

1	Contents	
2	1 Summary	2
3	2 Core team	3
4	3 Application questions	3
5	3.1 Research theme	3
6	3.2 Vision	4
7	3.2.1 Context	4
8	3.2.2 Focus areas	5
9	3.2.3 Cross fertilisation	5
10	3.3 Approach	7
11	3.3.1 Data collection & management	7
12	3.3.2 Sharing data and methods	8
13	3.3.3 Data visualisation	8
14	3.3.4 Spike sorting	8
15	3.3.5 Data analysis	8
16	3.3.6 Inference-driven experimentation	10
17	3.4 US applicants	11
18	3.5 Resources	12
19	References	12

1 Intention to submit document for the Work with
2 US researchers BBSRC-NSF/BIO lead agency
3 2024 funding opportunity

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

7
8 October 26, 2024

9 **1 Summary**

10 Word limit: 2 A summary is not required for this section, please write 'N/A' in
11 the textbox. Please still include a title for your project.
12 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

1 3.2 Vision

2 Word limit: 500

3 What are you hoping to achieve with your proposed work?

4 What the assessors are looking for in your response

5 Your vision should clearly address:

- 6 • one of the opportunity research themes (biological informatics, under-
7 standing host-microbe interactions, synthetic cells and cellular systems or
8 synthetic microbial communities)
- 9 • the remit of the BBSRC and the NSF/BIO division associated with your
10 chosen research theme

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

13 Images are not required for this section.

14 3.2.1 Context

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

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

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

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

36 Since the project started in 2021, our UK business partner, NeuroGEARS
37 Ltd. has been contracted by the SWC to lead the implementation of the NaLo-
38 DuCo experimental framework. It also provides services to the AIND.

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

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

7 Here we will address these complexities by sharing openly expertise and
8 building software infrastructure to enable this transformative type of experi-
9 mentation.

10 3.2.2 Focus areas

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

12 **Data Collection & Management** Efficiently gathering and organising mas-
13 sive datasets over extended periods.

14 **Data Sharing** Providing global access to large-scale datasets.

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

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

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

20 **Inference-Driven Experimentation** Creating a new type of experimenta-
21 tion driven by real-time behavioural and neural inferences.

22 3.2.3 Cross fertilisation

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

28 Currently, both GCNU and AIND are independently developing methods
29 to address the focus areas outlined in Figure 1. We will joint forces to co-
30 develop these areas and our foraging research programs, leveraging our combined
31 expertise for greater impact.

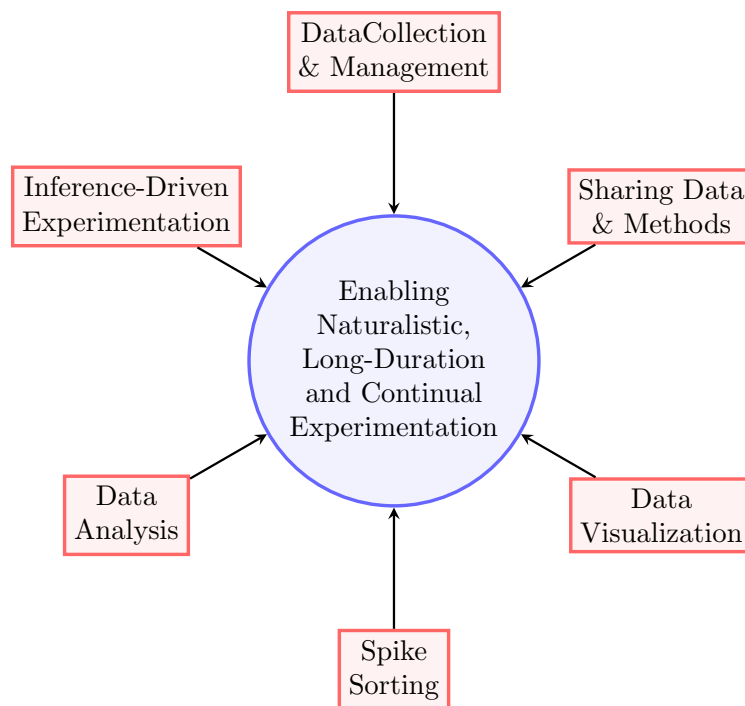


Figure 1: Project theme (blue) and focus areas (red).

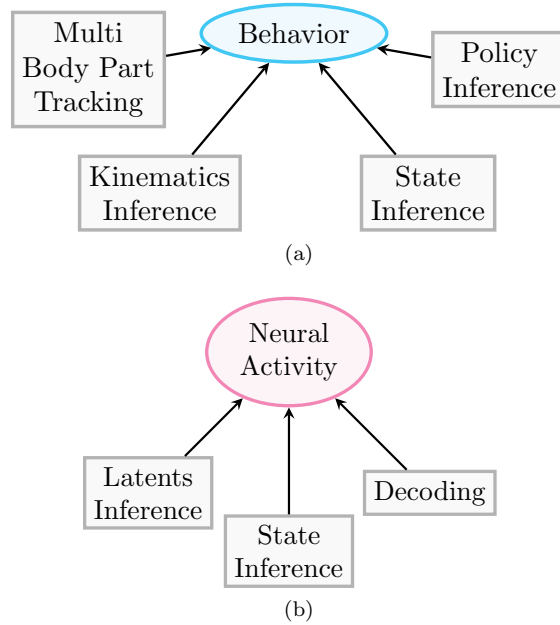


Figure 2: Behavioural (a) and neural (b) data analysis problems to address.

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

1 the entire brain depth. This setup has allowed us to collect several week-long
2 datasets with single and multiple mice per arena.

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

7 **3.3.2 Sharing data and methods**

8 NaLoDuCo experiments produce vast datasets, often exceeding hundreds of ter-
9 abytes, making traditional methods of data distribution inefficient and imprac-
10 tical. Instead, users will interact with the data directly where it is stored. The
11 maturation of cloud technologies now enables scalable and efficient solutions for
12 handling such large volumes of data.

13 We will leverage the Distributed Archives for Neuroscience Data Integra-
14 tion (DANDI), which utilizes Amazon S3 storage, for hosting the data. Addi-
15 tionally, we will provide software to visualize and analyze data using Amazon
16 EC2 instances, thereby minimizing the need for costly and time-consuming data
17 transfers.

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

24 **3.3.3 Data visualisation**

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

30 **3.3.4 Spike sorting**

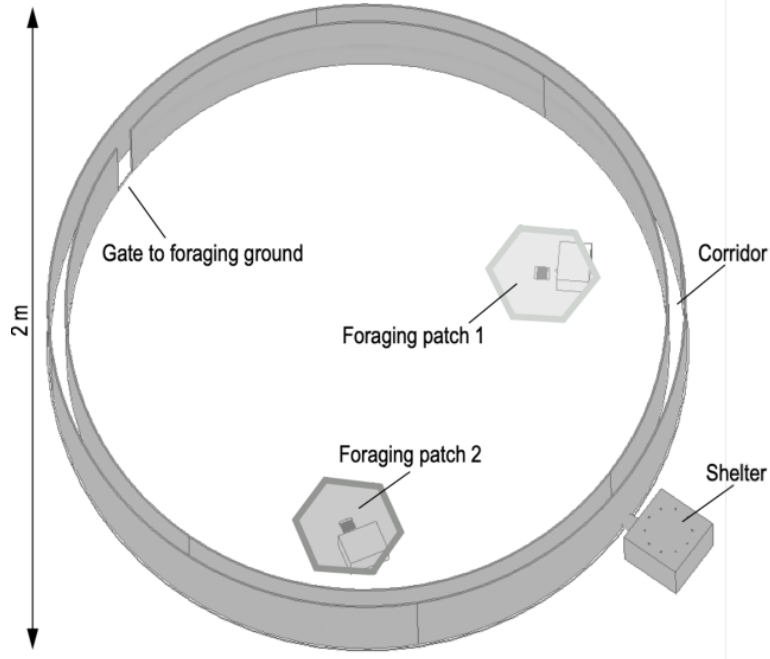
31 Spike sorting is specially challenging in NaLoDuCo experimentation since we
32 want to track individual neurons of freely moving mice for weeks to months.

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

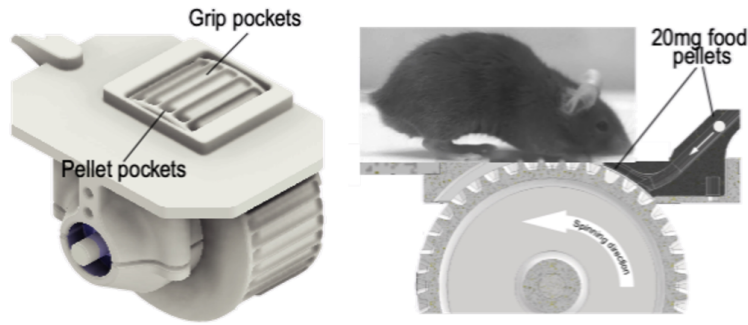
35 We will evaluate methods for tracking neurons over long periods of time [e.g.,
36 31, 27] and for online sorting [e.g., 23, 25].

37 **3.3.5 Data analysis**

38 The very large size of NaLoDuCo experimental data, the fact that the statis-
39 tics 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.

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

5 For behavioural data, we will evaluate methods to:

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

13 For neural data, we will evaluate methods to:

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

18 3.3.6 Inference-driven experimentation

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

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

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

1 3.4 US applicants

2 Word limit: 200

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

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

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

12 1. Sakia de Vries

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

17 2. David Feng

18 **institute** Allen Institute for Neural Dynamics
19 **job title** Sr. Director, Scientific Computing
20 **role in the project** project co-lead
21 **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 PSTRO33.03 / I26, 2024. Presented at the [Society for Neuroscience Annual Meeting](#).
- [3] Subhash Challa, Mark R. Moreland, Darko Mušicki, and Robin J. Evans. *Fundamentals of Object Tracking*. Cambridge University Press, 2011.

- 1 [4] Zhe Chen, Sujith Vijayan, Riccardo Barbieri, Matthew A Wilson, and
2 Emery N Brown. Discrete-and continuous-time probabilistic models and
3 algorithms for inferring neuronal up and down states. *Neural computation*,
4 21(7):1797–1862, 2009.
- 5 [5] Xinyi Deng, Daniel F Liu, Kenneth Kay, Loren M Frank, and Uri T Eden.
6 Clusterless decoding of position from multiunit activity using a marked
7 point process filter. *Neural computation*, 27(7):1438–1460, 2015.
- 8 [6] MDN Web Docs. Web workers. [https://developer.mozilla.org/
9 en-US/docs/Web/API/Web_Workers_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Workers_API), 2024. Web Workers makes it
10 possible to run a script operation in a background thread separate from
11 the main execution thread of a web application. The advantage of this is
12 that laborious processing can be performed in a separate thread, allow-
13 ing the main (usually the UI) thread to run without being blocked/slowed
14 down.
- 15 [7] Lea Duncker and Maneesh Sahani. Temporal alignment and latent gaussian
16 process factor inference in population spike trains. In *Advances in Neural
17 Information Processing Systems*, pages 10445–10455, 2018.
- 18 [8] Sean Escola, Alfredo Fontanini, Don Katz, and Liam Paninski. Hidden
19 markov models for the stimulus-response relationships of multistate neural
20 systems. *Neural computation*, 23(5):1071–1132, 2011.
- 21 [9] SWC Foraging Behavior Working Group. aeon_acquisition repository.
22 https://github.com/SainsburyWellcomeCentre/aeon_acquisition,
23 2024. Task control and acquisition systems for Project Aeon.
- 24 [10] SWC Foraging Behavior Working Group. aeon_mecha repository. [https:
25 //github.com/SainsburyWellcomeCentre/aeon_mecha](https://github.com/SainsburyWellcomeCentre/aeon_mecha), 2024. Project
26 Aeon’s main library for interfacing with acquired data. Contains modules
27 for raw data file input/output, data querying, data processing, data quality
28 control, database ingestion, and building computational data pipelines.
- 29 [11] Alexander I Hsu and Eric A Yttri. B-soid, an open-source unsupervised
30 algorithm for identification and fast prediction of behaviors. *Nature com-
31 munications*, 12(1):5188, 2021.
- 32 [12] Hueihan Jhuang, Estibaliz Garrote, Xinlin Yu, Vinita Khilnani, Tomaso
33 Poggio, Andrew D Steele, and Thomas Serre. Automated home-cage be-
34 havioural phenotyping of mice. *Nature communications*, 1(1):68, 2010.
- 35 [13] Fabian Kloosterman, Stuart P Layton, Zhe Chen, and Matthew A Wilson.
36 Bayesian decoding using unsorted spikes in the rat hippocampus. *Journal
37 of neurophysiology*, 111(1):217–227, 2014.
- 38 [14] NeuroGEARS Ltd. Bonsai. <https://bonsai-rx.org/>, 2024. A visual
39 language for reactive programming.

- 1 [15] Jakob H Macke, Lars Buesing, John P Cunningham, Byron M Yu, Kr-
2 ishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neu-
3 ral populations. *Advances in neural information processing systems*, 24,
4 2011.
- 5 [16] Dun Mao, Eric Avila, Baptiste Caziot, Jean Laurens, J David Dickman,
6 and Dora E Angelaki. Spatial modulation of hippocampal activity in freely
7 moving macaques. *Neuron*, 109(21):3521–3534, 2021.
- 8 [17] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe,
9 Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge.
10 Deeplabcut: markerless pose estimation of user-defined body parts with
11 deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.
- 12 [18] Thomas Mrsic-Flogel. Machine intelligence for neuroscience experi-
13 mental control. [https://gow.bbsrc.ukri.org/grants/AwardDetails.](https://gow.bbsrc.ukri.org/grants/AwardDetails.aspx?FundingReference=BB%2FW019132%2F1)
14 [aspx?FundingReference=BB%2FW019132%2F1](https://gow.bbsrc.ukri.org/grants/AwardDetails.aspx?FundingReference=BB%2FW019132%2F1), 2023. BBSRC award
15 BB/W019132/1.
- 16 [19] Chethan Pandarinath, Daniel J O’Shea, Jasmine Collins, Rafal Jozefow-
17 icz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T
18 Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial
19 neural population dynamics using sequential auto-encoders. *Nature meth-*
20 *ods*, 15(10):805–815, 2018.
- 21 [20] Talmo D Pereira, Nathaniel Tabris, Arie Matsliah, David M Turner, Junyu
22 Li, Shruthi Ravindranath, Eleni S Papadoyannis, Edna Normand, David S
23 Deutsch, Z Yan Wang, et al. Sleep: A deep learning system for multi-animal
24 pose tracking. *Nature methods*, 19(4):486–495, 2022.
- 25 [21] Joaquin Rapela. Linear dynamical systems in python. [https://github.](https://github.com/joacorapela/lds_python)
26 [com/joacorapela/lds_python](https://github.com/joacorapela/lds_python), 2024. Python code to estimate Gaussian
27 linear dynamical systems.
- 28 [22] Joaquin Rapela, Nicholas Guilbeault, and Goncalo Lopes. Bonsai ma-
29 chine learning. <https://bonsai-rx.org/machinelearning/>, 2024. Ma-
30 chine learning functionality for experimental control in Bonsai.
- 31 [23] Ueli Rutishauser, Erin M Schuman, and Adam N Mamelak. Online de-
32 tection and sorting of extracellularly recorded action potentials in human
33 medial temporal lobe recordings, in vivo. *Journal of neuroscience methods*,
34 154(1-2):204–224, 2006.
- 35 [24] Omid G Sani, Hamidreza Abbaspourazad, Yan T Wong, Bijan Pesaran,
36 and Maryam M Shanechi. Modeling behaviorally relevant neural dynam-
37 ics enabled by preferential subspace identification. *Nature Neuroscience*,
38 24(1):140–149, 2021.

- 1 [25] Gopal Santhanam, Maneesh Sahani, Stephen I Ryu, and Krishna V Shenoy.
2 An extensible infrastructure for fully automated spike sorting during online
3 experiments. In *The 26th Annual International Conference of the IEEE*
4 *Engineering in Medicine and Biology Society*, volume 2, pages 4380–4384.
5 IEEE, 2004.
- 6 [26] Ardi Tampuu, Tambet Matiisen, H Freyja Ólafsdóttir, Caswell Barry, and
7 Raul Vicente. Efficient neural decoding of self-location with a deep recur-
8 rent network. *PLoS computational biology*, 15(2):e1006822, 2019.
- 9 [27] Enny H van Beest, Célian Bimbard, Julie MJ Fabre, Sam W Dodgson,
10 Flóra Takács, Philip Coen, Anna Lebedeva, Kenneth D Harris, and Matteo
11 Carandini. Tracking neurons across days with high-density probes. *Nature*
12 *Methods*, pages 1–10, 2024.
- 13 [28] Benjamin Voloh, David J-N Maisson, Roberto Lopez Cervera, Indirah
14 Conover, Mrunal Zambre, Benjamin Hayden, and Jan Zimmermann. Hi-
15 erarchical action encoding in prefrontal cortex of freely moving macaques.
16 *Cell reports*, 42(9), 2023.
- 17 [29] William I Walker, Hugo Soulat, Changmin Yu, and Maneesh Sahani. Unsu-
18 pervised representation learning with recognition-parametrised probabilis-
19 tic models. In *International Conference on Artificial Intelligence and Statis-*
20 *tics*, pages 4209–4230. PMLR, 2023.
- 21 [30] Alexander B Wiltschko, Matthew J Johnson, Giuliano Iurilli, Ralph E Pe-
22 terson, Jesse M Katon, Stan L Pashkovski, Victoria E Abaira, Ryan P
23 Adams, and Sandeep Robert Datta. Mapping sub-second structure in
24 mouse behavior. *Neuron*, 88(6):1121–1135, 2015.
- 25 [31] Augustine Xiaoran Yuan, Jennifer Colonell, Anna Lebedeva, Michael Okun,
26 Adam S Charles, and Timothy D Harris. Multi-day neuron tracking in high-
27 density electrophysiology recordings using earth mover’s distance. *Elife*,
28 12:RP92495, 2024.
- 29 [32] Hao Zhu, Brice De La Crompe, Gabriel Kalweit, Artur Schneider, Maria
30 Kalweit, Ilka Diester, and Joschka Boedecker. L(m)v-iql: Multiple intention
31 inverse reinforcement learning for animal behavior characterization. *arXiv*
32 *preprint arXiv:2311.13870*, 2023.
- 33 [33] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al.
34 Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages
35 1433–1438. Chicago, IL, USA, 2008.