

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	6
11	3.3.1 Data collection & management	6
12	3.3.2 Sharing data and methods	6
13	3.3.3 Data visualisation	7
14	3.3.4 Spike sorting	7
15	3.3.5 Data analysis	7
16	3.4 US applicants	8
17	3.5 Resources	9
18	References	9

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 28, 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 subjects behaviours 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 While experiments in neuroscience that are naturalistic, long-duration, or
33 continuous have been conducted in the past (e.g., [1]), to the best of our knowl-
34 edge, we are the first to integrate all three of these features in a single experi-
35 mental paradigm.

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

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

1 The extremely large datasets—on the order of hundreds of terabytes—gathered
2 from experiments spanning weeks to months pose significant challenges in data
3 acquisition, visualisation, and analysis. Together, the GCNU, SWC, AIND and
4 NeuroGEARS will address these challenges, co-develop this new type of exper-
5 imentation, share expertise and build software infrastructure to help scientists
6 around the world perform NaLoDuCo experiments.

7 **3.2.2 Focus areas**

8 The focus areas of the proposed project are:

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

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

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

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

16 **Data Analysis** Characterising behavioural and neural recordings.

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

19 **3.2.3 Cross fertilisation**

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

25 Currently, both GCNU and AIND are independently developing methods
26 to address the previous focus areas. We will join forces to co-develop these
27 areas and our foraging research programs, leveraging our combined expertise
28 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 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 (online figure, [2]). We have openly shared software for supporting data acquisition [3] and management [4] 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

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, which utilises Amazon S3 storage, for hosting the 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.

1 3.3.3 Data visualisation

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

7 3.3.4 Spike sorting

8 Spike sorting is specially challenging in NaLoDuCo experimentation since we
9 want to track individual neurons of freely moving mice for weeks to months. In
10 addition, we need online spike sorting, to allow experiments driven by real-time
11 machine learning inference, as described below. We will evaluate methods for
12 tracking neurons over long periods of time (e.g., [5]) and for online sorting (e.g.,
13 [6]). If needed, we will develop new methods, as we are experienced on the
14 subject.

15 3.3.5 Data analysis

16 The very large size of NaLoDuCo experimental data, the fact that the statis-
17 tics of these data change across time, and the requirement for real-time and
18 close-loop inference create new challenges to conventional machine learning data
19 analysis methods. We will evaluate how existing methods targeting the focus
20 areas described above cope with these challenges and, if necessary, create new
21 ones.

22 For behavioural data, we will investigate methods to:

- 23 • track multiple body parts of animals (e.g., [7] and a switching-linear-
24 dynamical method using RFIDs that we will develop),
- 25 • infer kinematics of foraging mice (e.g., [8,9]),
- 26 • segment behaviour into discrete states (e.g., [10] and a hierarchical HMM
27 that we will develop),
- 28 • infer the rules that govern mice behaviour from behavioural observations
29 only (i.e., policy inference) (e.g., [11]).

30 For neural data, we will investigate methods to:

- 31 • estimate low-dimensional continual representations of neural activity (i.e.,
32 latents inference) (e.g., [12]),
- 33 • segment neural activity into discrete states (e.g., [13]),
- 34 • decode environment variables from neural activity (e.g., [14]).

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. Saskia 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

1 3.5 Resources

2 Word limit: 200

3 Please provide the following:

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

8 The overall budget should be below the maximum £2 million combined fun-

9 der contribution

10 If there is more than one UK or US team associated with the application,

11 please combine their estimates together.

12 A detailed calculation and breakdown of resources is not required at this

13 stage, nor is a justification of costs.

14 The following is an example of how this might look.

15 UK Resources:

16 Total cost estimate: £600,000

17 Research council contribution: £480,000

18 0.2 FTE time, 1.0 FTE PDRA, 0.5 FTE technician

19 US Resources:

20 Total cost estimate: \$300,000

21 1.0 FTE PDRA or 1.0 FTE doctoral researcher

22 Total funder contribution estimate:

23 £716,475 (£480,000 + £236,475 (\$300,000 at exchange rate 0.79))

24 UK Resources:

25 Total cost estimate: £xxx,xxx

26 Research council contribution: £yyy,yyy (i.e., $0.8 * £xxx,xxx$)

27 # 3 PI at 0.1 FTE

28 # 1 experimental postdoc at 0.25 FTE

29 # 2 RSE at 1.0 FTE

30 3 x 0.1 FTE PI, 1 x 0.25 FTE PDRA, 2 x 1.0 FTE RSE

31 US Resources:

32 Total cost estimate: \$www,www

33 # 1 AIND scientist1 at 0.5 FTE

34 1 x 0.5 FTE scientist1

35 Total funder contribution estimate:

36 £716,475 (£yyy,yyy + £ppp,ppp (\$www,www at exchange rate 0.79))

37 References