

This novel experimental approach will enable researchers to explore neural mechanisms underlying naturalistic behaviour over extended periods for the first time, offering the possibility of uncovering insights across a wide range of phenomena, including long-term behavioural adaptation, neural plasticity, and learning. The data generated from NaLoDuCo experiments represent an entirely new resource in neuroscience, with the potential to drive breakthroughs and discoveries that are beyond the reach of traditional experiments.

Our US collaborator, the Allen Institute for Neural Dynamics (AIND) is also investigating foraging, but using head-fixed mice. Key to their mission is distributing very large Neuroscience datasets, and providing functionality to process them on the cloud.

Our UK business partner, NeuroGEARS Ltd. has been a key business partner of the SWC for the implementation of the NaLoDuCo experimental framework since the project started in 2021, and also provides services to the AIND.

Together we aim at empowering research centres worldwide to adopt this groundbreaking approach. However, the extremely large datasets–on the order of hundreds of terabytes– gathered from experiments spanning weeks to months pose significant challenges in data acquisition, visualisation, and analysis.

While experiments in neuroscience that are naturalistic, long-duration, or continuous have been conducted in the past [e.g., 12, 16, 28], to the best of our knowledge, we are the first to integrate all three of these features in a single experimental paradigm. This combination introduces unprecedented complexities in data processing, as we aim to capture behaviour and brain activity in their most ecologically valid, extended, and uninterrupted forms.

Here we will address these complexities by sharing openly expertise and building software infrasctructure to enable this transformative type of experimentation.

3.2.2 Focus areas

The focus areas of the proposed project are (Figure 1):

Data Collection & Management Efficiently gathering and organising massive datasets over extended periods.

Data Sharing Providing global access to large-scale datasets.

Data Visualisation Developing efficient web-based tools to visualise very large behavioural and neural datasets.

Spike Sorting Assigning spikes to neurons reliably and tracking individual neurons across long-periods of time in real time.

Data Analysis Characterising behavioural and neural recordings (Figure 2).

Inference-Driven Experimentation Creating a new type of experimentation driven by real-time behavioural and neural inferences.

3.2.3 Cross fertilisation

The foraging experiments at the AIND are different from those at the SWC. They do not probe freely moving and naturalistic behaviour, but are able to perform electrophysiological recordings more densely than those at the SWC. These experimental approaches to foraging are complementary and this collaboration will greatly benefit both of them.

At the GCNU and the AIND we are independently developing methods to address several of the focus areas in Figure 1. We will join forces to codevelop these areas and our foraging research programs.

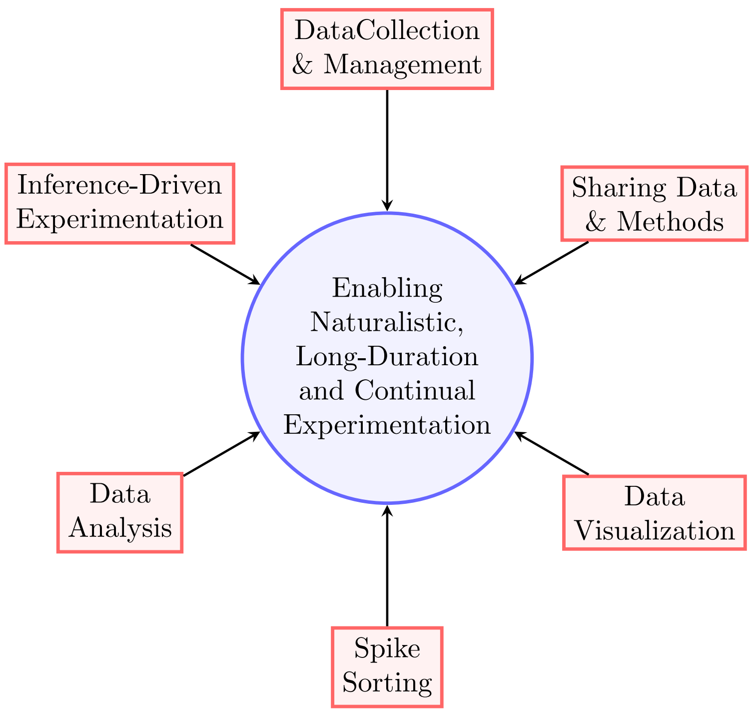


Figure 1: Project theme (blue) and focus areas (red).

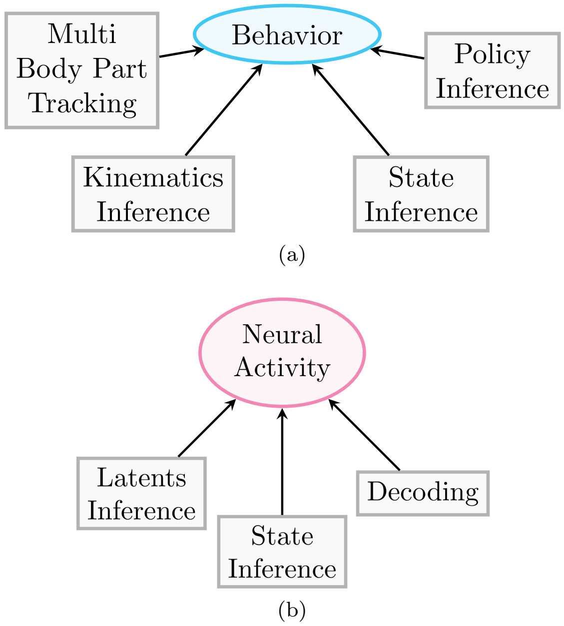


Figure 2: Behavioural (a) and neural (b) data analysis problems to address.

3.3 Approach

Word limit: 500

How are you going to deliver your proposed work?

What the assessors are looking for in your response

Your approach should give an overview highlighting:

- a clear description of the objectives and methodology for the proposed work, including the contributions of the UK and US teams
- the potential outputs and outcomes of the proposed work

References may be included within this section, but this will count towards your word count.

Images are not required for this section.

3.3.1 Data collection & management

We have developed an innovative platform for housing of mice in large arenas (>2m diameter) enabling precise behavioural manipulation and high-resolution monitoring [Figure 3, 2]. We have openly shared software for supporting data acquisition [9] and management [10] in this arena. Additionally, the platforms supports continuous, long term monitoring of neural activity with Neuropixels probes, capable of recording from thousands of neurons simultaneously spanning the entire brain depth. This setup has allowed us to collect several week-long datasets with single and multiple mice per arena.

To facilitate the replication of our experimental setup by other groups, we will share instructions for building foraging arenas, as well as specifications of hardware used in them, and we will improve the documentation of the software repositories for data acquisition and management.

3.3.2 Sharing data and methods

NaLoDuCo experiments produce vast datasets, often exceeding hundreds of terabytes, making traditional methods of data distribution inefficient and impractical. Instead, users will interact with the data directly where it is stored. The maturation of cloud technologies now enables scalable and efficient solutions for handling such large volumes of data.

We will leverage the Distributed Archives for Neuroscience Data Integration (DANDI), which utilizes Amazon S3 storage, for hosting the data. Additionally, we will provide software to visualize and analyze data using Amazon EC2 instances, thereby minimizing the need for costly and time-consuming data transfers.

Handling and sharing continuous behavioral and neural recordings of this scale presents unique challenges. For instance, analyzing such large datasets efficiently is a key concern. We will assess the computational performance of the methods used to address the problems in Figure 2. If we encounter unacceptable delays, we will explore advanced optimization strategies, such as parallel processing and resource-efficient cloud configurations.

3.3.3 Data visualisation

Our visualisation tools need to display very large datasets at different temporal scales, from milliseconds to weeks and months, and they need to be web based. We will use multi-resolution visualisation techniques, which store data at various resolutions, and use the appropriate resolution for each zoom level. Web-based visualisation will be optimised using web workers [6].

3.3.4 Spike sorting

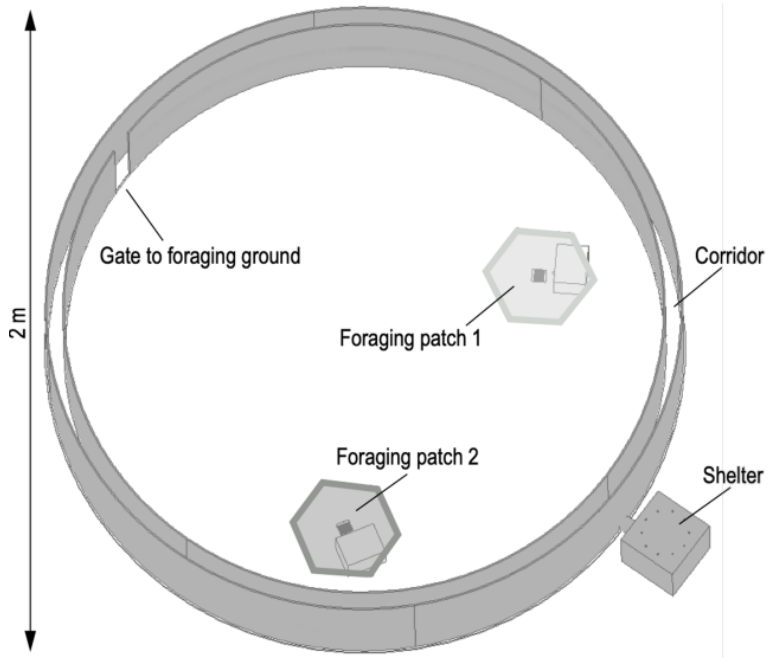
Spike sorting is specially challenging in NaLoDuCo experimentation since we want to track individual neurons of freely moving mice for weeks to months.

In addition, we need online spike sorting, to allow experiments driven by real-time machine learning inference, as described below.

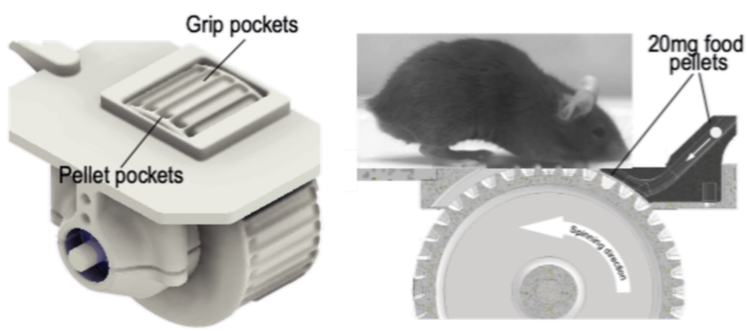
We will evaluate methods for tracking neurons over long periods of time [e.g., 31, 27] and for online sorting [e.g., 23, 25].

3.3.5 Data analysis

The very large size of NaLoDuCo experimental data, the fact that the statistics of these data change across time, and the requirement for real-time and



(a)



(b)

Figure 3: Foraging Arena (a) and Feeder (b). The arena is composed of tessellated hexagonal tiles (a), each featuring a newly designed underground feeder (b). Pellets are dispensed onto a foraging wheel once the mouse has spun it for a pre-defined programmable distance threshold using its forepaws (fictive digging). The arena contains up to six scale-equipped nesting modules that allows housing of mice in the arena and weight monitoring. Behavioural monitoring is achieved by an array of high-speed cameras (up to 15), by which mouse location, mouse identity and body parts can be track in real time.

close-loop inference create new challenges to conventional machine learning data analysis methods. We will evaluate existing methods targeting the experimental problems in Figure 2 and, if necessary, modify them, or create new ones, to address the previous challenges.

For behavioural data, we will evaluate methods to:

- track multiple body parts of animals [e.g., 17, 20, 1, and a switching-linear-dynamical method using RFIDs that we will develop],
- infer kinematics of foraging mice [e.g., 21, 3],
- segment behaviour into discrete states [e.g., 30, 11, and a hierarchical HMM that we will develop],
- infer the rules that govern mice behaviour from behavioural observations only (i.e., policy inference) [e.g., 33, 32].

For neural data, we will evaluate methods to:

- estimate low-dimensional continual representations of neural activity (i.e., latents inference) [e.g., 15, 7, 29, 19, 24],
- segment neural activity into discrete states [e.g., 4, 8],
- decode environment variables from neural activity [e.g., 5, 13, 26].

3.3.6 Inference-driven experimentation

We call inference-driven experimentation to a type of experimentation driven by machine learning inferences on neural or behavioural data, where the result of these inferences can change the experiment in real time.

We will apply inference-driven experimentation to test if patterns of neural activity are causally related to foraging behaviours. We would first check that a pattern of neural activity always precedes a given foraging behaviour. We would then detect the occurrence of the pattern and in real time optogenetically inactivate the neurons responsible for the pattern. If the behaviour disappears the causality argument would be supported.

For this we will use the Bonsai ecosystem for experimental control [14] and online machine learning functionality that we are adding to Bonsai [22], funded by a BBSRC award [18].

3.4 US applicants

Word limit: 200

Please provide the following details of the US applicants on this application:

1. name
2. institute
3. job title
4. role in project (for example, project lead or project co-lead)
5. email address

Please also indicate who the lead US applicant will be.
NSF will use this information to confirm applicant eligibility.
Please do not include details of US applicants in the 'Core team' section.

1. Sakia de Vries

   institute Allen Institute for Neural Dynamics
   job title Associate Director, Data and Outreach
   role in the project project lead
   email saskiad@alleninstitute.org

2. David Feng

   institute Allen Institute for Neural Dynamics
   job title Sr. Director, Scientific Computing
   role in the project project co-lead
   email david.feng@alleninstitute.org

3.5 Resources

Word limit: 200

Please provide the following:

- overall estimates for costings and staffing full time equivalent (FTE) for both the UK and US components
- clear separation of UK and US costings, in pounds sterling and US dollars (USD) respectively

The overall budget should be below the maximum £2 million combined funder contribution

If there is more than one UK or US team associated with the application, please combine their estimates together.

A detailed calculation and breakdown of resources is not required at this stage, nor is a justification of costs.

The following is an example of how this might look.

UK Resources:

Total cost estimate: £600,000

Research council contribution: £480,000

0.2 FTE time, 1.0 FTE PDRA, 0.5 FTE technician

US Resources:

Total cost estimate: $300,000

1.0 FTE PDRA or 1.0 FTE doctoral researcher

Total funder contribution estimate:

£716,475 (£480,000 + £236,475 ($300,000 at exchange rate 0.79))

References

[1] Dan Biderman, Matthew R Whiteway, Cole Hurwitz, Nicholas Greenspan, Robert S Lee, Ankit Vishnubhotla, Richard Warren, Federico Pedraja, Dillon Noone, Michael M Schartner, et al. Lightning pose: improved animal pose estimation via semi-supervised learning, bayesian ensembling and cloud-native open-source tools. *Nature Methods*, pages 1–13, 2024.

[2] D. Campagner, J. Bhagat, G. Lopes, L. Calcaterra, J. Ahn, A. Almeida, F. J. Carvalho, B. Cruz, A. Erskine, C. Lo, T. T. Nguyen, A. Pouget, J. Rapela, T. Ryan, J. Reggiani, and S. SWC Foraging Behaviour Working Group. Aeon: an open-source platform to study the neural basis of ethological behaviours over naturalistic timescales. In *Society for Neuroscience Abstracts*, page PSTR033.03 / I26, 2024. Presented at the Society for Neuroscience Annual Meeting.

[3] Subhash Challa, Mark R. Morelande, Darko Mušicki, and Robin J. Evans. *Fundamentals of Object Tracking*. Cambridge University Press, 2011.

[4] Zhe Chen, Sujith Vijayan, Riccardo Barbieri, Matthew A Wilson, and Emery N Brown. Discrete-and continuous-time probabilistic models and algorithms for inferring neuronal up and down states. *Neural computation*, 21(7):1797–1862, 2009.

[5] Xinyi Deng, Daniel F Liu, Kenneth Kay, Loren M Frank, and Uri T Eden. Clusterless decoding of position from multiunit activity using a marked point process filter. *Neural computation*, 27(7):1438–1460, 2015.

[6] MDN Web Docs. Web workers. https://developer.mozilla.org/en-US/docs/Web/API/Web_Workers_API, 2024. Web Workers makes it possible to run a script operation in a background thread separate from the main execution thread of a web application. The advantage of this is that laborious processing can be performed in a separate thread, allowing the main (usually the UI) thread to run without being blocked/slowed down.

[7] Lea Duncker and Maneesh Sahani. Temporal alignment and latent gaussian process factor inference in population spike trains. In *Advances in Neural Information Processing Systems*, pages 10445–10455, 2018.

[8] Sean Escola, Alfredo Fontanini, Don Katz, and Liam Paninski. Hidden markov models for the stimulus-response relationships of multistate neural systems. *Neural computation*, 23(5):1071–1132, 2011.

[9] SWC Foraging Behavior Working Group. aeon_acquisition repository. https://github.com/SainsburyWellcomeCentre/aeon_acquisition, 2024. Task control and acquisition systems for Project Aeon.

[10] SWC Foraging Behavior Working Group. aeon_mecha repository. https://github.com/SainsburyWellcomeCentre/aeon_mecha, 2024. Project Aeon's main library for interfacing with acquired data. Contains modules for raw data file input/output, data querying, data processing, data quality control, database ingestion, and building computational data pipelines.

[11] Alexander I Hsu and Eric A Yttri. B-soid, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nature communications*, 12(1):5188, 2021.

[12] Hueihan Jhuang, Estibaliz Garrote, Xinlin Yu, Vinita Khilnani, Tomaso Poggio, Andrew D Steele, and Thomas Serre. Automated home-cage behavioural phenotyping of mice. *Nature communications*, 1(1):68, 2010.

[13] Fabian Kloosterman, Stuart P Layton, Zhe Chen, and Matthew A Wilson. Bayesian decoding using unsorted spikes in the rat hippocampus. *Journal of neurophysiology*, 111(1):217–227, 2014.

[14] NeuroGEARS Ltd. Bonsai. https://bonsai-rx.org/, 2024. A visual language for reactive programming.

[15] Jakob H Macke, Lars Buesing, John P Cunningham, Byron M Yu, Krishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neural populations. *Advances in neural information processing systems*, 24, 2011.

[16] Dun Mao, Eric Avila, Baptiste Caziot, Jean Laurens, J David Dickman, and Dora E Angelaki. Spatial modulation of hippocampal activity in freely moving macaques. *Neuron*, 109(21):3521–3534, 2021.

[17] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.

[18] Thomas Mrsic-Flogel. Machine intelligence for neuroscience experimental control. https://gow.bbsrc.ukri.org/grants/AwardDetails.aspx?FundingReference=BB%2FW019132%2F1, 2023. BBSRC award BB/W019132/1.

[19] Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815, 2018.

[20] Talmo D Pereira, Nathaniel Tabris, Arie Matsliah, David M Turner, Junyu Li, Shruthi Ravindranath, Eleni S Papadoyannis, Edna Normand, David S Deutsch, Z Yan Wang, et al. Sleap: A deep learning system for multi-animal pose tracking. *Nature methods*, 19(4):486–495, 2022.

[21] Joaquin Rapela. Linear dynamical systems in python. https://github.com/joacorapela/lds_python, 2024. Python code to estimate Gaussian linear dynamical systems.

[22] Joaquin Rapela, Nicholas Guilbeault, and Goncalo Lopes. Bonsai machine learning. https://bonsai-rx.org/machinelearning/, 2024. Machine learning functionality for experimental control in Bonsai.

[23] Ueli Rutishauser, Erin M Schuman, and Adam N Mamelak. Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. *Journal of neuroscience methods*, 154(1-2):204–224, 2006.

[24] Omid G Sani, Hamidreza Abbaspourazad, Yan T Wong, Bijan Pesaran, and Maryam M Shanechi. Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nature Neuroscience*, 24(1):140–149, 2021.

[25] Gopal Santhanam, Maneesh Sahani, Stephen I Ryu, and Krishna V Shenoy. An extensible infrastructure for fully automated spike sorting during online experiments. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 4380–4384. IEEE, 2004.

[26] Ardi Tampuu, Tambet Matiisen, H Freyja Ólafsdóttir, Caswell Barry, and Raul Vicente. Efficient neural decoding of self-location with a deep recurrent network. *PLoS computational biology*, 15(2):e1006822, 2019.

[27] Enny H van Beest, Célian Bimbard, Julie MJ Fabre, Sam W Dodgson, Flóra Takács, Philip Coen, Anna Lebedeva, Kenneth D Harris, and Matteo Carandini. Tracking neurons across days with high-density probes. *Nature Methods*, pages 1–10, 2024.

[28] Benjamin Voloh, David J-N Maisson, Roberto Lopez Cervera, Indirah Conover, Mrunal Zambre, Benjamin Hayden, and Jan Zimmermann. Hierarchical action encoding in prefrontal cortex of freely moving macaques. *Cell reports*, 42(9), 2023.

[29] William I Walker, Hugo Soulat, Changmin Yu, and Maneesh Sahani. Unsupervised representation learning with recognition-parametrised probabilistic models. In *International Conference on Artificial Intelligence and Statistics*, pages 4209–4230. PMLR, 2023.

[30] Alexander B Wiltschko, Matthew J Johnson, Giuliano Iurilli, Ralph E Peterson, Jesse M Katon, Stan L Pashkovski, Victoria E Abraira, Ryan P Adams, and Sandeep Robert Datta. Mapping sub-second structure in mouse behavior. *Neuron*, 88(6):1121–1135, 2015.

[31] Augustine Xiaoran Yuan, Jennifer Colonell, Anna Lebedeva, Michael Okun, Adam S Charles, and Timothy D Harris. Multi-day neuron tracking in high-density electrophysiology recordings using earth mover's distance. *Elife*, 12:RP92495, 2024.

[32] Hao Zhu, Brice De La Crompe, Gabriel Kalweit, Artur Schneider, Maria Kalweit, Ilka Diester, and Joschka Boedecker. L(m)v-iql: Multiple intention inverse reinforcement learning for animal behavior characterization. *arXiv preprint arXiv:2311.13870*, 2023.

[33] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.