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 31, 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 subjects behaviours 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 approach will enable researchers to explore neural mechanisms underlying ethological behaviours in naturalistic environments over months, for the first time. The experiments will shed new light on a wide range of poorly understood neural mechanisms, including how the brain structures complex behavioural sequences as a function of the animal needs, learning and social dynamics. 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.

While experiments in neuroscience that are naturalistic, long-duration, or continuous have been conducted in the past (e.g., [1]), to the best of our knowledge, we are the first to integrate all three of these features in a single experimental paradigm. Experiments of this type have been advocated by experts in the field years ago ([2], p.19), yet they have not been implemented so far.

We anticipate that this new type of experimentation will become mainstream in the coming years. However, experiments spanning weeks to months generate extremely large datasets—often reaching hundreds of terabytes—which present substantial challenges across data acquisition, management, distribution, visualization, and analysis. Together, with our US partner, the Allen Institute for

Neural Dynamics (AIND), we will address these challenges building software infrastructure to help scientists around the world perform NaLoDuCo experiments.

3.2.2 Focus areas

Developing platform technologies for:

Experimental Control, Data Acquisition & Management Controlling sophisticated experiments and efficiently gathering and organising massive datasets over extended periods.

Data Sharing Providing global access to large-scale datasets.

Data Visualisation Building web-based visualisations for very large behavioural and neural data.

Data Analysis Characterising behavioural and neural recordings with advanced machine learning methods.

3.2.3 Synergistic developments

Our team is highly qualified to deliver this proposal, with world-class expertise in experimental and computational neuroscience, as well as machine learning. Both partners also include talented research software engineers, creating a solid foundation for achieving the project goals.

The AIND is also investigating foraging behavior, but using head-fixed mice. They do not probe freely moving and naturalistic behaviour, but are able to perform electrophysiological recordings more densely than the SWC. The experimental approaches to foraging by the AIND and SWC are complementary and this collaboration will greatly benefit both of them.

Currently, both GCNU and AIND are independently developing methods to address the previous focus areas. We will join forces to co-develop these methods and our foraging research programs, leveraging our combined expertise for greater impact.

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 Foundational progress: experimental control, data acquisition & management

We have developed an innovative platform for housing of mice in large arenas (>2m diameter) enabling precise behavioural manipulation and high-resolution monitoring (online figure, [2]). We have openly shared software for supporting data acquisition [3] and management [4] in this arena. Additionally, the platform supports continuous, long term monitoring of neural activity with Neuropixels probes. This setup has allowed us to collect several week-long datasets with single and multiple mice per arena.

3.3.2 Intelligent experimental control

Bonsai is an unparalleled software for experimental control that both the SWC and the AIND are using to control their foraging experiments. We are currently adding machine learning functionality to Bonsai, funded by BBSRC [5].

We will continue enhancing Bonsai with machine learning methods developed in this project. Our goal is to enable a new type of experimentation, driven by sophisticated inferences from behavioural and neural recordings. This innovative approach could be transformative for foraging research.

3.3.3 Data sharing

The very large datasets produced by NaLoDuCo experiments make traditional methods of data distribution impractical. Instead, users will interact with the data directly where it is stored. The maturation of cloud technologies now makes this possible.

We will leverage DANDI [6], which utilises Amazon S3 storage, for hosting data. Additionally, we will provide software to visualise and analyse data using Amazon EC2 instances, thereby minimising the need for time-consuming data transfers.

Handling and sharing continuous behavioural and neural recordings of this scale presents unique challenges. Runtime performance is one of them. If we encounter unacceptable delays, we will explore advanced optimisation strategies, such as parallel processing and resource-efficient cloud configurations.

3.3.4 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.

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 close-loop inference create new challenges to conventional machine learning data analysis methods. We will evaluate how existing methods targeting the focus areas cope with these challenges and, if necessary, create new ones.

For behavioural data, we will investigate methods to:

track multiple body parts of animals (e.g., [7] and a switching-linear-dynamical method using RFIDs that we will develop),

infer kinematics of foraging mice (e.g., [8,9]),

segment behaviour into discrete states (e.g., [10] 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., [11]).

For neural data, we will investigate methods to:

estimate low-dimensional continual representations of neural activity (i.e., latents inference) (e.g., [12]),

segment neural activity into discrete states (e.g., [13]),

decode environment variables from neural activity (e.g., [14]).

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. Saskia 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))

UK Resources:

Total cost estimate: £1,485,198.15

Research council contribution: £1,188,158.52

2 x 0.1 FTE PI, 1 x 0.5 FTE PDRA, 2 x 1.0 FTE RSE

US Resources:

Total cost estimate: $700.000

1 x 0.5 FTE scientist 1

Total funder contribution estimate:

£1,741,158.52 (£1,188,158.52 + £553,000 ($700.000 at exchange rate 0.79))