# Optimizing sensorimotor behaviors through information integration and mental simulation

by

Chia-Jung Chang

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational Neuroscience

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021



Author .....
Department of Brain and Cognitive Sciences
Jan 22, 2021

Certified by .....
Matthew Wilson
Sherman Fairchild Professor in Neurobiology
Thesis Supervisor

Accepted by .....
Rebecca Saxe
Associate Head, Department of Brain and Cognitive Sciences

# Optimizing sensorimotor behaviors through information integration and mental simulation

Chia-Jung Chang

Submitted to the Department of Brain and Cognitive Sciences
on Jan 22, 2021, in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Computational Neuroscience

## Abstract

To generate dynamic thoughts, make proper decisions and act appropriately, humans and animals need to make reliable estimates of the state of the world. Recent studies have shown that the brain reduces uncertainty associated with noisy measurements by strategies including incorporating prior knowledge with sensory cues, extracting low-dimensional manifolds from heterogeneous activities, and updating internal models through simulating upcoming events. However, it remains unclear whether the brain utilizes additional sources that might have been ignored in previous work. To address this question, my thesis starts by asking how implicit temporal rhythms are used during mental simulation of object trajectory with partial observations. By designing psychophysics experiments with varying spatial and temporal structures, I show that humans simulate temporal rhythms in addition to the kinematics when interacting with dynamic stimuli. Bayesian modeling further suggests that explicit kinematics and implicit timing are integrated optimally. Following this work, the neural mechanism of time reproduction is revealed by analyzing the dynamics of low-dimensional state-space from large-scale electrophysiology recordings. This approach is further applied to uncover mechanisms underlying fear observational learning, which would not be possible with a traditional single-cell analysis. In the next chapter, the idea of neural coding in calcium imaging studies is challenged by demonstrating that the background residuals represent additional behavioral information. By building a convolutional neural network, position and speed of the animals can be directly decoded from raw microendoscopic data. Critically, saliency maps of the model reveal emergence of video-decomposition, and identify neural clusters representing distinct behavioral aspects on original images. Finally, inspired by replays in the hippocampus, I design a reinforcement learning agent with mental simulation to approximate the relaxation of constrained optimization. The results reveal scenarios where simulating to break physical barriers can improve learning efficiency. Together, my thesis examines how additional information may be integrated with spatial and temporal simulation to optimize complex sensorimotor behaviors, and proposes efficient models for decoding and learning.

Thesis Supervisor: Matthew Wilson
Title: Sherman Fairchild Professor in Neurobiology

# Acknowledgments

This dissertation would not have been possible without support from many people. First, I would like to thank collaborators for work listed in this thesis. Stephen Allsop shared his electrophysiology recordings for the fear observational learning experiment. Wei Guo not only shared his calcium imaging data for the spatial navigation experiment, but also gave valuable feedback for my project. Jie Zhang and Jon Newman provided technical support for obtaining imaging data and helped me set up the hardware equipment. Shao-Hua Sun generously shared his codes with Tensorflow infrastructure for training a deep learning model. Yen-Ling Kuo and Andrei Barbu have been providing guidance for building and testing reinforcement learning models.

Specifically, I would like to thank Mark Harnett for serving on my thesis committee over many years, as well as giving me constructive advice for both graduate studies and thesis projects. I am also grateful to my committee members Ila Fiete and Susan Epstein for offering support and advice for computational modeling.

Most importantly, I cannot express my gratitude enough to my thesis advisor Matt Wilson, a brilliant scientist who I have been learning from his broad knowledge and deep insights. He has been giving me full freedom to pursue my own project, as well as support to explore my career options. Switching into his lab turns out to be the best thing that happened in my PhD.

I would like to heartily thank Cody Sicilano, a genuine collaborator who taught me microendoscopic imaging, a great mentor offering career advice, and a true friend giving support during tough situations. I would like to thank Eghbal Hosseini, Jakob Voigts, Lukas Fischer, and all the members in Wilson lab for frequent scientific discussions. I have to thank Mehrdad Jazayeri, Seth Egger, and Kay Tye for making me realize the importance of keeping a written record to protect myself in a work environment. I learn to appreciate more about what could go wrong from these experiences. On the other hand, I am grateful to Rebecca Saxe and Nancy Kanwisher for their extra support in my PhD.

This dissertation is dedicated to my family and my friends, without whose love and encouragement, I would not make it through rock bottoms from switching labs.

# Contents

**List of Figures** 11

**List of Supplementary Figures** 13

**1 Introduction** 15

1.1 Uncertainty in behaviors and neural codes . . . . . 17

1.1.1 Inherent noise in the nervous system . . . . . 17

1.1.2 Scenarios with partial observations and sparse rewards . . . . 18

1.2 Denoising strategies for better representations . . . . . 21

1.2.1 Population coding and Bayesian inference . . . . . 21

1.2.2 Mental simulation and feedback correction . . . . . 23

1.3 Balance between exploration and exploitation . . . . . 23

**2 Simulating temporal rhythms improves object interception** 27

2.1 Introduction . . . . . 28

2.2 Materials and Methods . . . . . 30

2.2.1 Experimental setup . . . . . 30

2.2.2 Statistical analysis . . . . . 34

2.2.3 Bayesian modeling . . . . . 35

2.3 Results . . . . . 38

2.3.1 Experiment 1: TTC estimation benefits from explicit timing cues 39

2.3.2 Hypotheses distinguished by cue conflicts . . . . . 41

2.3.3 Experiment 2: TTC estimation benefits from implicit timing cues 43

2.3.4 Experiment 3: TTC estimation improves with temporal rhythms 45
2.3.5 Bayesian integration of speed and timing estimates . . . . . 46
2.4 Discussion . . . . . 62

**3 Integrating heterogeneity in neuronal population activity 67**
3.1 Introduction . . . . . 67
3.1.1 Time reproduction task . . . . . 68
3.1.2 Fear observational learning task . . . . . 69
3.2 Materials and Methods . . . . . 71
3.2.1 Time reproduction task . . . . . 71
3.2.2 Fear observational learning task . . . . . 76
3.2.3 Latent estimation from the point process . . . . . 81
3.2.4 Latent estimation from the cell populations . . . . . 84
3.3 Results . . . . . 87
3.3.1 Time reproduction task . . . . . 87
3.3.1.1 Monkey integrates sequential measurements with prior knowledge to improve time estimates . . . . . 87
3.3.1.2 Hypothesis for neural codes using an internal model . 89
3.3.1.3 Temporal scaling of heterogeneous single-cell profiles 90
3.3.1.4 Ensemble dynamics reveals that the DMFC regulates speed and distance of manifolds for flexible timing . 90
3.3.2 Fear observational learning task . . . . . 93
3.3.2.1 Mice can learn associations through observing others 93
3.3.2.2 Heterogeneity in single-cell responses across learning 95
3.3.2.3 State-space model reveals that the ACC has an overall earlier rate-change than the BLA across learning . . 96
3.3.2.4 Ensemble dynamics reveals that the ACC transmits social information to the BLA whereas the BLA forms the association between cue and shock . . . . . 97

3.3.2.5 Optogenetic manipulation shows that the ACC to BLA pathway is necessary for observational learning . . . 99
3.4 Discussion . . . 125

**4 Integration of signals from cells and neuropil improves decoding 127**
4.1 Introduction . . . 128
4.2 Materials and Methods . . . 131
4.2.1 Data collection . . . 131
4.2.2 Behavioral Tasks . . . 132
4.2.3 Dataset preparation . . . 133
4.2.4 Decoder Modeling . . . 136
4.2.5 Manifold analysis . . . 143
4.2.6 Network saliency map . . . 143
4.2.7 Statistical analysis . . . 144
4.3 Results . . . 144
4.3.1 Consistency of behavioral variables across time . . . 145
4.3.2 Accurate decoding from raw microendoscopic data . . . 146
4.3.3 Effect of data representations and model architectures . . . 147
4.3.4 How behavioral states contributed to decoding errors . . . 148
4.3.5 Background residuals encode behavioral information . . . 150
4.3.6 Decoding bias toward previous behavioral states . . . 151
4.3.7 Incorporating multiple frames improves decoding . . . 152
4.3.8 Distinct functional ensembles are identified from raw microendoscopic images . . . 153
4.3.9 Differential encoding of the maze topology across layers . . . 154
4.4 Discussion . . . 178

**5 Simulating constraint-removal improves sparse reward navigation 183**
5.1 Introduction . . . 183
5.2 Related Work . . . 185
5.3 Methods and Materials . . . 187

5.3.1 Navigation task . . . 187
5.3.2 Reinforcement learning model . . . 189
5.3.3 Additional replays for constraint removal . . . 193
5.4 Results . . . 193
5.4.1 Wide variability in learning to solve gridworld mazes with DQN models. . . . 193
5.4.2 Simulating constraint-removal aids cue generalization . . . 194
5.4.3 Effect of state-space representations (input features) . . . 195

**6 Discussion and Future Direction** **203**

**Bibliography** **205**

# List of Figures


2.1 General task design and space of hypotheses for estimating TTC . . . 50

2.2 Behavioral task conditions . . . . . 51

2.3 TTC estimation using speed and explicit timing cues . . . . . 52

2.4 TTC estimation using speed and implicit timing information . . . . . 53

2.5 TTC estimation in distance identity and temporal identity conditions 54

2.6 The Bayesian observer model with integration of speed and timing . . 55

3.1 Behavioral characterization for time reproduction . . . . . 100

3.2 Integration of sequential timing cues in Bayesian and Kalman models 101

3.3 Hypothesis on neural activities based on an internal model . . . . . 103

3.4 Temporal scaling of hetereogenous single-cell activities in the DMFC 104

3.5 Temporal scaling of ensemble trajectories in low-dimensional subspace 105

3.6 Speed and distance between ensemble trajectories across time . . . . 106

3.7 Behavioral characterization for fear observational learning . . . . . 108

3.8 Single-cell responses within a trial in the ACC and the BLA . . . . . 110

3.9 Hetereogenous cue-evoked responses in the ACC and the BLA . . . . 111

3.10 Earlier rate-change in the ACC across observational learning . . . . . 112

3.11 Observational learning induces distinct ensemble trajectories . . . . . 113

3.12 Photoinhibition of ACC→BLA impairs observational learning . . . . 115

4.1 Typical and proposed analysis pipelines for one-photon imaging data 156

4.2 Behavioral paradigm and data preparation . . . . . 157

4.3 Accurate decoding of positions and running speeds from raw microendoscopic images . . . 159
4.4 Neuropil/background residuals contain behavioral information . . . . 160
4.5 Bias toward decoding previous behavioral states . . . . . 162
4.6 Incorporating multiple frames into the model input improves decoding performance . . . . . 163
4.7 Emergence of auto-decomposition and functional clusters . . . . . 164
4.8 Encoding of behavioral topology in manifolds across model layers . . 165

5.1 Deterministic gridworld navigation (the maze-solving task) . . . . . 196
5.2 Navigation performance with an agent simulating constraint removal 197
5.3 Visual cue generalization from navigating perfect mazes . . . . . 198
5.4 Visual cue generalization from navigating dungeon mazes . . . . . 199
5.5 Effect of state-space representation in learning visual cues . . . . . 200

# List of Supplementary Figures


2.1 Analysis of integration strategy in the V-F-O condition using a cue conflict paradigm . . . . . 57
2.2 Performance in the V-O condition with different visible lengths . . . . 58
2.3 Performance comparison between the V-O and O-F-O conditions . . . 59
2.4 Linear regression analysis to quantify TTCp biases in Experiment 1 . 60
2.5 A Bayesian model of behavior in the V-O condition augmented by a subjective prior of time intervals . . . . . 61

3.1 Variance explained by different polynomial fits to single-cell responses 117
3.2 Inverse relationship between produced times and neural speeds . . . . 118
3.3 Speed and distance of ensemble trajectories aligned to Go . . . . . 119
3.4 EO mice learn association through observation . . . . . 120
3.5 Baseline over learning and test-day responses . . . . . 122
3.6 Examples on state-space changed across trials . . . . . 123
3.7 Ensemble trajectories on test day . . . . . 124

4.1 Examples of T-maze and □-maze behavior . . . . . 167
4.2 Consistency of output joint distributions between training and test sets 168
4.3 Examples of decoded running traces in the T-maze and the □-maze . 169
4.4 Examples of decoded traces in STOP periods and vathe training set . 170
4.5 Decoding performance with log transformation on speeds . . . . . 171
4.6 Network performance with separate decoding streams . . . . . 172
4.7 Model selection with hyperparameter searching . . . . . 173

4.8 Possible factors contributing to decoded errors . . . . . 174
4.9 Earlier optimal decoded states in cell somata . . . . . 175
4.10 Optimal decoding using a 1-second input and convolutions independent of time . . . . . 176
4.11 Encoding of behavioral topology during STOP periods in manifolds across model layers . . . . . 177

5.1 Training with a random walk model and effect of a smaller step-limit to end an episode navigating perfect mazes. . . . . 201

# Chapter 1

# Introduction

Humans and animals need the flexibility to dynamically integrate new information and update their estimates to make proper decisions under ambiguous and fast-changing environments.

Imagine intercepting a moving ball on a pool table as it bounces around hitting different edges. Does one only process kinematic information such as distance, position and speed or does one additionally pay attention to when the ball hits the edges? At first glance, the answer seems trivial because these variables are directly related in a physics equation. What if the lights were too dim to clearly see the ball? In that case, one may choose to pay attention to when the ball hits the edges to improve accuracy. This example brings up an unresolved question in sensorimotor processing: do humans engage additional timing mechanisms when simulating the trajectory of a moving object?

Now let's switch the gear. Imagine planning a route to reach your workplace from home. If two places are spaced far apart, planning directly at a low level by replicating every action from the past beset by the curse of dimensionality. Instead, one can simulate a rough path as if there were no buildings in between and identify some landmarks to construct more detailed paths hierarchically. If there is a construction site blocking the route you take daily, you will need to integrate pieces of routes that were never experienced together to develop a novel path. These highlight the importance of replaying past experiences and simulating future events in navigation.

Motivated by above examples, this dissertation aims to ask how our internal model of the world can be improved through mental simulation, and whether this process additionally utilizes information which has been ignored in the field.

The roadmap of dissertation is presented as follows. In Chapter 1, I will review processes involved in a variety of sensorimotor behaviors, their sources of uncertainty, as well as common strategies proposed in the field to minimize inaccuracy.

In Chapter 2, by using an object interception task as an example, I will investigate whether there is a parallel process estimating timing in addition to simulating object kinematics. By designing a series of psychophysics experiments, several hypotheses can be teased apart. Experimental and modeling results will be presented as evidence for how kinematics and timing are integrated through mental simulation.

In Chapter 3, by using a timing task and a fear observational learning task as examples, I will examine neural representations underlying internal transformation required for simulating either a stimulus or an association. By extracting low-dimensional state space from electrophysiology recordings, population coding will be compared with a traditional single-cell analysis.

After demonstrating how information can be extracted by spiking activities from a population of single cells, Chapter 4 will extend the coding question to calcium imaging studies. It investigates whether neuropil represents behavioral variables in addition to cell somata, by using a spatial navigation task. A novel methodology for decoding and visualizing receptive fields of the model is proposed, which will bring interpretable machine-learning framework for the neuroscience community.

Chapter 5 will describe a novel reinforcement learning (RL) agent inspired by how hippocampal replays facilitate spatial navigation. Preliminary results on how this model compares with a classical RL model in both spatial navigation and concept learning are shown. Finally, Chapter 6 will summarize the findings and discuss the implications of this dissertation in neuroscience and artificial intelligence. New directions for future studies will be included.

## 1.1 Uncertainty in behaviors and neural codes

### 1.1.1 Inherent noise in the nervous system

Most biological systems have inherent stochasticity from chemical reactions. Neural signals, originally initiated by a series of biochemical reactions including opening and closing of ion channels, have inherent randomness at the single cellular level. Noise further gets built-up in the network circuit, and eventually translates into uncertainty and variability at the behavioral level (Faisal et al., 2008; van Beers et al., 2002).

Take sensorimotor control as an example, noise arises from multiple stages involved in the behavior, from receiving external stimuli to generating motor actions. The representations of sensory stimuli are intrinsically noisy, because chemical molecules or photons arrive at the receptor randomly on top of a precision limitation in the receptors (Bialek, 1987; Bialek and Setayeshgar, 2005). When transforming these sensory signals into discrete spikes by the transduction, extra noise is added (Lillywhite and Laughlin, 1979; Torre et al., 1995). With multiple modalities, uncertainty can arise from different noise across sensory systems. For example, proprioceptive localization is less precise in azimuth (Van Beers et al., 1998) whereas visual localization is less precise in depth (Foley and Held, 1972). When localizing an object simultaneously with visual and proprioceptive systems, uncertainty arises as the brain integrates information.

As sensory information is passed from the periphery to the cortex as well as the subcortical regions, biochemical processes involved in generating action potentials in a single neuron introduce extra noise. For example, random opening and closing of ion channels can cause variability in the resting membrane potential, which further induces fluctuations in spiking activities (Derksen and Verveen, 1966; Schneidman et al., 1998; White et al., 2000).

When information is transmitted between neurons, additional noise is added to the neural code due to the stochastic release of neurotransmitters as well as the spontaneous fusion of synaptic vesicles (Allen and Stevens, 1994; Calvin and Stevens, 1967; Lou et al., 2005; Tsodyks and Markram, 1997; Wu et al., 2007). In a network

circuit where neurons are connected with each other, chaotic dynamics (Van Vreeswijk and Sompolinsky, 1996) can further increase noise to the neural code.

Before generating motor commands, an internal model simulating the process and predicting future states (Miall and Wolpert, 1996; Wolpert et al., 1995) can have noise or errors, and thus the representation of an internal estimate also has limited accuracy. The idea of an internal model is not unique to a simple sensorimotor transformation. Many other high-level cognitive behaviors such as spatial navigation also require building an internal model. Despite that a cognitive process does not necessarily create motor outputs, eventually we need movements to interact with the environment. Motor variability (Harris and Wolpert, 1998; Jones et al., 2002) is mainly contributed by the noise in motor neurons which innervate muscle fibres.

Noise accumulated in all stages of a behavior contributes to uncertainty in the behavior. There are some computational strategies such as population coding to handle noise in the neural code, which will be described in section 1.2.

### 1.1.2 Scenarios with partial observations and sparse rewards

Noisy sensory representations and variability in the nervous system can contribute to uncertainty in behaviors, as described in section 1.1.1. However, these are not the only sources that account for uncertainty in high-level cognitive behaviors (e.g. decision-making, rule learning, causal learning, and spatial navigation). In many situations, information about the environment can be extremely limited or corrupted, such that it is difficult to make a judgement given a large space of possible states.

For example, we can easily identify our friend in a clear photograph. However, if the person in the picture turns his head down, it might get difficult to recognize who he is without relying on other features such as accessories he wears. Other possible sources of occlusion include hands on the face, use of sunglasses and a face mask, and poor camera view, can all pose challenges. Consider a more extreme case in which the person stands behind a frosted glass window, it will be impossible to figure out who is in the picture. In this scenario, the relationship between familiar faces and their features are already established, such that we can recognize people robustly. With a

growing degree of occlusion or image corruption, we lose access to the full features and uncertainty arises, leading to incorrect answers in facial recognition (Ekenel and Stiefelhagen, 2009; Martínez, 2002). In this example, we are dealing with a limited set of states (i.e., paired association between one's identity and facial features) that were already learned, yet different states can generate the same observation under occlusion (i.e., people sharing similar eyes are hard to distinguish when they wear face masks). A computational model that can optimally predict the identity given incomplete features is thus required.

Previous example describes how perceptual aliasing can arise with a single sample. For a given static image, there will be no distinct observation that can bring extra information. However, in many real-world scenarios, multiple observations under different views or at different timepoints are required to build the model of the state of the world. Continuous accumulation of evidence across time will update the functions and variables of the model, leading to better estimates (de Lange et al., 2010; Devine et al., 2019; Kira et al., 2015; Polanía et al., 2014). For example, when we watch a baseball game on television. At a given time, we can only see a subset of the players but we still have a rough idea of where all the players are (e.g., the base runner is probably standing on top of the base), even if they are not in the field of view. If we see the base runner standing somewhere between bases in the next frames, we update our model to predict a potential stolen base. Similarly, if we see both a moving baseball and a running outfielder, we predict whether it is a fly out based on our estimates of both trajectories. In this example, we are dealing with a dynamic world where the next state depends on both the observed states and the transition matrix.

Another source of uncertainty is associated with the identity of the current state in a novel environment when a model of possible states is unavailable. For example, if a driver drops you off at a wrong address in an unfamiliar city, it will be difficult to plan a route back to your hotel due to the lack of knowledge of the current state and the transitions between states, instead of inability to plan routes. This is a common challenge when searching for a known goal from an unknown location, which requires inferring the current location as well as updating the belief of the world

through navigation (Stankiewicz et al., 2006; Yoshida and Ishii, 2006). Similarly, in social behaviors when we decide whether we can trust a stranger or cooperate with new colleagues, a constant estimation of potential risks is required while continually updating the model of the person based on each interaction (Krueger et al., 2007).

In previous examples, uncertainty is mainly caused by partial observations whereas states and transitions are deterministic. When states or transitions are probabilistic, uncertainty associated with distribution estimation is inevitable (Behrens et al., 2007; Daw et al., 2005, 2011; Gläscher et al., 2010; Huettel et al., 2005). One example arises when the paired association between stimulus and outcome varies across time. Consider a two-alternative forced choice task with two stimuli. If one rule involves 70% reward outcomes for stimulus A and 30% reward outcomes for the stimulus B and the other rule inverses reward probabilities, volatility is achieved when two rules frequently alternate. In this case, complete information can only be approximated through accumulating obserations, but will never be achieved in finite time.

Learning the state-action-reward contingencies from complex environments can be even more challenging with sparse feedback (Kulkarni et al., 2016; Riedmiller et al., 2018; Salimans and Chen, 2018; Vecerik et al., 2017). For example, when a two-stage task extends to a N-stage task, where a feedback signal only returns after going through N consecutive states, we will have difficulty to figure out all the transitions. This situation is very typical in our daily environment, as we usually only receive a reward at the target goal after a series of actions. Learning in this setting requires representing knowledge at multiple levels of abstractions as well as an efficient exploration of the environment.

To deal with uncertainty and challenges in behavioral settings with partial observations and sparse feedback, different computational strategies including a Bayesian model and variants of a deep reinforcement learning model (e.g. hierarchical reinforcement learning, reinforcement learning with a auxiliary task) are developed, and will be discussed more in section 1.2 and section 1.3.

## 1.2 Denoising strategies for better representations

To deal with noise in the nervous system described in section 1.1.1, several studies proposed the following computational principles that the brain use.

### 1.2.1 Population coding and Bayesian inference

One idea is to simply average over activities from multiple units (e.g. neurons, sensory receptors, muscle fibres) which carry redundant information corrupted by independent noise. At the synaptic level, fluctuating membrane potentials can be integrated through graded synapses (Van Steveninck and Laughlin, 1996), which reduces noise in each presynaptic neuron. Apart from integrating multiple diverse inputs, one-to-many synaptic divergence can also support noise removal by sending redundant information over multiple axons and then combining these signals at the target neuron (Del Punta et al., 2002; Faisal et al., 2008; Glowatzki and Fuchs, 2002).

At the cellular level, individual neurons often have overlapping receptive fields to represent the environment including visual features (Bair, 2005; Fischer, 1973), auditory features (Kilgard and Merzenich, 1999; Linden et al., 2003), place locations (Hafting et al., 2005; Kjelstrup et al., 2008), and stimuli in other sensory modalities. If information is represented at a single source only, signals cannot be reconstructed with fluctuating noise in the neural code. A population of neurons subserving similar functions form a distributed representation of information, which is more robust to noise. In addition, it can also help sharpen the tuning of behavioral variables. For example, individual neurons in the motor cortex have broad tuning curves to direction, but humans and animals can still control movement precisely through linearly summing signals from the population (Georgopoulos et al., 1986; Lee et al., 1988). Noise can also be reduced through integration over time. For example, a coincidence detector detects simultaneous input activities and filters out stochastic signals, and evidence can be accumulated within a temporal window (de Lange et al., 2010; Devine et al., 2019; Kira et al., 2015; Polanía et al., 2014), leading to a reduced variability in representations.

Under most cases, noise is not completely independent between neurons, and a population coding strategy that takes noise correlation and signal correlation (Averbeck et al., 2006; Panzeri et al., 1999; Romo et al., 2003) into account can further help with information encoding.

At the system level, a weighted averaging mechanism can also help reduce variability in neuronal signals. For example, Haruno and Wolpert (2005) finds that movement variability can be reduced by coactivating multiple muscles controlling the same command. Similarly in the sensory system, sensory receptors are coupled to average sensory noise in individual receptor (Kozlov et al., 2007).

Other than simply averaging input signals, numerous strategies using Bayesian inference are found, including incorporating prior knowledge into poor measurements to guide inference (Brainard et al., 2006; Körding and Wolpert, 2004; Stocker and Simoncelli, 2006; Wei and Stocker, 2012), integrating multiple sensory cues presented simultaneously (Alais and Burr, 2004; Ernst and Banks, 2002; Hillis et al., 2002, 2004; Van Beers et al., 1999), and integrating multiple measurements over a same stimulus repeatedly (Shadmehr and Holcomb, 1997). These studies provide evidence on how Bayesian inference can be utilized in both perception, cognition, and action, as well as implementation in population coding (Deneve et al., 2001; Ma et al., 2006).

Homeostatic plasticity can globally attenuate noise effects through varying membrane currents and synaptic strengths in a flexible manner (Desai et al., 1999; Turrigiano and Nelson, 2004; Turrigiano et al., 1998). Specific network architectures can create attractors or coupled synchronization to stabilize network dynamics regardless of stochasticity in individual neurons and network parameters (Holcman and Tsodyks, 2006; Hopfield, 1982; Prinz et al., 2004; Wills et al., 2005). Information can also be encoded and maintained in a low-dimensional state space to overcome variability in individual neurons. Different dimensionality reduction and manifold learning techniques further reveal that behavioral variables and sensory stimuli can be represented more robustly in low-dimensional space (Archer et al., 2014; Churchland et al., 2012; Cueva et al., 2020; Petreska et al., 2011; Yu et al., 2005, 2009).

### 1.2.2 Mental simulation and feedback correction

In addition to computational strategies described above, correcting estimates from feedback can also update the model of the world. In the sensorimotor literature, inspired by Kalman filtering in the control theory, it is assumed that humans and animals have a forward model simulating a motor process and predicting future sensory states (Miall and Wolpert, 1996; Wolpert et al., 1995). This internal belief will be adjusted given the prediction error between sensory feedback and expected signals (Kilner et al., 2007; Tseng et al., 2007; Wolpert and Kawato, 1998). These error signals are found across the brain, including sensory cortices (Christensen et al., 2007; Sommer and Wurtz, 2008) and dopaminergic neurons (Bayer and Glimcher, 2005).

Similar ideas are also proposed in high-level cognitive studies, formulated as "imagery" or "mental simulation". For example, activities in the hippocampus system are found to support replaying previous experiences, forming episodic memory, and path planning (Buckner, 2010; Hassabis et al., 2007; Schacter et al., 2007). Likewise, egocentric representations or social mirror responses are also found in the medial parietal cortex (Formisano et al., 2002; Sack et al., 2008).

These simulation behaviors have several advantages. For one, it might be easier to generalize across contexts by evoking similar associations from the memory. For another, conceptual knowledge might be formulated. It might explain how a cognitive map can be built using pieces of state transitions. Overall, mental simulation enables humans and animals to recapitulate previous experiences into novel contexts and form better presentations of the world, which might benefit an artificial agent to learn more efficiently.

## 1.3 Balance between exploration and exploitation

As described in section 1.1.2, it is challenging to learn the appropriate state transitions and action-outcome contingencies in a complex environment. Because it is difficult to sample all the possibilities when the world may change rapidly, the learned model might not be optimal, and thus we constantly need to choose between committing to

the learned policy and exploring unfamiliar alternatives in every real-world decision we make (Addicott et al., 2017; Cohen et al., 2007; Mehlhorn et al., 2015). For example, do I stay at my current company or do I go? Do I order my favorite dish or try a new one on the menu? One one hand, if we solely exploit learned knowledge, we might not be able to adapt to changes in the environment or yield the optimal selection due to uncertainty in acquired knowledge. On the other hand, if we purely explore new options or randomly select, we will not be able to learn the associations and may reduce payoffs over time.

This exploration-exploitation dilemma also appears in many learning tasks (e.g. multi-armed bandit problems, robotic navigation) in artificial intelligence (Audibert et al., 2009; Auer, 2002; Dezza et al., 2017; Rickert et al., 2008; Sutton et al., 1998). To frame this problem under a reinforcement learning framework, the agent needs to learn values associated with each state and select actions based on the learned policy to generate maximal reward (or minimal cost). On the contrary, by performing sub-optimal actions different from past experience, the agent may discover new features or sample more states-action-outcome contingencies.

In the literature, different methods are proposed to deal with the trade-off between exploration and exploitation in a reinforcement learning model. For example, R-max (Brafman and Tennenholtz, 2002), a model learning method, uses the optimism in the face of uncertainty heuristic similar to the exploration mechanism in prioritized sweeping method in Moore and Atkeson (1993) and the interval exploration method in Kaelbling et al. (1996). This approach implicitly resolves the exploration-exploitation dilemma by assuming that the obtained reward in any situation the agent is not too familiar with is its maximal possible reward. In this case, the agent's model is either optimal or leads to the real model through exploration.

Thrun (1992) proposed an efficient counter-based exploration and a selective attention mechanism to switch between exploration and exploitation. In this case, actions are selected based on a linear combination of exploration and exploitation:

$$\hat{a}_c = \alpha \cdot f(a) + \frac{c(s)}{E[c \mid s, a]}$$

where $c(s)$ counts the occurrence of the state $s$, $E[c \mid s, a]$ is the expected counter value given an action $a$ at state $s$, and $\alpha$ is a constant for weighting exploration versus exploitation.

While the study by Thrun (1992) uses a direct approach, Ishii et al. (2002) controls the balance between exploitation and exploration by combining an indirect approach using the inverse-temperature $\beta$ with a direct approach using the exploration bonus. When $\beta$ is small, the policy randomly selects an action. When $\beta$ is large, the action maximizing the action-value function $Q(s, a)$ is selected. The exploration bonus is proportional to the entropy of the posterior of the state-transition $H_D(s, a)$. A small $H_D(s, a)$ suggests little information gained by taking action $a$ at state $s$, which encourages exploration because the probability to select this action $a$ decreases.

The above early-developed methods are less common nowadays compared to two other widely used methods: $\epsilon$-greedy and softmax (Sutton et al., 1998). For $\epsilon$-greedy, the agent selects a random action $a$ with a fixed probability $\epsilon$ at each time step, and selects the optimal learned action with probability $1 - \epsilon$:

$$\pi(s) = \begin{cases} \text{random action from } A(s) & \text{if } \xi \leq \epsilon \\ \text{argmax}_{a \in A(s)} Q(s, a) & \text{otherwise} \end{cases}$$

where $\xi$ is a random number drawn from a uniform distribution at each time step.

On the other hand, a softmax policy does not assume a uniform distribution between actions nor selects in a greedy manner, but selects an action probabilistically using a Boltzmann distribution:

$$\pi(a \mid s) = \frac{e^{\frac{Q(s,a)}{\tau}}}{\sum_b e^{\frac{Q(s,b)}{\tau}}}$$

where $\tau$ is temperature, a positive parameter. When $\tau$ is high, all actions become equally probable. When $\tau$ is low, action is selected probabilistically based on its value estimate. When $\tau$ is close to 0, this becomes a greedy policy essentially.

Recent efforts are developed to further approve $\epsilon$-greedy method, such as decaying

$\epsilon$ over time (Caelen and Bontempi, 2007; Kumar et al., 2019; Leibo et al., 2017; Sajedian et al., 2019), adaptive $\epsilon$-greedy strategy based on value differences (VDBE) in which $\epsilon$ decays based on the learning progress (Tokic, 2010), contextual $\epsilon$-greedy strategy where $\epsilon$ varies depending on context (Bouneffouf et al., 2012).

In addition to these variants, entropy regularization (Mnih et al., 2016; Williams, 1992) is also a popular alternative to control the degree of exploration. Other strategies leverage the recent advance in deep learning. For example, NoiseNet (Fortunato et al., 2017) adds parametric noise to its weights to induce stochasticity in the policy for efficient exploration.

All these methods require efforts in tuning the appropriate balance parameter and may still face challenges in searching the optimal solution efficiently. In this dissertation, we propose a new strategy inspired from replays recorded in the hippocampus, which is discussed inChapter 5.

# Chapter 2

# Simulating temporal rhythms improves object interception

To coordinate movements with events in a dynamic environment, the brain has to anticipate when those events occur. A classic example is the estimation of time-to-contact ($TTC$), i.e., when an object reaches a target. It is thought that $TTC$ is estimated from kinematic variables. For example, a tennis player might use an estimate of distance ($d$) and speed ($v$) to estimate $TTC$ ($TTC = d/v$). However, the tennis player may instead estimate $TTC$ as twice the time it takes for the ball to move from the serve line to the net line. This latter strategy does not rely on kinematics and instead computes $TTC$ solely from temporal cues. Which of these two strategies do humans use to estimate $TTC$? Considering that both speed and time estimates are inherently uncertain and the ability of the human brain to combine different sources of information, we hypothesized that humans estimate $TTC$ by integrating speed information with temporal cues. We evaluated this hypothesis systematically using psychophysics and Bayesian modeling. Results indicated that humans rely on both speed information and temporal cues, and integrate them to optimize their $TTC$ estimates when both cues are present. These findings suggest that the brain's timing mechanisms are actively engaged when interacting with dynamic stimuli.

## 2.1 Introduction

Movements of our body and of objects around us create temporal events that demand our attention and command appropriate behavioral responses. For example, to catch a bouncing ball, one must determine the moment the ball reaches the hand. To capture a tennis shot on camera, one must anticipate the moment the ball reaches the racket. To catch an escaping prey, the predator has to determine the time of the final leap. To shoot a flying disc, one must estimate the moment to pull the trigger. Anticipating and reacting to such movement-related temporal events require an ability to estimate time-to-contact ($TTC$), i.e., the time when a moving entity reaches a target location.

How does the brain estimate $TTC$? Early studies hypothesized that humans rely on variables derived from an object's visual angle and its rate of expansion on the retina, of which the so-called tau is a classic example (Bootsma and Oudejans, 1993; Lee et al., 1983; Savelsbergh et al., 1991). Later, this proposal was deemed inadequate as it failed to capture many empirical observations (Hecht and Savelsbergh, 2004; López-Moliner et al., 2007; Smeets et al., 1996; Tresilian, 1999; Zago et al., 2009). Most current theories posit that $TTC$ estimation results from computations that rely on kinematic information (Brenner and Smeets, 2015; Brouwer et al., 2002; Kwon and Knill, 2013; Mrotek et al., 2004; Soechting et al., 2009). Specifically, it is assumed that the brain uses information about distance, speed, and acceleration to determine when an object reaches a designated target point. In this view, if we denote the distance by $d$ and assume that the object moved with constant speed, $v$, $TTC$ can be derived as $TTC = d/v$. This seems like a natural solution and matches our intuition of how to compute time from kinematic variables. However, the algorithms the brain uses for computing $TTC$ need not match what is taught in physics classrooms. Here, we asked whether humans solely rely on kinematics (e.g., speed and distance), or if they additionally rely on temporal cues.

We use an example to demonstrate the potential relevance of temporal cues as an independent source of information for estimating $TTC$. Imagine catching an ap-

proaching bouncing ball. At first glance, it may seem that the $TTC$ can be readily inferred from kinematic variables and equations without any need to explicitly estimate when the ball bounces. However, if estimates of speed and position are unreliable, for example when it is too dim to see the ball, one may infer $TTC$ from the temporal structure of the sounds that the ball makes upon bouncing off the ground. This example highlights a general and unresolved question in sensorimotor processing: when estimating $TTC$, do we rely solely on kinematic equations or do we additionally rely on the timing information that can be derived from positional information associated with moving objects (Figure 2.1)?

Several decades of research in human psychophysics suggest that humans estimate behaviorally relevant variables by integrating information from multiple modalities (Ernst and Banks, 2002; Hillis et al., 2002; Jacobs and Fine, 1999). For example, to estimate the size of an object, humans optimally combine visual and tactile information (Ernst and Banks, 2002), and to reach for an object, subjects combine uncertain spatial cues with prior expectation (Körding and Wolpert, 2004). With these considerations in mind, we hypothesized that subjects estimate $TTC$ by combining kinematic variables derived from visual information (e.g., speed) with estimates of elapsed time derived from the times when an object appears at different locations (e.g., the time it takes for an object to move from one point in space to another). However, testing this hypothesis is challenging because when an object moves between two points, the brain can either directly estimate speed from visual motion (Mikami et al., 1986; Rodman and Albright, 1987), or it may infer speed from measuring the time it takes for the object to move between various locations along the movement path.

To investigate the complementary role of speed and timing mechanisms, we designed a series of experiments in which we varied the temporal structure between visible and occluded segments of the path in order to systematically manipulate the reliability of the speed and the temporal information independently (Figure 2.2). Consistent with our hypothesis, we found that subjects integrated temporal information that is available during both visible and occluded segments of the path with speed information that is only available during the visible segment to improve their esti-

mates of $TTC$. To better understand the nature of the underlying computations, we compared subjects' behavior to that of an ideal Bayesian observer who optimally integrates speed and timing information. Similar to work in other sensorimotor domains (Ernst and Banks, 2002; Faisal et al., 2008; Gray and Regan, 1998; Hillis et al., 2002; Knill and Pouget, 2004; Körding and Wolpert, 2004; Kording et al., 2004; Stocker and Simoncelli, 2006; Tassinari et al., 2006), the model was able to accurately capture subjects' estimation strategy, indicating that humans efficiently integrate prior statistics with measurements of both speed and elapsed time. These results highlight a hitherto unappreciated function of the brain's capacity to utilize time – independent of speed – to inform sensorimotor function while interacting with dynamic stimuli.

## 2.2 Materials and Methods

Work described in this chapter is published on Chang and Jazayeri (2018).

### 2.2.1 Experimental setup

All experiments were approved by the Committee On the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology, and all subjects provided informed consent before participation. We used three experiments to examine how people infer time-to-contact ($TTC$). Seven adult subjects participated in Experiment 1. A different group of seven adult subjects participated in Experiment 2. Another group of eight adult subjects participated in Experiment 3. All subjects had normal or corrected-to-normal vision. In all experiments, we quantified behavior by comparing the statistics of experimentally specified actual time-to-contact ($TTC_a$) with the subjects' time-to-contact ($TTC_p$).

#### Procedure

Subjects sat in a dark, quiet room at a distance of approximately 50 cm from a display monitor with a refresh rate of 60 Hz and a resolution of 1920 by 1200 on an Apple Macintosh platform. We used MWorks (http://mworks-project.org/) to implement

task contingencies. All stimuli were presented on a black background. Although eye movements were not monitored, all trials began with a central fixation spot that subjects were asked to hold their gaze on the fixation point throughout every trial. Responses were made on a standard Apple Keyboard connected to the experimental machine.

We used three experiments to examine how people infer time-to-contact ($TTC$). Each experiment consisted of conditions whose order was randomized across subjects. Each condition was tested twice in two different days: the first session was used for training, and the second was used for the main test session but the first 25 trials were considered as warm-up and were discarded from the main analysis.

Subjects were asked to press a key when the bar reached the target position. Feedback was provided to indicate the actual bar position along the path when the key was pressed. The target position and the stimulus feedback were shown in green when the produced time-to-contact ($TTC_p$) was within an experimentally defined window around the actual time-to-contact ($TTC_a$), and red otherwise. To account for scalar variability, the window width was scaled with $TTC_a$ with a constant of proportionality, $k$. The value of $k$ was controlled by an adaptive one-up one-down procedure during training condition, and eventually reached a stable value, $k_0$. We then set $k_0$ as a feedback accuracy window in the main test session.

### Experiment 1

The objective of this experiment was to test whether subjects could improve their estimate of $TTC$ by integrating motion and timing cues. The trials were structured as follows: the subject pressed a key to initiate a trial. After a variable delay drawn randomly from a truncated exponential distribution (0.3-0.6 sec), a bar started moving horizontally from a starting point along a 16-degrees long linear path toward a target position at the end of the path (Figure 2.2). In each trial, the speed of the bar ($v$) was drawn from a discrete uniform distribution (Figure 2.2d).

The experiment consisted of three different conditions (Figure 2.2a-c). In the first condition, the bar traveling path consisted of two sections: a section where the

stimulus movement was visible and a section where it was occluded. The target was placed at the end of the occluded section. We denote this condition by *V-O* as shorthand for Visible-Occluded. In the second condition, the motion was invisible throughout the path but the position of the stimulus was flashed at the beginning of the path and when it reached the central fixation point in the middle of the path. We denote this condition by *O-F-O* as shorthand for Occluded-Flashed-Occluded. In the third condition, the stimulus movement was visible early on, and additionally its position was flashed when it reached the central fixation spot. Accordingly, we denote this condition by *V-F-O*. In conditions where the position was flashed, the flash lasted 100 msec. The distribution of sample interval ($t_1$) between the start of the path and the time when the bar reached the fixation point was the same across the conditions.

We also tested subjects in a cue conflict version of the *V-F-O* in which the flash at the central fixation was jittered relative to when the moving bar reached the central fixation (Supplementary Figure 2.1). These three jitter values were randomized and presented with equal probability.

### Experiment 2

The objective of this experiment was to test whether subjects could take advantage of temporal structure in the absence of an explicit temporal cue to improve their performance in the *V-O* condition. A bar began to move from a starting point along a path. The bar was initially visible and then disappeared behind an occluder. Subjects pressed a key when the bar reached the target position at the end of the occluder. We tested subjects in three conditions (Figure 2.4a). In all conditions, the distance between the fixation spot and target (d2) was set at 8 degrees while the visible length between the starting and the fixation spots (d1) was varied between 12, 8, and 5 degrees. We expressed these conditions in terms of the ratio of $d_2$ over d1, which we define as a gain factor ($G$). The corresponding $G$ for the three conditions were 0.667, 1, or 1.6. We recruited a new set of subjects for this experiment to make sure that participants were not made sensitive to timing cues due to prior experience

with the *O-F-O* conditions.

To evaluate the relative importance of speed and timing information, we also tested the newly recruited subjects with the same gain factors but in the *O-F-O* condition (Figure 2.4d). However, all *O-F-O* conditions were tested after the subjects had completed the *V-O* conditions to avoid inadvertently sensitizing subjects to timing cues. Overall, Experiment 2 consisted of 6 conditions in total.

**Experiment 3**

The objective of this experiment was to test whether the improved performance in Experiment 2 in relation to the identity context was related to the distance identity or temporal identity. To facilitate the description of these conditions, let us define the visible interval ($t_1$) as the interval associated with the visible portion of the path, between the starting point and when the bar reached the central fixation ($d_1$). Similarly, we define the occluded interval ($t_2$) as the interval associated with the occluded part of the path ($d_2$). Since the experiment involved changing the relative distances and/or durations, we additionally defined a distance ratio corresponding to the ratio of the occluded length to the visible length ($G_d = d_2/d_1$), and a duration ratio ($G_t = t_2/t_1$) corresponding to the ratio of the occluded interval to the visible interval.

The experiment consisted of two variants of the *V-O* condition. In the first condition, we set $d_1$ to 8 degrees and $d_2$ to 10 degrees ($G_d = 1.25$), and in the second condition both $d_1$ and $d_2$ were 10 degrees ($G_d = 1$). In the training sessions, similar to experiment 1 and 2, the stimulus speed was constant throughout the interception path. In the test sessions, unbeknownst to the subjects, immediately after the stimulus entered the occluded segment, its speed was multiplied by 1.25. This manipulation changed $G_t$ to 1 and 1/1.25 in the first and second conditions respectively. In other words, in the first condition, $G_d = 1.25$ and $G_t = 1$, whereas in the second condition $G_d = 1$ and $G_t = 1/1.25$. This manipulation allowed us to tease apart the effects of distance and temporal identity (Figure 2.5a).

### 2.2.2 Statistical analysis

We defined the $TTC_a$ as the interval between when the bar passed the central fixation to when it reached the target position. The $TTC_p$ was defined as the time from when the bar passed the central fixation to when the subject pressed a key. Subjects that were not sensitive to the range of sample intervals during the training session or had unstable performance were excluded from the study. We considered a subject insensitive if the corresponding $TTC_p$ distribution for the longest $TTC_a$ was not significantly different from $TTC_p$ distribution for the shortest $TTC_a$ (paired t-test at $p = 0.05$ level). Performance was considered unstable if the mean and the variance of $TTC_p$ were different between the first and second halves of the session (paired t-test at $p = 0.05$ level).

#### Summary statistics

We characterized each subject's performance by computing the following summary statistics for $TTC_p$:

$$\text{BIAS}^2 = \frac{1}{5} \times \sum_{i=1}^{5} \text{BIAS}_i^2 = \frac{1}{5} \times \sum_{i=1}^{5} (\overline{TTCp}_i - TTCa_i)^2$$

$$\text{VAR} = \frac{1}{5} \times \sum_{i=1}^{5} \text{VAR}_i = \frac{1}{5} \times \sum_{i=1}^{5} \left( \frac{1}{N_i - 1} \sum_{k=1}^{N_i} (TTCp_{ik} - \overline{TTCp}_i)^2 \right)$$

$$\text{RMSE} = \sqrt{\text{BIAS}^2 + \text{VAR}}$$

BIAS and VAR represent the average deviations and average variance over the five intervals included in the prior distribution. $\text{BIAS}_i$ and $\text{VAR}_i$ represent the mean deviation and variance of produced times ($TTC_p$) for the i-th actual interval ($TTC_a$) with $N_i$ trials. It follows naturally that the overall root mean squared error (RMSE) is equal to the square root of the sum of $\text{BIAS}^2$ and VAR. To estimate the mean and variance of summary statistics for individual subject in each condition, we resampled data with replacement and repeated this resampling 100 times.

#### Effect size

It is known that a relatively large sample size could lead to a smaller p value. Since there were more than 100 trials in each session for each subject, we also measured the strength of difference between conditions for each subject. We used Hedges' g (Hedges and Olkin, 2014) to quantify the distance between two distribution means. $g = 0.2$ suggests small effect size, 0.5 suggests medium effect size, and 0.8 suggests large effect size.

$$g = J(n_1 + n_2 - 2) \cdot \frac{(\overline{x}_1 - \overline{x}_2)}{s*}$$

where

$$J(a) = \frac{\Gamma\left(\frac{a}{2}\right)}{\sqrt{\frac{a}{2}} \cdot \Gamma\left(\frac{a-1}{2}\right)} \quad \text{and} \quad s* = \frac{\sqrt{(n_1 - 1) \cdot {s_1}^2 + (n_2 - 1) \cdot {s_2}^2}}{n_1 + n_2 - 2}$$

### 2.2.3 Bayesian modeling

#### Bayesian observer model

We developed a Bayesian observer model to describe performance in the *V-O* condition (Figure 2.6a) based on previous work on interval reproduction (Jazayeri and Shadlen, 2010). Instead of fitting the Bayesian model to each dataset, we asked whether we could fit the model to some conditions and then use parameters of the fitted model to predict behavior in other conditions. We modeled the prior distribution over sample intervals ($t_1$) based on the ratio of the visible distance ($d_1$) to the bar's speed ($v$). To simplify derivations, we modeled the discrete prior distributions used in the experiment as a continuous uniform distribution ranging from the shortest to longest sample interval tests. The shortest and longest intervals were computed in terms of the smallest and largest speeds ($v^{\min}$ and $v^{\max}$).

$$p(t_1) = \begin{cases} \dfrac{v^{\min} \cdot v^{\max}}{d_1 \cdot (v^{\max} - v^{\min})} & \dfrac{d_1}{v^{\max}} \leq t_1 \leq \dfrac{d_1}{v^{\min}} \\ 0 & \text{otherwise} \end{cases}$$

We assumed that subjects made two conditionally independent measurements when the bar was visible in the *V-O* condition, one associated with the speed of the bar, and another associated with duration of the visible period. Following previous work on sensory measurements of time and speed (Gibbon et al., 1984; McKee et al., 1986; Rakitin et al., 1998; Stocker and Simoncelli, 2006; Welch, 1989), we assumed that both measurements were perturbed by scalar Gaussian noise. Specifically, we assumed that the standard deviation of measured speed ($v_m$) scales with speed ($v$) with constant of proportionality $w_{mV}$, and that the standard deviation of measured duration ($t_m$) scales with elapsed time ($t_1$) with constant of proportionality $w_{mT}$. From the perspective of the observer who makes a measurement $t_m$ and $v_m$, the problem can be written in terms of the corresponding likelihood functions $\lambda(t_m \mid t_1)$ in Eq (2.1) and $\lambda(v_m \mid v)$ in Eq (2.2):

$$\lambda(t_m \mid t_1) = \frac{1}{\sqrt{2\pi \cdot (w_{mT} \cdot t_1)^2}} \times e^{\frac{-(t_m - \frac{t_1}{t_1})^2}{2 \cdot (w_{mT} \cdot t_1)^2}} \tag{2.1}$$

$$\lambda(v_m \mid v) = \frac{1}{\sqrt{2\pi \cdot (w_{mV} \cdot v)^2}} \times e^{\frac{-(v_m - v)^2}{2 \cdot (w_{mV} \cdot v)^2}} \tag{2.2}$$

To be able to combine the two likelihoods, we rewrote the likelihood associated with speed in terms of distance and sample interval, as follows:

$$\lambda(v_m \mid t_1) = \frac{1}{\sqrt{2\pi \cdot \left(w_{mV} \cdot \frac{d_1}{t_1}\right)^2}} \times e^{\frac{-(v_m - \frac{d_1}{t_1})^2}{2 \cdot \left(w_{mV} \cdot \frac{d_1}{t_1}\right)^2}} \tag{2.3}$$

This formulation implicitly assumes that subjects have a perfect estimate of distance, which is generally not true. However, in our experiment, the main independent variables were speed and time, and the occluded distance was fixed. We therefore reasoned that noise in the measurement of distances is negligible and can be subsumed in other sources of uncertainty without loss of generality.

Assume the two measurements are conditionally independent, the posterior is:

$$\pi(t_1 \mid t_m, v_m) = \frac{p(t_1) \cdot p(t_m \mid t_1) \cdot p(v_m \mid t_1)}{\int p(t_1) \cdot p(t_m \mid t_1) \cdot p(v_m \mid t_1)\ dt_1} \tag{2.4}$$

We assumed that subjects' minimized expected loss using a quadratic loss function, and modeled the inferred duration of the visible period based on the Bayes least-squares (BLS) estimator. We assumed that the estimate of the visible duration was multiplied by the gain ($G$) to derive an estimate of time-to-contact ($TTC_e$). We allowed $w_{mT}$ to subsume the effect of this added uncertainty.

$$\begin{aligned} TTC_e = G \cdot f_{\text{BLS}}(t_m, v_m) &= G \cdot \underset{t_e}{\text{argmax}} \left[ \int (t_e - t_1)^2 \cdot \pi(t_1 \mid t_m, v_m)\ dt_1 \right] \\ &= G \cdot \int t_1 \cdot \pi(t_1 \mid t_m, v_m)\ dt_1 \end{aligned} \tag{2.5}$$

The model was augmented by post-estimation noise to account for motor variability in $TTC_p$. Following previous work (Gallistel and Gibbon, 2000; Gibbon et al., 1984; Rakitin et al., 1998), we assumed that the standard deviation of motor noise was proportional to $TTC_e$, with constant of proportionality of $w_p$. We included an offset term ($b_0$) in the fitting procedure to account for idiosyncratic stimulus- and prior-independent biases observed in responses.

$$p(TTC_p \mid TTC_e) = \frac{1}{\sqrt{2\pi \cdot (w_p \cdot TTC_e)^2}} \times e^{\dfrac{-(TTC_p - (TTC_e + b_0))^2}{2 \cdot (w_p \cdot (TTC_e + b_0))^2}} \tag{2.6}$$

Using chain rule and marginalization of hidden variables, we wrote the conditional probability of $TTC_p$ for a given $TTC_a$ as a function of the model parameters as follows:

$$\begin{aligned} p(TTC_p \mid TTC_a;\ w_{mT}, w_{mV}, w_p, b_0) = \iint & p(TTC_p \mid G \cdot f_{\text{BLS}}(t_m, v_m);\ w_p, b_0) \cdot \\ & p(t_m \mid (TTC_a/G);\ w_{mT}) \cdot \\ & p(v_m \mid (TTC_a/G);\ w_{mV})\ dt_m\ dv_m \end{aligned}$$

### Fitting procedure

For fitting procedure, we assumed that $TTC_p$ values associated with any $TTC_a$ were independent across trials and thus expressed the joint conditional probability of individual $TTC_p$ values across all the N trials by the product of their individual conditional probabilities.

$$\begin{aligned}&\log\, p(TTC_p^1, TTC_p^2, ..., TTC_p^N \mid TTC_a;\ w_{mV}, w_{mT}, w_p, b_0)\\ &= \sum_{i=1}^{N} \log\, p(TTC_p^i \mid TTC_a;\ w_{mV}, w_{mT}, w_p, b_0)\end{aligned}$$

We used the "fminsearch" function in MATLAB (MathWorks, Inc.) to find the model parameters that maximized the likelihood of model parameters across all $TTC_a$ and $TTC_p$ values measured psychophysically. Integrals were approximated numerically using the global adaptive quadrature (Shampine, 2008). We repeated the search with different initial values 10 times, and verified that the likelihood functions were stable with respect to initial values.

## 2.3 Results

To facilitate the presentation of the results, we first describe the general task design that we employed for all the experiments (Figure 2.2). On each trial, subjects held their gaze on a central fixation point (FP) at the origin of the screen ($X_{FP} = 0$) throughout the trial and viewed one stimulus to the left of FP ($X_{Init}$) and another to the right of FP ($X_{Tar}$). After a random delay, a bar began to move horizontally from $X_{Init}$ toward $X_{Tar}$ with a fixed speed, $v$. In each trial, $v$ was sampled from a discrete uniform prior distribution with 5 values ranging between 8 and 16 deg/s (Figure 2.2d). Subjects had to press a button the moment the bar reached $X_{Tar}$.

Across experiments, we tested subjects in three conditions (Figure 2.2a-c). In the first condition, the bar was initially visible and then occluded. The visible segment extended from $X_{Init}$ to a transition point, denoted by $X_{tran}$. The subsequent occluded segment extended from $X_{tran}$ to $X_{Tar}$. The distance of the visible and occluded

segments were denoted by $d_1$ and $d_2$, respectively and added up to the full length of the movement path ($L$). We refer to this condition as *V-O* (Visible-Occluded; Figure 2.2a). In the second condition, the moving bar was occluded throughout the path but the position of the bar was flashed at $X_{Init}$ and at FP. We refer to this condition by *O-F-O* (Occluded-Flashed-Occluded; Figure 2.2b). In the third condition, the bar was initially visible and was additionally flashed when it reached FP. We refer to this condition by *V-F-O* (Visible-Flashed-Occluded; Figure 2.2c).

### 2.3.1 Experiment 1: TTC estimation benefits from explicit timing cues

In the first experiment, the path was 16 deg, and $X_{Init}$ and $X_{Tar}$ were located symmetrically at a distance of 8 deg to the left and right of FP. In this experiment, subjects were tested in all three conditions (i.e., *V-O*, *O-F-O*, and *V-F-O*). As described by the space of hypotheses in Figure 2.1, in conditions in which the moving bar is visible (*V-O* and *V-F-O*), subjects could adopt one of two strategies to perform the task. First, they could decide when to press the button by relying on an estimate of the bar's speed , (hat denotes subjective estimate) derived from visual motion. We refer to this as the speed strategy (Figure 2.1, "Speed strategy"). Alternatively subjects could rely on timing information to perform the task. For example in the *V-O* condition, they could derive an estimate of the duration of the visible segment (hat denotes subjective estimate) and scale it by the ratio of the distance between the occluded and visible segments ($d_2/d_1$). We refer to this as the timing strategy (Figure 2.1, "Timing strategy"). This timing strategy can also be used in the *V-F-O* condition, and is the only strategy that can be used in the *O-F-O* condition, in which no explicit visual cue about speed is present.

Our aim was to compare behavior in these conditions to assess whether subjects combine both strategies to compute $TTC$ (Figure 2.1, "Integration strategy"). To quantify behavior, we compared the time it took for the bar to go from FP to $X_{Tar}$, which we refer to as actual time-to-contact ($TTC_a = X_{Tar}/v$) to the subjects' pro-

duced time-to-contact ($TTC_p$). In the *V-F-O* and *O-F-O* conditions in which the bar was flashed at FP, we defined $TTC_p$ as the interval between the flash at FP and button press. In the *V-O* condition, quantification of $TTC_p$ was less straightforward because the bar was not flashed at FP. In this condition, we first measured the interval between $X_{tran}$ and button press, and then scaled that interval by the ratio of $X_{Tar}$ to the occluded distance ($X_{Tar}/d_2$). This method of estimating $TTC_p$ for the *V-O* condition assumes that the internal estimate of the speed after $X_{tran}$ remains relatively stable.

As evident from the $TTC_p$ pattern for a typical subject (Figure 2.3a), subjects were able to perform the task in all three conditions with different degrees of sensitivity. $TTC_p$ values were variable and exhibited a characteristic regression to the mean in which the average $TTC_p$ for each $TTC_a$ was systematically biased away from the identity line and toward the mean of the average $TTC_a$ (750 ms). As numerous previous experiments have demonstrated, this bias toward the mean is indicative of a Bayesian estimation strategy in which subjects reduce uncertainty associated with their sensory measurements (of speed and/or time) by using their knowledge about the prior statistics of $TTC_a$ (Ernst and Banks, 2002; Gray and Regan, 1998; Hillis et al., 2002; Jazayeri and Shadlen, 2015; Knill and Pouget, 2004; Körding and Wolpert, 2004; Kording et al., 2004; Stocker and Simoncelli, 2006; Tassinari et al., 2006; van Beers et al., 2002).

We quantified this regression using a BIAS term that quantifies the overall deviation from the identity line (see Materials and Methods). When measurements are accurate, responses would be on average unbiased (i.e., near the identity line), and the corresponding BIAS would be small. On the other hand, when measurements are highly noisy, we expect stronger regression to the means and larger BIAS values. In our dataset, there was significant BIAS in all conditions. The magnitude of BIAS was significantly smaller in the *V-F-O* condition compared to both the *V-O* condition ($t_{198} = 26.6435$, $p < 0.001$, hedges' g $= 3.7537$), and *O-F-O* condition ($t_{198} = 27.4602$, $p < 0.001$, hedges' g $= 3.8687$) (Figure 2.3b). This reduction in BIAS was observed for all the subjects and was significant across subjects (Wilcoxon one-side signed-

rank test, statistics = 28, $p < 0.01$) (Figure 2.3c), suggesting that humans combine information gleaned from the visual motion (speed and possibly timing) with the additional explicit timing cue provided by the flash at FP to reduce uncertainty.

Although BIAS provides an overall estimate of deviations from veridical $TTC_a$, it does not specify the direction of bias. In other words, both positive and negative biases would lead to an overall increase in BIAS. To ensure that the direction of bias in the data were consistent with a regression toward the mean (i.e., overestimation of small $TTC_a$ and underestimation of large $TTC_a$), we additionally quantified the relationship between $TTC_p$ and $TTC_a$ using linear regression. In all conditions, the slope of the regression was significantly smaller than unity indicating that the BIAS was consistent with the hypothesized regression to the mean (Supplementary Figure 2.4).

### 2.3.2 Hypotheses distinguished by cue conflicts

While this result is consistent with subjects integrating the two cues, it is also possible that the flashed stimulus at FP was not used as an explicit timing cue, and instead was used to simply reset the internal estimate of the position of the bar along the path. To test this possibility, we tested a subset of subjects in a cue conflict version of the *V-F-O* condition in which the flash at FP occurred either at the correct time (i.e., when the bar reached FP) or 100 msec too early or 100 msec too late (Supplementary Figure 2.1). This experimental manipulation enables us to distinguish between multiple hypotheses.

H1. Position reset hypothesis. According to this hypothesis, at the time of flash, subjects reset the position of the bar to the location of FP without changing their estimate of bar speed and without using the time of the flash as an additional cue. Since we quantified $TTC_p$ from the time of the flash to the button press, we should see no change in the relationship between $TTC_p$ and $TTC_a$. Note that this hypothesis was not explicitly considered in Figure 2.1.

H2. Speed-only hypothesis. According to this hypothesis, subjects ignore the flashed stimuli entirely and therefore, $TTC_p$ measured with respect to the time of flash would be shifted by the same amount as the jitter but in the opposite direction.

In other words, $TTC_p$ should increase by 100 msec when the flash was presented 100 msec early, and decrease by the same amount when the flash was presented 100 msec late. The original experiment already rejects this hypothesis since there were clear differences between subjects' performance in the *V-O* and *V-F-O* conditions, but cue-conflict experiments could validate the importance of the external timing cue. This hypothesis was referred to as the 'Speed strategy' in Figure 2.1.

H3. Timing-only hypothesis. According to this hypothesis, subjects only rely on the timing cues and ignore the opportunity to estimate speed from the visual segment of the path. If subjects were only using the time of the flash to estimate $TTC$, we would expect the average $TTC_p$ to be shifted exactly by the same duration as the jitter and in the same direction. Therefore, $TTC_p$ should decrease by 100 msec when the flash was presented 100 msec early, and increase by the same amount when the flash was presented 100 msec late. Again, the original experiment already rejects this hypothesis since there were clear differences between performance in the *V-F-O* and *V-O* conditions. We expected cue-conflict experiments to also reject this hypothesis. This hypothesis was referred to as the 'Timing strategy' in Figure 2.1.

H4. Cue integration hypothesis. According to this hypothesis, the jittered flash time would alter the timing-based evidence and would therefore cause a concomitant bias in $TTC_p$. Importantly however, this bias should be less than the size of the jitter (i.e. less than 100 msec) since temporal cues only serves as part of the information that guides subject's behavior (with the other part being the speed information gleaned from the visual segment of the path). This hypothesis was referred to as the 'Integration strategy' in Figure 2.1.

We found that $TTC_p$ changed significantly in the presence of jittered flashes (t-test, $p < 0.001$ for subjects JT, CN, BS and $p = 0.377$ for subject MD) which rejected H1 and H2, and that the overall shift in $TTC_p$ was significantly smaller than 100 msec (t-test, $p < 0.001$ for all subjects) rejecting H3 (Supplementary Figure 2.1). Together with the main results of Experiment 1, the observations indicate that subjects integrated both speed and explicit timing information to estimate $TTC$.

### 2.3.3 Experiment 2: TTC estimation benefits from implicit timing cues

The performance improvement in *V-F-O* compared to *V-O* and *O-F-O* demonstrated that humans benefited from an explicit timing cue provided by the flash at FP. This raises the intriguing possibility that humans utilize timing information even if it is not in the form of an explicit flashed position. For example, it may be the case that even in the *V-O* condition where there are no flashed stimuli, subjects determine when to press the button by combining two cues, one coming from speed information (e.g., $X_{Tar}/v$), and the other from scaling the duration of the visible segment ($t_1$) by the ratio of the occluded to visible segments ($d_2/d_1$). The former follows directly from kinematic equations (e.g., "if speed is doubled, it would take half as long"), and the latter derives from an ability to scale time intervals (e.g., "if distance is doubled, it should take twice as along").

To validate if such implicit temporal cue is used for estimating $TTC$, we designed a variant of the *V-O* condition in which we varied the ratio of the visible and occluded segments ($d_1$ and $d_2$, respectively) reasoning that when the visible and occluded intervals have the same duration, subjects would find the timing information more reliable and give it more weight for estimating $TTC$. However, since our objective was to assess the importance of implicit timing, we needed to make sure that varying the visible segments would not cause an appreciable change in the subjects' estimate of speed. Therefore, we designed an experiment to quantify how changes in the visible segment influences the accuracy of subjects' speed estimate (Supplementary Figure 2.2).

The experiment was a variant of the *V-O* condition in which we changed the length of the visible segment parametrically from 0.625 to 5 deg in log scale while keeping the occluded distance fixed at 8 deg ($X_{Tar} = 8$ deg). We found that subjects' performance improved significantly as the visible lengths increased from 0.625 to 1.25 deg (paired-sample t-test, $t_{399} = 56.61$, $p < 0.001$) and saturated afterwards (paired-sample t-test, $t_{399} = 0.9031$, $p = 0.3670$) (Supplementary Figure 2.2). In other words, the fidelity of the speed estimate saturated at a visible length of 1.25 deg.

We then tested subjects' behavior in the *V-O* condition in a separate experiment where we varied the ratio of the visible and occluded segments while keeping the visible length longer than the empirically observed saturation point of 1.25 deg. This ensured that any change in performance could not be attributed to an improvement or degradation of speed estimates, and must therefore reflect a capacity to measure and scale time intervals. We tested subjects' performance in three conditions. In all conditions, the occluded length was fixed ($d_2 = 8$ deg). Across conditions, the ratio of the occluded to visible length ($d_2/d_1$) was varied by a gain factor ($G = d_2/d_1$). The three gain factors were 0.667, 1, and 1.6.

Figure 2.4a and b show the performance of a typical subject in the three conditions. Notably, the best performance was not associated with $G = 0.667$ when the visible length was longest. Instead, RMSE was smallest when the visible and occluded lengths were equal ($G = 0.667$, $t_{198} = 20.3981$, $p < 0.001$, hedges' g = 2.9308; $G = 1.6$, $t_{198} = 22.9261$, $p < 0.001$, hedges' g = 3.2299), which we refer to as the temporal identity condition. The same was true across subjects (Figure 2.4c; Wilcoxon one-side signed-rank test, statistics = 28, $p < 0.01$) revealing a systematic and consistent improvement of performance in the identity condition (i.e., when $G = 1$).

The same subjects were also tested in the *O-F-O* condition, and for the same three gain values. As evident from the behavior of the same typical subject, RMSE was smaller when the two segments (i.e., before and after the flash) were the same (Figure 2.4d and e) compared to when the first segment before the flash was longer ($G = 0.667$, $t_{198} = 29.7316$, $p < 0.001$, hedges' g = 4.1887), or shorter ($G = 1.6$, $t_{198} = 25.6390$, $p < 0.001$, hedges' g = 3.6122). This effect was present across subjects (Fig. 4f; Wilcoxon one-side signed-rank test, statistics = 28, $p < 0.01$) indicating that this temporal identity helped subjects improve their estimate of $TTC$.

One potential concern in the case of $G = 1$ is that subjects may have detected that the two segments were temporally identical and switched to a purely timing strategy. To evaluate this possibility, we compared subjects' performance in the *V-O* and *O-F-O* conditions in the specific case when the two segments are identical (Supplementary Figure 2.3). If subjects were only relying on a timing strategy, we

would expect subjects' performance in these two conditions to be the same (since both have the timing information with $G = 1$). On the other hand, if the presence of $G = 1$ only serves to make the timing cue more reliable, we would expect performance to be better in the *V-O* condition since that condition provides the additional speed-dependent information.

We found that RMSE was consistently and significantly smaller in the *V-O* condition (Wilcoxon one-side signed-rank test, statistics = 3, $p < 0.001$) ruling out the hypothesis that $G = 1$ motivated subjects to abandon the speed information, and rely only on the identity temporal structure. These results suggest that subjects exploited the temporal structure in addition to speed cue to improve their performance.

One question that we did not address in this experiment is why $G = 1$ provides a more reliable timing cue. Indeed, reproducing a time interval may be more reliable than producing an interval that is scaled by an arbitrary gain factor, but a detailed quantification of this factor requires additional experiments. Nevertheless, our results are consistent with a recent study that demonstrated that mentally multiplying time intervals reduces the reliability of timing cues (DeLucia and Liddell, 1998).

### 2.3.4 Experiment 3: TTC estimation improves with temporal rhythms

Experiment 2 clearly demonstrated that $TTC$ estimation was most accurate when the visible and occluded segments of the path were identical. This is consistent with our hypothesis that performance benefited from temporal identity (i.e., $G = 1$). However, it is also possible that when the visible and occluded segments are the same length ($d_1 = d_2$), performance improves because subjects can estimate the distances more accurately. In other words, it may be that subjects benefited from distance identity (same lengths) and not temporal identity (same durations). This seems unlikely given that $d_2$ was fixed throughout all experiments. Nonetheless, we conducted an additional experiment to assess which of the two properties helped subjects in estimating $TTC$.

Since distance and duration are related through speed, the only way to dissociate the two is to make the speed of the bar differ between the visible and occluded parts of the path. Therefore, we designed a variant of the *V-O* condition in which, unbeknownst to the subjects, the speed of the bar behind the occluder was made 1.25 times faster than the speed in the visible portion (Figure 2.5a). The non-identical speed ratio enables us to create conditions in which the distance and temporal identity were dissociated. In one condition, the visible and occluded distances were the same but the corresponding durations were not. We refer to this condition as the distance identity condition ($G_d = 1$, $G_t = 0.8$). In another condition, we matched the ratio of the distances to the ratio of the speeds so that the corresponding durations were the same, which was referred to as the temporal identity condition ($G_t = 1$, $G_d = 1.25$).

New subjects were recruited for this experiment to ensure that sensitivity to temporal identity could not be attributed to exposure to previous experiments. Since subjects were not aware of the speed change behind the occluder, they could only adjust their performance based on feedback. If an observer solely relies on the distance identity, he would have better performance (lower RMSE) in the $G_d = 1$ condition. In contrast, an observer who relies on the temporal identity would have a lower RMSE in the $G_t = 1$, despite the fact that the distances between the visible and occluded parts are not the same.

We found that RMSE was lower for the temporal identity compared to distance identity condition as shown for a typical subject (Figure 2.5b and c; $t_{198} = 25.6431$, $p < 0.001$, hedges' g $= 3.6127$) and across subjects (Wilcoxon one-side signed-rank test, statistics $= 34$, $p < 0.05$). This finding further substantiates our conclusion that subjects benefit from temporal identity. We note that this experiment does not rule out a potential complementary role for distance identity, but it reveals the importance of temporal structure in estimating $TTC$.

### 2.3.5 Bayesian integration of speed and timing estimates

Experiments 1 to 3 established that subjects rely on both speed and timing strategies to estimate $TTC$. Another salient feature of subjects' behavior across all conditions

was that the regression to the mean of $TTC_p$ across the range of $TTC_a$. This was true for the external timing cue conditions in Experiment 1, for the implicit timing condition in Experiment 2, and in the control condition in Experiment 3. This observation suggests that, in addition to speed and timing information, subjects rely on their prior knowledge of the range of $TTC_a$ they encounter in the experiment. Following previous work (Ahrens and Sahani, 2011; Jazayeri and Shadlen, 2010), we hypothesized that the subjects' responses may follow the prediction of a Bayesian model that optimally integrates both the speed and timing measurements with the prior distribution (see Materials and Methods) (Figure 2.6a).

To test this hypothesis rigorously, we developed a Bayesian observer model to explain subjects' behavior in the *V-O* condition. The observer model combined two conditionally independent measurements from the visible segment of the path, one associated with the speed of the bar ($v_m$), and another associated with the duration of the first visible segment ($t_{1m}$). Following previous work, we assumed that these measurements were subject to scalar variability (Brenner and Smeets, 2009; Gibbon et al., 1984; McKee et al., 1986; Rakitin et al., 1998; Stocker and Simoncelli, 2006; Welch, 1989). In particular, we assumed that the standard deviation of noise on speed scaled with the bar's speed ($v$) with constant of proportionality ($w_{mV}$) and standard deviation of noise on elapsed time scaled with visible duration ($t_1$) with constant of proportionality ($w_{mT}$). The ideal observer first computes the posterior from the product of the prior, $p(t_1)$, the likelihood of the bar speed, $\lambda(v_m \mid v)$, and the likelihood of the visible duration, $\lambda(t_{1m} \mid t_1)$, and then uses the mean of the posterior as the optimal estimate of $TTC$ ($TTC_e$ in Figure 2.6).

In other words, $TTC_e$ corresponds to the Bayes-least-squares (BLS) estimate. To compare the model to subjects' behavior, we augmented the ideal observer with a production stage by adding scalar noise to the $TTC_e$ with constant of proportionality ($w_p$) to generate $TTC_p$.

We first estimated $w_{mV}$, $w_{mT}$ for each subject. In most Bayesian models, the model is evaluated by assessing the quality of model fits to the data. A more powerful approach is to fit the model to a training dataset and examine how well it explains

a test dataset. An even more powerful approach is to fit the model to one set of conditions and ask whether it predicts data in another condition to which it was not fitted. We employed the last approach. For each subject, we estimated $w_{mT}$ from the *O-F-O* condition with $G = 1$ (Figure 2.6b, *left*), and $w_{mV}$ from the *V-O* condition with $G = 0.667$ (Figure 2.6b, *right*), and used those estimates to predict subjects' behavior in the *V-O* condition with $G = 1$ (Figure 2.6c).

To estimate $w_{mT}$, we developed a Bayesian observer for the *O-F-O* condition with $G = 1$. In this case, the sensory information provided was the interval between when the bar flashed at initial start and when it reached FP (halfway along the path), which we denote by $t_1$. We fitted subjects' behavior by a BLS estimator that only relied on the likelihood of $t_1$, $\lambda(t_m \mid t_1)$ and the prior distribution, $p(t_1)$. As shown for one subject (Figure 2.6b, *left*) and consistent with previous work in a similar task (Acerbi et al., 2012; Cicchini et al., 2012; Jazayeri and Shadlen, 2015; Miyazaki et al., 2005), the model accurately captured behavior.

Next, we estimated $w_{mV}$ from fits of the Bayesian model to the *V-O* condition with $G = 0.667$. For this fitting procedure, we used the corresponding $w_{mT}$ derived from the *O-F-O* condition with $G = 0.667$ (see Materials and Methods). As shown for the same subject (Figure 2.6b, *right*), the model successfully accounted for the behavior. Recall that in the *V-O* condition, we had made the visible length long enough so that subjects' estimate of speed had saturated and was thus no longer dependent on $G$ (Supplementary Figure 2.2). This allowed us to safely use the fit to $w_{mV}$ derived from the $G = 0.667$ condition to predict behavior in the $G = 1$ condition.

Finally, we used each subject's fits to $w_{mV}$, $w_{mT}$ to predict the behavior in the *V-O* condition for $G = 1$. The model was able to predict the observed $TTC_p$ values as shown for one example subject (Figure 2.6c) and captured the data's summary statistics (BIAS and VAR) across subjects (Figure 2.6d). This is remarkable considering that both $w_{mV}$ and $w_{mT}$ were estimated from other task conditions, and provides strong support that subjects integrate prior information, speed information, and timing information to optimize their estimate of $TTC$.

A point of potential concern in our modeling work is that we modeled both the

prior distribution over the speed and the time intervals as uniform. This formulation is inaccurate. Given that the objective prior distribution of speed was uniform and that duration was inversely proportional to speed, the objective prior on intervals cannot be uniform. In our original models, we made this approximation because both distributions were discrete and because the exact formulation of the prior was not relevant to our main conclusion about the integration of the likelihood functions associated with speed and duration. However, to ensure that our results did not depend on the specific assumption of a uniform prior over time intervals, we constructed another model in which the prior more accurately reflected the distribution used in the experiment. For this model, we followed previous work (Janssen and Shadlen, 2005) and derived a 'subjective' prior of time intervals by blurring (i.e., convolving) the objective distribution with a normal distribution whose standard deviation was proportional to elapsed time. This alternative formulation did not change our main conclusion about the integration of speed and timing information, but provided an overall better fit to behavior suggesting that subjects relied on the empirically observed samples to form their prior over time intervals (Supplementary Figure 2.5).

# Figures

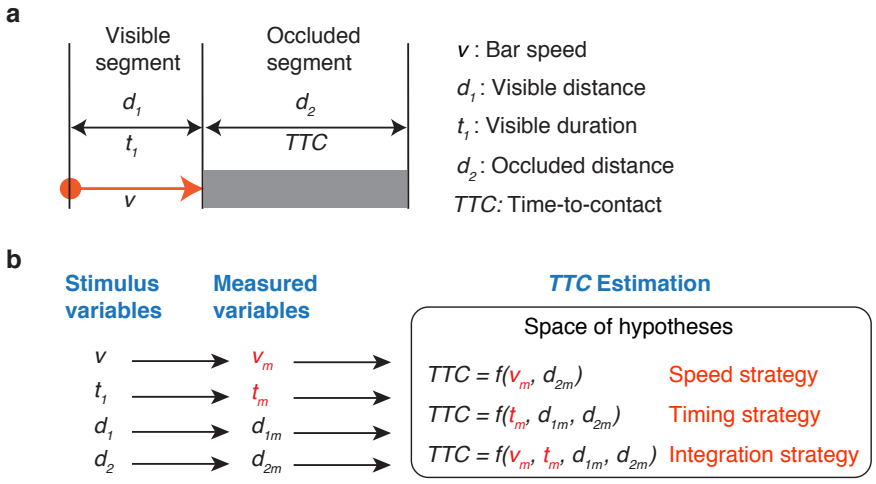



**Figure 2.1: General task design and space of hypotheses for estimating TTC.** (**a**) An illustration of the task design we used for investigating how subjects estimate time-to-contact ($TTC$). A bar moves along a path that is divided into two segments, a first segment of length $d_1$ where the bar is visible (orange arrow), and a second segment of length $d_2$ where the bar is occluded (gray rectangle). The bar moves at speed v, and the time it takes for it to reach the occluded segment is $t_1 = d_1/v$. (**b**) Three alternative strategies a subject can use to estimate the time it takes for the bar to get to the end of the occluded segment, which we denote as $TTC$. Left: The relevant stimulus variables for estimating $TTC$ are the distances of the two segments ($d_1$ and $d_2$), speed of the bar ($v$) and the visible duration ($t_1$). Middle: To estimate $TTC$, one has to rely on measured stimulus variables, which are denoted by subscript $m$ ($v_m, d_{1m}, d_{2m}$). Right: Three alternative strategies for estimating $TTC$. Speed strategy: Subjects estimate $TTC$ by combining information about the occluded distance and the measured speed; i.e., $f(v_m, d_{2m})$. Timing strategy: Subjects measure the visible duration and estimate $TTC$ by combining this timing cue with information about the distance of the two segments; i.e., $f(v_m, d_{1m}, d_{2m})$. Integration strategy: Subjects could combine both speed and timing cues to compute a more accurate estimate of $TTC$; i.e., $f(v_m, t_m, d_{1m}, d_{2m})$. The key variables that distinguish between strategies ($v_m, t_m$) are shown in red.

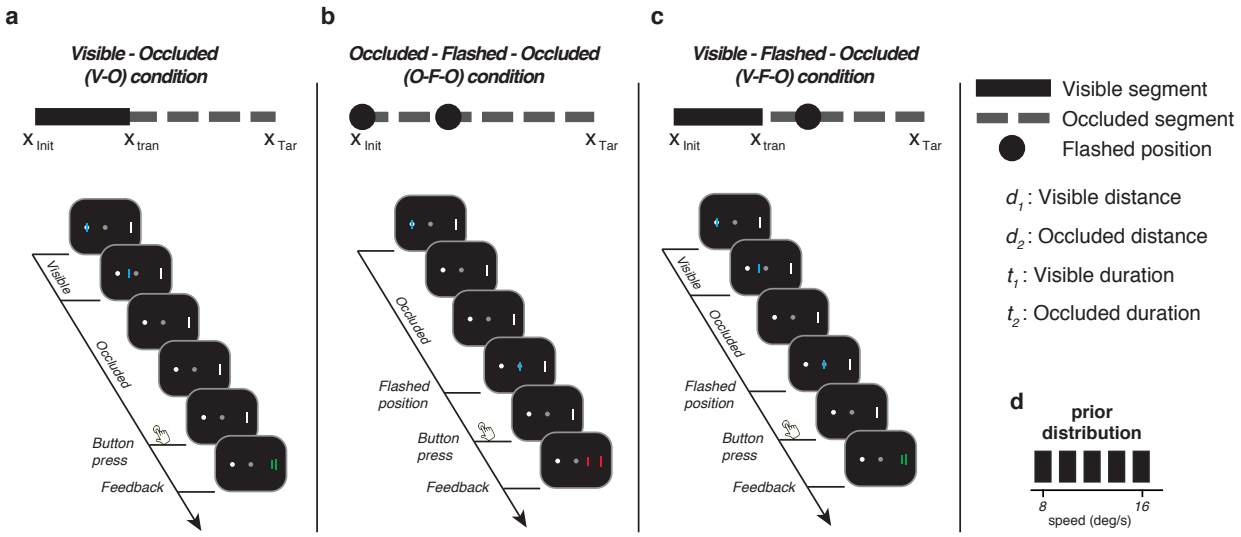



**Figure 2.2: Behavioral task conditions.** (**a**) Visible-Occluded (*V-O*) condition. Top: Schematic illustration of the *V-O* condition in which a bar moves along a path that is divided to two segments, a visible segments extending from an initial point ($X_{Init}$) to a transition point ($X_{tran}$), and an occluded segment (i.e., invisible) from $X_{tran}$ to a target point ($X_{Tar}$). Bottom: Trial structure for the *V-O* condition. Subjects were asked to fixate at the central fixation point (gray circle). Afterwards, a bar (blue) moved from $X_{Init}$ to the left of the fixation point (white circle) to $X_{Tar}$ to the right of the fixation point (white vertical line). The bar was visible initially and occluded afterwards (Top panel). Subjects had to press a key when they judged the moving bar to have arrived at $X_{Tar}$. When responses were sufficiently accurate, the moving bar and the target bar both turned green (shown). Otherwise, they both turned red. Details see Materials and Methods. (**b**) *O-F-O* (Occluded-Flashed-Occluded) condition. Top: Schematic illustration of the *O-F-O* condition in which the bar was only flashed at $X_{Init}$ and later at a position along the path (black circle). Bottom: Trial structure for the *O-F-O* condition in the format same as panel a. The example trial shows a case where the flashed position coincided with the fixed point, corresponding to the design in Experiment 1. In this example, the feedback is shown as red, indicating a hypothetical trial where the response was too early. (**c**) *V-F-O* (Visible-Flashed-Occluded) condition. This condition includes both an initial visible segment (from $X_{Init}$ to $X_{tran}$) and a flashed position some time after the visible segment (black circle). Bottom: Trial structure for the *V-F-O* condition in the format same as panel a. (**d**) Prior distribution of the bar speed ($v$). Speed was fixed in each trial but was sampled from a discrete uniform distribution ranging between 8 to 16 degrees per second across trials.

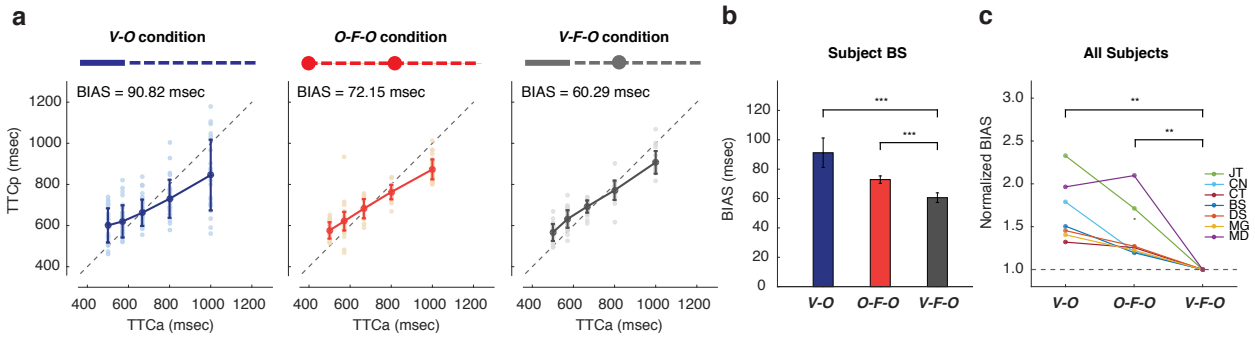



**Figure 2.3: TTC estimation using speed and explicit timing cues.** (**a**) Behavior of a typical subject for different conditions in Experiment 1. The left panel corresponds to the *V-O* condition (blue) where the bar was initially visible. The middle panel corresponds to the *O-F-O* condition (red) in which the bar was flashed at the starting point and the central fixation point. The right panel corresponds to the *V-F-O* condition (gray) with an initial visible segment and a flash at the fixation point. Performance was quantified by comparing produced time-to-contact ($TTC_p$) to actual time-to-contact ($TTC_a$). $TTC_a$ was defined as the time between when the bar reached the central fixation to when it reached the target. $TTC_p$ was defined as the time between when the bar reached the central fixation to when the button was pressed. Light dots and dark circles show $TTC_p$ in each trial and the corresponding averages for each $TTC_a$, respectively. The inset in each panel reports the overall BIAS in $TTC_p$. BIAS was quantified as the average error over the five distinct $TTC_a$, i.e., the root mean square of differences between five solid dark circles and the corresponding diagonal dash line on the plot (see Materials and Methods). (**b**) BIAS comparison across task conditions for a typical subject. We estimated the standard error through resampling data with 100 repetitions. BIAS was smaller for the *O-F-O* compared to *V-O*, and smallest in the *V-F-O* condition. See main text for statistics (***: $p < 0.001$). (**c**) Normalized BIAS across conditions for all subjects (N = 7) shown in different colors. Normalized BIAS was obtained by dividing BIAS in all conditions with BIAS in the *V-F-O* condition. Across subjects, BIAS patterns were similar to the typical subject in **b**. See main text for statistics (**: $p < 0.01$).

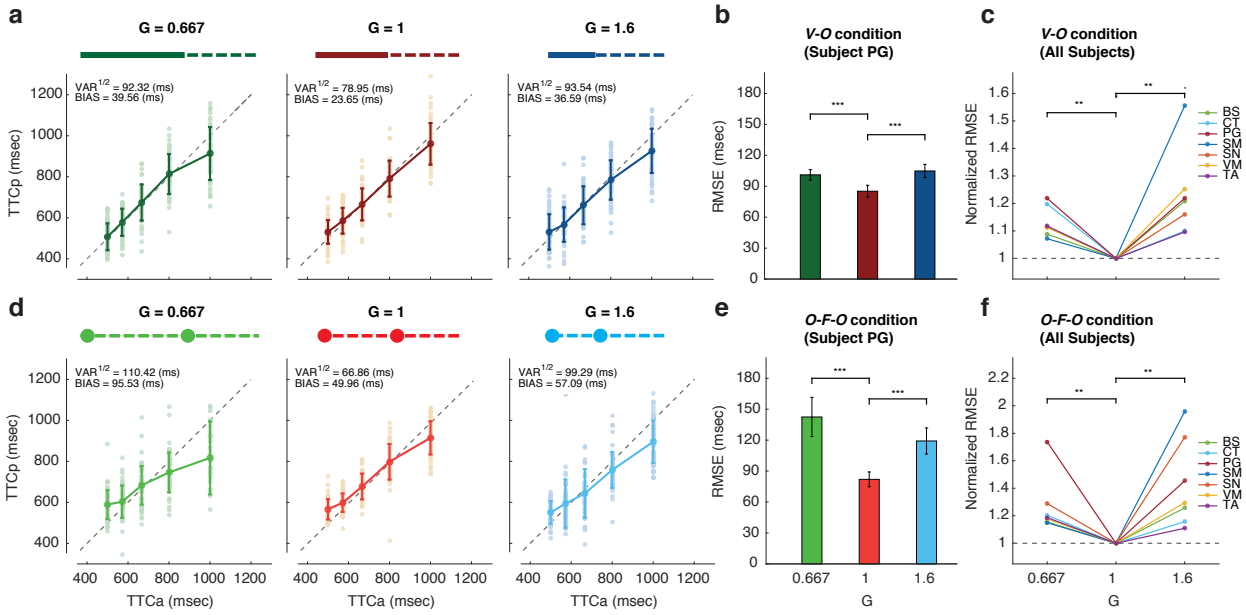



**Figure 2.4: TTC estimation using speed and implicit timing information.** (**a-c**) Three variants of the *V-O* condition with three different visible lengths and the same occluded length (Experiment 2). Each variant was identified by a gain factor (G) that quantified the ratio of the occluded to visible length. Subjects were tested for three $G$ values, $G = 0.667$ (green), $G = 1$ (red), and $G = 1.6$ (blue). Since the bar moved at a constant speed throughout each trial, the gain also reflected the ratio of the occluded to visible duration. (**a**) Produced time-to-contact ($TTC_p$) as a function of actual time-to-contact ($TTC_a$) of a typical subject. Light dots and dark circles showed $TTC_p$ in each trial and the corresponding averages for each $TTC_a$. BIAS was defined as described in Figure 3. VAR is the average variance of $TTC_p$ across all values of $TTC_a$ (see Materials and Methods). (**b**) Comparison of performance across gains in terms of root-mean-squared-error (RMSE) for the same subject in a. We estimated the standard error of RMSE through resampling data with 100 repetitions. RMSE was significantly smaller in the $G = 1$ variant of the *V-O* condition. (**c**) Normalized RMSE as a function of $G$ for the *V-O* condition across all subjects (N = 7). RMSE for each gain was divided by RMSE in the $G = 1$. Across subjects, RMSE were smallest for $G = 1$. (**d-f**) Behavioral analysis for the same three $G$ values in the *O-F-O* condition. (**d**) Behavior of the same subject in the *O-F-O* condition across gains. (**e**) Comparison across gains for the same subject using RMSE. (**f**) Normalized RMSE as a function of $G$ across all subjects (N = 7). Across all subjects for both *V-O* and *O-F-O* conditions, RMSE was significant smaller when $G = 1$. See main text for statistics (***: $p < 0.001$, **: $p < 0.01$).

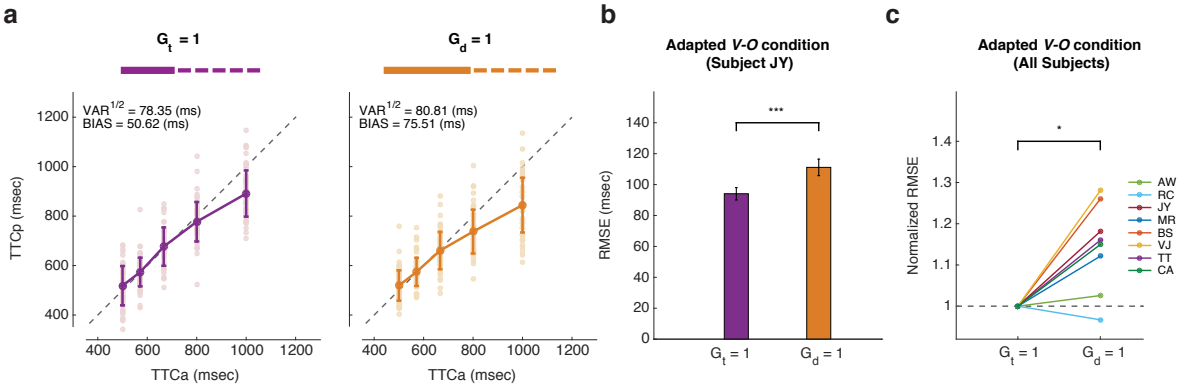



**Figure 2.5: TTC estimation in the distance identity and temporal identity conditions.** (**a**) Behavior of a typical subject for two variants of the *V-O* condition (Experiment 3), the temporal identity condition ($G_t = 1$; i.e., same visible and occluded durations) and the distance identity condition ($G_d = 1$; i.e., same visible and occluded distances). In both variants, unbeknownst to the subject, the speed in the occluded segment was multiplied by 1.25 (25% faster than the visible section). $G_t = 1$ (purple): The durations of movement in the visible and occluded sections were the same. Because of the speed difference between the two sections, the visible distance was shorter than the occluded distance (i.e., $G_d = 0.8$). $G_d = 1$ (orange): The visible distance was same as the occluded distance, but the corresponding durations were different (i.e., $G_t = 1.25$). (**b**) Comparison between these two conditions for a typical subject. We estimated the standard error of RMSE through resampling data with 100 repetitions. See main text for statistics (***: $p < 0.001$). (**c**) Normalized RMSE across all subjects (N = 8). Different colored lines represented different subjects. See main text for statistics (*: $p < 0.05$)

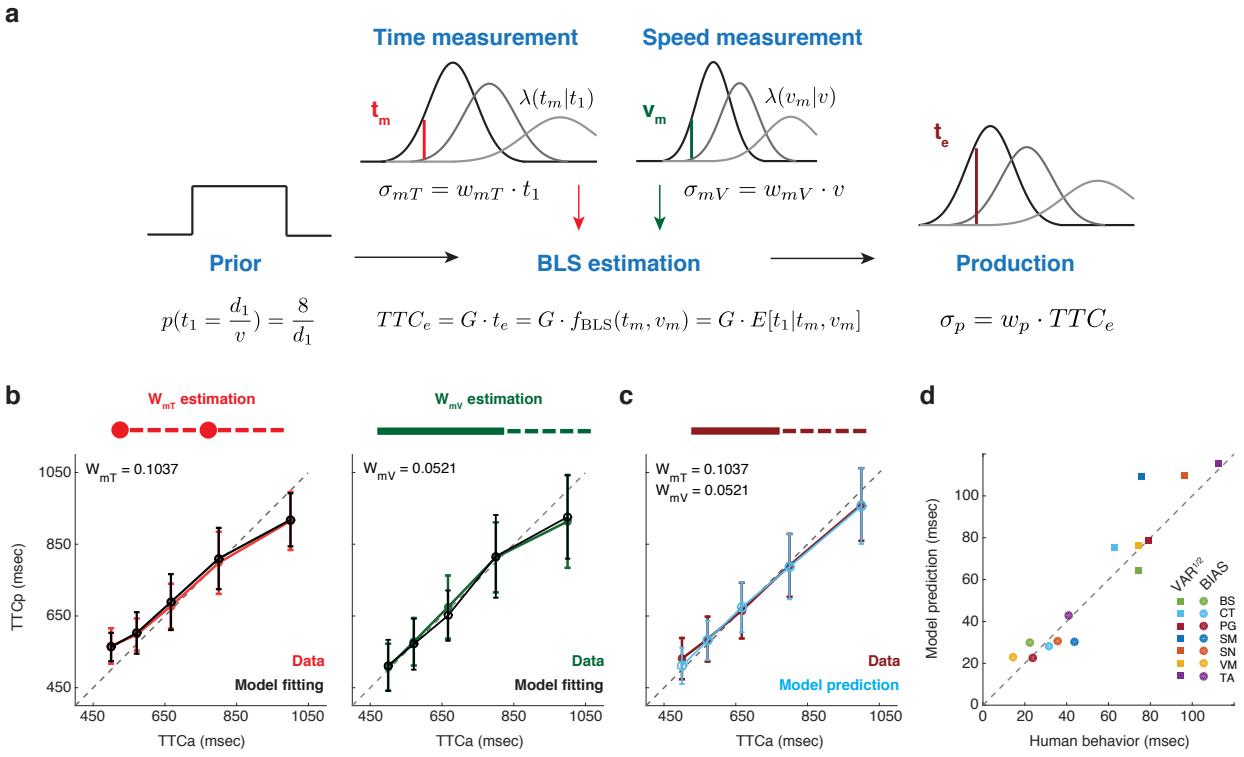
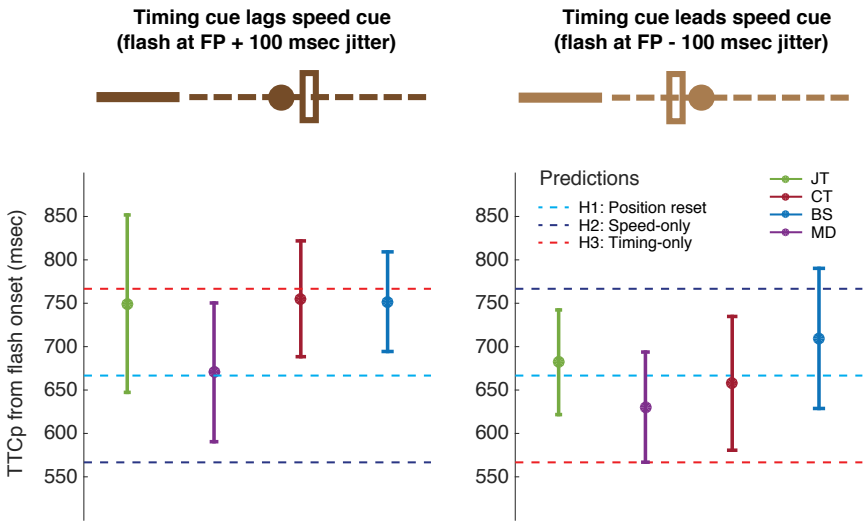


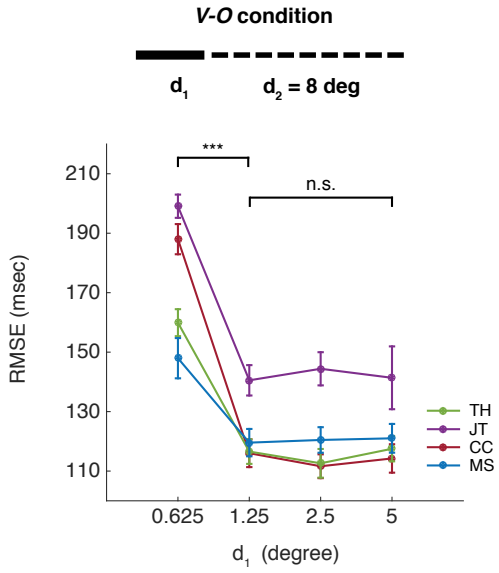

**Figure 2.6: The Bayesian observer model with integration of speed and implicit timing cues for the *V-O* condition.** (**a**) The Bayesian observer model for the *V-O* condition. On each trial, the speed ($v$) was drawn from a uniform prior distribution. We used the relationship between the distance of the visible section ($d_1$) and speed to express the prior in terms of the duration the bar is visible ($p(t_1)$; *left*) and assumed this to be uniform as well. We assumed that the observer makes two conditionally independent measurements of $v$ and $t_1$, which we denoted by $v_m$ (red vertical line) and $t_m$ (green vertical line), respectively. We assumed that $v_m$ and $t_m$ are perturbed by zero-mean Gaussian noise with standard deviations ($\sigma_{mV}$ and $\sigma_{mT}$) proportional to $v$ and $t_1$ (Gaussian curves; *top*) with constant of proportionality of $w_{mV}$ and $w_{mT}$, respectively. The Bayesian observer computes the posterior from the likelihood functions, $\lambda(v_m \mid v)$ and $\lambda(t_m \mid t_1)$, and the prior, and uses a Bayes-Least-Squares (BLS) estimator, $f_{\mathrm{BLS}}$, to infer the movement duration in the visible section, which we denoted by $t_e$ (brown vertical line). This estimate is then multiplied by the gain ($G$) to obtain an optimal estimated $TTC$ ($TTC_e$). Finally, the model incorporates motor variability via additional noise in the production stage. We modeled this noise as a sample from a zero-mean Gaussian with standard deviation scaling with $TTC_e$ with scaling factor $w_p$.

**Figure 2.6** ***(previous page)*:** (**b**) The left panel ($w_{mT}$ estimation) shows the behavior of a Bayesian observer model (black) fitted to the data (red) for a typical subject in the *O-F-O* condition with $G = 1$. Since the movement of the bar in the *O-F-O* condition was not visible, we estimated $w_{mT}$ from a Bayesian model that relies on the prior and $t_m$, but not $v_m$. The right panel ($w_{mV}$ estimation) shows the Bayesian model (black) fitted to its corresponding data (green) for the *V-O* condition with $G = 0.667$. In the *V-O* condition, the observer had access to both speed and timing cues. Therefore, we estimated $w_{mV}$ from a Bayesian model that uses the prior, $t_m$ and $v_m$ with $w_{mT}$ inferred from the *O-F-O* condition with $G = 0.667$. (**c**) Behavior (black) and model prediction (cyan) for a typical subject in the *V-O* condition with $G = 1$. The prediction was made based on a Bayesian model whose $w_{mT}$ and $w_{mV}$ were estimated from other experimental conditions (the left two panels in **b**). (**d**) Comparison of summary statistics (BIAS and $\text{VAR}^{1/2}$) between human behavior (abscissa) and predictions from a Bayesian model (ordinate) across subjects (N = 7). Summary statistics of the model were computed based on averages of 100 simulations of the Bayesian observer model. Different colors correspond to different subjects.

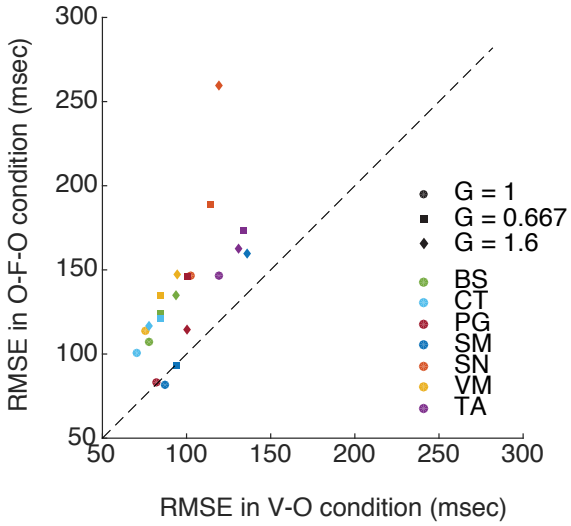
# Supplementary Figures



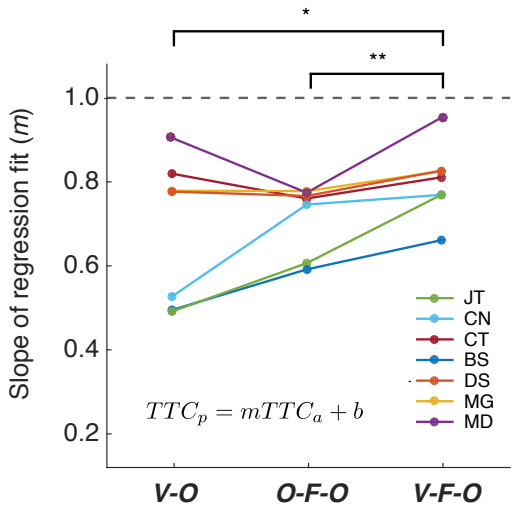

**Supplementary Figure 2.1: Analysis of integration strategy in the *V-F-O* condition using a cue conflict paradigm.** In Experiment 1 in the main text, we demonstrated that subjects' integrated speed information with explicit timing cues. Here, we used a cue conflict paradigm to rule out three alternative hypotheses about why an explicit timing cue might have influenced behavior: H1: subjects used the flash to reset the position of the bar without updating speed information ('Position reset'; cyan dashed line), H2: subjects ignored the timing cue ('Speed-only'; blue dashed line), and H3: subjects solely relied on the explicit timing cue ('Timing-only'; red dashed line). In the cue conflict paradigm, the flash at the fixation point was presented either 100 msec after the bar reached that position ('Timing cue lags speed cue'; *left*), or 100 msec before ('Timing cue leads speed cue'; *right*). Different colors represent different subjects (N = 4). Subjects' responses were variable and did not afford a definitive rejection of all three hypotheses for every individual subject. However, across subjects, responses were significantly different from all aforementioned alternative hypotheses (t-test, $p < 0.001$). We note that the observed variability is unsurprising given that subjects are expected to use different weights for the speed and timing cues in the integration process.



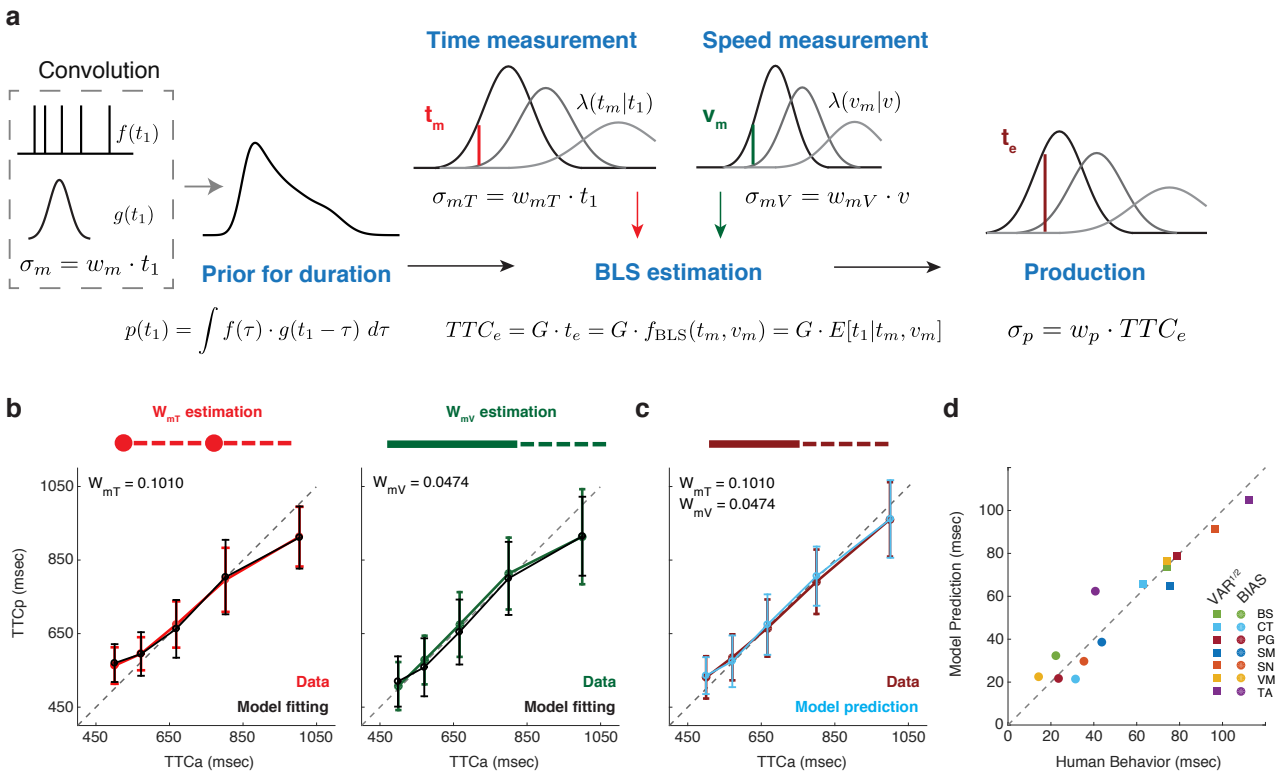

**Supplementary Figure 2.2: Performance in the *V-O* condition with different visible lengths.** RMSE was compared between different visible lengths for the *V-O* condition. The occluded distance was fixed at 8 deg and the distribution of bar speed was same as in Experiment 2. Different visible lengths ranging from 0.625 to 5 degrees (log-scale) were tested in randomized order across subjects (N = 4). For the speeds tested in our experiment, performance seems to saturate when the visible length reaches approximately 1.25 deg.




**Supplementary Figure 2.3: Performance comparison between the *V-O* and *O-F-O* conditions.** RMSE was compared between the *V-O* and *O-F-O* conditions for all subjects across all gains in Experiment 2. For nearly all subjects and gain values, RMSE was relatively smaller in the *V-O* condition where both speed and implicit timing information were available. Different colors represent different subjects (N = 7), and different symbols correspond to different gain values (circle for $G = 1$, square for $G = 0.667$, and diamond for $G = 1.6$).




**Supplementary Figure 2.4: Linear regression analysis to quantify TTCp biases in Experiment 1.** We used a linear regression model ($TTC_p = mTTC_a + b$) to assess the relationship between $TTC_p$ and $TTC_a$. We found that the slope of the regression line ($m$) was smaller than unity for all conditions (*V-O*, *O-F-O*, and *V-F-O*) and subjects (N = 7). This validated the inference we made in the main text from measuring the overall bias (BIAS, see Materials and Methods) that $TTC_p$ was biased toward the mean of the distribution of $TTC_a$. We also found that the slopes were higher in the *V-F-O* condition compared to *O-F-O* condition (Wilcoxon rank sum test, $p < 0.01$) as well as compared to the *V-O* condition (Wilcoxon rank sum test, $p < 0.05$), which is consistent with Bayesian integration in the presence of additional cues.




**Supplementary Figure 2.5: A Bayesian model of behavior in the *V-O* condition augmented by a subjective prior of time intervals.** (**a**) Left: Procedure for building a subjective prior of time intervals. In our experiments, durations were inversely proportional to speed. Since the speeds we used were governed by a discrete uniform prior distribution, the corresponding durations formed a non-uniform discrete prior distribution ($f(t_1)$). To construct a subjective prior distribution, we convolved $f(t_1)$ with a normal distribution whose standard deviation was proportional to the mean with a constant of proportionality of 0.1 ($g(t_1)$), which is close to the Weber fractions observed in our data. (**b-d**) Using the augmented Bayesian model to capture subjects' behavior. Results are shown with the same organization and format as in **Figure 2.6** in the main text. For all subjects except subject TA, the augmented Bayesian model provided better fits to performance compared to the model with a uniform prior in sample time.

## 2.4 Discussion

Current models assume that estimation of $TTC$ between the body and an object or between two objects depends on measurements of kinematic variables such as speed, distance or depth (Bosco et al., 2012; Brenner and Smeets, 2015; Brouwer et al., 2002; Dubrowski et al., 2000; Fooken et al., 2016; Gray and Regan, 1998; Hecht et al., 1996; Kwon and Knill, 2013; Lacquaniti and Maioli, 1989; Lee et al., 2014; Mrotek and Soechting, 2007; Mrotek et al., 2004; Soechting et al., 2009; Zago et al., 2009). Our work reveals that humans additionally exploit timing information gleaned from the temporal structure of events in the environment as an alternative source of information to estimate $TTC$. Moreover, we show that humans automatically combine this timing strategy with speeds to derive more accurate estimates of $TTC$.

We demonstrated the role of timing in two complementary sets of experiments. In the first set of experiments, we presented subjects with a task in which estimation of $TTC$ could benefit from either timing or speed information. Results indicated that when explicit timing cues were available, subjects integrated timing information with their measurements of speed to derive more accurate estimates of when a moving bar would reach a target position. This result extends a large body of evidence showing that humans fuse information from multiple modalities to improve their performance (Ernst and Banks, 2002; Hillis et al., 2002; Jacobs and Fine, 1999).

In the second set of experiments, we removed the explicit timing cue and instead asked whether subjects could exploit implicit timing cues in the environment. In our experiment, we varied the ratio between the intervals when the bar was visible and occluded. Based on recent work (Remington and Jazayeri, 2017), we hypothesized that when the visible and occluded epochs have the same duration, subjects would automatically make use of this temporal identity and rely more on the timing information to estimate $TTC$. Results validated that subjects relied on the temporal identity structure to improve their performance. Notably, performance in the temporal identity context was even better than when the occluded length was the same and the visible length was made longer. In other words, prolonging the visible portion,

which could only improve subjects' estimate of speed, was harmful to performance when it broke the temporal structure conferred by the identity context. This result powerfully demonstrated that the key factor driving the performance improvement was the presence of temporal identity. This conclusion was reinforced by control experiments showing that the result was due to temporal – not distance – identity. Finally, it is also important to note that the role of timing strategy in our experiment cannot be attributed to apparent motion because the distances and time intervals we used in our experiment were well outside the range that typically induce an apparent motion percept (de la Malla and López-Moliner, 2015).

Our work also intersects with the body of work revealing subjects' ability to integrate sensory information with prior expectations (Ernst and Banks, 2002; Gray and Regan, 1998; Hillis et al., 2002; Knill and Pouget, 2004; Körding and Wolpert, 2004; Kording et al., 2004; Stocker and Simoncelli, 2006; Tassinari et al., 2006). This integration is often characterized in the context of Bayesian models that formalize how prior knowledge and sensory cues must be integrated to optimize performance. We found that a Bayesian model that optimally integrates the prior distribution of $TTC$ with evidence derived from both speed and temporal cues accurately captured subjects' behavior. This result suggests that the human brain is optimized to combine speed and time information for object interception. Note that the integration of speed and time information is distinct from the indirect role that time would play by improving one's estimate of speed (Cavallo and Laurent, 1988; Krukowski et al., 2003; Mason and Carnahan, 1999). As we demonstrated in the control experiment (Supplementary Figure 2.2), the improvement of speed estimate saturates rapidly as viewing time increases and cannot account for our finding in Experiment 2. In other words, our results reveal that humans actively integrate elapsed time with speed information to estimate $TTC$.

These experiments lead to a simple and novel conclusion that humans actively engage timing mechanisms during estimation of $TTC$. To put this finding in context, it is important to distinguish between the role of time during the visible and occluded regions of the path. When an object moves behind an occluder, subjects could no

longer measure the object's speed and thus have no choice but to rely on their sense of time. This idea was formalized by Tresillian and others in relation to human's ability to extrapolate an object's location behind an occluder (DeLucia and Liddell, 1998; Tresilian, 1995). This is fundamentally different from what we propose; our findings indicate that humans actively integrate information about temporal contexts and events even when the object is visible. In other words, timing seems to be an integral component of how we interact with dynamic stimuli, both to better estimate where they are when they are visible, and to infer where they might be when they are occluded.

One important implication of our work is for studies of object interception. Real world object interception involves a decision to initiate a movement followed by ongoing adjustments based on sensorimotor feedback. Although successful interception requires a tight coordination between the initiation and the subsequent adjustments, the two processes typically involve different computations (Resulaj et al., 2009). The decision of when to initiate is, by and large, determined by a prediction of how long it would take to reach the object, which is directly related to our work on $TTC$ estimation. While our work does not address any potential role of timing for the sensorimotor coordination after movement initiation, it does invite a revision of the computational models that specify how the brain computes the movement initiation time. In particular, it suggests that the cognitive and/or motor planning stage of interception behavior (Brenner and Smeets, 2010; Pellizzer and Hedges, 2003; Scheidt et al., 2001) may be particularly sensitive to preceding temporal events in the environment as recent physiology experiments suggest (Jazayeri and Shadlen, 2015). It is also consistent with numerous imaging and electrophysiological studies that find an important role for premotor and supplementary motor areas in timing (Cui et al., 2009; Harrington et al., 2010; Kunimatsu and Tanaka, 2012; Merchant et al., 2013; Mita et al., 2009). In contract, temporal cues may not play an active role during the adjustments that follow movement initiation when the brain has access to movement related state-dependent information (Conditt and Mussa-Ivaldi, 1999; de la Malla and López-Moliner, 2015; Diedrichsen et al., 2007).

It is worthwhile considering why the role of time was not noted in prior research on object interception. We think that answer has to do with the simplicity of behavioral tasks used in laboratory settings (but see some studies using more naturalistic paradigms or done with virtual reality (Diaz et al., 2013; Fooken et al., 2016)). Many previous experiments lacked a rich spatiotemporal context that could reveal the relevance of temporal structure. However, real world examples of object interception take place in the presence of temporal statistics, spatial landmarks, and temporal events such as collisions and/or reflections, all of which make knowledge about time highly informative. A notable observation in our experiment was that subjects' estimate of $TTC$ was more accurate in the temporal identity context, suggesting that non-identity transformations are associated with higher sensorimotor noise.

We speculate that the improvement of performance we found in the temporal identity context may be an instance of a more general principle related to the temporal structures for which the human brain has a strong internal prior. If so, we would expect stronger effects of timing information in the presence of sounds that form rhythms or for integer ratios for which strong internal priors have been reported (Jacoby and McDermott, 2017). A real world example of this conjecture applies to intercepting a bouncing ball. According to our results, we predict that subjects benefit from the bounce sound, especially when visual information is uncertain (e.g., dribbling a basketball without looking at the ball). These considerations highlight the need for future research to move beyond simple behavioral tasks and examine object interception and $TTC$ estimation in more naturalistic settings where the underlying dynamics are governed by richer spatiotemporal contexts. Exploration of behavior in more naturalistic settings may further substantiate the importance of temporal events and contexts in processing dynamic stimuli.

# Chapter 3

# Integrating heterogeneity in neuronal population activity

## 3.1 Introduction

Single-cell analysis has brought insights to how sensory information is represented in the neuroscience field. For example, different groups of neurons in the primary visual cortex (V1) are tuned to different orientations (Hubel and Wiesel, 1959, 1968) and spatial frequencies (Tootell et al., 1988). Similarly, neurons in the primary auditory cortex (A1) are often categorized based on their temporal profiles to a sound stimulus (Feng and Wang, 2017; Wang, 2007). Brain regions are often characterized based on these single-cell profiles (e.g., excitatory vs. inhibitory, monotonic vs. non-monotonic, phasic vs. tonic). However, complex and hetereogenous profiles are not well-captured in this type of analysis. As discussed in Chapter 1, noise in the nervous system introduces uncertainty in information processing. How the brain integrates information across the population holds the key question. For example, neural decoding is found to be more accurate using populations than single neurons (Georgopoulos et al., 1986; Lee et al., 1988). Here, by using a time reproduction task and a fear observational learning task as examples, we compared the low-dimensional ensemble dynamics and temporal profiles of single cells, and demonstrated how population coding helps us understand mechanisms underlying mental simulation.

### 3.1.1 Time reproduction task

Humans and animals can generate motor plans in response to sensory signals from the environment. However, due to noise in the nervous system as well as uncertainty associated with the environment, either the internal estimate could have large errors or the generated actions deviate from the desired target. For example, if you tried to serve a shuttlecock in an outdoor badminton court, you would take wind into consideration to adjust your direction and strength to avoid a service long (before the wind) or service short (against the wind). Consequently, the ability to dynamically correct the internal estimates based on sensory feedback is critical in daily life.

Inspired by the control theory, numerous theoretical and psychophysics studies have proposed that humans and animals use an inverse model to compute input for motor system to reach the desired state, and a forward model simulating a motor process and predicting future states (Ito, 2008; Miall and Wolpert, 1996; Scott, 2004; Shadmehr and Krakauer, 2008; Todorov, 2004; Wolpert et al., 1995). These can be viewed as the building blocks of fundamental operations in the brain including state estimation, sensory cancellation, learning, adaptation, and motor control (Angelaki et al., 2004; Golub et al., 2015; Kennedy et al., 2014; Sabes, 2000; Sommer and Wurtz, 2002; Stavisky et al., 2017).

In the literature, a common behavioral paradigm (e.g., arm reaching) often includes a visual or haptic input that is constantly available throughout the motor generation. Therefore, it has been challenging to dissociate the internal estimate from sensory information. Some tasks (e.g. eye saccade) comprise a transient action with sensory feedback provided precisely at time when action completes, which can disentangle the timing of internal estimate and motor initiation from sensory correction, but these events are discrete on a trial structure whereas many behaviors require adaptive representation of dynamics (Shadmehr and Mussa-Ivaldi, 1994).

To address these challenges in studying the neural mechanism underlying mental simulation and state correction, a behavioral task in which monkeys need to reproduce a time interval that is presented twice in a row was designed. In this task, time to

initiate a movement has to be adjusted based on uncertain and discrete sensory cues. This paradigm not only segregates the inverse model and the forward model in time, but also enables us to tackle how cortical dynamics are regulated upon multiple feedback signals prior to movement initiation.

Here we trained a macaque monkey (*Monkey G*) to perform this time reproduction task, and collected in vivo electrophysiology recordings of single-unit activity from the frontal cortex. Here we proposed to estimate ensemble dynamics through a state-space model, and characterized both single-cell and ensemble dynamics from the initial cue presentation to eye saccade. It is found that individual neurons demonstrate heterogeneous within-trial activities, yet across-trial responses seem to be scaled by target intervals. Integrating diverse response profiles from all the neurons, similar temporal scaling is also evident in trial-averaged trajectories in a low-dimensional state-space (i.e., manifold). By analyzing the geometry of these manifolds, we showed that the brain uses the speed of cortical dynamics to adjust the time estimation, and updates the speed based on the error between the time of sensory feedback (i.e., flash) and the anticipated timing derived from a simulated motor plan.

### 3.1.2 Fear observational learning task

Humans and animals can acquire the contingencies between cues and their associated negative outcome. However, learning negative associations through direct experience can be life-threatening. For example, if you saw a friend get chased by a neighborhood dog, you would learn to stay away from the dog without having to undergo that experience yourself. Consequently, the ability to extract predictive information about potential threats through observing the experiences of others is critical to evolution.

Observational learning can be used to acquire the knowledge of the world through mental simulation (Baeyens et al., 1996; Burke et al., 2010; Lindström et al., 2018; Olsson and Phelps, 2004). It depends on detection and integration of social signals to achieve adaptive behaviors in fast-changing environments (Chen et al., 2009; Dulac and Torello, 2003; Isogai et al., 2011; Twining et al., 2017). This basic ability to perceive and mimic the behavioral responses of others has been demonstrated across

species, including birds (Adriaense et al., 2019), non-human primates (Ferrucci et al., 2019; Mineka et al., 1984; Tomasello et al., 1987), and rodents (Atsak et al., 2011; Chen et al., 2009; Jeon et al., 2010; Kim et al., 2012; Knapska et al., 2006, 2010; Pereira et al., 2012; Twining et al., 2017).

Among different observational learning behaviors, fear observational learning is the acquisition of an association between a conditioned stimulus and a punishment that may have each been directly experienced, but never temporally paired in a contingent manner, except through observation of another animal.

In the literature, both the anterior cingulate cortex (ACC) and the basolateral amygdala (BLA) have been considered to account for fear observational learning. For example, Olsson and Phelps (2004) found that both the ACC and the BLA were recruited when human subjects acquired fear of unseen faces through observation. Likewise, Jeon et al. (2010) showed an increased theta-frequency synchronization between both regions during observational fear conditioning, which is interrupted by pharmacological inhibition. However, it remains unknown how each brain region represents the observational learning process. It is also unclear if both brain regions work cooperatively to acquire implicit values or they are parallel redundant pathways.

To address these questions, in vivo electrophysiology recordings of single-unit activity in the ACC and the BLA were performed in mice observing a demonstrator undergoing a fear conditioning paradigm, in which an auditory cue predicted electric shock. Data was initially analyzed through examining the proportion of single neurons showing an excitatory or inhibitory cue-evoked response in their peri-stimulus time histograms (PSTHs). This traditional analysis, however, ignored the heterogeneity between cells and complex temporal dynamics across time.

Here we proposed to estimate spiking activities and ensemble dynamics through state-space models, and characterized both single-cell and ensemble dynamics across fear observational learning. With these computational models, an intuitive framework is provided to explain how the ACC and the BLA contribute to fear observational learning. Specifically, ACC neurons represent socially-derived information of cue, and this contextual information is transmitted to the BLA, where the association between

the cue and the demonstrator's distress response is formed. This circuit model would not have been uncovered if we solely relied on a traditional single-cell analysis. Optogenetic manipulation was later used to demonstrate that the ACC→BLA pathway is necessary for the acquisition of observational, but not classical, fear conditioning.

## 3.2 Materials and Methods

### 3.2.1 Time reproduction task

The behavioral task was designed by Seth Egger, and my major contributions included collection and analysis of behavior and electrophysiology recordings from monkey G. This work is published on Egger et al. (2019). In this chapter, results and figures are adapted to focus on my contributed data. Single-cell and ensemble dynamics are compared and evaluated on how they can support a mental simulation model with information integration.

#### Experimental setup

Two awake male monkeys (*Macaca mulatta*) trained to perform a time reproduction task were used in the study, but data presented in this chapter came from monkey G. All experiments were conducted in accordance with NIH guidelines and were approved by the Committee of Animal Care at Massachusetts Institute of Technology.

#### Stereotactic surgery procedures

Surgeries were conducted under aseptic conditions using a stereotaxic instrument for non-human primates, where anesthesia administration and monitoring were conducted by vet staff in the MIT Department of Comparative Medicine. Prior to surgery, the animal already acquired the behavioral task through a year of training. Animals were anesthetized initially with ketamine and then sevofluorane (1.75%–2% for maintenance). Custom-made headposts and pins for head-fixation were inserted at 4 to 5mm into the skull. A rectangle chamber was implanted where the target recording

region was placed centrally inside the chamber. The animal was given a post-surgical recovery time of at least 2 weeks prior to the start of electrophysiology recordings.

### Behavioral task

During experimental sessions, monkeys were seated comfortably in a dark and quiet room. We used MWorks (http://mworks-project.org/) to implement task contingencies. All stimuli were presented on a black background. Eye positions were tracked with an infrared camera (Eyelink 1000; SR Research) and sampled at 1 kHz. Animals were trained to perform simple experiments including eye fixation and target-directed saccade, and tasks were gradually modified to a time reproduction task with reward shaping. The final behavioral paradigm (Figure 3.1a) is described as following:

Each trial began with the presentation of a central fixation point (FP; circular, red, 0.5 deg in diameter). Following fixation drawn from a uniform hazard (min = 100 ms, mean = 400 ms), a white circular target was presented 10 deg to the left or right of the FP. After another uniform hazard period (min = 500 ms, mean = 1000 ms), three flashed visual stimuli (S1, S2 and S3) were presented. Each flash was presented for 100 ms around the FP (white, inside diameter = 2.5 deg, outer diameter = 3 deg). The time between consecutive flashes was fixed within a trial and was drawn at random from a discrete uniform distribution across trials, ranging from 600 to 1000 ms (Figure 3.1b). We refer to the time between flashes as the sample interval $t_s$. After S3, monkeys had to initiate a saccade to the target to produce an interval equal to $t_s$. The produced interval $t_p$ was measured as the time between S3 and when the eye of the animal entered a circular window (diameter = 7 deg) around the saccadic target.

Early fixation breaks (before 200 ms post S3) or extremely long production times (e.g., above 1000 ms) were discarded. We additionally removed outlier trials whereby the value of $t_p$ was more than three standard deviations from the mean $t_p$ for a given $t_s$. Animals received trial-by-trial visual feedback on the magnitude and sign of error by a visual stimulus presented immediately after the saccade along the line connecting the FP to the saccadic target. The magnitude of the relative error was

represented by the distance of the feedback stimulus to the saccadic target, and positive (negative) errors were associated with locations farther away (closer to) the FP. When the magnitude of the relative error was smaller than a threshold, both the saccadic target and the feedback stimulus turned green and the animal received a juice reward. The magnitude of the reward decreased linearly with the magnitude of the relative error (Figure 3.1c). If the relative error was larger than the threshold, the target stimulus and analog feedback remained white and no reward was delivered.

Performance was evaluated according to the relative error to accommodate the scalar variability of time interval production. The reward threshold was initialized at 0.15 at the start of every session and was adjusted adaptively and on a trial-by-trial basis in a one-up one-down fashion with a fixed step size of 0.001 for unrewarded/rewarded trials, respectively. With this scheme, animals received reward in approximately 50% of the trials.

### Cue-conflict experiments

To test whether animals integrated both measurements, we ran a separate experiment in which one-third of the trials with $ts = 800$ ms were modified such that either the first measurement S1-S2 or the second measurement S2-S3 were 50 ms longer or shorter. We tested whether $tp$ was biased in a direction depending on which epoch was jittered. Next, we asked which of the following three models best explained the behavior in the cue conflict trials: (1) monkeys used both S1-S2 and S2-S3 to guide production; (2) monkeys used S1-S2 only; or (3) monkeys used S2-S3 only. We used a maximum likelihood procedure to fit the models to the data without conflict and compared the likelihood of the conflict data under each model.

### In vivo electrophysiology recordings

Prior to acute recordings, a craniotomy was made above the dorsaomedial frontal cortex (DMFC), which comprises the supplementary eye field (SEF) (Huerta and Kaas, 1990; Schlag and Schlag-Rey, 1987; Shook et al., 1991) and the presupplementary motor area (Pre-SMA) (Fujii et al., 2002; Matsuzaka et al., 1992), when the animal

was awake and head-fixed. Each craniotomy allowed about 3 to 5 recording sessions, and the whole experiment consisted of about 10 craniotomies. Anatomical locations of DMFC were determined by reported stereotactic coordinates, histological plots and MRI scan.

We sampled the DMFC using a quasi-systematic fashion: starting from the center of the stereotaxically identified region and outward to regions with strong modulation during either epoch of the task. For each recording session, neural activities were recorded with a 24-channel laminar V-probes (Plexon Inc.; Dallas Texas, USA) through the craniotomy, between 0.9 mm to 4.6 mm lateral of the midline and 1.8 mm to 9.3 mm anterior of the genu of the arcuate sulcus. Recordings were pre-processed with a Cerebus Neural Signal Processor (Blackrock Microsystems). Spike-sorting was performed online, followed by offline sorting using the MKsort package (https://github.com/ripple-neuro/mksort).

### Bayesian observer model for behavior

Similar to the Bayesian observer model in Chapter 2, two independent measurements $m_{\text{S1-S2}}$ and $m_{\text{S2-S3}}$ were used to compute the likelihood function $\lambda(m \mid t_s)$. The BLS estimate $e_{\text{S3-Go}}$ is derived with the posterior $\pi(t_s \mid m_{\text{S1-S2}}, m_{\text{S2-S3}})$ as below:

$$\pi(t_s \mid m_{\text{S1-S2}}, m_{\text{S2-S3}}) = \frac{p(t_s) \cdot \lambda(m_{\text{S1-S2}} \mid t_s) \cdot \lambda(m_{\text{S2-S3}} \mid t_s)}{p(m_{\text{S1-S2}}, m_{\text{S2-S3}})} \tag{3.1}$$

$$e_{\text{S3-Go}} = f_{\text{BLS}}(m_{\text{S1-S2}}, m_{\text{S2-S3}}) = \int t_s \cdot \pi(t_s \mid m_{\text{S1-S2}}, m_{\text{S2-S3}})\ dt_s \tag{3.2}$$

These are further used to compute the posterior for production time $t_p$ in (Eq. 3.3)

$$p(t_p \mid e_{\text{S3-Go}}) = \frac{1}{\sqrt{2\pi \cdot (w_p \cdot e_{\text{S3-Go}})^2}} \times e^{\dfrac{-(t_p - e_{\text{S3-Go}} - b)^2}{2 \cdot (w_p \cdot e_{\text{S3-Go}})^2}} \tag{3.3}$$

where $b$ is the offset term, and $w_p$ is the Weber fraction to account for scalar variability.

### Kalman filter model for behavior

The Bayesian model ignored that two time measurements were taken sequentially not simultaneously, so it is highly possible that two measurements were not conditional independent. To deal with the sequential updating, we implemented a model based on Kalman filter (Egger and Jazayeri, 2018). The estimated time right after S1 flash equals the the mean of the prior distribution.

$$e_{\text{S1-S2}} = \int t_s \cdot p(t_s)\ dt_s \tag{3.4}$$

The estimated time right after S2 flash, it combines the measurement $m_{\text{S1-S2}}$ of the first interval with previous estimate $e_{\text{S1-S2}}$. Similarly, the estimated time after S3 flash, it combines the second measurement $m_{\text{S2-S3}}$ with previous estimate $e_{\text{S2-S3}}$.

$$e_{\text{S2-S3}} = e_{\text{S1-S2}} + \alpha \cdot f(m_{\text{S1-S2}} - e_{\text{S1-S2}}) \tag{3.5}$$

$$e_{\text{S3-Go}} = e_{\text{S2-S3}} + \beta \cdot f(m_{\text{S2-S3}} - e_{\text{S2-S3}}) \tag{3.6}$$

where $f(m)$ is the mean of the posterior $\pi(ts \mid m)$, $\alpha = 1$ and $\beta = 0.5$. This estimate $e_{\text{S2-S3}}$ is later used for producing $t_p$.

### Fitting procedure

Similar to the fitting procedure in Chapter 2, we assumed that $t_p$ values associated with any $t_s$ were independent across trials and thus expressed the joint conditional probability of individual $t_p$ values across all the N trials by the product of their individual conditional probabilities.

$$\log\, p(t_p^1, t_p^2, ..., t_p^N \mid t_s;\ w_m, w_p, b_0) = \sum_{i=1}^{N} \log\, p(t_p^i \mid t_s;\ w_m, w_p, b_0)$$

We derived the animal's internal estimate $t_{e(\text{S1})}$, $t_{e(\text{S2})}$ and $t_{e(\text{S3})}$ after each flash S1, S2, and S3 from both Bayesian and Kalman models. We assume the only available information to the animal at S1 is prior, and thus $t_{e(\text{S1})}$ is set as the mean of the prior.

### 3.2.2 Fear observational learning task

The behavioral data and electrophysiology recordings were provided by Stephen Allsop and his colleagues, and my major contributions were computational methodology, analysis and interpretation. This work is published on Allsop et al. (2018). In this chapter, results and figures are adapted to focus on the comparison between how single-cell and ensemble dynamics can support a mental simulation model in which associations are learned through observations.

#### Experimental setup

Adult (8-12 weeks) male C57BL/6J were used for all experiments. All mice were pair-housed in the Picower Institute on a reverse 12 hour light-dark cycle with food and water ad libitum. All experiments were conducted in accordance with NIH guidelines and with approval of the MIT Institutional Animal Care and Use Committee and the MIT Department of Comparative Medicine.

#### Stereotactic surgery procedures

All surgeries were conducted under aseptic conditions using a digital small animal stereotaxic instrument (David Kopf Instruments, Tujunga, CA, USA). Mice were anaesthetized with isoflurane (5% for induction, 1%–2.5% for maintenance). Injections were performed using a beveled 33-gauge microinjection needle. A 10-$\mu l$ microsyringe (nanofil; WPI, Sarasotam FL, USA) was used to deliver virus at a rate of 0.1 $\mu l$ per min using a microsyringe pump (UMP3; WPI) and controller (Micro4; WPI). Mice were given a post-surgical recovery time of at least 7 days prior to start of any experimental procedures.

#### Surgery for in vivo recordings

To target the basolateral amygdala (BLA) for in vivo recordings, a craniotomy was made in the right hemisphere at anteroposterior (AP) -1.6 mm and mediolateral (ML) +3.35 mm. Two to three skull screws were implanted around the site of the

craniotomy. One layer of adhesive cement (C&B Metabond; Parkell, Edgewood, NY, USA) followed by cranioplastic cement (Dental cement; Ortho-Jet, Lang Dental, Wheeling, IL, USA) was used to stabilize screws to the skull. A 16-channel multi array electrode (Innovative Neurophysiology) was then lowered at approximately 0.01 mm/s to -4.75 dorsoventral (DV) as measured from bregma. A ground wire was placed in the contralateral posterior hemisphere at an approximate depth of 1mm. An additional layer of cranioplastic cement (Ortho-Jet, Lang Dental, Wheeling, IL, USA) was applied to the skull as well as around the wires. The electrode was then lowered to - 4.9 DV and stabilized with additional layers of cement.

In order to record from anterior cingulate cortex (ACC) neurons in a circuit-specific manner an adeno-associated virus serotype 5 carrying a construct for expression of ChR2 fused to a enhanced yellow fluorescent protein, under the control of a double-inverted open reading frame (AAV5-DIO-ChR2-eYFP) (1 $\mu l$) was injected into the ACC (AP: +1.0 mm, ML: -0.3 mm, DV: -2.1 mm) and 1 $\mu l$ of the retrogradely traveling canine adenovirus carrying cre-recombinase CAV2-EF1$\alpha$-Cre was injected into the BLA (AP: -1.6 mm, ML: 0.35 mm, DV: -4.9 mm). 5-8 weeks later, a second surgery was performed to implant an optrode. Using the same surgical conditions and anesthesia as previously described, one craniotomy was drilled over the ACC (AP: +1.0 mm, ML: -0.3 mm). Cranioplastic cement was placed around the optrode and ground wire and the optrode was then lowered to -2.1 mm DV.

**Surgery for optogenetic experiments**

In order to inhibit ACC input to the BLA, an Adeno-associated virus carrying the gene for a fusion protein comprised of enhanced Halorhodopsin and enhanced yellow fluorescent protein under the calmodulin kinase II promoter (AAV5-CaMKII$\alpha$-eNpHR3.0-eYFP) or eYFP alone was bilaterally injected into the ACC (AP: +1 mm, ML: $\pm$0.25 mm, DV: -2.1 mm). After waiting for 5 minutes the needle was raised to -2.0 mm and another 300 $\mu l$ of virus was injected at the same rate. After waiting 5 additional minutes the needle was raised to -1.9 mm for 10 minutes before being slowly withdrawn. After 4-8 weeks mice underwent a second surgery in which two

optical fibers were implanted bilaterally over the BLA (AP: -1.6 mm, ML: ±3.35 mm, DV: -4.5 mm). Fibers were lowered at approximately 0.01 mm/s and were secured using a thin layer of adhesive cement followed by dental cement. In a subset of mice instead of optical fibers an optrode was implanted in the BLA following the same surgical procedure as described above.

**Behavioral task**

Mice were placed in a soundproof conditioning chamber (Med Associates, St Albans, VT, USA) with a shock floor side and a plastic floor side separated by a transparent plastic divider containing holes.

Mice in the Experienced Observers (EO) group received a "shock experience" by being placed on the shock floor side of the chamber and allowed to freely explore. After 5 minutes they received 1 unpredicted, un-cued, footshock (all mice greater than 30 g were shocked with 1.5 mA, while mice less than 30 g were shocked with 1mA) and were immediately transferred to the plastic floor side of the chamber. The cage mate of the observer was then placed into the shock side of the chamber as the demonstrator for "observational conditioning". After 5 minutes of habituation, demonstrators underwent 30 trials that occurred at random intervals (60, 90, 120, 150, 180 seconds) in which a 20 second compound cue (light and 10 kHz tone) predicted the delivery of a 2 second shock (1 mA - 1.5 mA) 10 seconds after the onset of the cue. Directly after, mice were placed back into their home cages. 24 hrs later, observer mice were placed back into the shock side of the chamber and 30 cues were delivered to the chamber in the absence of shock.

Mice in the Unpaired Observers (UO) group received a "shock experience", however during observational conditioning cues and shocks delivered to the demonstrator were explicitly unpaired. Experienced Solo mice (ES) received a "shock experience" but then no demonstrator was placed into the shock side of the chamber. Paired cues and shocks were delivered to the empty side of the chamber.

Naive Observers (NO) did not receive a "shock experience" but instead were placed on the plastic side of the chamber where they observed demonstrators in the same

way as EO. Lastly, Naive Solo mice (NS) also did not receive a "shock experience" but instead were placed on the plastic side of the chamber. Paired cues and shocks were delivered to the shock floor side of the chamber in the absence of a demonstrator.

Behavioral performance was recorded by digital video cameras. All videos were manually analyzed offline by an experimenter blind to experimental conditions. Freezing behavior of the observer was scored on both conditioning and test day as the amount of freezing during the cue minus the amount of freezing in the 20 seconds directly preceding the cue. Freezing was defined as absence of movement, with the exception of respiration. Based on preliminary behavioral data, analysis of freezing behavior for the training day was performed on trials 5-20. Analysis of freezing behavior on test day was performed on the first 5 trials. Additional other stereotyped behaviors shown by the observer such as grooming and escaping, were quantified by manual scoring. Observer mice were said to be mimicking when they displayed escape behaviors in response to demonstrators showing escape behaviors.

Behavior of electrode-implanted observer mice was recorded throughout observational conditioning and test day, with videos manually scored offline as described above to identify bouts of freezing, grooming and escape behaviors. Raster plots of behavior were generated for each individual mouse and presented in 0.1s bins. Raster plots of average freezing behavior for each group were calculated and presented in 1s bins, indicating the average proportion of time spent freezing during each 1s bin by all mice in each group.

### In vivo electrophysiology recordings

Electrophysiological recordings were performed in the ACC and the BLA during either paired or unpaired observational training (i.e., EO or EU, respectively), as described above for the initial behavioral experiments. Recordings were performed from an initial "habituation" session shortly prior to the observational conditioning. During habituation, only the cue was delivered at random intervals, but there was no demonstrator across nor shock delivery. This enables comparisons of neural cue responses before and during observational learning.

Mice that expressed halorhodopsin (NpHR) bilaterally in the ACC and had an optrode implanted in the right BLA were plugged into the spike acquisition system and an optical patch cord. Mice were then placed into the observational conditioning chamber described above, received an initial shock experience, and were then transferred to the plastic side of the chamber. The demonstrator was placed on the shock floor and 15 cue-shock pairings were delivered to the chamber. The next 30 trials had a subset of trials were the laser delivered yellow light (593 nm) 1 s before the onset of the cue and stayed on until 1 s after the cue. This was done in a pseudorandom order. Overall, there were 25 trials in which no laser stimulation was delivered and 20 trials where laser stimulation was delivered. Neural activity was recorded continuously throughout the experiment.

The response of individual neurons to the cue was examined using the non-parametric Wilcoxon signed-rank test. Two signed-rank tests were performed per neuron to capture phasic and sustained responses to the cue. For phasic cue responses, neural activity was binned in 100 ms epochs, and statistical comparisons were made between the firing frequency within a baseline window of 1000 ms prior to cue onset and the firing frequency within an experimental window of 500 ms after cue onset. For sustained cue responses, neural activity was binned in 1 s epochs, and statistical comparisons were made between the firing frequency within a baseline window of 20 s prior to cue onset and the firing frequency within an experimental window of 9 s after cue onset. Both tests were performed with 1000 bootstraps and Bonferroni corrections were performed to control for multiple comparisons.

Neurons were deemed as cue responsive if they exhibited statistical significance on either the phasic or sustained response test. To determine whether cells exhibited training-induced changes in the magnitude of cue response, we compared activities during the habituation and observational conditioning phases. Since there was a higher number of trials during the observational conditioning phase, only a subset of 16 conditioning trials were used for comparisons against 15 habituation trials. Conditioning trials 5-20 were chosen for this analysis as they corresponded to the trials that we used to measure learning behaviorally.

The response of neurons to shock delivery to demonstrator mice was also examined using the signed-rank test. Neural activity was binned in 100 ms epochs, and statistical comparisons were made between the firing frequency within a baseline window of 5 s prior to shock onset and the firing frequency within the 2 s of shock delivery to the demonstrators. To account for potential electrical contamination during shock delivery, statistical comparisons were also made between shock delivery with and without the presence of demonstrator mice, while observer mice freely behaved on the "safe" plastic floor compartment. As for cue response analysis, shock response analysis was performed with 1000 bootstraps and Bonferroni corrections for multiple comparisons.

### 3.2.3 Latent estimation from the point process

To estimate firing rates for individual neurons and describe variability across trials, a state-space generalized linear model is developed based on (Czanner et al., 2008; Smith and Brown, 2003; Smith et al., 2010). The spike train is assumed to be a Poisson point process (Brown et al., 1998), i.e., a time-series of random binary events that occur in continuous time. We designate spikes as 1's times at which no spikes occur as 0's. Given a trial interval $(0, T]$, we define $N(t)$ as the number of spikes recorded in interval $(0, T]$ for $t \in (0, T]$ in a trial. A point process model is completely defined by its conditional intensity function $\lambda(t \mid H_t)$:

$$\lambda(t \mid H_t) = \lim_{\Delta \to 0} \frac{\Pr(N(t+\Delta) - N(t) = 1 \mid H_t}{\Delta} \tag{3.7}$$

where $H_t$ is the spike history up to time $t$. Given the history up to time $t$, the approximate probability of one spike in an interval $(t, t + \Delta]$ is defined in Eq (3.7). The conditional intensity function generalizes the Poisson process rate function by allowing history dependence.

Assume an experiment has $K$ trials of the same task ($k = 1, \ldots, K$) and $T$ is the length of each trial where neural spiking activities are simultaneously recorded with the task. The observation interval is $(0, T]$. We develop a discrete time representation of the conditional intensity function with subintervals $\Delta = T/L$, where $L$ is large

enough so that each subinterval ($l = 1, \ldots, L$) contains at most one spike. $n_{k,l}$ is an indicator depending on whether there is a spike in subinterval $((l-1)\Delta, l\Delta]$ on trial $k$. $n_k = n_{k,1}, \ldots, n_{k,L}$ are the spikes on trial $k$ and $N_{1:k} = n_1, \ldots, n_k$ are the spikes from trials 1 to $k$. $H_{k,l}$ denotes the spike history in trial $k$ up to time $l\Delta$. We assume that the conditional intensity function for the spike trains at time $l\Delta$ of trial $k$ may be written as

$$\lambda_k(l\Delta \mid \theta_k, \gamma, H_{k,l}) = \lambda^S(l\Delta \mid \theta_k) \cdot \lambda^H(l\Delta \mid \gamma, H_{k,l}) \tag{3.8}$$

where the component $\lambda^S(l\Delta \mid \theta_k)$ describes the task effect on the neural spiking and the component $\lambda^H(l\Delta \mid \gamma, H_{k,l})$ defines the spike history effect on neural spiking.

We model the task effect on the neural spiking activity by assuming that the first component of the conditional intensity function in Eq (3.8) has the form:

$$\log\lambda^S(l\Delta \mid \theta_k) = \sum_{r=1}^{R} \theta_{k,r} g_r(l\Delta) \tag{3.9}$$

where $g_r(l\Delta)$ is a set of unit pulse functions (Czanner et al., 2008) for modeling the within-trial task specific-effect on spiking activity parameterized by $\theta_k = \theta_{k,1}, \ldots, \theta_{k,R}$ for each trial $k = 1, \ldots, K$. Eq (3.9) defines for each trial how the task modulates the spiking activity. We assume a different $\theta_k$ for each trial. A state-space model is used to define the dynamics across trials.

To model the effect of spike history on current neural spiking activity, we assume that the second component on the right of Eq (3.8) can be defined as

$$\log\lambda^H(l\Delta \mid \gamma, H_{k,l}) = \sum_{j=1}^{J} \gamma_j n_{k,l-j} \tag{3.10}$$

where $\gamma = (\gamma_1, \ldots, \gamma_J)$ is the vector of parameters which define the dependence of the current spiking activity on recent spike history, for trial $k = 1, \ldots, K$ . In this model, we take the history to be $H_{k,l} = n_{k,l-J}, \ldots, n_{k,l-1}$, which is the spiking activity during the preceding $J$ time intervals prior to time $l\Delta$.

Together, Eqs (3.8) - (3.10) define the observation equation of our state-space model of neural activity:

$$\lambda_k(l\Delta \mid \theta_k, \gamma, H_{k,l}) = \exp\left(\sum_{r=1}^{R} \theta_{k,r} g_r(l\Delta)\right) \cdot \exp\left(\sum_{j=1}^{J} \gamma_j n_{k,l-j}\right) \tag{3.11}$$

Eq (3.11) defines a generalized linear model formulation of the conditional intensity function. The probability of a spike event on trial $k$ at time $l\Delta$ is

$$\Pr(n_{k,l} \mid H_{k,l}) = (\lambda_k(l\Delta \mid \theta_k, \gamma, H_{k,l})\Delta)^{n_{k,l}} \cdot e^{-\lambda_k(l\Delta|\theta_k,\gamma,H_{k,l})\Delta} \tag{3.12}$$

which is the state equation to specify the relation of spiking activity between trials. Stochastic dependence is defined by the random walk model $\theta_k = \theta_{k-1} + \epsilon_k$ for $k = 1, \ldots, K$, where $\epsilon_k$ is a R-dimensional Gaussian random vector with zero mean and covariance $\Sigma$. The initial vector $\theta_0$ is assumed unknown.

Eq (3.12) with a random walk model defines our state-space generalized linear model. For unknown parameters $\psi = (\gamma, \theta_0, \Sigma)$, we use an EM algorithm to compute their estimates $\hat{\psi}$ by maximum likelihood (Czanner et al., 2008; Smith and Brown, 2003), with a penalized log likelihood function below:

$$\begin{aligned}\log \Pr(N, \theta \mid \psi) = \sum_{k=1}^{K}\sum_{l=1}^{L} n_{k,l} \log\left(\lambda_k(l\Delta \mid \theta_k, \gamma, H_{k,l})\Delta\right) - \lambda_k(l\Delta \mid \theta_k, \gamma, H_{k,l})\Delta \\ - \frac{1}{2}\sum_{k=1}^{K}(\theta_k - \theta_{k-1})^T \Sigma^{-1}(\theta_k - \theta_{k-1}) \\ + K \cdot \log\left((2\pi)^{-R/2} \mid \Sigma \mid\right)^{-1/2}\end{aligned}$$

The task-effect coefficients impose a stochastic regularization constraint on the conditional intensity function.

### rate-change detection

To evaluate how activity of a single neuron changes across learning, a trial response window from the observation interval is used to determine how the firing rate during

that window changes across time, using latent-estimation model for point-process. In order to find the trial during observational conditioning at which the firing rate during the cue was significantly different than the firing rate observed during habituation, we modified the trial-to-trial comparison algorithm in Smith et al. (2010). In Smith et al. (2010), samples at trial $i$ were compared with samples at trial $j$ and the probability that these distributions were different was computed. Here, we compare samples from the average of the 15 habituation trials (i.e., the covariance matrix from trials 1-15) with samples from trial $j$. When the samples from trials $j$ and $j+1$ are different than the combined samples from trials 1-15 with 95% probability, we designate trial $j$ as the rate-change trial.

We performed these analyses on all neurons. We chose a 1 sec response window instead of 500 ms because it enabled estimations with greater confidence through data accumulation. We also performed the analysis with 9 sec response windows and used the earliest trial estimate. The algorithm successfully provided estimates for 87% of BLA and 92% of ACC neurons, whereas the rest failed to converge due to extreme sparsity or fast-spiking. These estimates were used to generate the histogram shown in Figure 3.10 and Supplementary Figure 3.6.

### 3.2.4 Latent estimation from the cell populations

#### Time reproduction task

We computed peri-stimulus time histograms (PSTHs) for each unit to estimate its trial-averaged activity over time for each target interval (or sample interval). We aligned trials to different timepoints (i.e., S1, S2, S3 and Go). Spiking activities were smoothed with a 150 msec box filter and the square root of spike counts were used to reduce the effect of Poisson noise (Yu et al., 2009). Prior to dimensionality reduction, we removed neurons that had extremely sparse spiking. We then z-scored the PSTH for each unit through normalizing by its standard deviation across time and conditions, so that all the neurons used for the analysis have similar overall variability across conditions (Ames et al., 2014).

We then performed principal component analysis (PCA) on these PSTHs for each condition to find the representative features in the face of the heterogeneity in neural dynamics. We preserved the first 10 dimensions that represented more than 90% of the total variance of the original data, and visualized them in a 3D space formed by PC1, PC2, and PC3. We also performed PCA individually on activity within each epoch (e.g., S1-S2, S2-S3, and S3-Go).

In this study, to calculate the distance between neural trajectories in the low-dimensional state-space, we cannot simply compare the distance between them at each corresponding timepoint. It is because different trials and conditions do not contain the same number of time points, i.e., trials with a longer sample interval $t_s$ or a longer produced interval $t_p$ have more timepoints. Therefore, we adapted a geometric analysis described in Ames et al. (2014) to calculate the distance between high-dimensional trajectories, by selecting relevant timepoints on a reference trajectory and finding the closest point on other trajectories.

Similarly, we adapted the analysis described in Afshar et al. (2011) to calculate the speed of projected neural trajectories, we a preliminary speed can be calculated from the difference between neural states in the PCA space at two timepoints, and then projected onto the speed of the reference trajectory (i.e., the mean trajectory across all trials and all conditions), timepoint by timepoint.

By combining both methods for a timing task (Remington et al., 2018), we used the mean trajectory from trials with a medium sample interval (i.e., $t_s = 800$ msec) to build a reference trajectory $\mathbf{r}^{[800]}$. This reference trajectory is sampled with a resolution of 1 msec, so there are 800 points on the time vector in each measurement epoch (S1-S2 and S2-S3), and $\langle t_p \mid t_s = 800 \rangle$ points in the production epoch (S3-Go). We denote the vector of timepoints as $\mathbf{t}^{[800]}$, with a total length of $T$.

For each time point $t$ sampled from $\mathbf{t}^{[800]}$, the corresponding neural states on the reference trajectory, $\mathbf{r}^{[800]}$, form a 10-dimensional vector. These vectors can be arranged into a 10-by-$T$ matrix $\mathbf{R}^{[800]}$ to represent the reference trajectory. We then find a state vector $\mathbf{r}^{[t_s]}$ on a non-reference trajectory associated with $t_s$ that has minimum Euclidean distance to $\mathbf{r}^{[800]}$.

The difference vector $\delta^{t_s}$ quantifies the distance between neural states on non-reference trajectory and the reference trajectory as below:

$$\delta^{t_s} = \mathbf{r}^{[t_s]} - \mathbf{r}^{[800]}$$

Across all timepoints, the set of states form a matrix $\mathbf{R}^{[t_s]}$. The difference between time point $t$ and each corresponding time in $\mathbf{R}^{[t_s]}$ describes how much leading or lagging in other trajectories relative to the reference trajectory, i.e., the time it takes for the state variable to reach a point in $\mathbf{r}^{[t_s]}$. We can thus approximate the speed of trajectory $\mathbf{V}^{[t_s]}$ by:

$$\mathbf{V}^{[t_s]} = \mathbf{V}^{[800]} \cdot \mathbf{r}^{[800]}/\mathbf{t}^{[t_s]}$$

where $\mathbf{r}^{[800]}/\mathbf{t}^{[t_s]}$ is the slope of the regression line relating $\mathbf{t}^{[t_s]}$ to $\mathbf{t}^{[800]}$.

### Fear observational learning task

We computed peri-stimulus time histograms (PSTHs) for each unit to estimate its trial-averaged activity over time for each condition. We aligned trials to the cue, i.e., conditioned stimulus (CS), onset, and each trial was sampled from 2 s before the cue onset to 5 s after the cue onset. We binned spike trains with a 50-msec window, and convolved these with a 150-msec Gaussian kernel to smooth the PSTHs. For each condition, a high-dimensional neural activity space was created, where each axis is the firing rate for each unit. PSTHs from all the units will form a trajectory evolving in an n-dimensional space (n = the number of units). Prior to dimensionality reduction, we removed neurons that had extremely low firing rates over the 7 s window. We normalized the PSTH for each unit by its maximum variance across conditions, so we can avoid being biased by neurons with higher firing-rates and ensure that neurons have similar overall variability across conditions (Ames et al., 2014).

We then performed principal component analysis (PCA) on these PSTHs for each condition to find the representative features in the face of the heterogeneity in neural dynamics. We preserved the dimensions that represented more than 70% of the total variance of the original data. As the first two principal components (PCs) are the

most salient features of the population, we projected the trajectory on a 2D plane formed by PC1 and PC2.

The Euclidean distance between the neural trajectories for the habituation and observational conditioning conditions at each time point was derived and visualized across time (from 2 sec before the cue onset to 5 sec after the cue onset). The average distance during baseline period before the cue onset (2 sec, 40 data points where each data point is separated by 50 msec) and average distance after the cue onset (2 sec, 40 data points) were compared. We calculated their difference to quantify how far apart the neural state at the habituation and observational conditioning training phases were during the baseline and CS presentation periods.

## 3.3 Results

### 3.3.1 Time reproduction task

#### 3.3.1.1 Monkey integrates sequential measurements with prior knowledge to improve time estimates

To study the neural mechanism underlying time reproduction with mental simulation, we designed a behavioral paradigm in which a monkey observes three isochronous flashes (S1, S2, and S3) presented around the fixation point on the screen and initiates a delayed saccade to a target (Go) with interval matching the inter-flash interval (Figure 3.1a). For each trial, the sample interval ($t_s$) between consecutive flashes is drawn from a uniform distribution (Figure 3.1b), and the animal needs to estimate $t_s$ and produce a matched interval ($t_p$) after S3 presentation. A reward is delivered if responses are sufficiently accurate, i.e., the relative error $\mid (t_p - t_s) \mid / t_s$ is below the threshold. Maximal reward was delivered when $t_p = t_s$ and decreased linearly with the relative error (Figure 3.1c).

By plotting the produced intervals ($t_p$) as a function of sample interval ($t_s$), we found that the animal learned to time flexibly based on $t_s$ (Figure 3.1e). Consistent with scalar variability in timing tasks in the literature (Acerbi et al., 2012; Brenner

and Smeets, 2009; Gibbon et al., 1984; Rakitin et al., 1998), the variability of $t_p$ also increased with $t_s$ (n = 699, 724, 243, 685 and 643 trials for $t_s$ from 600 to 1000 msec respectively). We also observed systematic biases of $t_p$ toward the mean of the prior (BIAS = 29.89 ± 1.54 msec) with a regression slope less than unity (two-tailed t-test: $t_{(2993)} = -987.8966$, $p < 0.001$).

Similar to our model in Chapter 2, this behavior could be explained by a Bayesian observer model integrating the prior knowledge with two measurements (Figure 3.2a). Different from the previous chapter, the two measurements came from the same modality in this study. Here we assumed that two measurements were conditional independent, but this assumption might not be realistic considering two time samples were measured sequentially instead of simultaneously. To deal with the sequential updating, we implemented a Kalman filter model to better capture how the animal's estimate is updated in this task (Figure 3.2c), through the following steps: Immediately after S1 cue, the estimated time, $t_{e(\mathrm{S1})}$, was set to the mean of the prior, given no other available information prior to cue presentation. After S2 cue was presented, the model updated its estimate to $t_{e(\mathrm{S2})}$ by correcting the error between $t_{e(\mathrm{S1})}$ and measurement during the S1-S2 epoch. Similar process was repeated to obtain the final estimate $t_{e(\mathrm{S3})}$ by correcting the error between $t_{e(\mathrm{S2})}$ and measurement during the S2-S3 epoch. In these steps, error correction was governed by a nonlinear function (see Materials and Methods). Finally, $t_{e(\mathrm{S3})}$ was corrupted with production noise to generate $t_p$. Indeed, the animal's behavior was better captured by the Kalman filter model compared to the Bayesian observer model (Figure 3.2e).

To test whether the animal attended to both measurements instead of one of them, an additional cue-conflict experiment was performed. In this experiment, a subset of trials corresponding to $t_s = 800$ msec were made either 50 msec longer or shorter, in either the S1-S2 interval or S2-S3 interval. When either the S1-S2 interval or S2-S3 interval was 850 msec, the average $t_p$ was significantly longer than both intervals being 800 msec. Similarly, when either the S1-S2 interval or S2-S3 interval was 750 msec, the average $t_p$ became significantly shorter (Figure 3.2f). In addition, whether the jittering was introduced in the first or the second interval did not make

a difference statistically (two-tailed t-test: $t_{(213)} = -1.54$, $p = 0.125$), suggesting that both measurements were combined with comparable weights. By comparing the log-likelihood of different models, we showed that a model integrating both measurements was more likely than using one measurement (Figure 3.2g). These results together suggest that the animal integrated both measurements with the prior to derive a time estimate. Note that these results examined performance across trials, so we could not rule out the possibility that the animal switched attention to either measurement on a trial-by-trial basis.

#### 3.3.1.2 Hypothesis for neural codes using an internal model

Behavioral and modeling data have revealed that the animal can update his internal estimate sequentially to reproduce a sample interval, which may be interpreted under an internal model framework. As described in other sensorimotor behaviors (Miall and Wolpert, 1996; Shadmehr and Krakauer, 2008; Wolpert et al., 1995), the brain may comprise a controller driving the motor commands based on the estimated state from a state estimator, a simulator predicting outputs, and a state estimator integrating sensory inputs with predictions from simulation (Figure 3.3a). Previous studies also have shown that the neural dynamics are temporally scaled to control movement initiation (Afshar et al., 2011; Russo et al., 2020; Wang et al., 2018; Xu et al., 2014).

We thus hypothesized that the controller and the simulator are implemented by two populations, jointly regulating the neural dynamics to update the time estimate (Figure 3.3b). Under this hypothesis, activities of the controller ($r_1$) determine the speed of neural dynamics of the simulator ($r_2$), whereas the simulator signal the controller to update the speed based on the prediction error at the time of flash (Figure 3.3a, b). Specifically, the speed of $r_2$ dynamics in the S1-S2 epoch should be fixed to reflect the initial estimate $t_{e(\mathrm{S1})}$. Deviations from the expected state at S2 are used to update $r_1$ signals, which in turn adjusted the speed of $r_2$ dynamics in the S2-S3 epoch. Same procedure is repeated to adjust $t_{e(\mathrm{S3})}$ for the final motor plan. To generalize, we predicted that both the $r_1$ activities (controller) and the slope of ramping activities in $r_2$ (simulator) can reflect the underlying time estimates.

#### 3.3.1.3 Temporal scaling of heterogeneous single-cell profiles

To test the above hypothesis, we collected single-unit recordings (n = 118) in the dorsomedial frontal cortex (DMFC) when the animal performed the task. We found that individual neurons had heterogeneous response profiles across three different epochs (S1-S2, S2-S3, and S3-Go) throughout the trial (Figure 3.4a). In addition, we observed that response profiles from the same neuron were similar across different sample intervals $t_s$, i.e. responses seemed to be temporally scaled, regardless of response characteristics.

Similar scaling effect has been reported from the literature (Afshar et al., 2011; Merchant et al., 2011; Murakami et al., 2014; Russo et al., 2020; Xu et al., 2014), but many studies focused on analyzing neurons with monotonic ramping profiles. To further evaluate temporal scaling with non-monotonic responses, we normalized the firing rates by the animal's time estimate for each epoch, i.e., $t_{e(\mathrm{S2})}$ for the S2-S3 epoch (Figure 3.4b, *top*) and $t_{e(\mathrm{S3})}$ for the S3-Go epoch (Figure 3.4b, *bottom*). We then fit a 6th order polynomial to the normalized responses (Supplementary Figure 3.1) and compared with the fittings of original data. The difference in explained variance, i.e. $\Delta R^2$, reflected the degree of temporal scaling. By comparing $\Delta R^2$ in different periods within a trial, we found stronger temporal scaling in the S2-S3 epoch than the S1-S2 epoch (one-tailed Wilcoxon: $\mathrm{Z} = 4.62$, $p < 0.001$), with this effect being more evident in the S3-Go epoch (one-tailed Wilcoxon: $\mathrm{Z} = 7.08$, $p < 0.001$) (Figure 3.4c). These results suggest that the speed of activities were regulated across measurements to obtain a better time estimate, and support the internal model hypothesis at a single neuron level.

#### 3.3.1.4 Ensemble dynamics reveals that the DMFC regulates speed and distance of manifolds for flexible timing

Despite the scaling effect revealed in single-cell analyses, it is challenging to unify these heterogeneous profiles across neurons. Through discussions in Chapter 1, population coding can extract more information without assuming specific response char-

acteristics in individual neurons. Single cells whose responses that did not seem to be modulated by the task might embed information recruited at a population level. Moreover, population coding can help reduce uncertainty associated with ambiguous environments. With these considerations, we proposed to analyze the dynamics of the entire neural population using a state-space method (see Materials and Methods).

With this approach, high-dimensional response profiles of all the neurons were projected to a low-dimensional manifold (Churchland et al., 2012; Cueva et al., 2020; Mante et al., 2013; Petreska et al., 2011; Yu et al., 2005, 2009). To better visualize the ensemble dynamics across time within a single trial (i.e., neural trajectories), we projected them to a 3D space spanned by the first 3 principal components (PCs) as a function of $t_s$. We found that neural trajectories terminated in distant states at S2 but terminated in nearby states at S3 and Go across different $t_s$ (Figure 3.5a). In addition, we found that activities in different epochs resided in distinct subspaces. By examining whether PCs built to explain variance for one epoch could explain comparable variance in other epochs, we found that 10 PCs in each epoch were sufficient to capture 90% of the variance within that epoch, but needed a lot more PCs to explain the same amount of variance in another epoch (Figure 3.5b). This verified that different epochs were distinctly encoded in low-dimensional manifolds. Furthermore, we found that neural trajectories shared similar shapes irrespective of $t_s$, i.e., low-dimensional profiles were also temporally scaled (Figure 3.5c). This implied that the ensemble dynamics might evolve at speeds that reflect the animal's estimate of time, e.g., $t_{e(S2)}$ and $t_{e(S3)}$.

To quantify the speed of a neural trajectory, we initially computed the distance between neural states along the trajectory aligned to Go (Supplementary Figure 3.2a), and found that produced times $t_p$ were inversely proportional to the speed along each trajectory in the S3-Go epoch on a trial-by-trial basis (Supplementary Figure 3.2b). However, this approach could create biases in estimation with respect to trajectories with different lengths. For example, a trajectory associated with a longer $t_p$ or $t_s$ contained more timepoints. Moreover, to estimate instantaneous speeds at each timepoint, it is important to avoid any misalignment in time. Therefore, we defined

the neural trajectory associated with $t_s = 800$ msec as the reference trajectory, and selected states on other trajectories that were closest to the reference states to infer their distances and speeds (see Materials and Methods)

We denoted the vector of timepoints along the reference trajectory as $\mathbf{t}^{[800]}$, and neural states on the reference trajectory at any given time as $\mathbf{r}^{[800]}$. The closest neural states on other trajectories to $\mathbf{r}^{[800]}$ were denoted as $\mathbf{r}^{[t_s]}$ (Figure 3.6a), with the distance $\delta^{t_s}$ (Figure 3.6c). $\mathbf{t}^{[t_s]}$ represented a set of timepoints required to reach $\mathbf{r}^{[t_s]}$ (Figure 3.6b). By mapping $\mathbf{t}^{[t_s]}$ to $\mathbf{t}^{[800]}$, we obtained the speed along each trajectory.

We analyzed the relative speed and distance between neural trajectories with different $t_s$, and found neural trajectories in the S1-S2 epoch shared a same constant speed, i.e., the slope of $\mathbf{t}^{[t_s]}$ as a function of $\mathbf{t}^{[800]}$ was fixed (Figure 3.6d, top *left*). As S2 was presented, the animal was able to update its internal estimate based on the measurement of S1-S2 period. The sets of $\mathbf{t}^{[t_s]}$ started to diverge across $t_s$ in the S2-S3 epoch, demonstrating $t_s$-dependent speeds. Specifically, trajectories associated with longer $t_s$ had larger slopes of $\mathbf{t}^{[t_s]}$ to $\mathbf{t}^{[800]}$, i.e., it took longer to reach neural states $\mathbf{r}^{[t_s]}$ compared to the reference trajectory (Figure 3.6d, top *middle*). Similarly, the ratio of $\mathbf{t}^{[t_s]}$ with respect to $\mathbf{t}^{[800]}$ in the S3-Go epoch also demonstrated $t_s$-dependent speeds (Figure 3.6d, top *right*). To further quantify the speeds, we fitted a piecewise linear model to trajectories across trials, by assuming a linear relationship between $\mathbf{t}^{[t_s]}$ and $\mathbf{t}^{[800]}$. The distributions of speeds across trajectories, sorted by $t_s$, got wider in the S2-S3 epoch and mildly adjusted in the S3-Go epoch (Figure 3.6e). These suggest that the animal's time estimate after the first measurement, $t_{e(\mathrm{S2})}$, was good enough and only got slightly improved by the second measurement to reach $t_{e(\mathrm{S3})}$.

The differences in speeds of neural trajectories were reflected in their distances. In the S1-Se epoch, trajectories were overlapped due to the same neural speeds as well as the same initial states (Figure 3.6d, bottom *left*). As S2 was presented, the neural trajectories were separated by initial states that were determined by the time of flash (Figure 3.6d, bottom *middle*). Distance between trajectories, i.e., $\delta^{t_s}$, remained approximately constant during each epoch, suggesting that neural dynamics evolved with constant speeds in each epoch. We additionally found that neural trajectories

became closer in the S3-Go epoch (Figure 3.6d, bottom *right*), indicating that $\delta^{t_s}$ got adjusted by additional measurements.

One potential concern was that production times $t_p$ were different for the same$t_s$, and averaging firing rates across these trials might create biases in obtaining a trial-averaged trajectory during the S3-Go epoch. We thus aligned the trials to the time of sacade (Go) and still observed $t_s$-dependent speeds of the trajectories. Despite the complex heterogeneous response profiles in individual neurons, the ensemble dynamics revealed a constant speed in the S1-S2 epoch and systematic modulations in the subsequent epochs.

Taken together, these findings indicated that the DMFC represented the animal's estimate of time, which got improved sequentially with more measurements presented. Both the behavioral data and the ensemble dynamics supported the internal model hypothesis.

## 3.3.2 Fear observational learning task

### 3.3.2.1 Mice can learn associations through observing others

To study the neural mechanism underlying observational fear learning, we designed a behavioral paradigm in which a mouse observes a demonstrator mouse undergo a classical fear conditioning through a transparent divider allowing for the observation of auditory, visual, and olfactory information (Figure 3.7a).

To further examine the learning effect, different task procedures were developed for different groups of the animals.

In the Experienced Observers (EO) group, animals first explored the shock floor side of the chamber, received a single shock, and were immediately placed on the plastic "safe" floor side of the chamber. On the other hand, another mouse was placed on the shock floor side, and 30 cue-shock pairings were delivered. We referred this animal as the demonstrator. On next day (i.e., Test day), observers were placed on the shock floor side to see if they acquire fear conditioning despite they never experienced the direct cue-shock pairing, by measuring freezing responses to the cue in the absence

of both shock and the demonstrator. It is found that EO mice demonstrated increased freezing in response to the cue during observational conditioning as well as on test day (Figure 3.7b-d and Supplementary Figure 3.4a, b).

In the Experienced Unpaired (EU) group, shock-experienced mice observed a demonstrator receiving shocks explicitly unpaired to the cue (Figure 3.7b). EU mice did not display conditioned freezing in response to the cue during conditioning or the test day (Figure 3.7b–d and Supplementary Figure 3.4a, b).

In the Experienced Solo (ES) group, animals had an initial shock experience but then observed the delivery of cues and shocks to an empty chamber (i.e., no demonstrator present). ES mice did not demonstrate conditioned responses to the cue during conditioning or the test day (Figure 3.7b-d and Supplementary Figure 3.4a, b).

As for animals in the Naive Observer (NO) group, they never directly experienced the shock, but observed that the demonstrator received cue-shock pairings. Similar to the initial EO group, NO mice showed increased freezing to the cue during conditioning (Figure 3.7b, c and Supplementary Figure 3.4a). However, NO mice did not show significantly higher freezing on test day than did "naive solo" (NS) mice that never experienced the shock but observed cues and shocks were delivered to an empty chamber (i.e., no demonstrator present), and NS mice did not display cue-elicited freezing during conditioning or test (Figure 3.7b–d and Supplementary Figure 3.4a, b). In addition to conditioned freezing responses, EO and NO mice both mimicked the demonstrator by displaying distress-related behaviors or interruption of grooming (Supplementary Figure 3.4c).

To rule out the possibility that EO mice learned the predictive value of the cue by context, we updated the behavioral paradigm in which animals were tested in a novel context instead on the Test day (Figure 3.7e). Results showed that these mice had significantly greater freezing during the cue relative to baseline (Figure 3.7f), indicating context-independent memory of the cue-shock association. We also examined avoidance of the shock floor in the EO, ES, NO, and NS groups before and after observational conditioning (Supplementary Figure 3.4d-f) and found that EO mice showed increased avoidance of the shock floor, whereas NO mice did not. Addi-

tionally, ES mice showed shock-floor avoidance, demonstrating that shock experience can drive place avoidance but is not sufficient to drive freezing during the cue as observed in the EO group (Supplementary Figure 3.4d-f). Finally, both EO and NO mice showed increased social interactions with the demonstrators after observational fear conditioning (Figure 3.7g and h). This suggested the increase in interaction was driven by the observation of distress in the demonstrator, potentially reflecting empathy (Burkett et al., 2016; Pisansky et al., 2017) or the social buffering of stress (Beery and Kaufer, 2015; Taylor, 1981).

Overall, these behavioral results suggest that the animals can acquire association between a stimulus and an outcome through observation, which was not driven by non-specific sensitization induced by the prior shock (Poulos et al., 2015).

#### 3.3.2.2 Heterogeneity in single-cell responses across learning

After characterizing behaviors, single-unit recordings (Supplementary Figure 3.4g-i) were performed in EO (paired) and EU (unpaired) mice to examine how ACC and BLA neurons encode information during observational learning (Figure 3.8a).

To examine information encoding, perperi-stimulus time histograms (PSTHs) of individual neurons are often clustered based on manually-defined criteria to characterize a brain region. Using this approach, we examined single-cell activities across observational learning. To compare responses to the cue before, during, and after conditioning, an initial block of trials (i.e., Habituation) where cues were delivered in the absence of shock delivery prior to conditioning was added (Figure 3.8a). We obtained trial-averaged responses of individual neurons in the ACC (Figure 3.8b) and the BLA (Figure 3.8c).

These responses were further classified based on their direction and magnitude to the cue (Figure 3.8d, e). The proportion of cue-responsive ACC neurons in the paired group (47.5%, n = 112/236 neurons; N = 16 mice) was similar to the unpaired group (46.7%, n = 56/120; N = 7)(Figure 3.9a), and the proportion in which further showing conditioning-dependent responses was not significantly different between both groups (paired group: 41.1%, n = 46/112; unpaired group: 35.7%, n = 20/56). Interestingly,

we found significantly more ACC neurons with an increased response magnitude in the paired group (74%, n = 34/46) relative to the unpaired group (40%, n = 12/20).

On the contrary, the proportion of BLA neurons showing conditioning-dependent changes was significantly greater in the paired group (38.3%, n = 18/47) compared to the unpaired group (5.3%, n = 2/38) (Figure 3.9b), whereas the proportion of cue-responsive BLA neurons in the paired group (42.7%, n = 63/110; N = 7) was again similar to the unpaired group (36.9%, n = 38/103; N = 7).

Taken together, the BLA single-cell encoding of the cue reflects the implicit cue value acquired by observation, whereas the ACC encodes the cue-shock contingency through the magnitude of cue-evoked responses (Figure 3.9f) as well as its baseline firing rates (Supplementary Figure 3.5a).

However, do these analyses faithfully reflect the coding principles of observational learning in both the ACC and the BLA? As shown in the Figure 3.8, both the ACC and the BLA activities are heterogeneous between neurons. However, both areas were characterized based on proportions of cells that were classified based on their average rates around the cue, and thus the difference in the temporal dynamics (e.g., a large sharp transient vs. a small persistent potentiation) were not captured. Furthermore, the hetereogenous dynamics across observational learning were ignored.

#### 3.3.2.3 State-space model reveals that the ACC has an overall earlier rate-change than the BLA across learning

Given potential caveats in previous PSTH-based analyses, we proposed to characterize individual neurons across observational learning using a latent estimation approach based on point process, i.e., a state-space model (see Materials and Methods). In this case, the underlying firing rate of neurons are more accurately estimated, such that changes in their responses across trials can be identified in a statistically robust manner (Barbieri et al., 2004; Czanner et al., 2008; Smith and Brown, 2003)

We applied this model to examine temporal dynamics of single neurons across observational conditioning in the ACC and the BLA (Figure 3.10a-d), and identified the trial when each neuron began to encode the cue value, i.e., "the rate-change trial"

of the neuron, (Figure 3.10a, b and Supplementary Figure 3.6a, b). Specifically, the rate-change trial is the first trial during Observational Conditioning whose firing rate and its subsequent trial both differ significantly from the average rate in Habituation. We found that the ACC had a significantly earlier distribution of rate-change trials compared to the BLA (Figure 3.10c, d). We also found similar changes in behavior of these animals around the cue presentation across trials (Figure 3.10e).

These results led us to hypothesize that the observational learning initiates in the ACC whereas the learned value is passed to the BLA afterwards. Alternatively, it is possible that the ACC simulates the observed behavior around the cue presentation and signals a context signal to the BLA for value assignment. Unfortunately, these two hypotheses might be difficult to tease apart. For example, if inhibiting the ACC→BLA projections repairs the ability to learn an association by observation whereas the ability to acquire a new association through direct experience is unaffected, it still remains unknown whether the ACC only encodes social context for observational learning, because value assignment mechanism might be different between direct experience and mental simulation. Another challenge arises with the assumption that information is encoded in individual cells independently through significant changes in their firing-rates, which ignores population coding.

#### 3.3.2.4 Ensemble dynamics reveals that the ACC transmits social information to the BLA whereas the BLA forms the association between cue and shock

Given challenges in interpreting hetereogeneity in single cells, we proposed to analyze the dynamics of the entire neural population using another state-space method (see Materials and Methods). With this approach, diverse activities of neurons were projected to a low-dimensional manifold (Churchland et al., 2012; Cueva et al., 2020; Mante et al., 2013; Petreska et al., 2011; Yu et al., 2005, 2009), allowing visualization of the ensemble dynamics across time (i.e., neural trajectories) within a single trial.

We compared the neural ensemble dynamics in the ACC and the BLA during habituation and observational conditioning, for both paired (EO) and unpaired (EU)

groups (Figure 3.11a, b). It is found that the ACC trajectory during habituation shared a similar shape to the trajectory during observational conditioning, regardless of the cue-shock contingency (Figure 3.11a). The distance between both trajectories remained unchanged across time from the baseline to after cue presentation, suggesting that the neural trajectory during observational conditioining is basically a translation of the neural trajectory during habituation (Figure 3.11c). Moreover, the results of the EO group were consistent with the EU group, suggesting that the ACC populations did not represent value assignment in observational learning.

On the contrary, the BLA trajectory during observational conditioning in the EO group revealed a distinct pattern relative to the trajectory during habituation (Figure 3.11b). The distance between two trajectories changed significantly from the baseline to cue presentation in the EO group (Figure 3.11d). However, these patterns were not observed in the EU group. Instead, its BLA trajectory during observational conditioning shared a similar geometric translation of the trajectory during habituation (Figure 3.11c, d). The distinction between the paired and unpaired groups further indicated the observed association was learned in the BLA.

To assess what memories were stored in the ACC and BLA, we also examined the neural ensemble dynamics on Test day, when no demonstrator was present and no shocks were delivered (Supplementary Figure 3.7). We found that the BLA ensemble response changed significantly upon cue presentation in the EO group but not in the EU group, whereas the ACC ensemble response remained almost unchanged upon cue presentation. Similar findings also reflected in the single-cell analysis. The proportion of ACC neurons being cue-responsive on Test day in the EO group (30%) was similar to the EU group (29%), but the proportion of BLA neurons being cue-responsive on Test day was significantly greater in the EO group (31%) than in the EU group (17%). (Supplementary Figure 3.5b). These results indicated that the cue value was represented in the BLA instead of the ACC

Consistent with these results, the baseline firing rates of the ACC changed significantly more than the BLA across observational learning (Supplementary Figure 3.5a), suggesting a major force driving the geometric translation between two manifolds.

Together, these data prompted us to revise our initial hypothesis that the ACC neurons were representing the association between the cue and shock to the demonstrator earlier than the BLA neurons. Instead, the baseline changes suggested that the ACC neurons are rapidly changing basal firing rates in the face of demonstrator distress, thereby potentiating responses to salient stimuli, but that the acquisition of the predictive value of the cue is mediated by the BLA. Thus, the ACC transmits socially-derived information to the BLA during the cue presentation, allowing the BLA to form an association between the cue and shock to the demonstrator.

#### 3.3.2.5 Optogenetic manipulation shows that the ACC to BLA pathway is necessary for observational learning

To test the hypothesis derived from above analyses, we expressed NpHR bilaterally in the ACC of the EO mice and placed optical fibers over the BLA in both hemispheres (Figure 3.12e) to inhibit the ACC→BLA pathway during cue presentations throughout observational fear conditioning (Figure 3.12a).

Results showed that mice receiving this manipulation showed no differences in freezing during the training session on day 1 (Figure 3.12b, *left*), but when tested the next day, in the absence of photoinhibition or a demonstrator, they showed less freezing compared to eYFP control mice (Figure 3.12b, *right*). On the other hand, when the ACC→BLA pathway was inhibited only upon expression of observational fear memory, during cue presentations on Test day (Figure 3.12c), we found no changes in the freezing behavior (Figure 3.12d). Importantly, when inhibiting the ACC→BLA pathway during a classical fear conditioning paradigm, i.e., associations were formed through direct experience, there was no difference in freezing behavior between NpHR mice and eYFP control mice (Figure 3.12f).

Taken together, these experiments demonstrate that the representation of social context in the ACC transferred to the BLA is necessary for observational learning, but not for subsequent memory expression (Figure 3.12a-e), suggesting that the the ACC→BLA pathway is only required for observational learning, but not associative learning in general.

# Figures

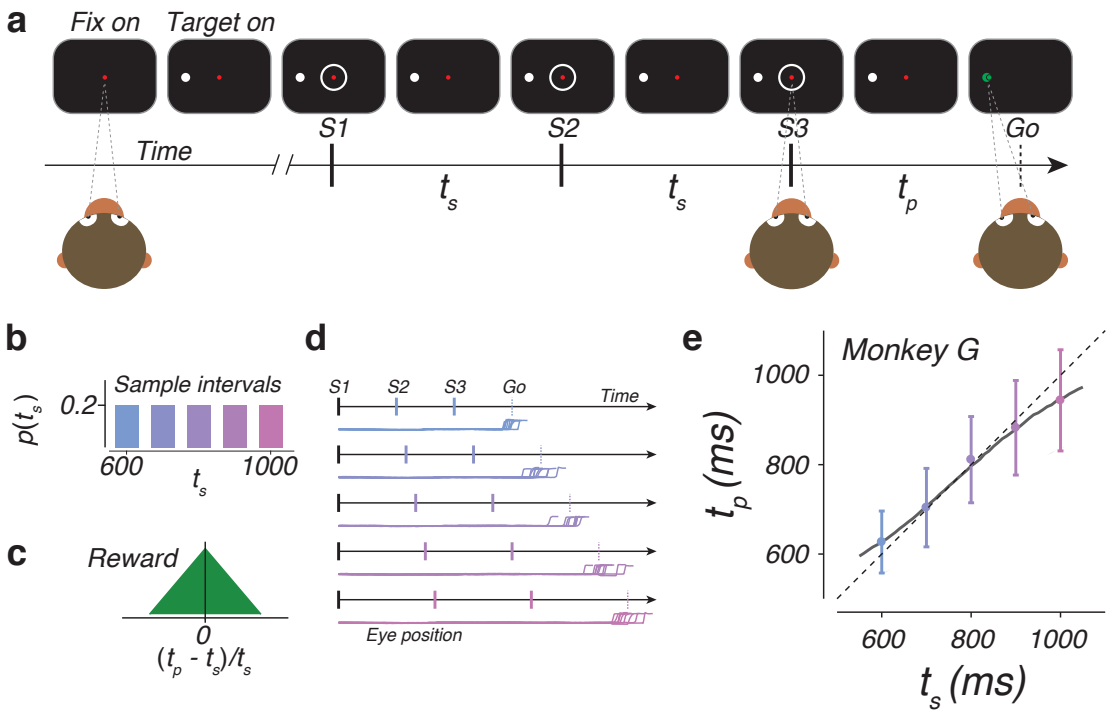



**Figure 3.1: Behavioral characterization for time reproduction.** (**a**) The task schematic diagram. The monkey needs to fixate a central point (*Fix on*). After the target is presented (*Target on*), three isochronous flashed circles (S1, S2 and S3) are presented. The animal needs to measure the sample interval $t_s$ between consecutive flashes and produce a matched interval $t_p$ after S3 by making a saccade to the target. If responses are sufficiently accurate, a juice reward is delivered along with a green dot on screen. Details see Materials and Methods. (**b**) The sample interval $t_s$ is drawn from a discrete unifrom distribution with values ranging from 600 to 1000 msec. (**c**) Reward schedule for production time $t_p$. Maximal reward was delivered when $t_p = t_s$ and decreased linearly to 0 with increasing relative error due to scalar variability. (**d**) Example eye saccade behavior for time reproduction from 10 randomly selected trials for each sample interval $t_s$. (**e**) Produced intervals ($t_p$) as a function of sample interval ($t_s$). Colored circles and bars showed the corresponding averages and stand deviations for each $t_s$. Smooth black curve plots a model fit.

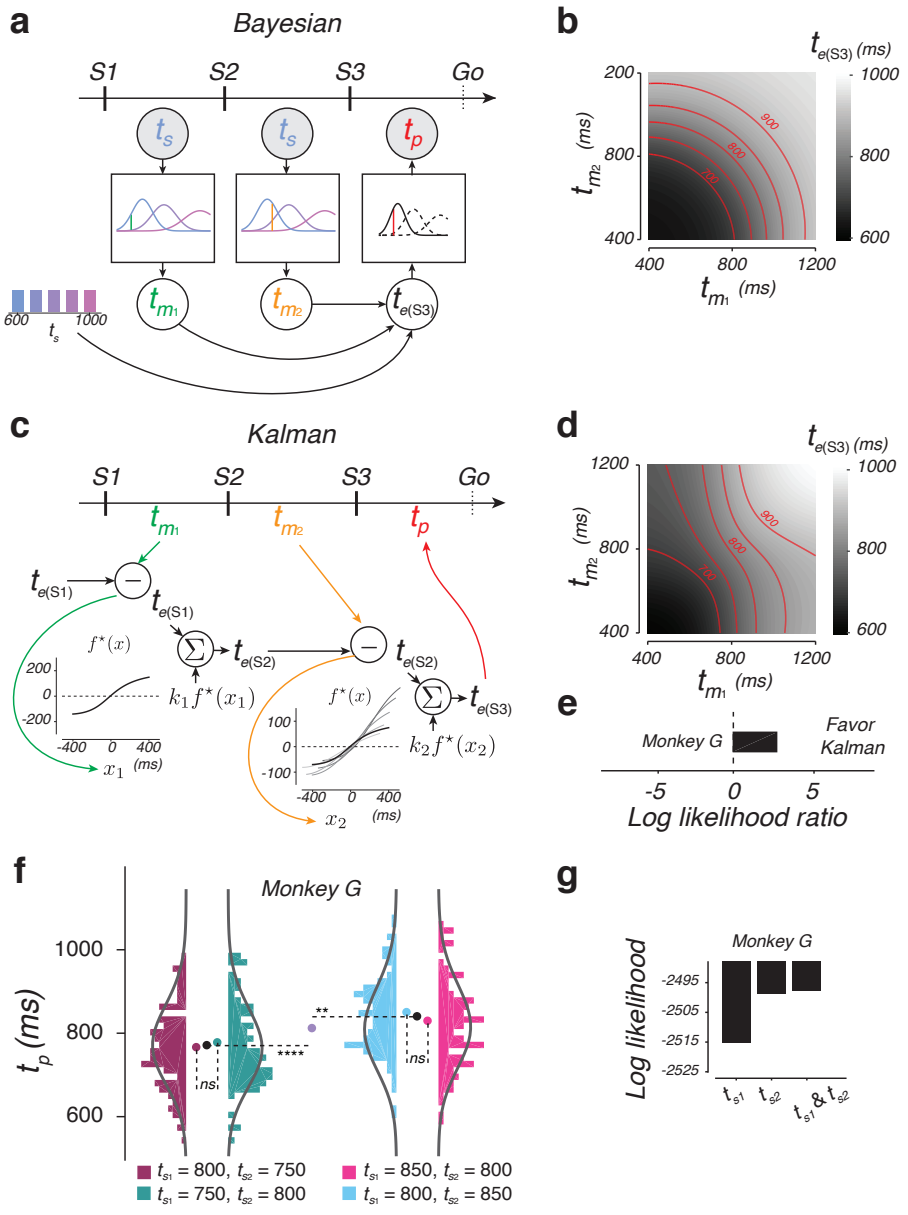



**Figure 3.2: Integration of two sequential timing cues in the Bayesian and Kalman models.** (**a**) The Bayesian model. A Bayes-least-squares (BLS) estimator is used to compute an estimate $t_{e(S3)}$ by integrating the prior with measurements in both epochs (S1-S2 and S2-S3), by assuming two measurements were conditionally independent. A production interval $t_p$ is generated with motor noise from $t_{e(S3)}$. Both measurement and production noise are modeled as Gaussian with standard deviations that scale $t_s$ with Weber constants $w_m$ and $w_p$ respectively. The green, orange, and red vertical lines in the boxes are corresponding measurements and produced time for an example trial. Details see Materials and Methods. (**b**) Plot of the final estimate $t_{e(S3)}$ as a function of two measurements $t_{m_1}$ and $t_{m_2}$. The red lines show various combinations that lead to the same estimation. The estimated values are colored in grayscale.

**Figure 3.2** ***(previous page)*****: Integration of two sequential timing cues in the Bayesian and Kalman models.** (**c**) The Kalman filter model. At each cue (S1, S2, and S3), an interval estimate is derived. At initial stage, $t_{e(\mathrm{S1})}$ is set as the average of the prior distribution because there is no other available information before S1 presents. When S2 is presented, the model updates its estimate with a nonlinear operation on the error between original estimate $t_{e(\mathrm{S1})}$ and new measurement $t_{m_1}$ from the S1-S2 epoch, resulting in $t_{e(\mathrm{S2})}$. When S3 is presented, the model further updates its estimate based on the error between the previous estimate $t_{e(\mathrm{S2})}$ and new measurement $t_{m_2}$ from the S1-S2 epoch, resulting in $t_{e(\mathrm{S3})}$. Produced time $t_p$ is corrupted with motor noise from $t_e$. Details see Materials and Methods. (**d**) Plot of the final estimate $t_e$ as a function of two measurements $t_{m_1}$ and $t_{m_2}$, as in **b**. (**e**) Ratio of the log-likelihood of the data given the Kalman model to the log-likelihood given the Bayesian model (n = 415 trials). (**f**) Behavior in the cue-conflict experiments. Each histogram shows a distribution of produced time $t_p$ in different cue-conflict conditions (brown, green, cyan and magenta), where there are 117, 83, 111, and 104 trials respectively. Pairs of neighboring histograms corresponds to the same average $t_s$ (e.g., $t_{s_1} = 800$ ms and $t_{s_2} = 750$ ms vs. $t_{s_1} = 750$ ms and $t_{s_2} = 800$ ms). Colored circles represent the average $t_p$ across conditions with the same $t_s$. Black circle represents the average $t_p$ across conditions with same average $t_s$, i.e., either 775 or 825. Purple circle represents the average $t_p$ for trials when two measurements are identical, i.e., 800. Two-sided t-test is performed (ns: not significant; $^{**}p < 0.01$; $^{****}p < 0.0001$). (**g**) Different log-likelihoods based on assuming either only the measurement $t_{m_1}$ from S1-S2 epoch is used, or only the measurement $t_{m_2}$ from S2-S3 epoch is used, or both are used in a Kalman model (n = 415 trials).

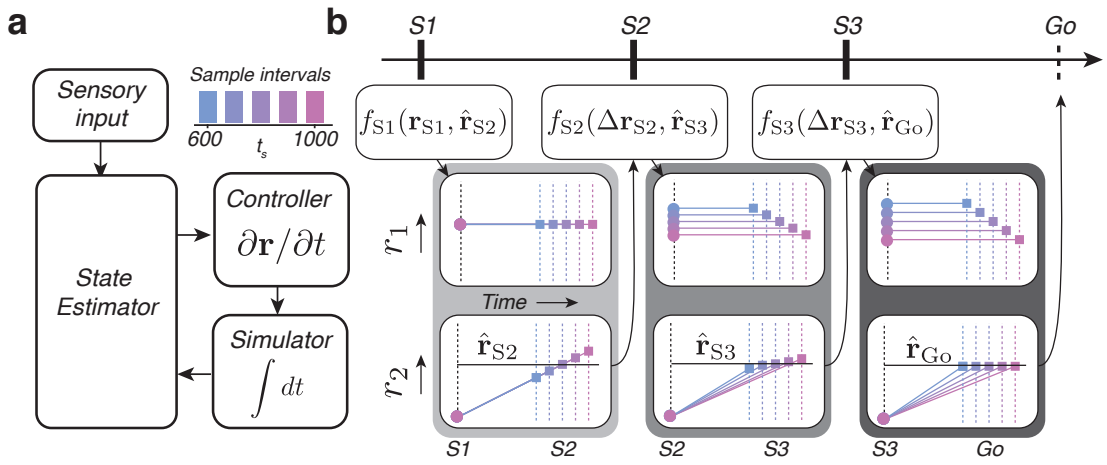



**Figure 3.3: Hypothesis on neural activities based on an internal model.** (**a**) Hypothesis of an internal model. The model comprises a controller driving the output based on the estimated state from a state estimator, a simulator predicting outputs, and a state estimator integrating sensory inputs with predictions from simulation. (**b**) Hypothesis of how two populations $r_1$ and $r_2$ can implement an internal model. With a given input $t_s$ drawn from a uniform distribution, one population $r_1$ may act as a controller sending signals to regulate the dynamics of the other population $r_2$, whereas $r_2$ may act as a simulator integrating the speed signals from $r_1$ to generate a state estimate. Specifically, after a cue is first presented, $r_1$ signals a constant speed to $r_2$, and path integration leads to ramping activities of $r_2$ during the S1-S2 epoch. Deviations from the expected state at S2 are used to update $r_1$ signals, which further adjusted the slope of ramping activities of $r_2$ in the S2-S3 epoch. The same process is repeated in the S3-Go epoch.

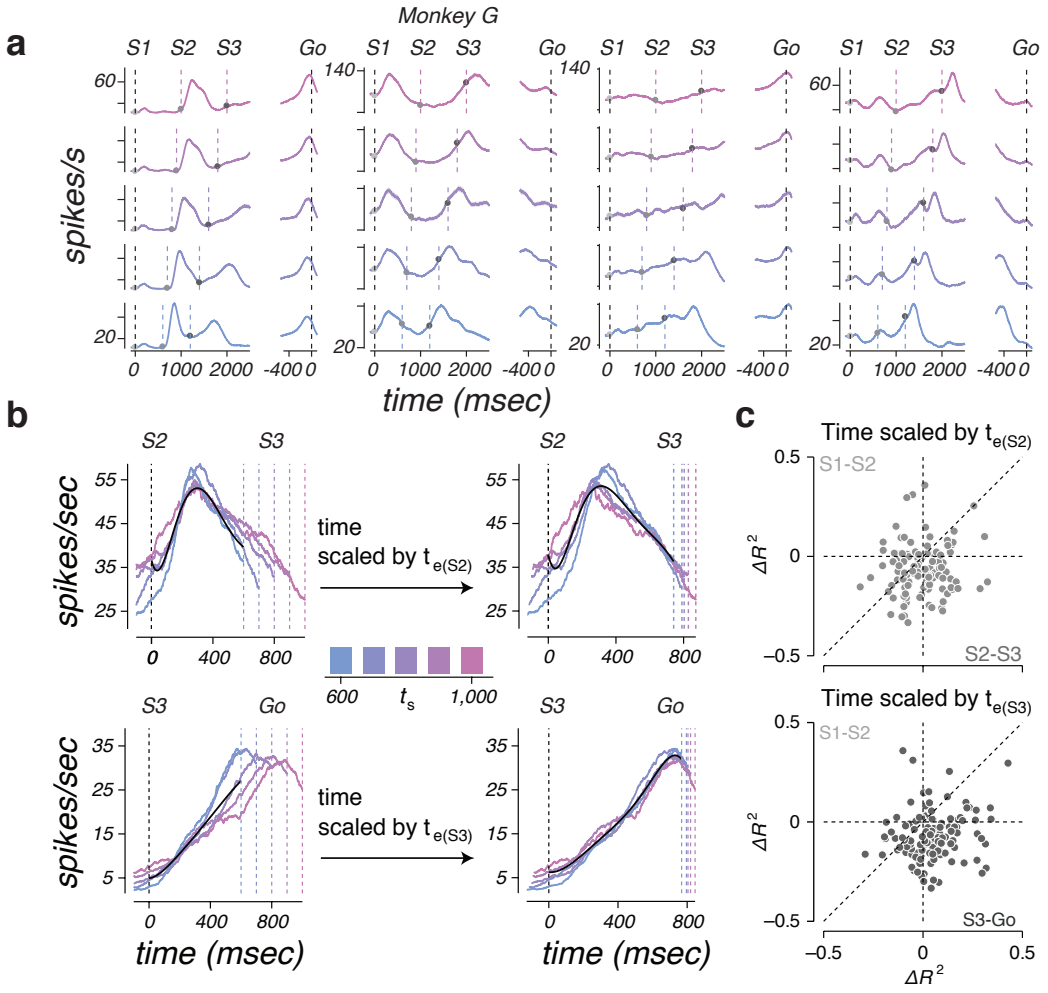



**Figure 3.4: Temporal scaling of hetereogenous single-cell activities in the DMFC.** (**a**) Average response profiles from four example neurons in the DMFC. Responses are aligned to S1 and Go sorted according to $t_s$ (gradually-colored). Gray circles indicate standard errors at each flash cue (S1, S2, and S3). Vertical dash lines indicate flash (S1, S2, and S3) and saccade (Go) times. (**b**) Scaled temporal responses from two example neurons in the DMFC. Top: (*left*) Average firing rates aligned to S2 (black dash line) according to different $t_s$ (gradually-colored) from an example neuron. Black lines represent the polynomial fit. Colored dash lines represent the time of S3. (*right*) Firing rates are temporally scaled by the animal's estimate at S2 inferred from the Kalman model, i.e., $t_{e(\mathrm{S2})}$. Bottom: (*left*) Average firing rates aligned to S3 according to different $t_s$ from another example neuron. Colored dash lines represent the time of Go. (*right*) Firing rates are temporally scaled by $t_{e(\mathrm{S3})}$. (**c**) Scatter plot of the change in goodness of fit ($\Delta R^2$) using a sixth-order polynomial to the scaled versus unscaled data across neurons. Top: $\Delta R^2$ in the S2-S3 epoch compared with the S1-S2 epoch. Bottom: $\Delta R^2$ in the S3-Go epoch compared with the S1-S2 epoch. Vertical and horizontal broken lines indicate no change in goodness of fit, and the diagonal in the unity line.

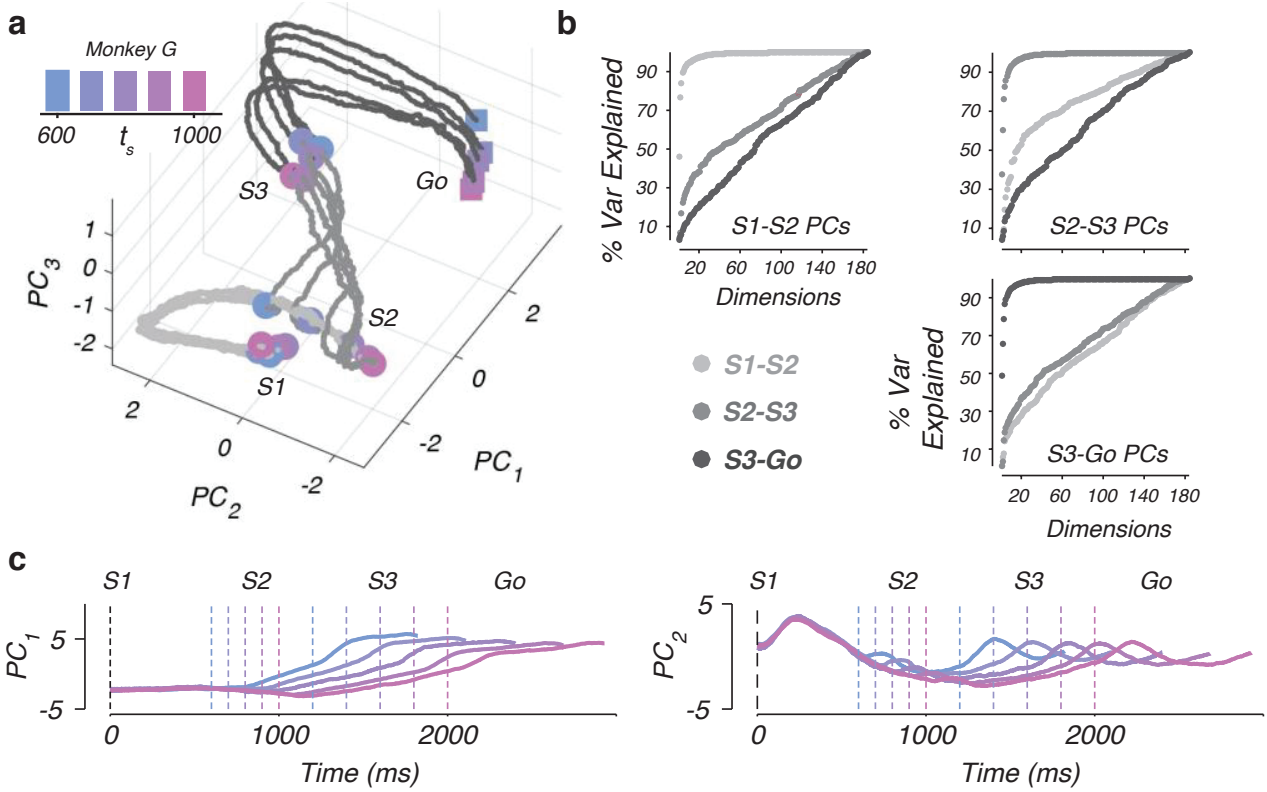



**Figure 3.5: Temporal scaling of ensemble trajectories in the subspace spanned by PCs separating different epochs (S1-S2, S2-S3, and S3-Go).** (**a**) Neural trajectories plotted in a subspace spanned by top 3 principal components (PCs) from 118 neurons across time, from S1 to Go. Circles indicate states at S1, S2 or S3 and squares denote states at 100 ms before the expected $t_p$ matching each $t_s$ (colored). (**b**) Percentage of variance explained as a function of the number of PCs developed independently for each epoch. Left: Reconstruction of data from the PCs derived from activities in the S1-S2 epoch. Right: Same as Left with the PCs derived from activities in the S2-S3 (*top*) and S3-Go (*bottom*) epochs, respectively. S1-S2 epoch is colored in light-gray, S2-S3 epoch is colored in medium-gray, and S3-Go epoch is colored in dark-gray. (**c**) Scaled average ensemble trajectories for each sample interval $t_s$ are aligned to S1 according to $t_s$ (gradually-colored). Vertical dash lines indicate flash (S1, S2, and S3) and saccade (Go) times.

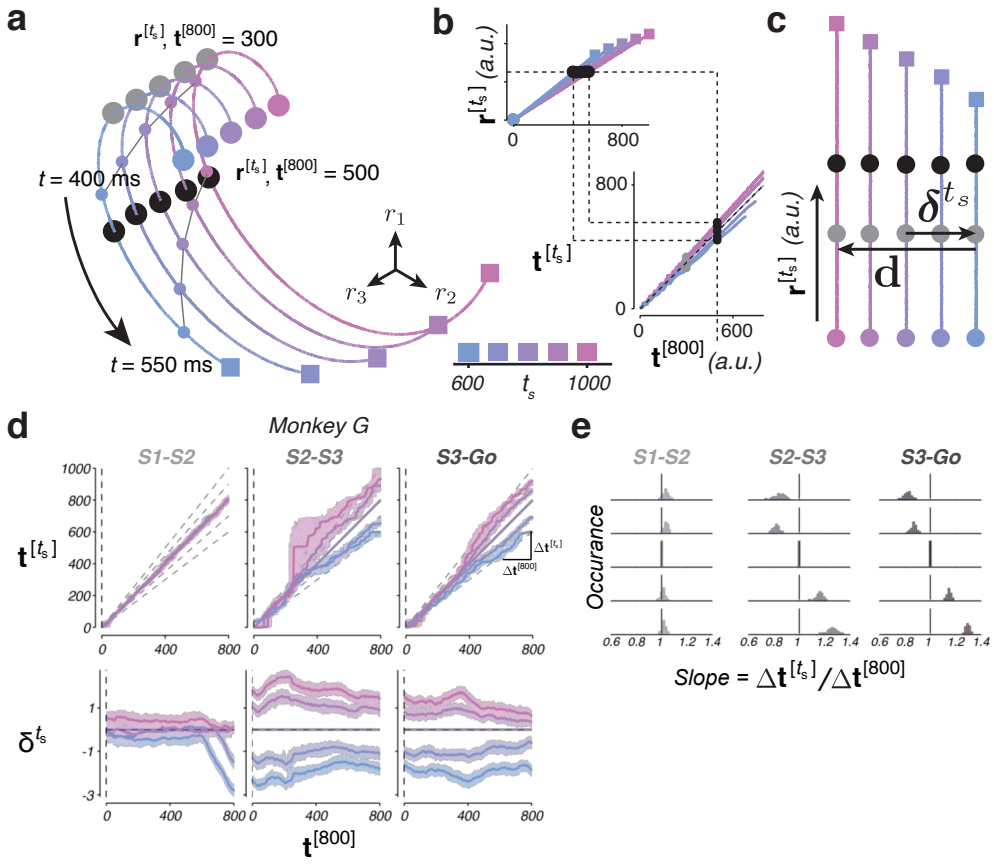



**Figure 3.6: Speed and distance between ensemble trajectories across time.** (**a-c**) Schematic diagram of speed and distance analysis on ensemble trajectories. (**a**) Cartoon illustration of neural trajectories in a 3D subspace. Gray and black circles are the set of states $\mathbf{r}^{[t_s]}$ representing closest points to $\mathbf{r}^{[800]}$ at 300 msec (gray) and 500 msec (black), respectively. Small circles connected by thin gray lines are states on the trajectories at 400 msec and 550 msec after S1. (**b**) Top: States $\mathbf{r}^{[t_s]}$ for each trial-averaged trajectory as a function of $\mathbf{t}^{[800]}$. The time required to reach $\mathbf{r}^{[t_s]}$ is denoted as $\mathbf{t}^{[t_s]}$. Bottom: By mapping $\mathbf{t}^{[t_s]}$ to $\mathbf{t}^{[800]}$, we can get the speed along each trajectory from the slope. Black points represent $\mathbf{t}^{[t_s]}$ for $\mathbf{t}^{[800]} = 500$, and gray points represent $\mathbf{t}^{[t_s]}$ for $\mathbf{t}^{[800]} = 300$. (**c**) Cartoon illustration of relative positions of trajectories corresponding to different $t_s$. $\delta^{t_s}$ represents distance between states on other trajectories, i.e., $\mathbf{r}^{[t_s]}$, and states on the reference trajectory, i.e., $\mathbf{t}^{[800]}$. Black and gray circles correspond to the neighboring states, as in **a**.

**Figure 3.6** ***(previous page)*****:** (**d-e**) Relative speed and distance between neural trajectories. (**d**) Top: Speed of each neural trajectory relative to the reference trajectory, by comparing the elapsed time $\mathbf{t}^{[800]}$ and the elapsed time to reach $\mathbf{r}^{[t_s]}$, i.e., $\mathbf{t}^{[t_s]}$. The unity line indicates the same speed to the reference trajectory. Dashed lines represent expected relationship between $\mathbf{t}^{[t_s]}$ and $\mathbf{t}^{[800]}$ assuming an observer has perfect knowledge of $t_s$. Bottom: Distance between neighboring states ($\delta^{t_s}$) as a function of elapsed time $\mathbf{t}^{[800]}$. The horizontal dashed line indicates a complete overlap to the reference trajectory. Shading areas indicate the median ± 95% CIs (n=100 bootstrap samples). (**e**) Histograms of speed distribution of trial-by-trial trajectories. Speed is estimated by assuming a linear relationship between $\mathbf{t}^{[t_s]}$ and $\mathbf{t}^{[800]}$.

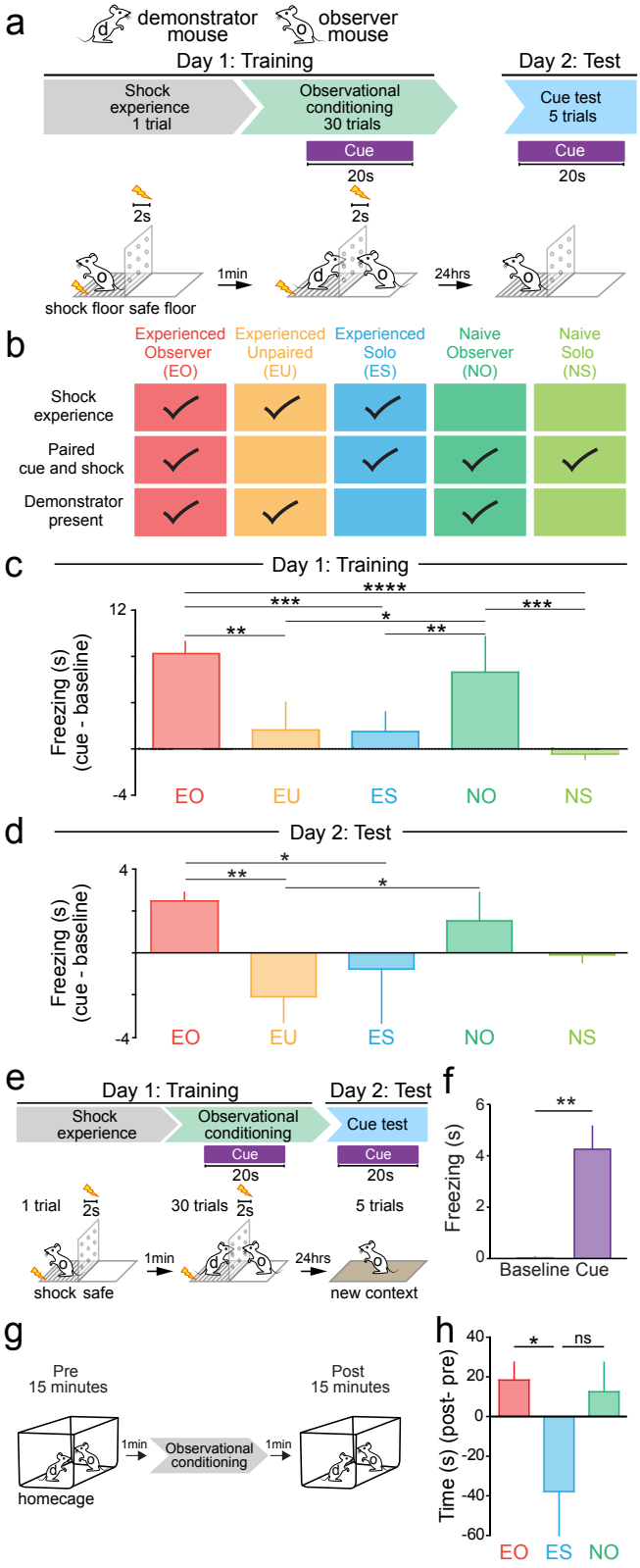



**Figure 3.7: Behavioral characterizations for fear observational learning.**

**Figure 3.7** ***(previous page)*:** (**a**) Observational fear conditioning paradigm. (**b**) Conditioning paradigms for all behavioral groups (EO, N = 7; EU, N = 5; ES, N = 6; NO, N = 7; NS, N = 4 mice). (**c**) On day 1: training, EO and NO mice showed a significantly higher difference in freezing between cue and baseline epochs (cue - baseline) than did EU, ES, and NS mice (one-way ANOVA, $F_{(4,24)} = 13.76$). (**d**) On day 2: test, EO mice showed significantly higher freezing (cue - baseline) than did EU and ES mice (one-way ANOVA, $F_{(4,24)} = 5.964$, $p < 0.01$). (**e**) Modified observational conditioning paradigm to test for context independent cue learning in EO mice. (**f**) EO mice showed significantly higher freezing during cue presentation in a novel context (two-tailed, unpaired Student's t test, t = 4.535, degrees of freedom [df] = 10). (**g**) Testing for affiliative interactions between demonstrator and observer mice before and after observational conditioning. (**h**) Time interacting with the demonstrator after observational conditioning was statistically higher for EO mice (one-way ANOVA, $F_{(2,16)} = 3.779$, $p < 0.05$). All error bars indicate SEM. Significance using Bonferroni post hoc analysis is marked as *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

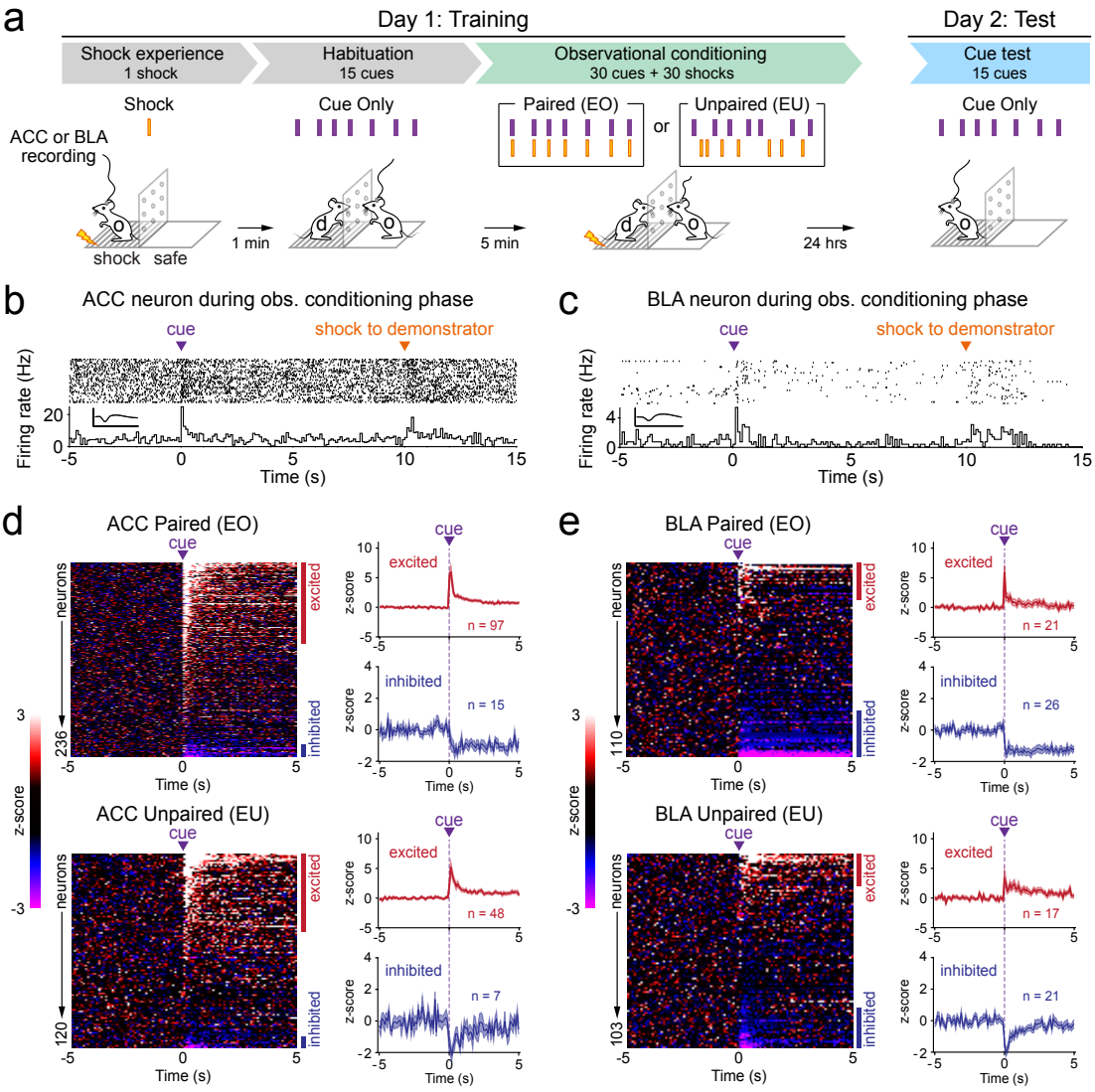



**Figure 3.8: Single-cell responses within a trial in the ACC and the BLA.** (**a**) Observational fear conditioning paradigm used for in vivo single-unit recordings in the ACC (paired group, N = 16; unpaired group, N = 7 mice) or BLA (paired group, N = 6; unpaired group, N = 6 mice). (**b** and **c**) Representative ACC (**b**) and BLA (**c**) neuron responses to cue and shock delivery during paired observational conditioning. Raster plots depict neural spikes (1 trial per row) and each peri-stimulus time histogram (PSTH) depicts the average firing rate across all trials, relative to cue onset (100-ms bins). Insets show the average waveform recorded for each neuron (y axis, 200 $\mu V$; x axis, 1 ms). (**d** and **e**) Cue responses for paired and unpaired groups in ACC (**d**) and BLA (**e**). Heatmap rows represent the Z-score-transformed average PSTH for individual neurons, and columns represent time bins relative to cue onset (100-ms width). Blue and red bars indicate statistically significant cue-responsive cells. Plots to the right show the average Z score responses for cue-excited and cue-inhibited cells.

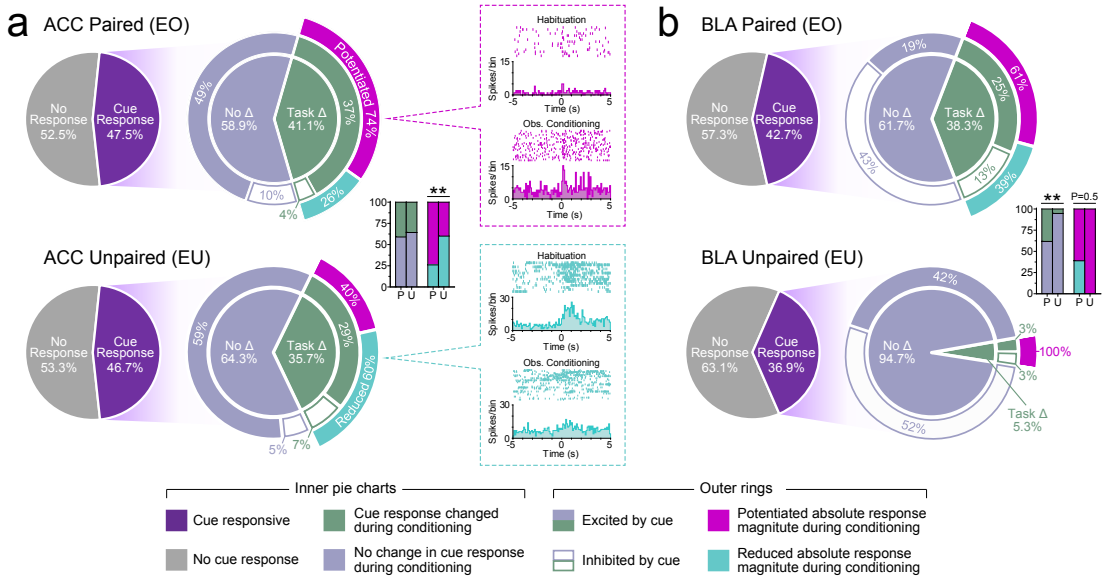



**Figure 3.9: Hetereogenous cue-evoked responses in the ACC and the BLA.** (**a**) ACC subpopulations clustered by PSTH. Task-modulated neurons in the paired group showed a greater proportion of potentiated responses during conditioning than the unpaired group (bar graph inset; $\chi^2 = 6.93$, **: $p < 0.01$). On the right, each PSTH shows example ACC neurons with training-induced potentiated or reduced cue responses. (**b**) BLA subpopulations clustered by PSTH. A significantly greater proportion of cue-responsive neurons showed task-modulated responses in the paired group (P) than in the unpaired group (U) (bar graph inset; $\chi^2 = 8.27$, **: $p < 0.01$).

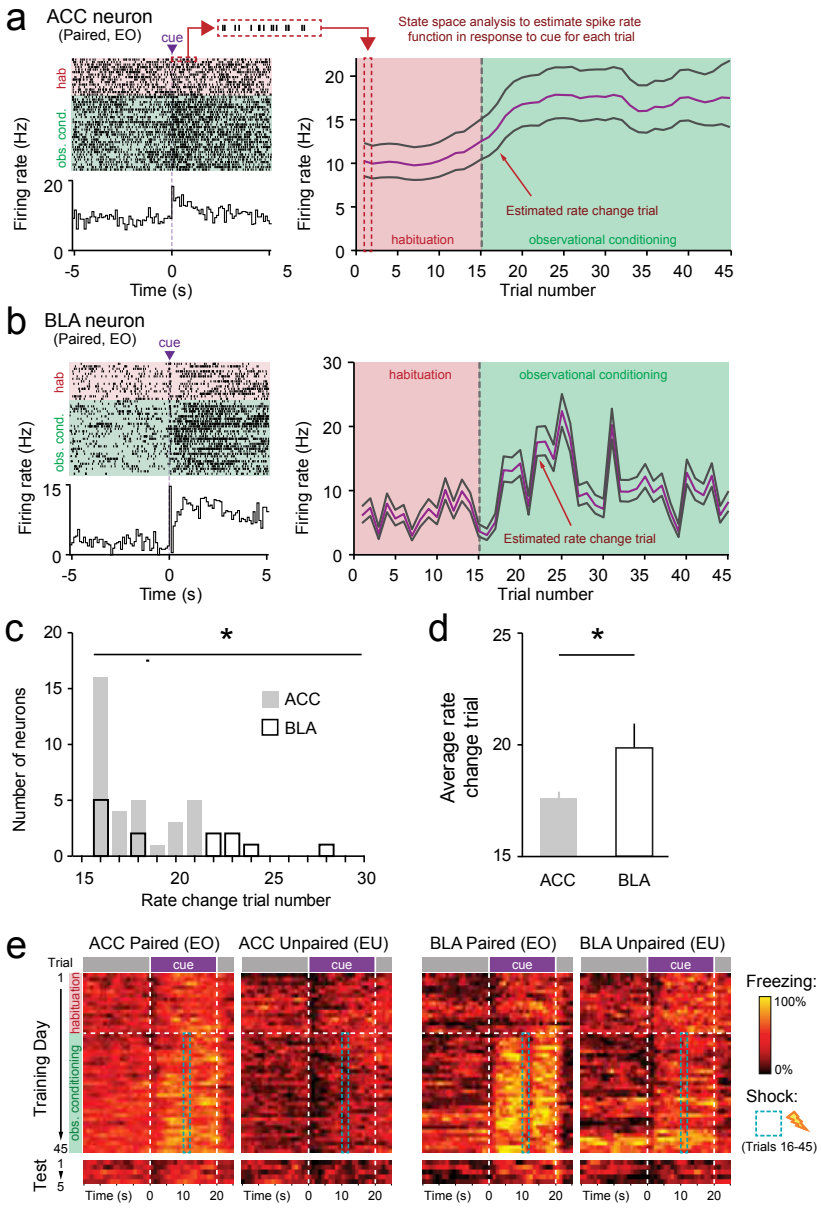



**Figure 3.10: Earlier rate-change in the ACC across observational learning.** (**a** and **b**) Raster and PSTH (100-ms bins) of an ACC (**a**) and BLA (**b**) neuron identified as having a significant change in cue response during conditioning. State-space analysis provides a probabilistic estimate of the trial at which the neuron undergoes a rate change. (**c**) The distribution of rate change trials calculated by state-space analysis of conditioning-dependent neurons was significantly earlier in ACC than BLA neurons (Kolmogorov-Smirnov test, *: $p < 0.05$). (**d**) The average rate change trial of neurons in the ACC was earlier than those in the BLA (two-tailed, unpaired Student's t test, t = 2.622, df = 45, *: $p < 0.05$). (**e**) Behavioral rasters (1-s bins) of the average freezing for all paired (EO) and unpaired (EU) mice across day 1 training and day 2 test in both ACC and BLA groups.

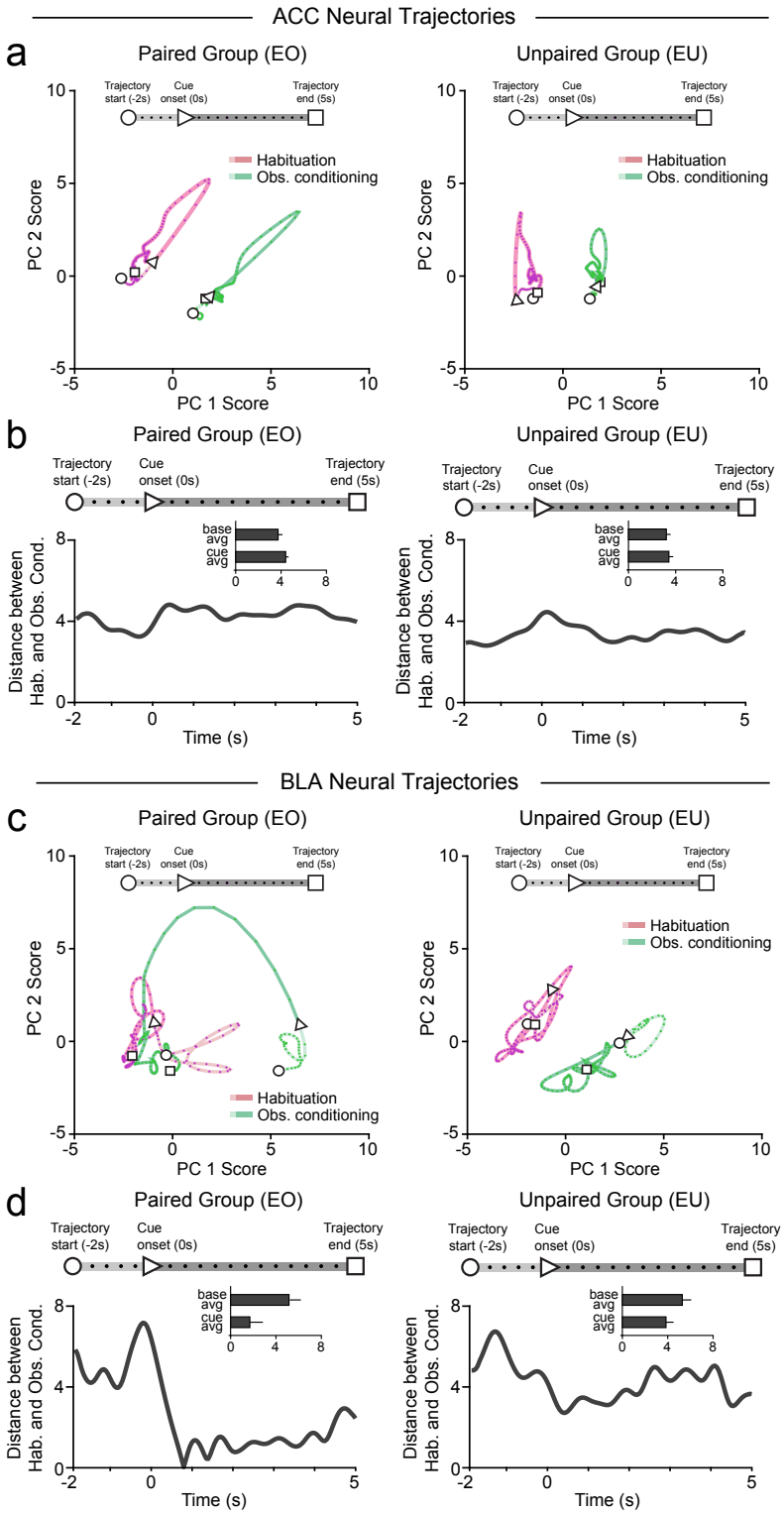



**Figure 3.11: Observational learning induces distinct ensemble trajectories.**

**Figure 3.11** ***(previous page)*:** (**a** and **b**) Neural ensemble dynamics in ACC and BLA across habituation and conditioning trials. Trial-averaged neural trajectories projected on a 2D space formed by first (PC1) and second (PC2) principal components for ACC neurons (**a**) (paired, n = 201 neurons, N = 12 mice; unpaired n = 93 neurons, N = 7 mice) and BLA neurons (**b**) (paired, n = 106, N = 6 mice; unpaired n = 97 neurons, N = 6 mice). Dots on the trajectories represent timestamps (50 ms). (**c** and **d**) Calculated Euclidean distance between trajectory for habituation and trajectory for observational conditioning in paired and unpaired mice in the ACC (**c**) and BLA (**d**) plotted as distance across time (-2 s to +5 s from CS onset). Insets show averaged values for baseline and cue period. The distance between baseline and cue epochs in the BLA paired group was significantly different from other groups (Pearson's Chi-square test $\chi^2 = 4.953$, df = 1, *: $p < 0.05$).

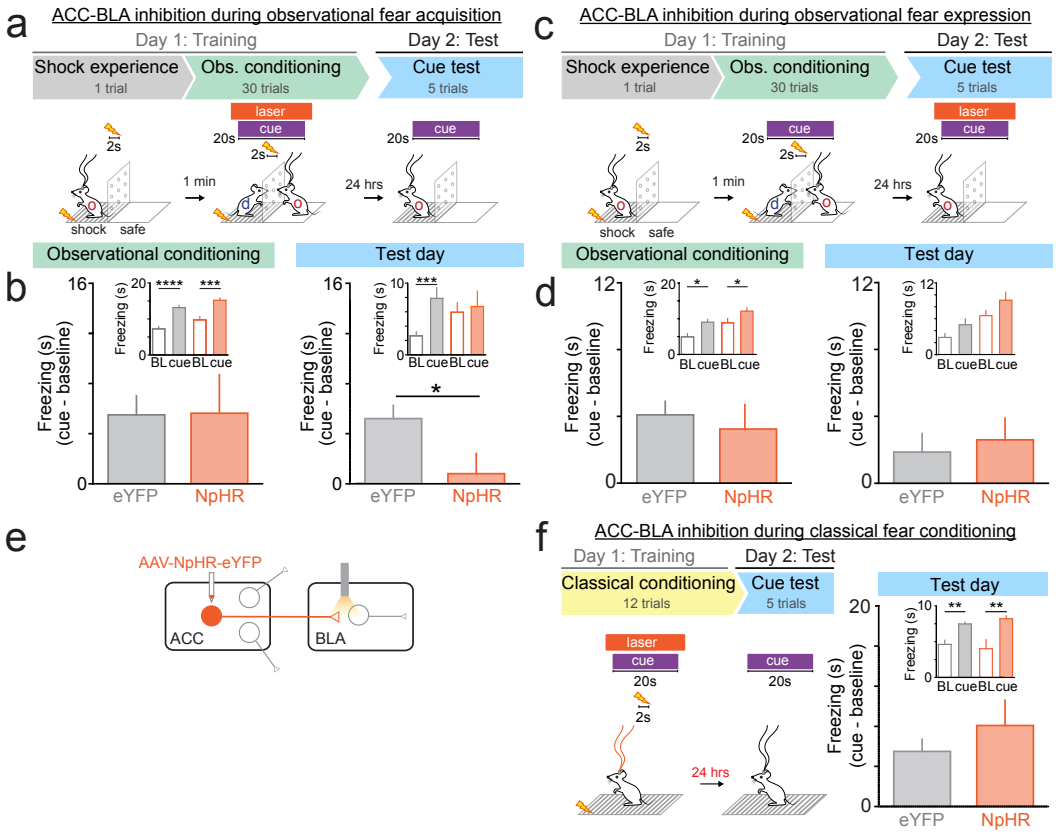
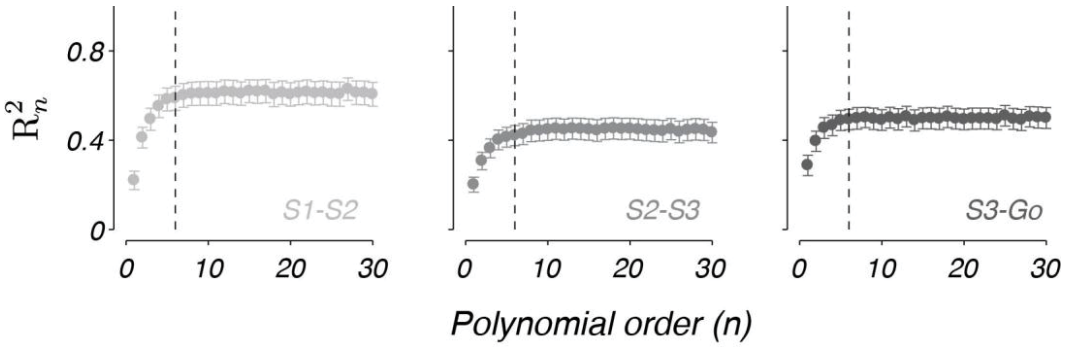


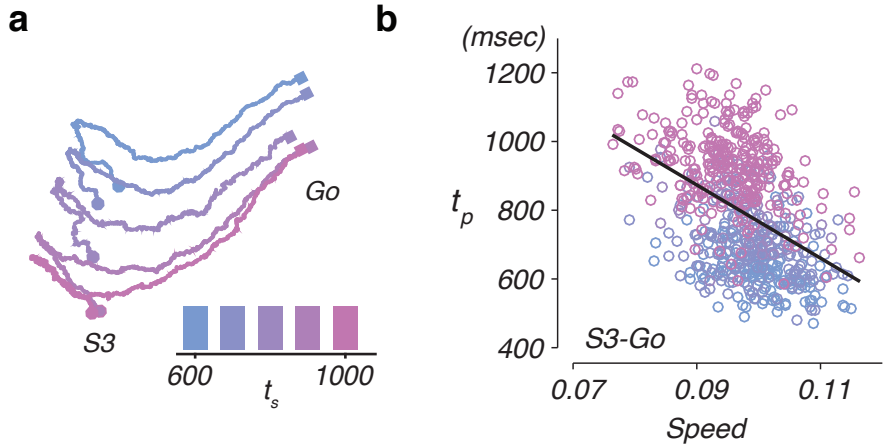

**Figure 3.12: Photoinhibition of ACC→BLA impairs observational learning.** (**a**) Behavioral and light delivery paradigm for inhibition of ACC→BLA circuit during cue presentations during acquisition (Day 1) of observational conditioning. (**b**) During observational conditioning, there were no significant differences in freezing between NpHR (N = 7) and eYFP (N = 12) mice (unpaired, two-tailed t test, t = 0.0785, df = 17, $p = 0.9383$). However, on Test day, cue driven freezing was impaired in NpHR compared to eYFP mice (unpaired, two-tailed, t test, t = 2.378, df = 17). Insets show cue and baseline (20 s prior to cue onset) freezing values for observational conditioning and test day (BL = baseline; Observational conditioning: two-way repeated-measures (RM) ANOVA, group effect, $F_{(1,17)} = 8.286$, $p < 0.05$, epoch effect, $F_{(1,17)} = 66.26$, $p < 0.0001$, group X epoch interaction, $F_{(1,17)} = 0.0829$, $p = 0.7769$; Test day: two-way RM ANOVA, group effect, $F_{(1,17)} = 0.3596$, $p = 0.5566$, epoch effect, $F_{(1,17)} = 10.64$, $p < 0.01$, group X epoch interaction, $F_{(1,17)} = 5.657$, $p < 0.05$).

**Figure 3.12** ***(previous page)*:** (**c**) Behavioral and light delivery paradigm for inhibition of ACC→BLA circuit during cue presentations during expression (Day 2) of observational conditioning. (**d**) There were no significant differences in cue driven freezing between NpHR (N = 9) and eYFP (N = 8) mice during observational conditioning (unpaired, two-tailed, t test, t = 0.4916, df = 15, $p = 0.6301$) or Day2: Test (unpaired, two-tailed, t test, t = 0.2615, df = 15, $p = 0.7973$). Insets show cue and baseline (20 s prior to cue) freezing values during conditioning and test day (Observational conditioning: two-way RM ANOVA, group effect, $F_{(1,15)} = 10.46$, $p < 0.01$, epoch effect, $F_{(1,15)} = 18.17$, $p < 0.001$, group X epoch interaction, $F_{(1,15)} = 0.2416$, $p = 0.6301$; Test day: two-way RM ANOVA, group effect, $F_{(1,15)} = 12.30$, $p < 0.01$, epoch effect, $F_{(1,15)} = 6.778$, $p < 0.05$, group X epoch interaction, $F_{(1,15)} = 0.06837$, $p = 0.7973$). (**e**) Viral injection and optic fiber placement for selective inhibition of ACC→BLA circuit. (**f**) Inhibition of ACC→BLA circuit during classical fear conditioning. No significant differences were detected between NpHR (N = 7 mice) and eYFP (N = 10) mice in cue driven freezing on test day (unpaired, two-tailed, t test, t = 1.02, df = 15, $p = 0.3237$). Inset shows cue and baseline (20 s prior to cue) freezing values (two-way RM ANOVA, group effect, $F_{(1,15)} = 0.0061$, $p = 0.9389$, epoch effect, $F_{(1,15)} = 28.48$, $p < 0.0001$, group X epoch interaction, $F_{(1,15)} = 1.041$, $p = 0.3237$). All error bars indicate SEM. Significance using Bonferroni post hoc analysis is marked as *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

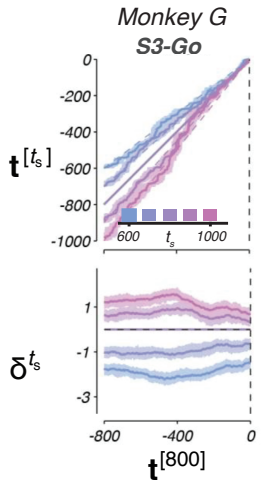
# Supplementary Figures



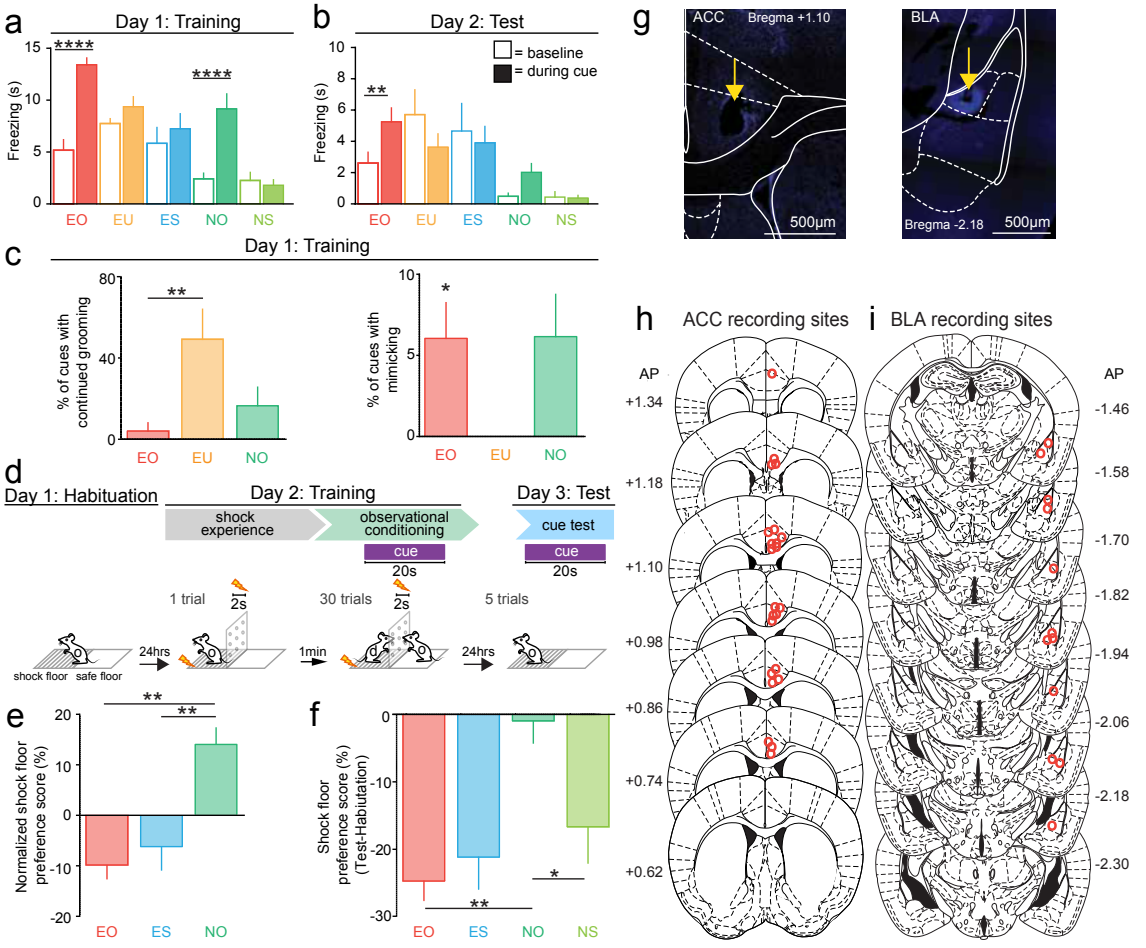

**Supplementary Figure 3.1: Variance explained by different polynomial fits to single-cell responses.** Average variance explained by polynomials of increasing order. The degree of polynomial required to best describe the temporal response profiles of individual neurons was determined by fitting a polynomial to firing rates computed from a subset of training trials (50% of trials, sampled randomly without replacement) during the S1-S2, S2-S3, and S3-Go epochs. Polynomials were fit across ts for the first 600 ms of the temporal response for each epoch. The goodness of fit was assessed by the explained variance, $R_n^2$, between the polynomial of order $n$ and the firing rates in validation trials. Each column above plots the mean $\pm$ 95% confidence intervals. Vertical dashed line indicates the point at which increasing polynomial order no longer increased the quality of fit, i.e., a 6th order polynomial in this case.




**Supplementary Figure 3.2: Inverse relationship between produced times and neural speeds.** **a** Trial-averaged neural trajectories during the S3-Go epoch projected on first two PCs, sorted by $t_s$ (colored). **b** Open circles denote trial-by-trial produced times $t_p$ as a function of intrinsic speed along each trajectory during the S3-Go epoch, colored by $t_s$. The solid black line denotes the regression line.

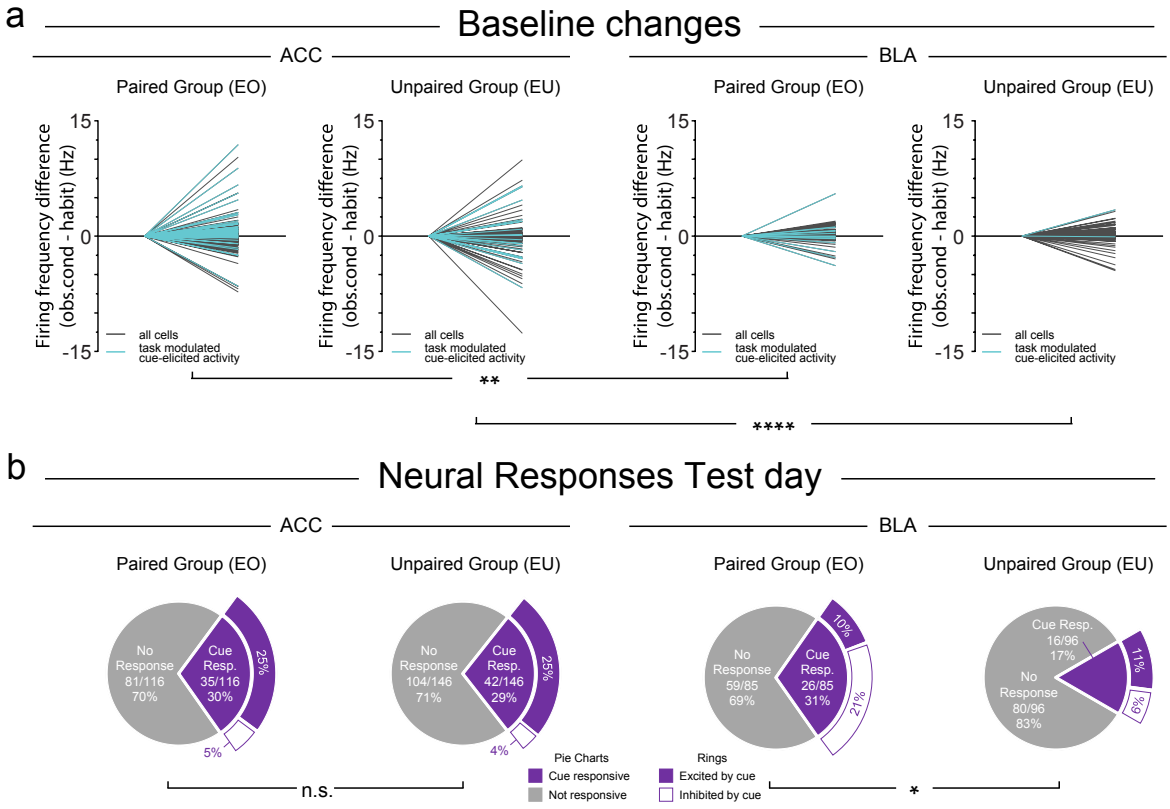



**Supplementary Figure 3.3: Speed and distance of ensemble trajectories aligned to Go.** Similar to **Figure 3.6**, but trajectories are aligned to Go instead. Top: Speed of each neural trajectory relative to the reference trajectory, by comparing the elapsed time $\mathbf{t}^{[800]}$ and the elapsed time to reach $\mathbf{r}^{[t_s]}$, i.e., $\mathbf{t}^{[t_s]}$. The unity line indicates the same speed to the reference trajectory. Dashed lines represent expected relationship between $\mathbf{t}^{[t_s]}$ and $\mathbf{t}^{[800]}$ assuming an observer has perfect knowledge of $t_s$. Bottom: Distance between neighboring states ($\delta^{t_s}$) as a function of elapsed time $\mathbf{t}^{[800]}$. The horizontal dashed line indicates a complete overlap to the reference trajectory. Shading areas indicate the median ± 95% CIs (n=100 bootstrap samples).

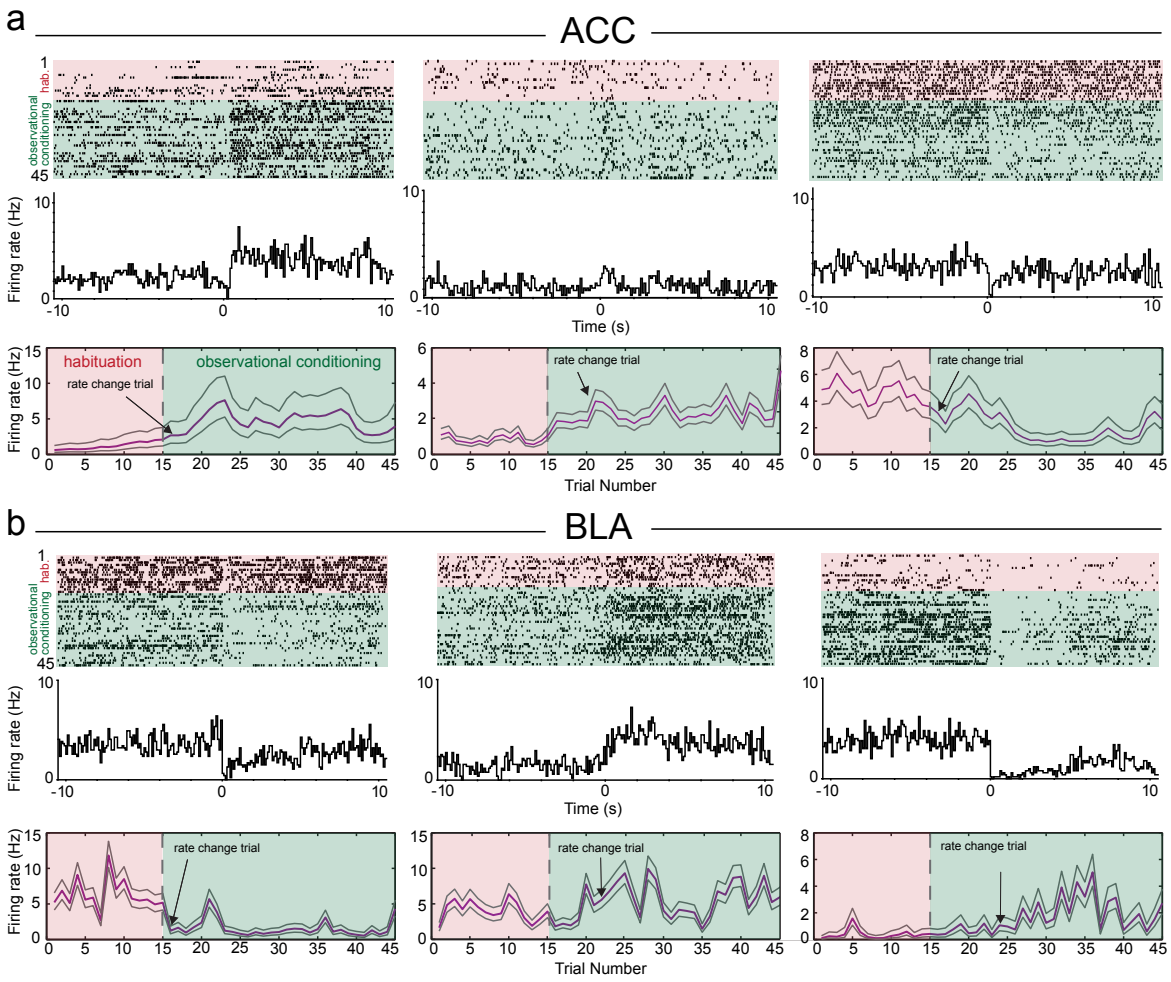



**Supplementary Figure 3.4: EO mice learn association through observation.** (**a**) On Day 1: Training; only mice that observed a cue that was paired to shock delivered to the demonstrator (EO and NO) showed increased freezing during the cue relative to baseline (20 s prior to cue onset) (two-way repeated measures (RM) ANOVA, group effect, $F_{(4,24)} = 7.37$, $p < 0.001$, epoch effect, $F_{(1,24)} = 57.51$, $p < 0.0001$, group X epoch interaction, $F_{(4,24)} = 13.76$, $p < 0.0001$). (**b**) On Day 2: Test; only EO mice showed significantly increased freezing during the cue relative to baseline (20 s prior to cue onset) (two-way RM ANOVA, group effect, $F_{(4,24)} = 4.368$, $p < 0.01$, epoch effect, $F_{(1,24)} = 0.4843$, $p = 0.4931$, group X epoch interaction, $F_{(4,24)} = 5.524$, $p < 0.01$). (**c**) Additional behaviors observed during observational conditioning. Left: EU mice show continued grooming after cue onset compared to EO (one-way ANOVA, $F_{(2,14)} = 6.273$, *: $p < 0.05$, EO versus EU, $p < 0.001$, EO versus NO, p = 0.7027). Right: Only EO mice significantly mimicked behavior of the demonstrator (one sample t test, EO: t = 2.750, df = 6, *: $p < 0.05$), while NO mice showed a trend (NO: t = 2.359, df = 4, $p = 0.0777$).

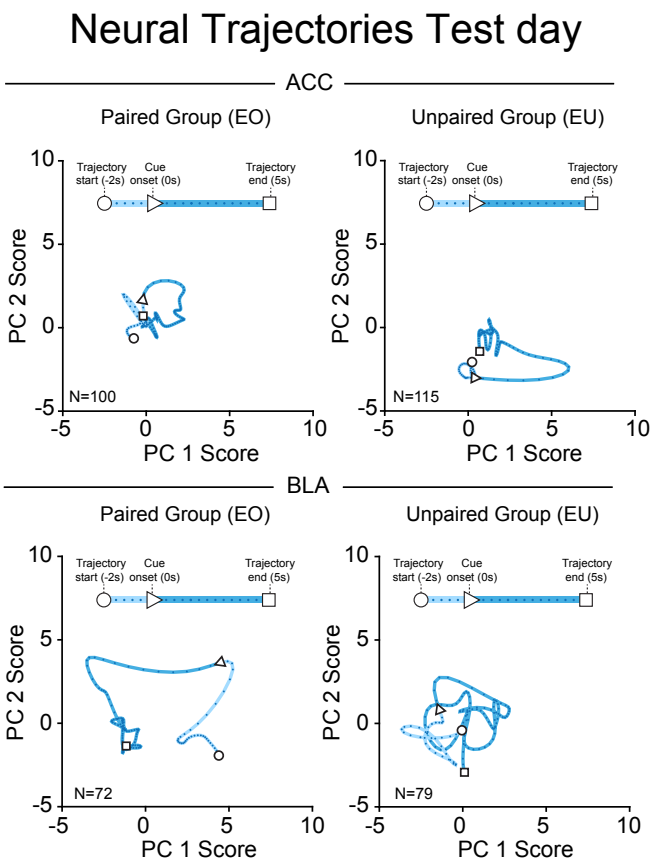
**Supplementary Figure 3.4** ***(previous page)*****:** (**d**) Outline of adapted conditioning paradigm to test for place preference after conditioning. On Day 1, mice were allowed to explore a chamber with a shock floor side and a safe floor side without a barrier. On Day 2, EO (N = 7), ES (N = 8), NO (N = 7), and NS (N = 7) mice underwent observational conditioning as outlined above. On Day 3 mice were once again placed into the chamber without the barrier and allowed to explore for 30 minutes. (**e**) On the Test day, NO mice show an increase in time spent on the shock floor compared to both EO and ES groups (one-way ANOVA, $F_{(3,25)} = 5.663$, $p < 0.01$). (**f**) On the Test day, all groups of mice spent less time on the shock floor in comparison to the Habituation day. However, mice in the different groups were significantly different in the amount of time they spent on the shock floor relative to habituation (one-way ANOVA, $F_{(3,25)} = 5.802$, $p = 0.0037$, NO versus NS, $^{*}p < 0.05$, EO versus NO, $^{**}p < 0.01$). (**g**) Representative confocal images of electrolytic lesions in ACC (left) and BLA (right). Lesions indicated by yellow arrowhead (blue = DAPI). (H and I) Histologically verified electrode placements in ACC (**h**) and BLA (**i**) for all subjects included in the study. Symbol represents electrolytic lesion site for each individual mouse. All error bars indicate SEM. Significance using Bonferroni post hoc analysis is marked as *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.




**Supplementary Figure 3.5: Baseline over learning and test-day responses.** (**a**) ACC baseline firing rates changed more from habituation to conditioning than BLA baseline firing rates. The average firing frequency of each neuron was calculated for 5 s preceding cue presentations during the habituation phase (trials 1-15) and the observational conditioning phase (trials 5-20). Then a firing frequency difference score was calculated for each neuron to evaluate training-induced changes. Black lines in the line plots represent all recorded neurons per experimental group, whereas teal lines represent the neurons that showed significant training-induced changes in cue responsiveness. Overall, the ACC exhibited a significantly larger variance than the BLA on training-induced changes in baseline activity (Kruskal-Wallis non-parametric test: ACC-EO versus BLA-EO groups, $^{**}p < 0.01$; ACC-EU versus BLA-EU groups, $^{****}p < 0.0001$). (**b**) Cue response analysis for the test day. Central pie charts show the proportions of neurons per neural recording group that exhibited either no cue responsiveness (gray) or significant cue responsiveness (purple). Semi-rings illustrate the proportions of neurons that exhibited either cue-induced excitation or cue-induced inhibition. Chi-square tests were performed to compare Paired (EO) and Unpaired (EU) groups. For the ACC ($\chi^2 = 0.0615$, and $p = 0.8041$; n.s. = not significant). For the BLA, we detected a significant difference in the proportion of neurons that were cue-responsive on test day between the EO and EU groups, ($\chi^2 = 4.9032$, $^{*}p < 0.05$).




**Supplementary Figure 3.6: Examples on state-space changed across trials.** (**a** and **b**) Representative PSTHs (100 ms bins) of conditioning dependent neurons in the ACC (**a**) and BLA (**b**) responding to the cue and the state-space analysis of these neurons showing their respective rate change trial.




**Supplementary Figure 3.7: Ensemble trajectories on test day.** Neural trajectories for the test day. Each plot illustrates the trial-averaged low-dimensional neural trajectory projected on a 2D space formed after PC1 and PC2 per neural recording group. The light blue color represents the neural trajectory during the baseline period (i.e., prior to cue onset), whereas the darker blue color represents the trajectories during cue presentation. The white-filled circle represents the starting point of the trajectory during a baseline period (-2 s relative to cue onset), the white-filled triangle represents cue onset, and the white-filled square represents the trajectory end at 5 s after cue onset. Smaller dots inside the trajectories represent timestamps with 50 ms apart from each other, indicating the speed of trajectory evolvement across time.

## 3.4 Discussion

Heterogeneous responses of individual neurons were observed in both studies, i.e., the time reproduction and fear observational learning experiments, across different brain regions (e.g., the DMFC, the BLA, and the ACC). By building state-space models, we were able to extract the ensemble dynamics. Through analyzing speeds and distances of the neural trajectories, we were able to further uncover the mechanisms underlying an internal model for generating actions or acquiring associations.

### Neural dynamics across time reproduction

We found that during time reproduction, the neurons in the DMFC showed temporally scaled responses across trials with different intervals. By projecting these complex profiles to a low-dimensional state-space, we found that the speed of a neural trajectory was modulated after each flash based on the error between the timing of the flash and the anticipated timing, and that the distance between neural trajectories remained approximately constant between consecutive flashes but was adjusted upon cue presentation. Consistent with behavioral data, how trajectories evolved through time also revealed sequential updating on a simulated motor plan. However, it is still unclear how the DMFC modulates the ensemble dynamics, i.e., the source of speed commands. It is possible that these neurons receive persistent inputs across time. It is also possible that these distinct dynamics were set by transient signals from neuromodulatory system given at each sensory feedback (e.g., predictive errors encoded by dopaminergic neurons), or determined simply by initial conditions of the dynamical system (e.g., trajectories terminated at different states after each measurement). These findings provide evidence for generating internal estimates through control of cortical dynamics.

### Neural dynamics across fear observational learning

We found that during observational learning, neurons in both the ACC and BLA show rapid conditioned responses to the cue during observational fear learning, as in

classical fear conditioning (Maren, 2000; Quirk et al., 1995; Steenland et al., 2012). However, we revealed several surprising features of neural dynamics in these regions during observational conditioning. In both the ACC and BLA, the overall proportion of cue-responsive neurons alone was not indicative of learning the predictive value of the cue, as it was similar for the paired and unpaired groups, perhaps due to latent inhibition. Rather, the proportion of neurons undergoing significant changes in firing rate across learning (and the direction of those changes) proved to be a greater differentiator between the paired and unpaired groups. In addition, our state-space analysis of neural trajectories showed that ACC neurons show baseline changes in the context of the demonstrator's distress and that some BLA neurons were dependent on ACC input during the cue in order to respond appropriately to the cue. These data suggest that during observational conditioning, the ACC encodes the demonstrator's distress response, thereby enabling the acquisition of the aversive value of the cue by BLA neurons and subsequent behavioral output.

# Chapter 4

# Integration of signals from cells and neuropil improves decoding

*In vivo* calcium imaging using head-mounted miniature microscopes enables tracking activities from neural populations over weeks in freely behaving animals. Previous studies focus on inferring behavior from a population of neurons, yet it is challenging to extract neuronal signals given out-of-focus fluorescence in endoscopic data. Existing analysis pipelines include regions of interest (ROIs) identification, which might lose relevant information from false negatives or introduce unintended bias from false positives. Moreover, these methods often require prior knowledge for parameter tuning and are time-consuming for implementation. Here, we develop an end-to-end decoder to predict the behavioral variables directly from the raw microendoscopic images. Our framework requires little user input and outperforms existing decoders that need ROI extraction. We show that neuropil/background residuals carry additional behaviorally relevant information. Video analysis further reveals an optimal decoding window of 1 second and dynamics between residuals and cells. Critically, saliency maps reveal the emergence of video-decomposition across our decoder, and identify distinct clusters representing different behavioral aspects. Together, we present a framework that is robust and efficient for decoding behavior from microendoscopic imaging, and may help discover functional clustering.

## 4.1 Introduction

Tracking the activity of large neuronal populations in awake and behaving animals is crucial to understand the neural computation underlying cognitive functions. Recent advances in *in vivo* calcium imaging technology have nurtured studies of the neural circuits underlying sensory perception (Glas et al., 2019; Mittmann et al., 2011; Rothschild et al., 2010; Wang et al., 2020; Yoshida and Ohki, 2020), motor control (Ebina et al., 2018; Huber et al., 2012; Klaus et al., 2017), innate behaviors (Betley et al., 2015; Evans et al., 2018; Jennings et al., 2015), and a wide variety of cognitive behaviors such as decision making (Harvey et al., 2012; Pinto and Dan, 2015; Tanimoto et al., 2017), as well as learning and memory (Grewe et al., 2017; Roberts et al., 2017; Yu et al., 2017; Ziv et al., 2013). Specifically, a miniaturized head-mounted microscopy combined with a microendoscopic lens enables deep brain imaging of freely moving animals (Cai et al., 2016; Flusberg et al., 2008; Ghosh et al., 2011). Although two-photon miniature microscopes (Helmchen et al., 2001; Sawinski et al., 2009; Zong et al., 2017) are developed to achieve optical sectioning, they are not widely adopted in neuroscience studies due to problems including slower acquisition speed, significant motion artifacts, and optical limitations of the fibers which cannot deliver 920nm femtosecond laser pulses without significant distortion. On the other hand, microendoscopic imaging using a single-photon light source suffers from large background fluctuations due to out-of-focus fluorescence. Toward achieving the goal to extract external observables like sensory stimuli or behavioral variables from the recordings, how to properly extract cellular signals from these low-contrast images has been a key step in analysis pipelines for calcium imaging (Lu et al., 2018; Pnevmatikakis, 2019; Zhou et al., 2018).

A typical analysis pipeline (Figure 4.1a) includes the following steps: First, registering image frames by correcting motion artifacts. Oftentimes, brain motion within the field of view (FOV) can be non-rigid along all directions and lead to frame distortions. To correct these motion artifacts, either a rigid template matching method (Dubbs et al., 2016; Kaifosh et al., 2014) or a non-rigid registration method (Lu et al.,

2018; Pnevmatikakis and Giovannucci, 2017) is applied.

Second, extracting neuronal signals through estimating spatial filters and their corresponding temporal traces. Common methods can be divided into three classes: semi-manual region of interest (ROI) analysis (Klaus et al., 2017; Pinto and Dan, 2015), principal component analysis/independent component analysis (PCA/ICA) (Mukamel et al., 2009; Reidl et al., 2007), and constrained nonnegative matrix factorization (CNMFe) (Pnevmatikakis et al., 2016; Zhou et al., 2018). Next, inferring spike times from extracted fluorescence traces. To simplify spike inference, fluorescence traces can be viewed as spike trains convolved with a kernel with asymmetrical rising and falling kinematics. Various algorithms (Jewell and Witten, 2018; Pnevmatikakis et al., 2016; Vogelstein et al., 2010; Yaksi and Friedrich, 2006; Zhou et al., 2018) have been proposed to solve this deconvolution problem.

Finally, decoding behavioral variables or sensory stimuli from either extracted fluorescence traces or inferred spike trains. Common decoders including population vector reconstruction (Georgopoulos et al., 1986; Salinas and Abbott, 1994), Wiener filter (Carmena et al., 2003), template matching (Wilson and McNaughton, 1993; Zhang et al., 1998)), Bayesian paradigm (Brown et al., 1998; Sanger, 1996; Wu et al., 2006; Zhang et al., 1998), and machine-learning models (e.g., support vector machine (SVM), k-nearest neighbors (kNN), Fisher's linear discriminant analysis (LDA), multilayer perceptron (MLP), recurrent neural network (RNN)) have been implemented in various studies (Glaser et al., 2020; Pereira et al., 2009; Quiroga and Panzeri, 2009).

In a typical analysis pipeline, how neuronal signals are extracted and structured is a key determinant with single-photon microendoscopic data. Unfortunately, existing source extraction methods have several limitations. Taking semi-manual ROI analysis (Klaus et al., 2017; Pinto and Dan, 2015) as an example, neuropil contamination from out-of-focus and scattered fluorescence is often inadequately removed and require explicit assumptions on neuropil structure. Additionally, signals of spatially overlapped neurons cannot be demixed. Moreover, manual ROI annotation can be highly variable with inter-labeler disagreement and laborious for large-scale recordings. As for PCA/ICA analysis (Mukamel et al., 2009; Reidl et al., 2007), it ignores

signal correlations and noise correlations in the neuronal populations (Averbeck et al., 2006). Moreover, it is a linear method and cannot demix signals of spatially overlapped neurons. Recently, CNMFe (Zhou et al., 2018) has gained popularity given its superior ability in demixing and denoising neuronal signals from microendoscopic data. However, it depends on sophisticated hyperparameter tuning which requires prior knowledge on the number of neurons and their sizes. Additionally, it is sensitive to parameter initialization and has non-unique solutions for ROI identification. Specifically, all these methods have detection errors, which lead to either introducing extra bias or throwing away relevant information.

A recent study developed a new signal extraction pipeline for single-photon microendoscopic imaging (MIN1PIPE) (Lu et al., 2018) to improve ROI identification. By incorporating a two-component Gaussian mixture model (GMM) and a trained recurrent neural network (RNN), fluctuations unlikely to be calcium transients are removed, leading to a cleansed set of ROI seeds. This method reduces the need to set many hyperparameters heuristically. However, it is computationally expensive and requires extra manual judgement for training to classify calcium transients. Recognizing the challenge of obtaining highly preprocessed data for decoding, another study developed Deepinsight (Frey et al., 2019), a deep-learning framework, to infer behaviors from wide-band neural data (e.g., raw electrophysiology recordings without spike sorting). However, when it comes to calcium imaging data, this study still requires extracting temporal traces from identified ROIs, before converting them into frequency representations which are provided to the decoder as input.

Here we propose an end-to-end decoding paradigm to directly extract behavioral information from raw single-photon microendoscopic data (Figure 4.1b), and assess using the problem of predicting kinematics from calcium imaging of hippocampal CA1 in freely foraging mice (Figure 4.2a). We hypothesize that the neuropil/background, which was discarded in most analysis pipelines, may encode the animal's position and/or speed. Instead of making assumptions on the format of relevant signals, our paradigm leverages recent progress in deep learning to determine which imaging components are most informative without a need for sophisticated preprocessing.

Our decoder outperforms decoding methods that use extracted neuronal signals. The results compare encoding capacity between putative neurons and the background residuals, and identify contribution across timepoints. By examining the saliency maps, we find video-decomposition naturally emerges in our decoder. Our findings suggest neuropil as additional information encoder in microendoscopic recordings, and provide a new way to identify neural ensembles representing distinct aspects of spatial navigation.

## 4.2 Materials and Methods

### 4.2.1 Data collection

A total of 8 male mice of C57/B6 background were used in the experiments. All procedures were approved by the MIT Committee on Animal Care and followed the guidelines established by the National Institutes of Health for the care and use of laboratory animals.

#### Virus-mediated gene delivery

Animals were first anesthetized with 5% Isoflurane in oxygen. Throughout the procedure, the animals were anesthetized at a stereotaxic surgical frame (Kopf) with continuous 1.5% to 2% Isoflurane in oxygen. A burr hole is made at -1.5mm medial-lateral and 2.1mm rostral-caudal to the Bregma landmark. Using a motorized injector at a speed of 0.05 $\mu$L/min, 0.5$\mu$L of pGP-AAV1-syn-jGCaMP7f-WPRE (titer: 1.9 x 1012 gc/ml) was injected at 1.5mm under the skull surface in the burr hole. Afterwards, the scalp was sutured and analgesic was administered (Buprenex, 0.05 mg/kg) prior to the animals' recovery.

#### Implant protocol

After virus injection, GRIN lens implant and baseplate implant were performed separately. During implant, the animals were anesthetized following the same protocol.

**GRIN lens** The skull surface was roughened with etchant gel (C&B Metabond). A craniotomy 2 mm to 2.2 mm in diameter was made on the skull, with the burr hole from virus injection procedure in the center. Cortical tissue was aspirated out, exposing the rostral-caudal running striation pattern right above CA1. A GRIN lens (1.8 mm diameter, 0.25 pitch, 670 nm, Edmund) was perpendicularly placed on top of CA1 and fixed to the skull with adhesive (Loctite 454). Afterwards, the exposed skull and the surrounding of the GRIN lens were covered with transparent dental cement (C&B Metabond), followed by black orthodontic acrylic resin to block out ambient light. The top of the GRIN lens was covered with silicone adhesive (Kwik-Sil).

**Baseplate** With a baseplate attached to a microendoscope (UCLA miniscope), the optimal field of view (FOV) was searched through imaging and manually adjusting the angle of the baseplate attachment. After the optimal field of view was reached, the baseplate was first fixed in place with adhesive (Loctite 454), and then affixed to the skull with dental cement (M&B metabond).

#### Imaging setup

One-photon calcium imaging was performed through a head-mounted miniscope. The imaging power was between 1 mW and 10 mW approximately, adjusted for each animal. The excitation LED is filtered by a 470/40 band-pass filter. The emitted photons were collected by a CMOS sensor, filtered by a 525/50 band-pass filter, sampled at 30 Hz. The imaging FOV was 480 by 752 pixels, and the optimal plane was reached during the baseplate implant.

### 4.2.2 Behavioral Tasks

Once the animals were fully recovered from previous procedures (approximately 3 days after the baseplate implant), *in vivo* recordings were performed on awake animals.

**Miniscope acclimation in the home cage** The animals were attached to the imaging cables, allowing them to acclimate to the miniscope and tethering for a few

sessions in the home cage, with each session lasting for 30 minutes. This procedure repeated for about 3 days, until the animals were comfortable with the imaging setup and exhibit normal behavior.

**Maze exploration experiments** After miniscope acclimation sessions, the animals underwent maze exploration experiments in which maze enclosures are made of corrugated plastic sheets (Home Depot). The shapes of these mazes can be categorized as T-shape, H-shape, or □-shape. Each maze has a size of about 1 m by 1 m. For each maze, the animals underwent multiple sessions, with each session lasting for about 30 minutes. In our study, we excluded the first few sessions when the maze configuration was still novel to the animals, and only used data in later sessions when the animals were already familiar with the environment.

### 4.2.3 Dataset preparation

#### Behavioral data

To track the animals' behavior, we used a webcam sampled at 60 Hz and synchronized with imaging using a package (https://github.com/jonnew/Bonsai.Miniscope) for Bonsai (Lopes et al., 2015). Red tape on the head-mounted miniscope was tracked as a proxy for the animal's locations, reported as position coordinates $(x, y)$.

**Temporal down-sampling and smoothing** Raw x, y coordinates were downsampled to 10 Hz, and smoothed with a Savitzky-Golay finite impulse response smoothing filter of quadratic order. The speed $v$ was further estimated from these smoothed x, y coordinates.

**Coordinate orthogonalization** Linear dependencies between different dimensions of the output labels were minimized by orthogonalizing the smoothed x, y coordinates. For the T-maze, x, y coordinates were centered at the intersection of two perpendicular arms. For the H-maze and the □-maze, x, y coordinates were centered in the middle of the maze. After centering, a rigid rotation matrix (R) was estimated

through singular value decomposition (SVD) and applied to orthogonalize the x, y coordinates.

**Label scaling** All behavioral labels $(x, y, v)$ were scaled to the range [-1, 1].

### Calcium Imaging

**Movement artifact removal** To correct non-uniform movement artifacts caused by brain motion, NoRMCorre, a fast non-rigid registration method (Pnevmatikakis and Giovannucci, 2017) was applied. This algorithm operates by splitting the FOV into overlapping spatial patches, which are further registered at a subpixel resolution for rigid translation against a template. Afterwards, a smoothed motion field for each frame was created by up-sampling the estimated alignments.

**Spatial down-sampling and cropping** Images were down-sampled in each dimension by a factor of 2, leading to 240 by 376 pixels. The FOV was cropped to 224 by 224 pixels, by removing surrounding regions that did not contain calcium signals.

**Source decomposition and denoising** This step was only applied to generate data for the baseline decoder (i.e., it was skipped for our end-to-end decoder). To separate different sources such as single neurons and out-of-focus fluorescence from neuropil, the CNMFe algorithm (Zhou et al., 2018) was applied. Hyperparameters such as the estimated diameter of a neuron were optimized for each animal's dataset. The average fluorescence traces C from regions of interest (ROIs) were extracted and denoised through spike deconvolution. ROIs with weak (peak $C$ lower than 100% dF/F) or sporadic activity (less than 3 calcium transients) were eliminated. The remaining ROIs were assumed to be putative neurons. Temporal activities $C$ were used as input for the baseline decoder.

**Temporal down-sampling** Calcium images were down-sampled to 10 Hz.

**Pixel scaling** Raw image pixel amplitudes (0-255) were scaled to the range [-1, 1].

**Experiment on residual information**

**Raw** This set consisted original images ($Y$) after movement artifact removal, FOV cropping, and pixel scaling.

**Clean** A video matrix was obtained after the CNMFe algorithm by multiplying the spatial footprint matrix $A$ with the temporal dynamics matrix $C$. Image frames from this video matrix only contained putative neurons, and thus were called Clean images.

**Residual** Residual images were obtained by subtracting the Clean images from the Raw images, i.e. $Y - AC$. Theoretically, behavioral-relevant information in these images came from the background residuals, as any spike-triggered temporal fluctuations in the putative neurons were extracted to the temporal dynamics matrix $C$. Ideally, Residual images lack information from putative neurons. However, the CNMFe algorithm updates parameters through alternating iterations such that spike-triggered temporal fluctuations might not be fully captured in the Clean images. To further rule out spike-triggered information left in the Residual images, locations at putative neurons were occluded, through either Hollow ROI or Hollow A methods.

**Hollow ROI** The Hollow ROI images were obtained after occluding the putative cells after setting a local adaptive threshold. The threshold was computed for each pixel according to the image characteristics within a moving circular window. The local threshold was computed using the Niblack method:

$$Thr_{\text{Niblack}} = \mu + k\sigma - c$$

where k = 0.2, c = 0, $\mu$ is the mean pixel value within the window, and $\sigma$ is the standard deviation of the pixel values within the window.

**Hollow A** The Hollow A images were obtained after occluding the entire spatial footprint matrix $A$ from the Residual images. This further removed any local signals associated with each putative neuron.

**Scramble images** Each pixel in each image was randomly re-positioned per row. The pixel distributions remained identical, but the spatial patterns were destroyed.

**Random images** Image frames were synthesized by sampling each pixel from a uniform distribution that had the same minimum and maximum as the original image frames. The synthesized images were basically random noise. Decoders trained on random images were used as the chance level.

### 4.2.4 Decoder Modeling

#### Dataset evaluation

**Train/val/test dataset split** Dataset was separated into the 'RUN' and the 'STOP' periods based on running speed with a separation threshold at 5 cm/s. For most of our model training and evaluation, the 'STOP' period was excluded from analysis unless otherwise indicated. Each animal's dataset was segmented into 5 consecutive blocks. Following a 80-20 split rule, one of these blocks was selected to be the test set whereas the rest were used as the training set. During the model development phase, the training set was segmented into 5 consecutive blocks, with one of them randomly selected to be the validation set.

**Similarity metrics for marginal distribution** To quantify similarity between marginal distribution of each label variable $(x, y, v)$ between the training set and the test set, the Jensen–Shannon divergence (JSD) was used.

$$\mathrm{JSD}(P\|Q) = \frac{1}{2}D(P\|M) + \frac{1}{2}D(Q\|M)$$

where $P$ is the probability distribution of the training set, $Q$ is the probability distribution of the test set, and $M = (P + Q)/2$. If JSD = 0, the two distributions completely overlap. A larger JSD indicates more differences between two sets.

**Similarity metrics for joint distribution** To quantify similarity between joint distribution for paired variables $((x, y), (x, v), (y, v))$ between the training set and the test set, structural similarity index measure (SSIM) (Wang et al., 2004) was performed with *scikit-image* library. The joint distribution of the training set was viewed as an image ($\mathrm{Im}_{\mathrm{train}}$), and so was the joint distribution of the test set ($\mathrm{Im}_{\mathrm{test}}$). Each image was convolved with a normalized Gaussian kernel of width $\sigma = 1.5$, resulting in multiple patches. SSIM was calculated based on various patches from two images.

$$\mathrm{SSIM}(\mathrm{Im}_{\mathrm{train}}, \mathrm{Im}_{\mathrm{test}}) = \frac{(2\mu_{\mathrm{train}}\mu_{\mathrm{test}} + c_1) \cdot (2\sigma_{\mathrm{train,\ test}} + c_2)}{(\mu^2_{\mathrm{train}} + \mu^2_{\mathrm{test}} + c_1) \cdot (\mu^2_{\mathrm{train}} + \mu^2_{\mathrm{test}} + c_2)}$$

where $c_1$ and $c_2$ are constant, $\mu$ and $\sigma$ are statistical attributes of all these patches. If SSIM $= 1$, two images of joint distributions have the same textures. A lower SSIM suggests more differences two sets.

**Similarity metrics for average occupancy map and average speed map** The occupancy map, i.e., x-y Density map, is a 2D hexagonal binning plot quantifying normalized occupancy on the maze. Average speed map, i.e., AvgSpeed map, is a 2D hexagonal binning plot quantifying average speed on the maze. 25 bins were used in x-direction and y-direction. To quantify similarity of the occupancy map and the average speed map between the training set and the test set, the 2D correlation coefficient ($r$) was used. If $r = 1$, it suggests that the training set and the test set have the same maps. A lower $r$ indicates larger differences between two sets.

### Baseline Decoder

For the baseline decoder, we identified ROIs as putative neurons from the microendoscopic images, and built a decoder based on temporal activities from these ROIs. Average fluorescence traces from N identified ROIs formed a N-dimensional input, and the decoder output the animal's positions and running speed. Given that decoding performance depended on input representations, we prepared different representations as the following.

**Smoothed traces of ROIs after background removal (ROI_trace)** We first obtained an average intensity projection from the image stacks, and then blurred the projection image using a moving 2D Gaussian filter with size of about 25% of the FOV, using ImageJ (Schindelin et al., 2012). The blurred projection was considered as the global background and is subtracted from all the raw images. Subsequently, an anisotropic diffusion operation (Perona and Malik, 1990) was applied to further denoise the images, followed by a morphological opening operation using a circular structure element with size similar to a neuron in the FOV. Afterwards, we obtained a maximum intensity projection from the denoised images whose backgrounds are removed, and annotated ROIs using the ImageJ Cell Magic Wand tool. The average fluorescence trace from each ROI is smoothed with a Savitzky-Golay smoothing filter.

**Globally normalized denoised traces from CNMFe (C_normG)** Another input representation was created by denoised temporal traces through applying the CNMFe algorithm (Zhou et al., 2018). The average fluorescence trace $C$ from each ROI was normalized by the global maximum across all the traces. With this procedure, the correlations between the neurons were informative and thus relative magnitudes between them were preserved.

**Feedforward neural network** The decoder was a feedforward neural network with a regression head. The number of dense layers matched the number in the ConvNet decoder (see section below).

### End-to-end decoder with image inputs

**Regression task** The end-to-end decoder took each image frame as an input and predicted the animal's behavioral attributes (positions and/or speed) at the corresponding time point. The decoding problem could be formulated as a multi-class classification task by binning the behavioral variables, but we chose to formulate the problem as a regression task for the following reasons. First, the animals were freely exploring the maze, so the maze occupancy and running speed distribution

were highly imbalanced. However, the data size was not large enough for undersampling data to create a uniform distribution, yet oversampling minority classes using techniques like SMOTE (Chawla et al., 2002) introduced extra bias. In addition, the common loss function in a classification task could not keep the ordinal relationship between positions and speeds, i.e., prediction error between neighboring positions and error between distant positions were same using cross entropy.

The output dimensions (or number of attributes) depended on the decoding strategy (simultaneous or separate).

**Simultaneous decoding** For simultaneous decoding strategy, convolutions and feed forward layers were shared to extract features for position and speed encoding, and different weights were used only in the final layer to predict position and speed. In this case, there are 3 output labels $(x, y, v)$ in the regression head.

**Separate decoding** For separate decoding strategy, there were two sets of convolutions and feed forward layers separately trained to predict position and speed. In this case, one decoder has 1 output label $(v)$, and the other has 2 output labels $(x, y)$.

**Vanilla convolutional neural network (ConvNet)** The model took a grayscale (single channel) image as input and used convolutional layers connected to a regression head to decode continuous behaviour. A kernel size of 3 was used for all the 2D convolutions. The number of filters was doubled for each subsequent layer and saturated at 512. For downsampling, we used a stride of 2 to replace pooling layers for a better computational efficiency. After convolutions, features were flattened and fed into a series of dense layers. The number of units was halved for each subsequent dense layer. To encourage network sparsity and accelerate learning without a risk of introducing dying units, Leaky-Relu was used as the activation function except for the final layer. The final layer was applied with a Tanh activation function to ensure the output values match the range of target labels.
The number of convolutional layers, the number of fully connected layers, and the

number of units in the first fully connected layer are hyperparameters. Evaluated on the validation set, the selected model used in this study had 8 convolutional layers and 5 fully connected layers (with number of units = [256, 64, 32, 16, 3] respectively). In our task, adding batch normalization made the prediction performance worse, so we did not use batch normalization in our decoder.

**Residual neural network with 50 layers (ResNet-50)** The ResNet model (He et al., 2016) was originally designed to take a RGB image input for object recognition tasks, so the grayscale images were converted into RGB images, with the per-pixel mean subtracted. We kept the feature extraction architecture that was composed of convolutional blocks and identity blocks, and replaced the classification architecture with the dense layers whose activation functions were Leaky-Relu (the first few layers) and Tanh (the last layer).
For transfer learning, the parameters in feature extraction architecture were pre-trained by ImageNet and frozen during the model training, whereas the parameters in the dense layers are updated throughout training. Afterwards, the last few layers of the feature extraction architecture were unfrozen and jointly trained with the new dense layers. For learning from scratch, all the parameters in the model were initialized with Xavier initialization (Glorot and Bengio, 2010).

**Efficient convolutional network for mobile applications (MobileNet)** MobileNet (Howard et al., 2017) was developed to achieve a lightweight architecture by replacing a convolutional layer with a depthwise convolutional layer, followed by a pointwise convolutional layer that combined these filtered values to create new features. It used a stride of 2 to reduce the spatial dimensions instead of pooling layers. Same preprocessing steps used in the ResNet model were applied to this decoder. Similarly, both the transfer learning and learning from scratch procedures were performed in this study.

**Feedforward neural network (FFNN)** We built a feedforward neural network to learn features directly through dense connections. The model first applied a 2D

max-pooling on each grayscale image for spatial downsampling, and then flattened these features for a series of dense layers. The number of units, activation functions, and procedure for hyperparameter searching were the same as in the vanilla ConvNet.

### End-to-end decoder with video inputs

We also created an end-to-end decoder that took a short video with a window size of $N$ to predict behavior at the corresponding time centroids. Two architectures were developed, depending on whether convolutional weights were shared across time.

**Convolutions independent of time** The input dimension expanded from a single channel to multiple channels, each corresponding to an image frame. The decoder architecture kept the original vanilla convolutional neural network with 2D kernels. Features from different frames were weighted and integrated after the first layer. There was no constraint on convolutional kernels.

**Convolutions shared across time** The input dimension expanded with a temporal dimension for $N$ frames. Each frame went through a block of convolutional layers whose weights were shared across time, generating a frame-specific feature. These frame-specific features were linearly weighted by an attention layer, and then integrated into a final feature $F_{post}$ before feeding into the dense layers of the model.

$$F_{post} = \sum_{k=-\Delta t}^{\Delta t} F_{T+k} \cdot W_k \quad \text{where} \quad \sum_{k=-\Delta t}^{\Delta t} W_k = 1, \quad N = 2\Delta t + 1 \tag{4.1}$$

### Training procedure

Mini-batch gradient descent was used for training the model. For each mini-batch, a subset of input images and output labels were randomly sampled from the training set. Each input image had corresponding normalized output labels $(x, y, v)$. The batch size (i.e., number of samples per step) was 64 and the order of samples were further shuffled for each epoch. Hyperparameter searching was achieved by grid search for learning rate, number of the convolutional layers, number of the dense layers,

number of units, and dropout. Adam optimization was used to accelerate learning. Kernel weights and biases are initialized by Xavier initialization (Glorot and Bengio, 2010), and updated by backpropagating the L1 loss, i.e., absolute differences between estimated and target values.

$$L(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \left( \frac{1}{d} \sum_{j=1}^{d} \left| y_{i,j} - \hat{y}_{i,j} \right| \right) \tag{4.2}$$

where $y_i$ is the ground truth, $\hat{y}_i$ is the predicted value for sample $i$, $m$ is the batch size, and $d$ is the dimension of the model output ($d = 3$ for simultaneous decoding and $d = 1$ for separate decoding).

The training was stopped when the number of iteration steps reached 50,000 or when the validation performance remained unchanged or increased over 5 epochs. The model variables and training hyperparameters were saved in a checkpoint every 500 iteration steps. The training was performed on GeForce RTX 2080 Ti using Tensorflow, and parallel computation was utilized for grid search.

### Evaluation metrics

After a model was trained, its performance was evaluated on the test set. Given that the model output normalized variables, the predicted values were transformed back to the original scales in [cm] and [cm/s]. Afterwards, we calculated the Euclidean distance between each ground truth and its predicted position, and used the median among these errors as a performance metric for position decoding. For speed decoding, we calculated the absolute difference between each ground truth and its predicted value, and used the median error as a performance metric. For an experiment or a specific model evaluated on several datasets, the distribution of the median decoding error was described as mean ± s.e.

**Chance level calculation** The chance level was defined as decoding performance using Random images (see previous section Experiment on residual information).

### 4.2.5 Manifold analysis

To visualize the ensemble representation at each network layer in our decoder, ensembles at different time points were projected to a low-dimensional subspace.

**Isomap** Isomap is a nonlinear dimensionality reduction algorithm for seeking embeddings that maintains geodesic distances between all data points. We used scikit-learn (Pedregosa et al., 2011) to implement this algorithm and extracted the low-dimensional embeddings of each layer. The optimal number of nearest neighbors ($K_{\text{opt}}$) was selected by finding the elbow at which the reconstruction error (ER) curve ceased to decrease significantly with more nearest neighbors.

$$E_R = \left\| \frac{K(D_x) - K(D_y)}{N} \right\|_F \tag{4.3}$$

where $N$ is the number of samples, $D_x$ and $D_y$ follow the same definitions above, and $K$ is the Isomap kernel.

#### Manifold similarity metrics

To measure geometric similarity between the manifold and the behavioral topology formed by position, we calculated the Pearson correlation between pairwise distances in the 2D manifold coordinates and behavioral coordinates (Low et al., 2018), which was used as a metric to evaluate learning across different network layers.

### 4.2.6 Network saliency map

To better interpret how the encoder network makes the predictions, we used a saliency map to indicate the pixels whose changes have the most impact on the prediction. Instead of directly propagating the output gradient to the input (Simonyan et al., 2013), we modified the gradient-weighted class activation mapping (Grad-CAM) method (Selvaraju et al., 2017) to create a smoother saliency map.

Grad-CAM is originally designed to backpropagate the gradients of target class score $y^c$ to the last convolutional layer whose feature maps preserve spatial information,

and generate a localized saliency map $L^c$ for class $c$, by a weighted combination of activation maps $A^k$ followed by ReLU operation, as in Eq (4.4).

$$L^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \quad \text{where} \quad \alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \tag{4.4}$$

We modified Grad-CAM because our study used a continuous regression task. We visualized salience over input that either increased or decreased the output by computing the absolute gradients. We also adapted this algorithm to visualize saliency maps of different filters in the network model, by computing gradients from the filter at a specific layer with respect to the closest convolutional layer whose feature maps had large enough spatial resolution.

### 4.2.7 Statistical analysis

Statistical analyses were performed using SciPy (Virtanen et al., 2020). All values were reported as mean ± s.e. unless otherwise indicated. Wilcoxon signed-rank tests were used when paired samples were not normally distributed. For cases matching the normality requirement, Student's t-test was used. Bonferroni correction was used for multiple comparisons. Pearson correlation was used to obtain correlations.

## 4.3 Results

To evaluate how well behavioral information can be directly extracted from microendoscopic imaging data by our proposed decoding paradigm (Figure 4.1b), we expressed $Ca^{2+}$ indicator jGCaMP7f in the dorsal hippocampus of mice, implanted a GRIN lens right above the CA1 region (Figure 4.2a), and recorded calcium activities of the CA1 through a miniature head-mounted single-photon microscope when the animal was freely navigating mazes of different geometric shapes (Figure 4.2b and Supplementary Figure 4.1a). Imaging data and the animal's behavior were simultaneously recorded, synchronized at a sampling rate of 30 Hz.

To generate required inputs and labels for the decoder model, we first preprocessed

behavioral data in the following steps. Both images and animal positions $(x, y)$ were downsampled to 10 Hz to reduce memory consumption. A Savitzky-Golay filter was applied to smooth position coordinates and obtain running speeds $(v)$. Afterwards, position coordinates were rotated to encourage orthogonalization between output labels. Finally, all the variables $(x, y, v)$ were scaled to [-1, 1] (Figure 4.2c). As for calcium imaging data, the field of view (FOV) was cropped into 224 by 224 pixels, with values being scaled as well (see Materials and Methods).

### 4.3.1 Consistency of behavioral variables across time

Before model training, we split each dataset into a training set and a test set. By segmenting each session into five sequential periods, we selected one of them as the test set (Figure 4.2d), because a random split would not make sense for our recordings. In the literature, a rolling forecasting strategy for train/test split is often applied to time-series data, by assuming that distributions are time-invariant and that a generative model can be learned from history. This strategy might not be appropriate for our data if the animals change their behaviors across time.

To evaluate the consistency of behavioral variables $(x, y, v)$ between training and test sets, we computed their cumulative density functions (Figure 4.2d and Supplementary Figure 4.1b) and compared these distributions by Jensen–Shannon divergence (JSD). In a typical H-maze dataset with middle period selected as the test set (Figure 4.2d), we found the distributions between the training and test sets were similar ($\text{JSD}(x) = 0.14$, $\text{JSD}(y) = 0.10$, and $\text{JSD}(v) = 0.06$). Similar results were observed across datasets in different mazes (Figure 4.2e, top). We further evaluated the effect of test-set selection across five test-folds (Figure 4.2e, bottom), and found that these variables $(x, y, v)$ were more separated at the beginning and the ending of an experiment, but remained mostly similar throughout the session (average JSD in T-maze: $0.18 \pm 0.02$, $0.13 \pm 0.01$, $0.13 \pm 0.02$, $0.13 \pm 0.02$, $0.14 \pm 0.01$; H-maze: $0.09 \pm 0.01$, $0.11 \pm 0.03$, $0.13 \pm 0.02$, $0.12 \pm 0.02$, $0.13 \pm 0.04$; □-maze: $0.10 \pm 0.03$, $0.08 \pm 0.02$, $0.07 \pm 0.01$, $0.12 \pm 0.02$, $0.10 \pm 0.02$).

We also quantified the similarity of each paired variable by computing structural

similarity index measure (SSIM) between the training and test sets (Supplementary Figure 4.2). If two joint distributions have the same structures, SSIM = 1. We found that the joint distributions in the T-maze experiments were highest, and joint distributions remained similar throughout the session in all datasets (average SSIM in T-maze: 0.82 ± 0.03; H-maze: 0.72 ± 0.02; □-maze: 0.64 ± 0.04).

Furthermore, we compared the average occupancy and average speed in the space between the training and test sets (Figure 4.2d and Supplementary Figure 4.1b), by computing their 2D correlations ($r$). Results showed that occupancy and average speed were mostly consistent throughout the session, but dropped near the end of the session (average $r$ in T-maze: 0.71 ± 0.06, 0.74 ± 0.05, 0.78 ± 0.05, 0.76 ± 0.06, 0.71 ± 0.03; H-maze: 0.70 ± 0.03, 0.71 ± 0.02, 0.68 ± 0.06, 0.62 ± 0.04, 0.46 ± 0.13; □-maze: 0.75 ± 0.03, 0.75 ± 0.03, 0.72 ± 0.03, 0.66 ± 0.05, 0.70 ± 0.06).

Overall, these analyses reveal a general consistency of behavioral variables across time, with a higher dissimilarity in both the beginning and the ending of a session. These findings support our strategy for train/test splits.

### 4.3.2 Accurate decoding from raw microendoscopic data

To directly decode positions and running speeds of the animal from raw imaging data, we built a convolutional regression model (ConvNet) (Figure 4.3a). At each timepoint, a single frame was fed to a network consisting of 2D convolutional layers with leaky-ReLU activations, followed by dense layers. A Tanh operation was applied to the last layer. The hyperparameters of the model were searched and determined based on the validation set (Supplementary Figure 4.7c). Results showed that this model accurately decoded the positions of the animals as well as their running speeds from unprocessed microendoscopic images, across different maze exploration experiments (Figure 4.3b-e and Supplementary Figure 4.3a-c).

By comparing decoding performance with a feedforward neural network (FFNN) trained on CNMFe-denoised neuronal signals, i.e., FFNN(normG), and a model trained on background-removed ROI signals, i.e., FFNN(ROI) (see Materials and Methods), we found that in terms of predicting positions (Figure 4.3d, left), our ConvNet model

had less median decoding error (10.17 ± 0.43 cm) than the FFNN(ROI) model (12.32 ± 0.72 cm; Wilcoxon signed-rank test: stats = 54, $p < 0.001$) and the FFNN(normG) model (10.74 ± 0.49 cm; stats = 135, $p < 0.05$). These models were all significantly better than the chance level (48.57 ± 2.67 cm; $p < 1e-5$). As for predicting running speeds (Figure 4.3d, right), the median decoding error of our ConvNet model (3.39 ± 0.14 cm/s) was less than the FFNN(ROI) model (3.62 ± 0.18 cm/s; stats = 121, $p < 0.05$), and comparable to the FFNN(normG) model (3.36 ± 0.14 cm/s; stats = 219, $p = 0.78$). These models were also better than the chance level (4.66 ± 0.19 cm/s; $p < 1e-5$).

### 4.3.3 Effect of data representations and model architectures

After showing that a ConvNet model can directly extract behavioral information from raw microendoscopic data, we evaluated how data representations affect decoding performance. Ideally, we would like output labels to approximate a uniform distribution for training a deep-learning model. However, the running speeds of the animals were heavily skewed toward lower values. We applied a log-transformation on speeds (Supplementary Figure 4.5a), and re-trained a subset of models. Results showed that the original ConvNet trained with linear speeds (11.49 ± 0.13 cm) had significantly less median error in predicting positions than the same architecture trained with log speeds (12.45 ± 0.14 cm; $p < 1e-5$) (Supplementary Figure 4.5b, left). Similarly, the median error in predicting running speeds was significantly lower in the original model trained with linear speeds (3.76 ± 0.03 cm/s) than the model trained with log speeds (3.95 ± 0.04 cm/s; $p < 1e-5$) (Supplementary Figure 4.5b, right). This result matched with studies suggesting neurons in the hippocampus encoded running speeds in a linear manner (Góis and Tort, 2018; Kropff et al., 2015), different from log-scale representations in the sensory cortex (Nover et al., 2005).

We further evaluated how different model architectures could affect the decoding performance with our datasets. The original ConvNet model predicted positions and speeds simultaneously (Supplementary Figure 4.6a, top). We hypothesized that speed decoding benefited from a joint training paradigm, and a separate decoding strategy

could remove such benefits. We modified the model to have two separate decoders, one for position and the other for speed (Supplementary Figure 4.6a, bottom), and compared their performance in predicting positions and running speeds of the animal (Supplementary Figure 4.6b). Results showed that a model with two separate decoders improved position prediction (9.85 ± 0.42 cm) compared to the original model with simultaneous decoding (10.17 ± 0.43 cm; stats = 133, $p < 0.05$), but came at a cost in speed decoding, i.e., its median error (3.53 ± 0.16 cm/s) was more than the original model (3.39 ± 0.14 cm/s; stats = 131, $p < 0.05$).

We also did transfer learning with a ResNet-50 model and a MobileNet model (Supplementary Figure 4.7a), whose feature extraction parameters were pre-trained by the ImageNet and frozen during model training, whereas the parameters in the dense layers were updated throughout training (see Materials and Methods). We also built a feedforward neural network to extract features from raw microendoscopic images after a max pooling (Supplementary Figure 4.7a). These architectures with different combinations of hyperparameters were evaluated on the validation set. It is found that a ConvNet model had a less average L1 loss (0.1361) than transfer learning with a ResNet-50 (0.2370) or a MobileNet (0.2617) (Supplementary Figure 4.7b). This was probably because the ImageNet had very different image characteristics from our recordings. Another reason was that these computer vision models were developed to detect objects, but our goal was not to segment neurons but to extract embedded temporal information. Surprisingly, while batch normalization has been suggested to enhance model performance in the literature, we found our model without batch normalization had less L1 loss (0.1361) than using batch normalization (0.1450). We suspected that batch normalization removed fluctuations across samples, which might embed behavioral information, and thus compromised the decoding performance.

### 4.3.4 How behavioral states contributed to decoding errors

When comparing the decoding performance, we focused on periods when the animals were running at speeds above 5 cm/s (RUN), because previous studies have demonstrated that the hippocampal neurons had very different response profiles when the

animal was resting or moved at very slow speeds (STOP) (Ahmed and Mehta, 2012; Davidson et al., 2009; Geisler et al., 2007; Ólafsdóttir et al., 2017). Indeed, when a trained ConvNet was evaluated on STOP periods, the decoded trajectories were very different from the true trajectories, as if the animal was mentally simulating running (Supplementary Figure 4.4a, b). While this partly explained large prediction errors during STOP periods, whether we could use it as a tool to study hippocampal replays was beyond the scope of this study. Nevertheless, this result demonstrated how internal states of the animal (e.g., whether the animal was running or not) could contribute to decoding errors.

Several factors could contribute to decoding errors in our experiments. One possibility was that the internal running-speed threshold varies across animals, such that some of the STOP periods were mistaken as the RUN periods, leading to decoding errors. If this was the case, we would observe a negative correlation between decoding errors and running speeds. Another possible source for decoding errors originated from limited temporal precision in calcium imaging. In this case, we would observe a positive correlation between decoding errors and running speeds. The other possible factor came from sampling bias in the training set. If there was a bias in the spatial occupancy, i.e., certain positions were visited more by the animal, we would observe a bias in predictions toward these frequently-sampled positions.

To evaluate these factors, where decoding errors occurred in the maze as well as where they pointed to were visualized (Supplementary Figure 4.8a), and compared with the average occupancy (x-y Density) and speed (AvgSpeed) maps in both training and test sets (Supplementary Figure 4.8b). Take a session from T-maze exploration as an example (Supplementary Figure 4.8a, b), where decoding error pointed to was mostly correlated with AvgSpeed map of the training set ($r = 0.61$), relative to the test set ($r = 0.51$) and x-y Density (training set: $r = 0.38$; test set: $r = 0.34$). On the contrary, where decoding error occurred was mostly correlated with AvgSpeed map of the test set ($r = 0.81$), relative to the training set ($r = 0.62$) and x-y Density (training set: $r = 0.50$; test set: $r = 0.52$). This result was consistent across all the models (Supplementary Figure 4.8c). Where wrong predictions located were more

correlated with the AvgSpeed map of the training set ($r = 0.56 \pm 0.02$; stats = 134, $p < 0.05$), whereas where error occurred was mostly correlated with the AvgSpeed map of the test set ($r = 0.78 \pm 0.01$, blue; Wilcoxon signed-rank test: stats = 1, $p < 1e-5$)

### 4.3.5 Background residuals encode behavioral information

Given that our ConvNet model trained on raw microendoscopic recordings was able to decode positions and speeds better than a decoder trained on CNMFe-extracted neuronal signals, we hypothesized that neuropil/background residuals, often discarded in imaging studies, encode additional behavioral information than cell somata.

To evaluate how much behavioral information is embedded in different components in microendoscopic imaging, we generated several image sets (Figure 4.4a), including Clean images composed with cell somata only, Residual images contained signals after subtracting neuronal signals from raw data, and Hollow A images where spatial footprints of detected cells were occluded (see Materials and Methods). As revealed by an example test set (Figure 4.4b), both models trained on Raw and Clean images were able to decode behavioral variables $(x, y, v)$ across time. Surprisingly, models trained on Residual images were still able to predict behavioral traces (Figure 4.4b, c). This result was not unique to a single test set. We evaluated all the datasets from different maze exploration experiments (Figure 4.4d), and found decoding positions from Residual images had a median error ($14.93 \pm 1.04$ cm) comparable to decoding from Clean images ($11.03 \pm 0.57$ cm; stats = 165, $p = 0.16$). Models trained on Hollow ROI ($13.46 \pm 0.64$ cm) and Hollow A ($15.13 \pm 0.69$ cm) images had significantly less median decoding error than the chance level ($48.57 \pm 2.67$ cm; $p < 1e-5$).

If there was no extra information embedded in residuals compared to identified neurons, we would have observed a similar decoding performance between models trained on Raw and Clean images. Instead, decoding positions from Raw images, with no information excluded, had less median error ($14.93 \pm 1.04$ cm) than decoding from Clean images ($11.03 \pm 0.57$ cm; stats = 102, $p < 0.01$). In some cases, a model trained on Residual images was able to decode positions even better than using Clean

images (e.g., across H-maze exploration datasets: Residual: 10.87 ± 0.42 cm, Clean: 12.25 ± 0.43 cm; stats = 4, $p < 0.05$). These results suggest that additional position information can be embedded in neuropil/background residuals.

### 4.3.6 Decoding bias toward previous behavioral states

In previous experiments, predicted trajectories sometimes demonstrated a temporal mismatch to true trajectories. To quantify the amount of time-shifts, we computed cross-correlation between predicted and true trajectories (Figure 4.5a), and defined optimal frame-shift as the required number of frames to shift from predicted to true traces in order to reach largest correlation. Across H-maze and □-maze experiments, there was a significant negative frame-shift in decoding $x$ (H-maze: -2.00 ± 0.56, $p < 0.05$; □-maze: -1.60 ± 0.35, $p < 0.01$), $y$ (H-maze: -1.70 ± 0.54, $p < 0.05$; □-maze: -3.30 ± 0.41, $p < 0.01$), and $v$ (H-maze: -1.70 ± 0.99, $p < 0.05$; □-maze: -3.30 ± 0.73, $p < 0.01$) (Figure 4.5b). For T-maze experiments, there was an evident frame-shift in speed decoding, but not in position decoding ($x$: -0.20 ± 0.23, $p = 0.52$; $y$: -0.10 ± 0.22, $p = 0.65$; $v$: -0.90 ± 0.26, $p < 0.05$).

Given the nonlinear calcium dynamics, cross-correlation might not fully capture temporal bias in decoding. We further conducted a series of experiments by using the original images to predict behaviors at other timepoints (Figure 4.5c). Results showed that positions in H-maze, □-maze and T-maze were best decoded at 5 frames Before (7.68 ± 0.22 cm), 10 frames Before (8.77 ± 0.26 cm) and 1 frame Before (6.86 ± 0.39 cm), with median errors significantly less than decoding positions synchronized to image sampling (H-maze: 10.65 ± 0.26 cm, $p < 0.01$; □-maze: 12.39 ± 0.49 cm, $p < 0.01$; T-maze: 7.47 ± 0.36 cm; $p = 0.07$). These results indicated an overall bias in decoding toward previous behavioral states.

Next, we asked whether such decoding bias was similar between cell somata and neuropil residuals. If neuropil residuals are composed primarily of dendritic structures, we expect to decode more recent behaviors from the residuals compared to the cell somata, given that information is passed from the dendrites to cell somata. To test this hypothesis, we conducted a similar series of experiments in models trained

on Clean and Hollow ROI images (Supplementary Figure 4.9a, b). We found that the decoding lag of Clean images was earlier than the optimal decoding lag of Hollow ROI images across H-maze (Clean: -1.2 sec, Hollow ROI: -0.5 sec), □-maze (Clean: -1.0 sec, Hollow ROI: -0.8 sec), and T-maze (Clean: -0.1 sec, Hollow ROI: 0.0 sec). Recall that Clean images are composed of cell somata whereas Hollow ROI images are composed with neuropil residuals. These results support our original hypothesis that position information is present earlier in background residuals than cell somata.

### 4.3.7 Incorporating multiple frames improves decoding

The frame-shift experiment (Figure 4.5c) suggest that behavioral information at each timepoint is embedded across multiple frames. Therefore, we hypothesize that incorporating multiple frames into the input improves decoding performance. To test this hypothesis, we modified the ConvNet architecture to incorporate multiple frames in the input (see Materials and Methods for details).

When there was no constraint on convolutional kernels, we found that inputs with 51 frames (Supplementary Figure 4.10a, right) corrected temporal mismatch in predictions compared to single-frame input (Supplementary Figure 4.10a, left). Across all datasets, position and speed decoding improved significantly by incorporating multiple frames into the input (Supplementary Figure 4.10b). Specifically, using 11 frames (1-sec time window) had least median error in position prediction (single-frame: 10.17 $\pm$ 0.43 cm; 500-msec time window: 9.96 $\pm$ 0.41 cm; 1-sec time window: 7.50 $\pm$ 0.60 cm; 5-sec time window: 7.62 $\pm$ 0.19 cm; stats = 5, $p < 1e-5$).

We further modified the architecture by adding a constraint on convolutions across the temporal dimension (Figure 4.6a). Specifically, each frame went through the same block of convolutional layers whose weights were shared across frames. By learning attention weights attributed to these frames, information about the behavioral state was extracted from different timepoints. A similar decoding improvement by incorporating multiple frames into the input was observed (Figure 4.6b). By examining the distribution of learned attention weights for an input with 51 frames, we found that the imaging frame contributed most to decoding the behavioral state was delayed

(Figure 4.6c), suggesting future imaging frames carry more information about the behavioral states of the animal. These results are aligned with previous finding about decoding bias toward previous behavioral states when using a single-frame input.

### 4.3.8 Distinct functional ensembles are identified from raw microendoscopic images

One challenge in using any deep-learning model is to interpret what features are learned from the data. Our ConvNet model was able to decode the animal's positions and running speeds, but it was unclear how behavioral information was extracted from the microendoscopic images by the decoder. A decoder may use two possible strategies: randomly sampling pixels across the image, and identifying clusters with distinct temporal profiles.

To tease apart two strategies, we modified the Grad-CAM algorithm (Selvaraju et al., 2017) to identify saliency maps across different layers in the model (Figure 4.7a). Specifically, the saliency map of a target unit in a specific layer was estimated by a gradient-weighted combination of feature maps (see Materials and Methods for details). If the ConvNet decoder utilized the former learning strategy, the saliency maps would demonstrate uniform scattered patterns. On the other hand, if the latter strategy was used, the saliency maps were likely to identify ensembles composed with distinct combinations of cells and background features.

We found that in the early stage of the training, global background fluorescence were already segmented from and neuronal signals across layers, and the edges of cell clusters were partially detected in the middle convolutional layers (Figure 4.7b, top). When the model was well-trained (i.e., after 26000 steps), auto-decomposition was emerged across different layers (Figure 4.7b, bottom). Specifically, global background signals were segmented out in early layers of the model, but distinct ensembles of cells and background features were identified in the later layers of the model.

One might argue that a convolutional model can naturally learn object identification by extracting morphological features in an image, so the emergent decomposition

across model layers is simply a byproduct of a convolutional neural network. To test against this hypothesis, we compared saliency maps in a model where output labels were randomly assigned. If the hypothesis was true, we would expect similar saliency maps. Instead, a different decomposition process was formed when the relationship between input images and output labels were removed (Figure 4.7c), suggesting auto-decomposition that emerged from the ConvNet decoder was functionally relevant.

### 4.3.9 Differential encoding of the maze topology across layers

Given that distinct functional ensembles were identified by the decoder, we next examined how the maze topology was represented in these ensembles. As shown in Chapter 3, the intrinsic dynamics of a neural ensemble often occupy a low-dimensional manifold within the high-dimensional state space. Here, the neural ensemble consist of the activity states of artificial units in the decoder. To extract the representations from attended units in each layer of the trained model, we projected their activity into a 2-dimensional state-space using Isomap (see Materials and Methods).

To further explore how the maze topology was learned across layers in the ConvNet model, we mapped each neural state to the animal's true positions in the maze (Figure 4.8a, b). When the model was well-trained (Figure 4.8a), the projected activity accurately captured the intersection and each arm of the maze in the middle layers of the model, despite the distances between these neural states being unable to capture the actual maze topology. However, in the later layers of the model, projected neural states started to reflect the actual maze topology. We also visualized the ensemble representations in an early training stage whose validation error was twice of the fully trained model (Figure 4.8b).

To evaluate the similarity between each neural manifold and the maze topology, we computed the Pearson's correlation between pairwise distances in the neural manifold and pairwise distances in the physical maze. As shown in Figure 4.8c and d, the similarity between each neural manifold and the maze increased from the early layers to the later layers of a trained model (Early: $0.03 \pm 0.01$; Conv6: $0.29 \pm 0.03$; Conv7: $0.67 \pm 0.01$; Last: $0.87 \pm 0.01$; Dense: $0.91 \pm 0.01$), and was significantly

higher compared to models in an early stage of the training (Early: $0.03 \pm 0.01$, stats $= 98$, $p < 0.05$; Conv6: $0.26 \pm 0.03$, stats $= 117$, $p = 0.05$; Conv7: $0.57 \pm 0.03$, stats $= 102$, $p < 0.05$; Last: $0.70 \pm 0.03$; Dense: $0.74 \pm 0.03$, stats $= 1$, $p < 1e-5$) across different maze exploration experiments.

As discussed in previous sections, neural activity in the STOP periods could be very different from activity in the RUN periods. Therefore, we examined the neural manifolds of the decoder during STOP periods to evaluate how the maze topology was represented (Supplementary Figure 4.11a, b). Surprisingly, the maze topology was not completely lost during the STOP periods. Instead, a global maze topology was still captured with fuzzy local structures, with reduced similarities relative to manifolds during the RUN periods (Early: $0.07 \pm 0.02$, stats $= 62$, $p = 0.87$; Conv6: $0.31 \pm 0.03$, stats $= 26$, $p < 0.05$; Conv7: $0.67 \pm 0.02$, stats $= 27$, $p < 0.05$; Last: $0.80 \pm 0.02$, stats $= 1$, $p < 0.001$; Dense: $0.86 \pm 0.02$, stats $= 5$, $p < 0.01$). Overall, these results reveal that the decoder represents the animal's internal representations of the maze, by decomposing the maze topology across layers.

# Figures

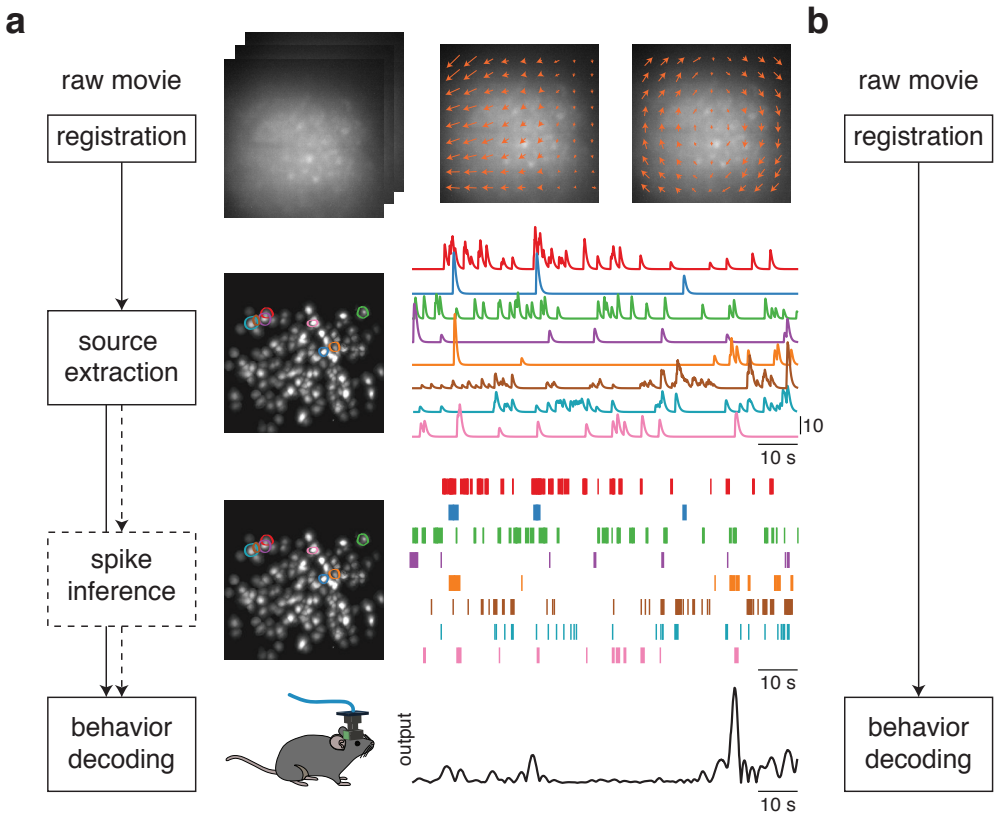



**Figure 4.1: Typical and proposed analysis pipelines for one-photon imaging data.** (**a**) Typical analysis pipeline. The data is first processed to remove motion artifacts by estimating a motion field from each data frame to a template (*top*). Due to limited axial resolution and low contrast in microendoscopic imaging, neurons can appear spatially overlapped and often require a video decomposition algorithm such as CNMFe (see Materials and Methods). After removing background signals, the locations of the neurons in the field of view (FOV) and their activities are extracted (*middle*). Spiking activity can then be inferred from each fluorescence trace. Finally, either spikes or fluroscence traces were used to decode the animal's behavior (*bottom*). Different steps of the pipeline are displayed on one animal in our experiment. (**b**) Proposed pipeline. Animal behavior is directly decoded from raw movies after removing motion artifacts.

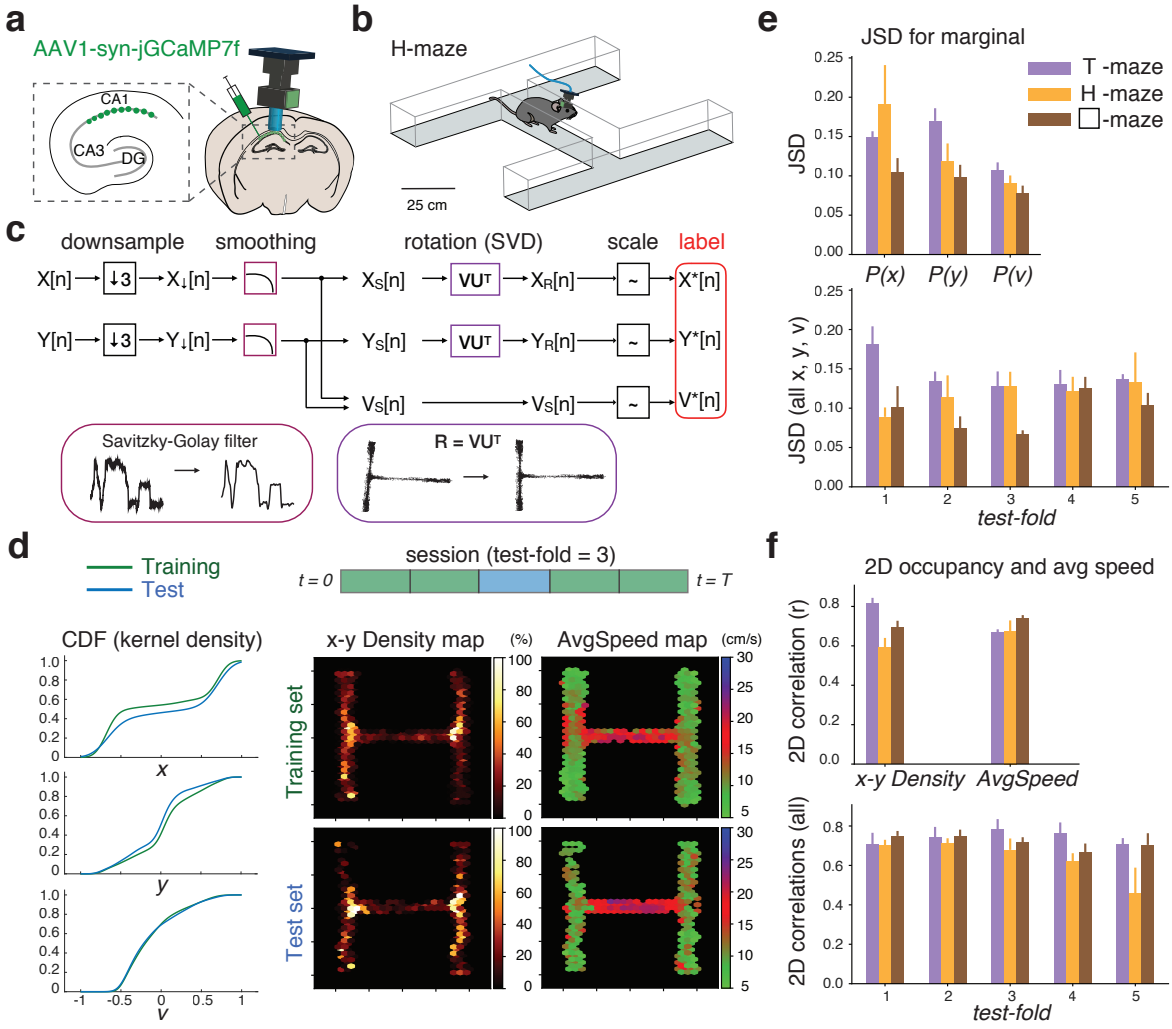



**Figure 4.2: Behavioral paradigm and data preparation.** (**a-c**) The schematic diagram of data preparation. Details see Materials and Methods. (**a**) Microendoscopic imaging in the CA1 expressing jGCaMP7f. (**b**) Diagram of the H-maze. (**c**) Label generation. Animal positions $(x, y)$ were aligned to imaging data sampled at 30 Hz, and downsampled to 10 Hz. A Savitzky-Golay filter was applied to smooth position coordinates and obtain speeds $(v)$. Position coordinates were rotated using a SVD method to reduce correlations between x and y. Finally, they were scaled to [-1, 1]. (**d**) The train/test split. Each dataset was split into 5 continuous periods, with one selected as the test set. The distributions of labels $(x, y, v)$ between the training (green) and test (blue) sets were compared. Left: The cumulative density function (CDF) and Jensen–Shannon divergence (JSD) between train and test distributions of each label for a typical dataset. JSD$(x)$ = 0.1428, JSD$(y)$ = 0.0951, JSD$(v)$ = 0.0619, n = 5265 (train) and 1692 (test). Right: Average occupancy (x-y Density) and speed (AvgSpeed) maps of both for a typical train/test split. r(x-y Density) = 0.7230, r(AvgSpeed) = 0.7678

**Figure 4.2** ***(previous page)*****:** (**e**) Similarity of marginal distributions between the training and test sets. Top: JSD for each label $(x, y, v)$ in different mazes. Bottom: JSD $(x, y, v)$ across test-folds in different mazes. (**f**) Similarity of average occupancy (x-y Density) and average speed (AvgSpeed) between the training and test sets. Top: 2D correlations of occupancy and speed maps in different mazes. Bottom: All 2D correlations (x-y Density, AvgSpeed) across test-folds.

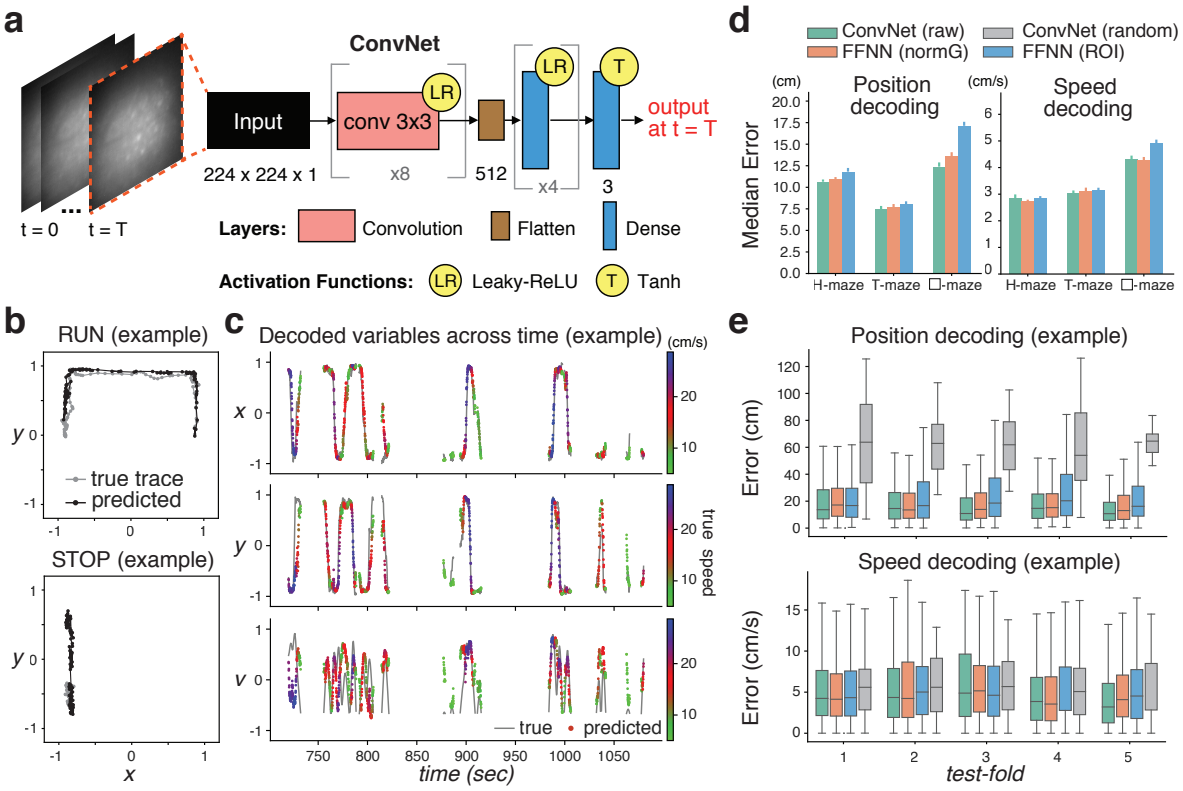



**Figure 4.3: Accurate decoding of kinematics from raw microendoscopic images.** (**a**) Schematic of ConvNet architecture. At each timepoint, an image was fed to a deep network consisting of 2D convolutional layers with leaky-ReLU activations, followed by dense layers with a regression head to decode position $(x, y)$ and speed $(v)$. (**b**) Example 10-sec trajectories of true (gray) and decoded (black) positions from an animal in the □-maze exploration experiment. Top: A 10-sec trajectory when the animal was running. Bottom: A 10-sec trajectory when the animal was stationary or running at a speed below 5 cm/s. (**c**) Decoded variables across time in an example test set. The gray line denotes true values. Dots denote predicted values, colored based on running speeds. The average decoding errors for x, y, and v were 7.84 cm, 12.70 cm, 6.23 cm/s, respectively. (**d**) Model performance across all datasets. Left: Our ConvNet model had less median error in predicting positions (10.17 ± 0.43 cm, green) than baseline decoders. They were all better the chance level (48.57 ± 2.67 cm; $p < 1e-5$). Right: Speed decoding in our ConvNet model (3.39 ± 0.14 cm/s, green) was comparable to a decoder trained on neuronal signals (3.36 ± 0.14 cm/s, blue; $p = 0.78$), and better than a decoder trained on ROI signals (3.62 ± 0.18 cm/s, orange; $p < 0.05$). All were better than the chance level (4.66 ± 0.19 cm/s; $p < 1e{-}5$). (**e**) Error distribution of position (*top*) and speed (*bottom*) decoding in an example dataset across test-folds. All p-values were obtained by Wilcoxon signed-rank test.

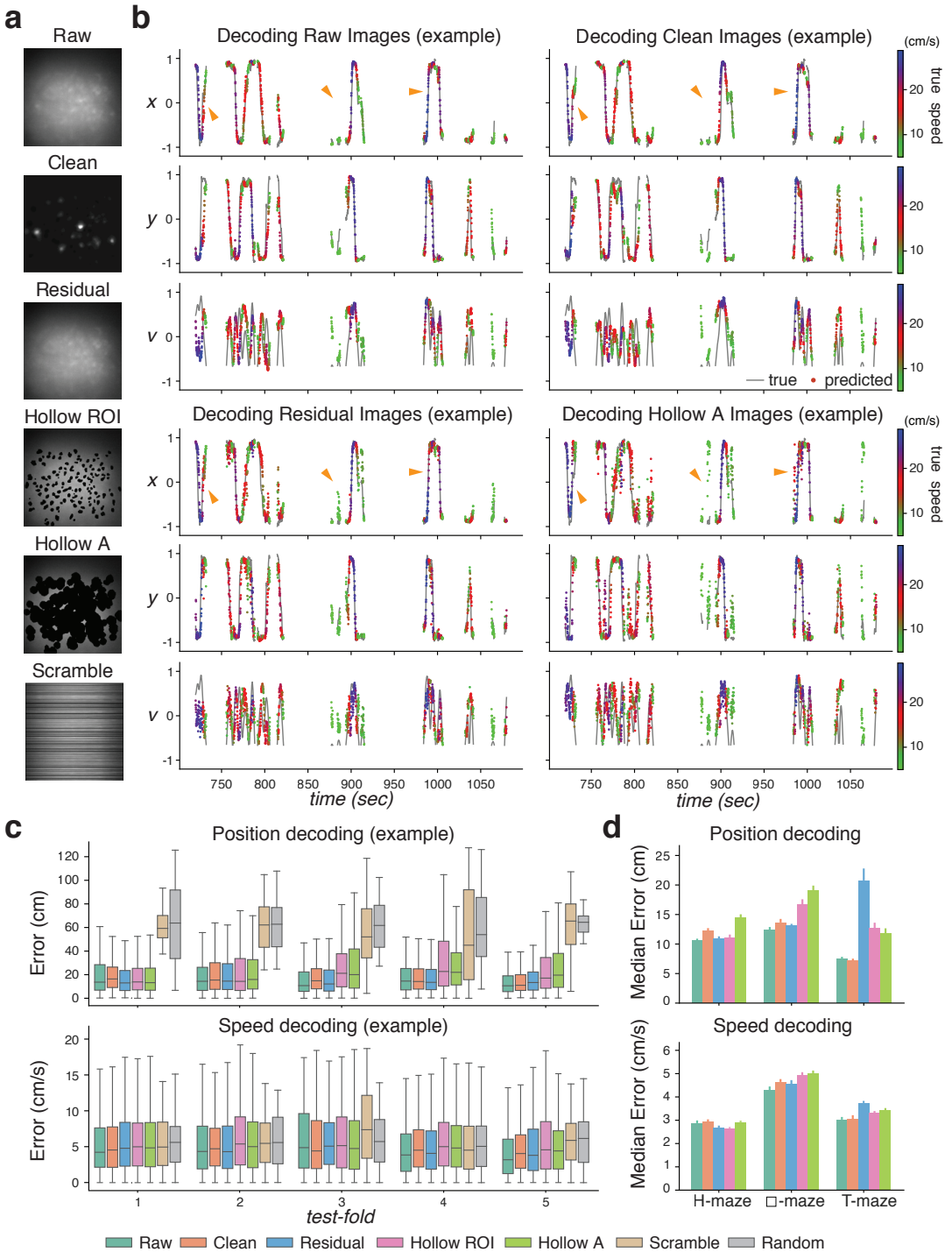



**Figure 4.4: Neuropil/background residuals encode behavioral information.** (**a**) Data generation using different components in calcium imaging data. Clean: Images composed with denoised neuronal signals only. Residual: Images after subtracting Clean from Raw. Hollow ROI: Images after removing ROIs from Residual. Hollow A: Images after removing CNMFe-detected cell footprints from Residual. Scramble: Pixels were randomly re-positioned per row. Details see Materials and Methods.

**Figure 4.4** ***(previous page)*****:** (**b**) Examples of variables decoded by using images prepared in **a** from the same dataset. Orange triangles mark timepoints where we can easily see prediction differences between models. The gray lines denote true traces. Dots denote predicted traces, colored based on running speeds. (**c**) Error distribution of position (*top*) and speed (*bottom*) decoding in an example dataset (same as **b**) across test-folds. (**d**) Model performance across all datasets. Top: Decoding positions using Residual images (14.93 ± 1.04 cm, blue) had more median error than using Raw images (10.17 ± 0.43 cm, green; $p < 0.01$), but comparable to using Clean images (11.03 ± 0.57 cm, orange; $p = 0.16$). Both Hollow ROI (13.46 ± 0.64 cm, pink) and Hollow A (15.13 ± 0.69 cm, lime) images had significantly less median decoding error than the chance level (48.57 ± 2.67 cm; $p < 1e-5$). Bottom: Decoding speeds using Residual images (3.66 ± 0.15 cm/s, blue) had more median error than using Raw images (3.39 ± 0.14 cm/s, green; $p < 0.05$), but similar to using Clean images on average (3.54 ± 0.16 cm/s, orange; $p = 0.46$). Both Hollow ROI (3.63 ± 0.18 cm/s, pink) and Hollow A (3.78 ± 0.17 cm/s, lime) images had significantly less median decoding error than chance level (4.66 ± 0.18 cm/s; $p < 1e-5$). All p-values were obtained by Wilcoxon signed-rank test.

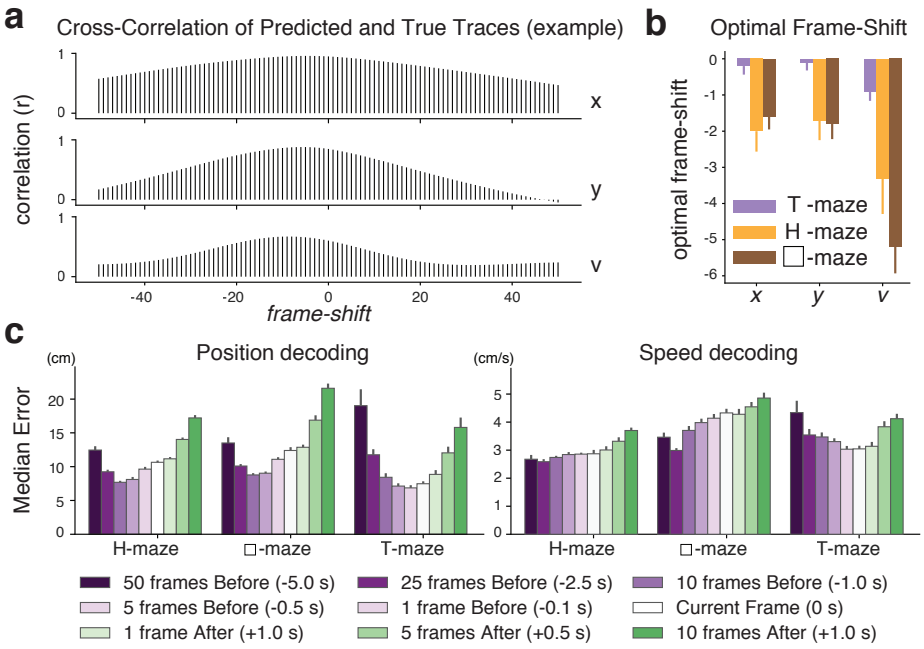



**Figure 4.5: Bias toward decoding previous behavioral states.** (**a-b**) Negative optimal frame-shift between predicted and true variables across time. (**a**) Cross-correlation of predicted and true behavioral traces in an example test set. The optimal frame-shift refers to the required number of frames to shift from predicted to true traces in order to reach largest correlation. The optimal frame-shift in $x, y$, and $v$ was -5, -6, -9, respectively. (**b**) Optimal frame-shift across all datasets. (**c**) Model performance using images to decode behaviors at other timepoints across all datasets. Left: Positions in H-maze, □-maze and T-maze were best decoded at 5 frames Before (7.68 ± 0.22 cm), 10 frames Before (8.77 ± 0.26 cm) and 1 frame Before (6.86 ± 0.39 cm), significantly different from no frame-shift (H-maze: 10.65 ± 0.26 cm, $p < 0.01$; □-maze: 12.39 ± 0.49 cm, $p < 0.01$; T-maze: 7.47 ± 0.36 cm; $p = 0.07$). Right: Speeds in H-maze, □-maze and T-maze were best decoded at 25 frames Before (2.57 ± 0.09 cm/s), 25 frames Before (2.97 ± 0.08 cm) and 1 frame Before (3.01 ± 0.11 cm/s), but mostly no significant difference from no frame-shift (H-maze: 2.85 ± 0.14 cm/s, $p = 0.11$; □-maze: 4.30 ± 0.15 cm/s, $p < 0.01$; T-maze: 3.03 ± 0.11 cm/s; $p = 0.80$). All p-values were obtained from Wilcoxon signed-rank test compared to no frame-shift.

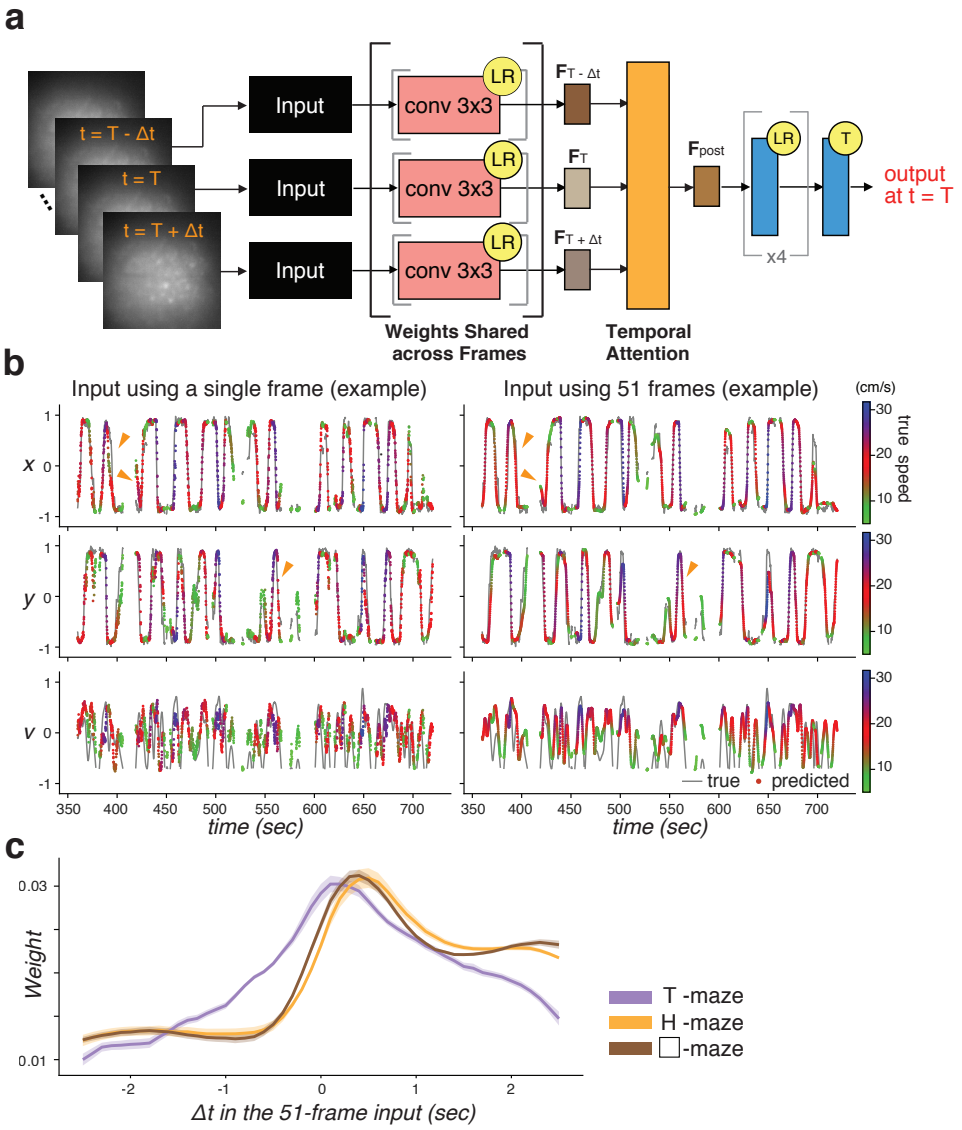



**Figure 4.6: Incorporating multiple frames into the model input improves decoding performance.** (**a**) Schematic diagram of ConvNet architecture modified to incorporate multiple frames into the input. Each frame in the input goes through a block of convolutional layers whose weights are shared across frames, and generates a frame-specific feature (e.g., $F_{T-\Delta t}$, $F_T$, and $F_{T+\Delta t}$). The attention layer (orange block) learns weights attributed to frames at different timepoints in the input. The weighted-average feature ($F_{post}$) is fed into the dense layers. (**b**) Examples of variables decoded from the same dataset, using a single-frame input (*left*) and 51 frames (*right*) when convolutions were shared across frames. Orange triangles mark timepoints where we can easily see prediction differences between models. The gray line denotes true traces. Dots denote predicted traces, colored based on running speeds. (**c**) The distribution of learned attention weights for an input with 51 frames ($\Delta t = \pm$ 2.5 seconds). Across all the datasets, the imaging frame contributed most to decoding each behavioral state was biased toward the future.

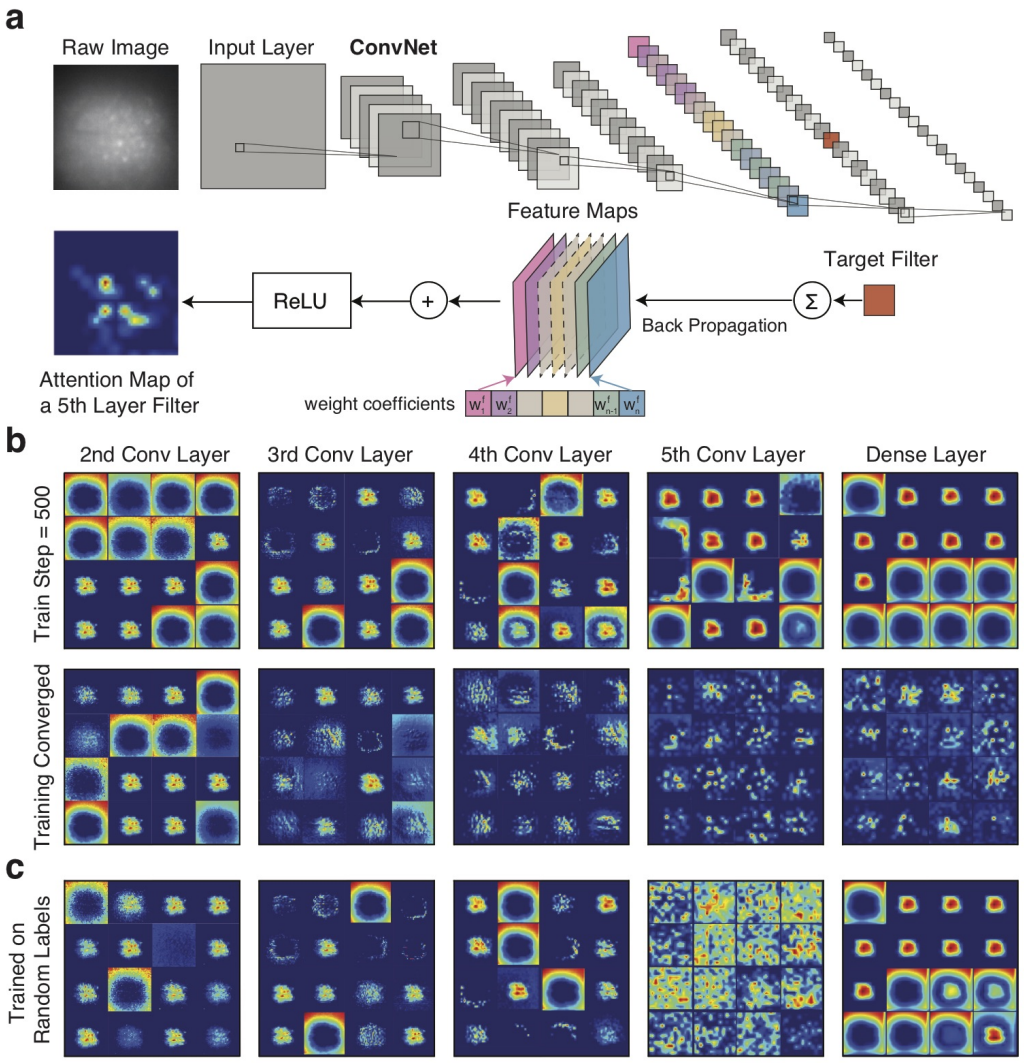



**Figure 4.7: Emergence of auto-decomposition and functional clusters.** (**a**) Schematic diagram of adapted Grad-CAM algorithm. We first back-propagate gradients of a target filter at a specific layer to the closest previous convolutional layer. The gradient-weighted combination of feature maps followed by a ReLU operation gives the estimated saliency map. Details see Materials and Methods. (**b**) saliency maps across different model layers. Each square box corresponds to saliency map of a typical unit/filter in that layer. For each layer, a example subset of Saliency maps were shown. Top: When the model was trained with 500 steps only, global background fluorescence and signals from cell somata or local neuropils were already segmented across layers, and object edges were detected based on brightness in the middle convolutional layers. Bottom: When the model was well-trained, auto-decomposition emerged across different layers. In early convolutional layers, similar global background signals were segmented out, whereas distinct ROIs were identified in the later convolutional layers. (**c**) Saliency maps across different model layers when output labels were randomly assigned. A different decomposition process was formed.

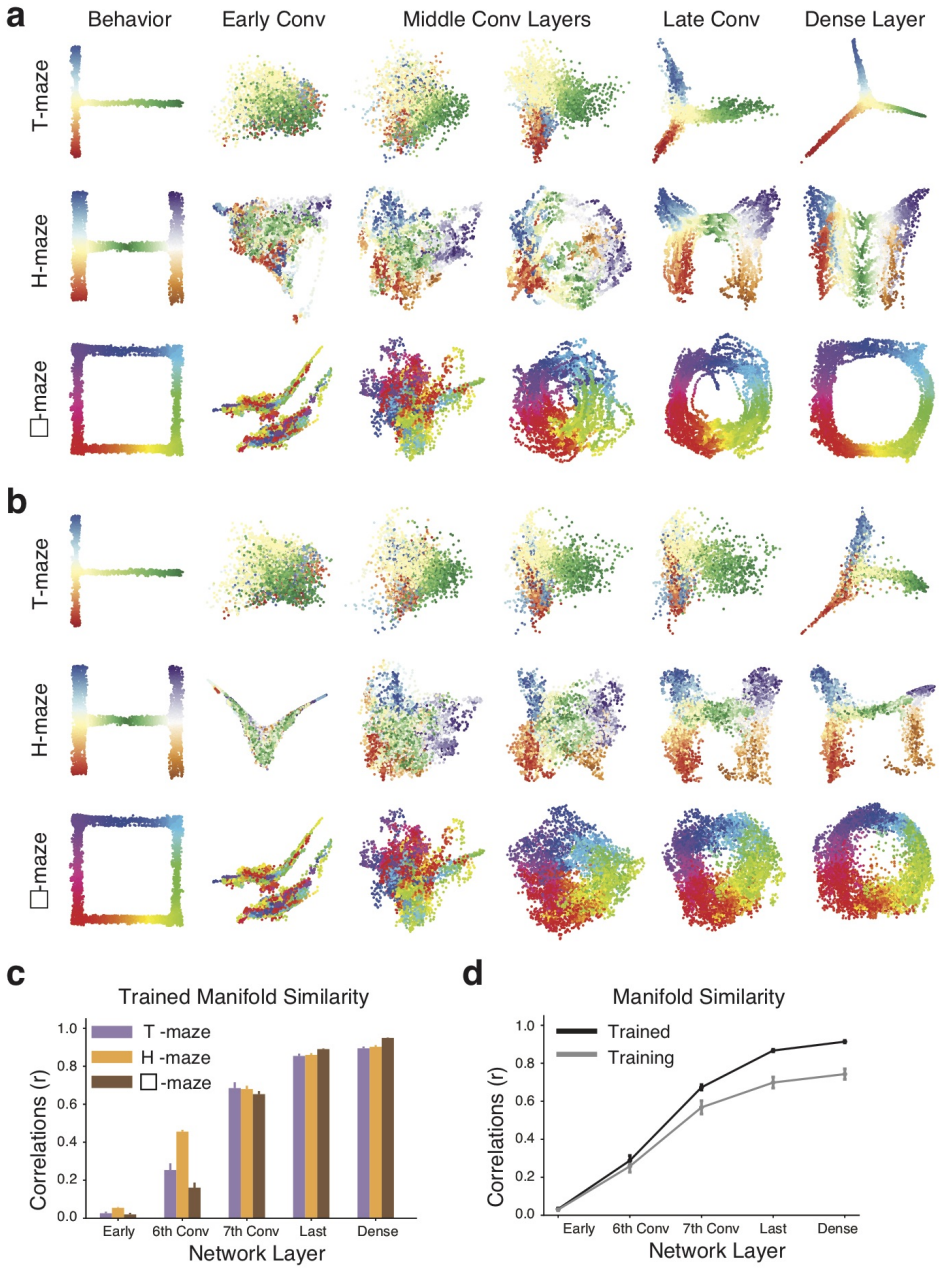
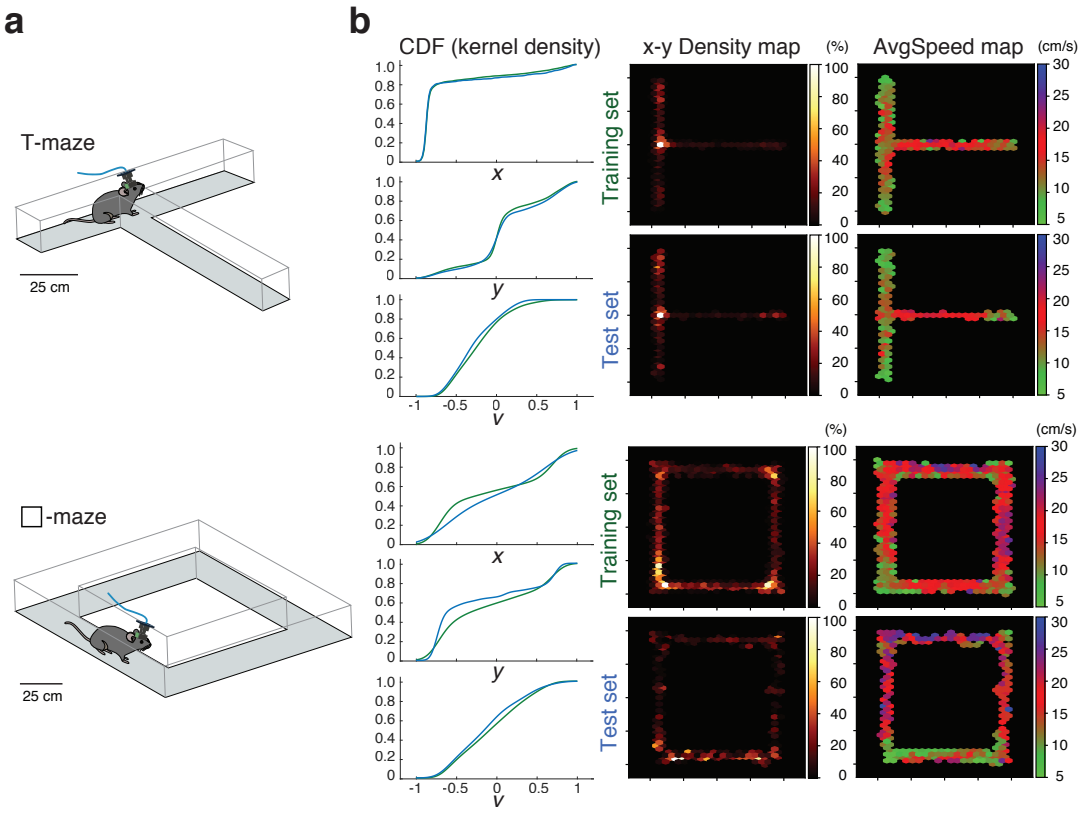


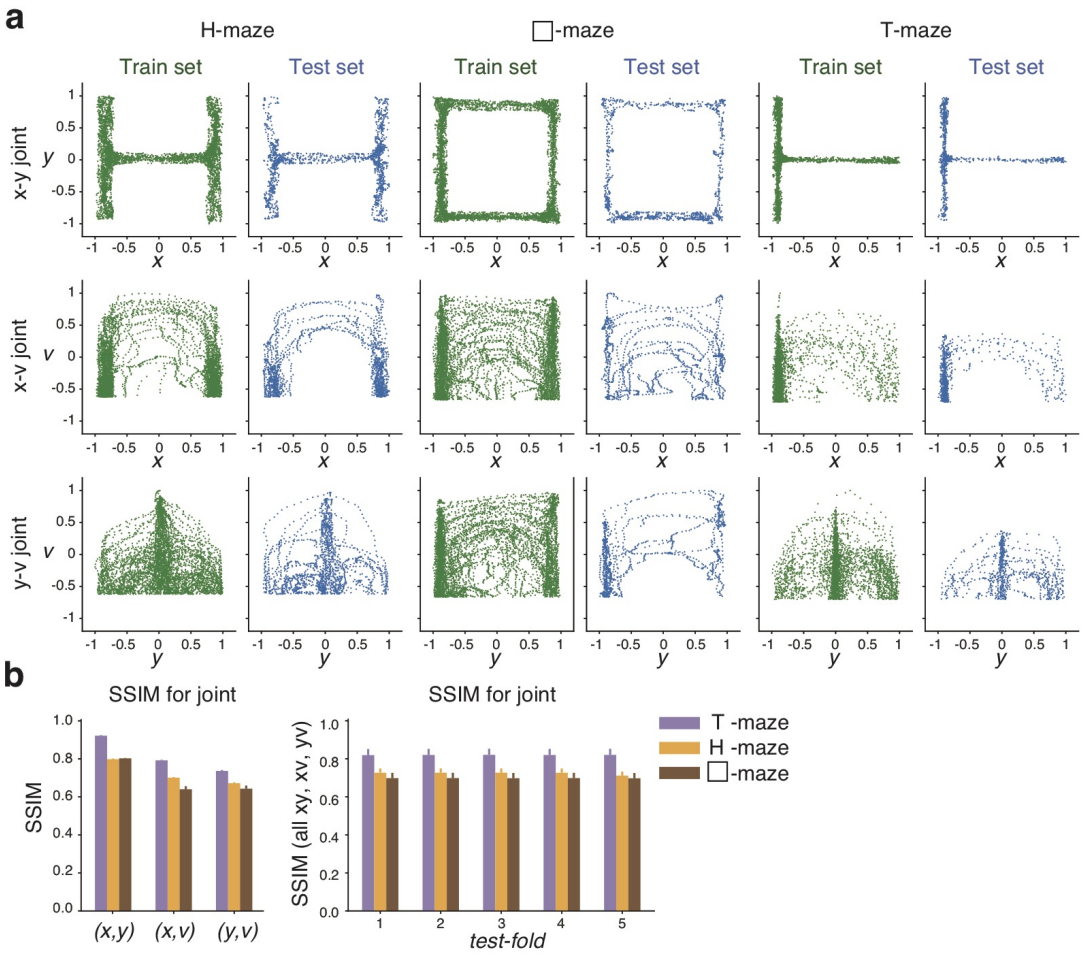

**Figure 4.8: Encoding of behavioral topology in manifolds across model layers.** (**a**) Ensemble representations from different layers of a trained model on a typical dataset in T-maze (*top*), H-maze (*middle*), and □-maze (*bottom*) experiments. We extracted the representations from attended units in each layer of the trained model, and projected their activities to a 2-dimensional state-space using Isomap. Each dot represents a behavioral state, colored by the animal's true positions in the maze. The 1st Column: Position distributions from a typical animal. Other columns: The topology of the animal's positions started to form at later layers. Details see Materials and Methods. (**b**) Ensemble representations from different layers of a model in an early training stage whose validation error was twice of the fully trained model.

**Figure 4.8** ***(previous page)*:** (**c**) Similarity between neural manifolds and behavioral topology. Similarity was defined as the correlation between pairwise distances in a neural manifold and pairwise distances in behavioral topology. Similarity increased significantly from the 6th convolutional layer to the 7th convolutional, and further increased in the last convolutional layer as well as the dense layer, across T-maze (purple), H-maze (gold) and □-maze (brown). (**d**) Comparison between neural manifolds in well-trained models (black) and models whose validation errors were about twice (gray) of the well-trained model (black). Across all datasets, manifold similarity was significantly higher in the trained models (black) than models under training (gray) at later layers (Wilcoxon signed-rank test: stats $= 1$, $p < 1e-5$, whereas such difference was less yet still significant at earlier layers (Wilcoxon signed-rank test: stats $= 98$, $p < 0.05$).

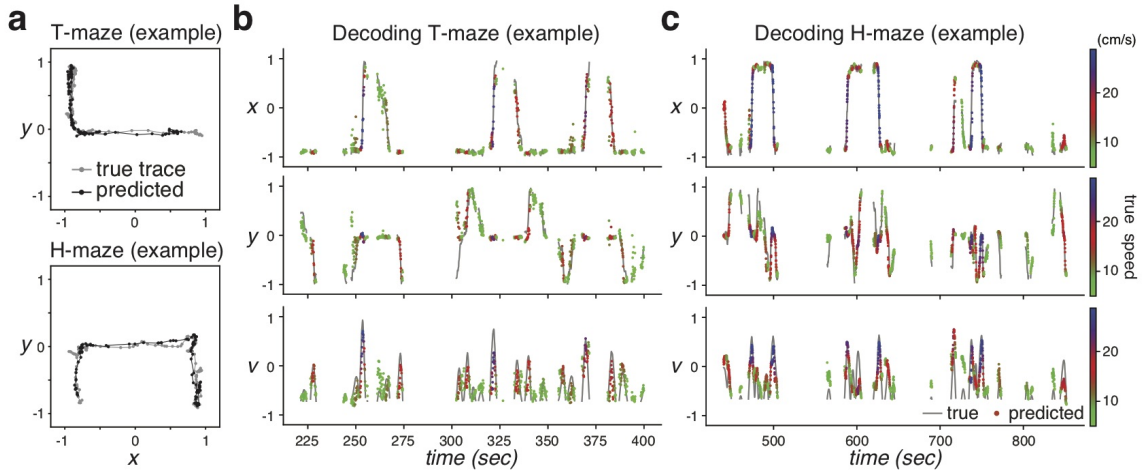
# Supplementary Figures



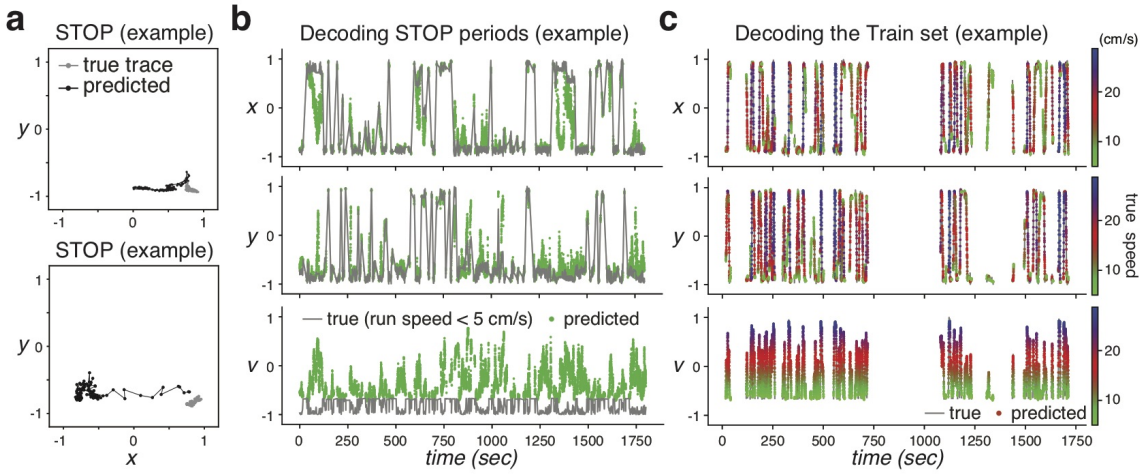

**Supplementary Figure 4.1: Examples of T-maze and □-maze behavior.** (**a**) Schematic diagrams for T-maze *(top)* and □-maze *(bottom)* exploration experiments. (**b**) Evaluation of train/test split for T-maze *(top)* and □-maze *(bottom)* exploration. Left: The cumulative density function between train and test distributions of each label. A typical T-maze dataset: JSD($x$) = 0.1207, JSD($y$) = 0.0998, JSD($v$) = 0.1179, n = 3320 (train) and 903 (test). A typical □-maze dataset: JSD($x$) = 0.1747, JSD($y$) = 0.2187, JSD($v$) = 0.0727, N = 5899 (train) and 1221 (test). Right: Average occupancy (x-y Density) and speed (AvgSpeed) maps of both for a typical train/test split. Same T-maze dataset: r(x-y Density) = 0.9154, r(AvgSpeed) = 0.7379. Same □-maze dataset: r(x-y Density) = 0.5232, r(AvgSpeed) = 0.6612.



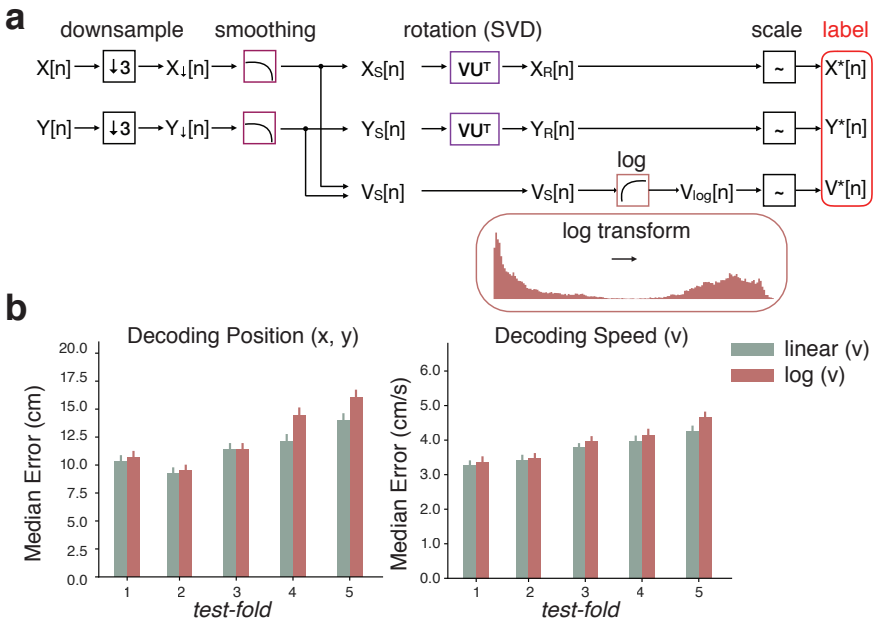

**Supplementary Figure 4.2: Consistency of joint distributions of output labels between training and test sets.** (**a**) Evaluation of train/test split using structural similarity index measure (SSIM) on joint distributions of output labels. Left: A typical H-maze dataset. SSIM$(x, y)$ = 0.7975, SSIM$(x, v)$ = 0.7006, SSIM$(y, v)$ = 0.6576. Middle: A typical □-maze dataset. SSIM$(x, y)$ = 0.7949, SSIM$(x, v)$ = 0.5862, SSIM$(y, v)$ = 0.5859. Right: A typical T-maze dataset. SSIM$(x, y)$ = 0.9222, SSIM$(x, v)$ = 0.8031, SSIM$(y, v)$ = 0.7229. (**b**) Similarity of joint distributions between the training and test sets. Left: SSIM for each paired variables $(x, y)$, $(x, v)$ and $(y, v)$ in different mazes. Right: SSIM ($(x, y)$, $(x, v)$ and $(y, v)$) across test-folds in different mazes.



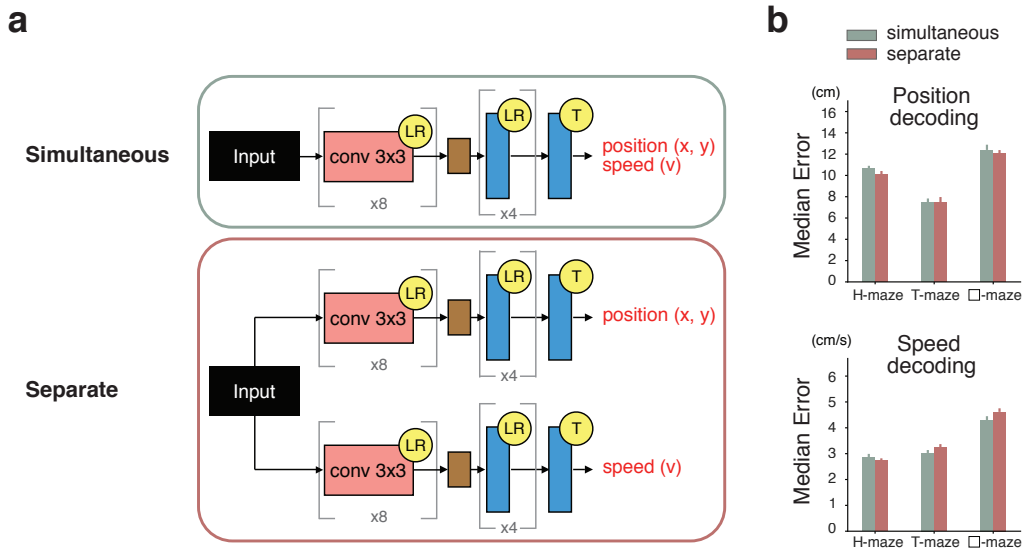

**Supplementary Figure 4.3: Examples of decoded running traces in the T-maze and the □-maze.** (**a**) Top: Example 10-sec trajectories from an animal in a T-maze exploration experiment. Bottom: Example 10-sec trajectories from an animal in a H-maze exploration experiment. (**b**) Decoded variables across time in an example T-maze test set. Average error in decoding x, y, and v: 2.95 cm, 6.15 cm, 3.12 cm/s, respectively. (**c**) Decoded variables across time in an example H-maze test set. Average error in decoding x, y, and v: 5.53 cm, 9.94 cm, 3.65cm/s, respectively.



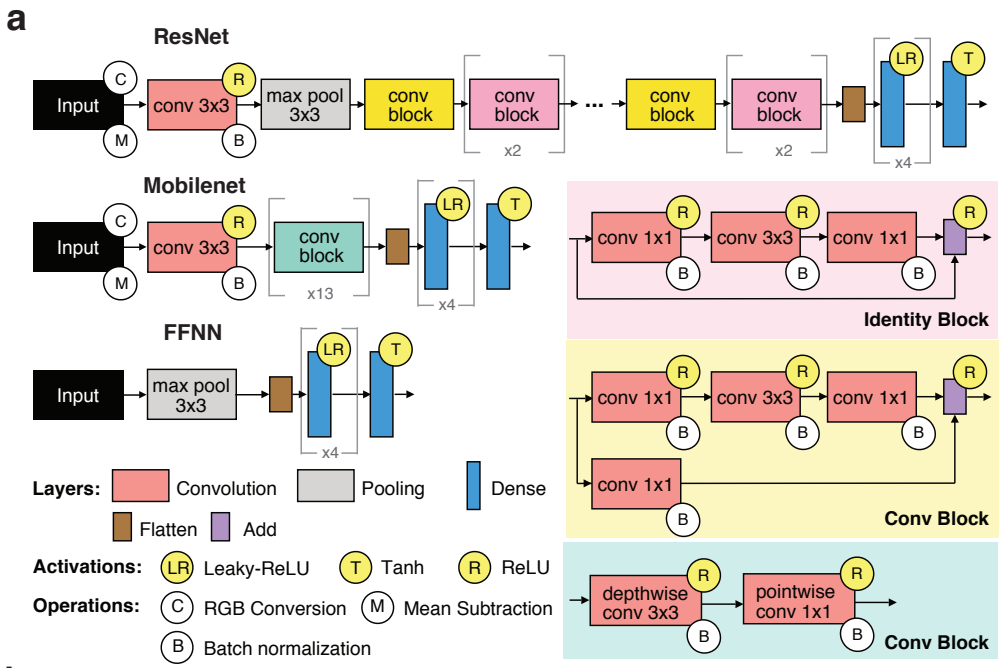

**Supplementary Figure 4.4: Examples of decoded traces in STOP periods and the training set.** (**a**) Example 10-sec trajectories of true (gray) and decoded (black) positions when the animal was mostly stationary. Decoded traces formed a sequence as if the animal was simulating running. (**b**) Decoded variables across time during STOP periods in an example test set. Gray line denotes true traces. Green dots denote predicted traces. (**c**) Decoded variables across time in an example training set. Dots denote predicted traces, colored in actual running speeds.




**Supplementary Figure 4.5: Decoding performance with log transformation on speeds.** (**a**) Schematic diagram of data preparation with speeds being log-transformed to create a log-uniform distribution. (**b**) Model performance across a subset of datasets across different test-folds. Left: The original model trained with linear speeds (11.49 ± 0.13 cm, sage) had significantly less median error in predicting positions than the same model trained with log speeds (12.45 ± 0.14 cm, cardinal; Wilcoxon signed-rank test: stats = 15274374, $p < 1e-5$). Right: The original model trained with linear speeds (3.76 ± 0.03 cm/s, sage) had significantly less median error in predicting speeds than the same model trained with log speeds (3.95 ± 0.04 cm/s, cardinal; Wilcoxon signed-rank test: stats = 15506009, $p < 1e-5$).




**Supplementary Figure 4.6: Network performance with separate decoding streams.** (**a**) Schematic diagram of ConvNet architectures with original simultaneous (top) and separate (bottom) decoding strategies. (**b**) Model performance across all datasets. Top: The original model with simultaneous decoding (10.17 ± 0.43 cm, sage) had significantly more median error in predicting positions than the model with separate decoding streams (9.85 ± 0.42 cm, cardinal; Wilcoxon signed-rank test: stats = 133, $p < 0.05$). Bottom: The original model (3.39 ± 0.14 cm/s, sage) had significantly less median error in predicting speeds than the model with separate decoding streams (3.53 ± 0.16 cm/s, cardinal; Wilcoxon signed-rank test: stats = 131, $p < 0.05$).



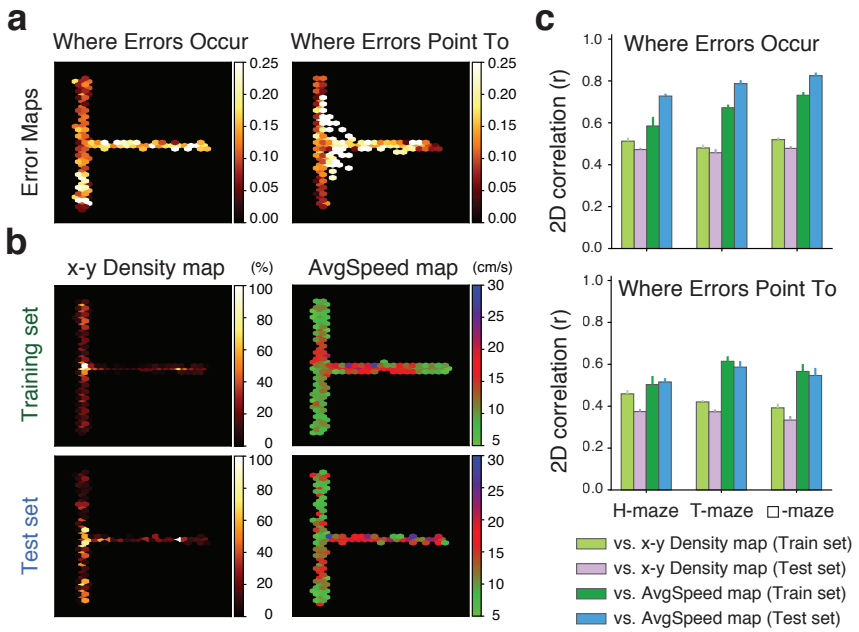

**b** **Model Comparison on Validation set**

| Models | #Dense Layers | Avg *L1* Val Loss |
|---|---|---|
| **ConvNet** | 2 | 0.1366 |
| | **5** | **0.1361** |
| ResNet-50 | 2 | 0.2558 |
| | 5 | 0.2370 |
| MobileNet | 2 | 0.2931 |
| | 5 | 0.2617 |
| FFNN | 2 | 0.2116 |
| | 5 | 0.1606 |

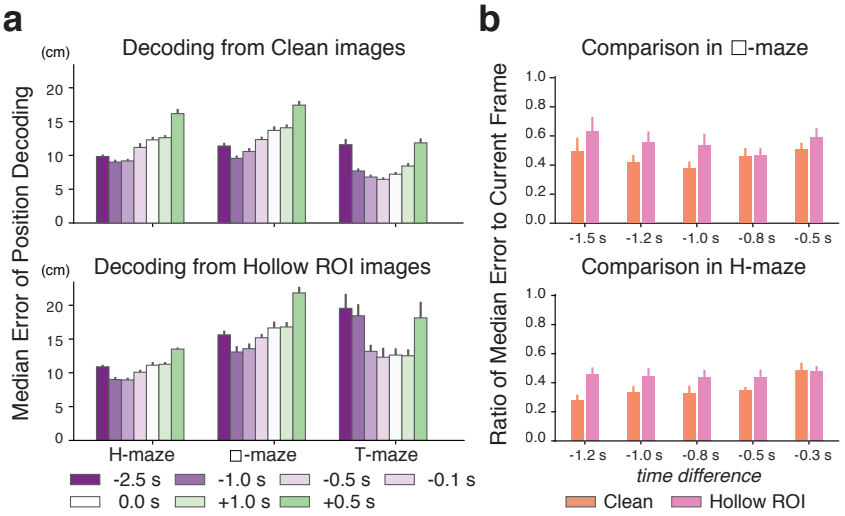
**c** **Hyperparameter Searching of ConvNet (8 Convolution + 5 Dense Layers)**

| Learning Rate | Norm | Avg *L1* Val Loss |
|---|---|---|
| 0.1 | Batch Norm | 0.1523 |
| | None | 0.5472 |
| 0.01 | Batch Norm | 0.1420 |
| | None | 0.5375 |
| 0.001 | Batch Norm | 0.1366 |
| | None | 0.1370 |
| **0.0001** | Batch Norm | 0.1450 |
| | **None** | **0.1361** |

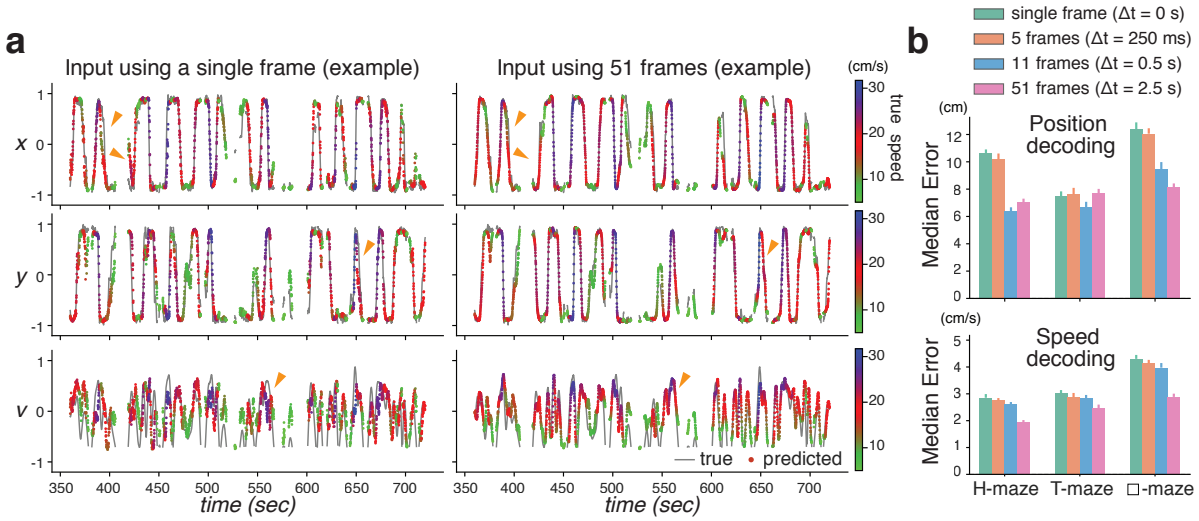
**Supplementary Figure 4.7: Model selection with hyperparameter searching.** (**a**) Schematic diagram of different model architectures. We did transfer learning with a ResNet model (top) and a Mobilenet model (middle) pre-trained on Imagenet. We also designed a feedforward neural network, where max pooling was applied on input images before feeding flattened features to the dense layers. Hyperparamter searching was applied to all models and the best combination was selected based on validation performance. Details see Materials and Methods. (**b**) Model comparison on validation set. We compared all models with their best hyperparameter sets, and found the ConvNet model had the least average L1 loss. (**c**) Subsets of hyperaparameter searching on validation set. For a ConvNet model with 8 convolution with 8 convolutional layers and 5 dense layers, we found training model with learning rate at 1e-4 without batch normalization had least average L1 loss. These hyperparameters were used for models in all the experiments.



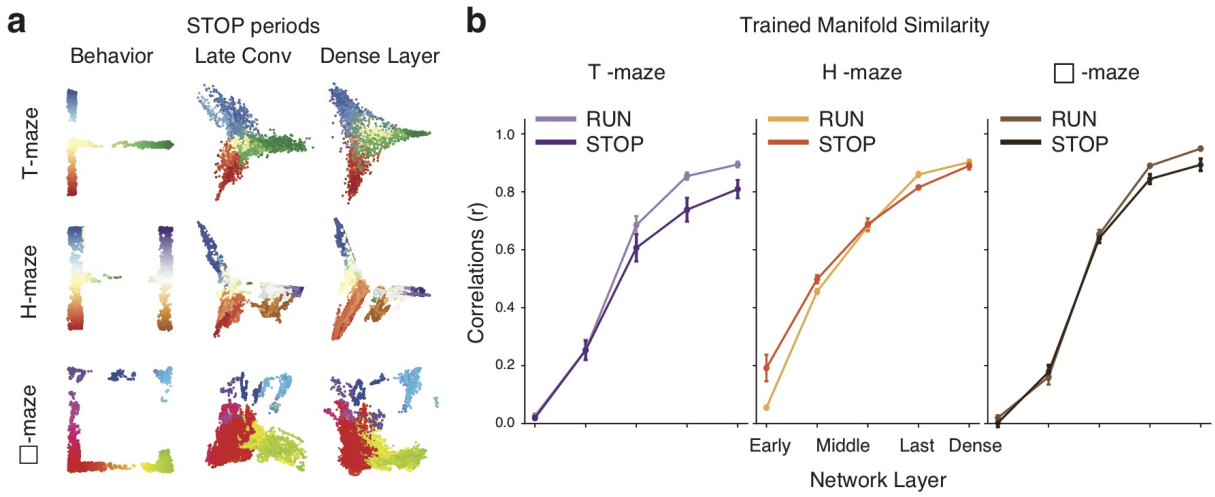

**Supplementary Figure 4.8: Possible factors contributing to decoded errors.** (**a**) Error visualizations in the space from an example T-maze exploration dataset. Heatmaps represent decoded errors (a.u.) before converting back to the original scales (cm and cm/s). Left: Spatial maps of where decoded errors occurred. Right: Spatial maps of where errors pointed to (i.e., where wrong predictions were). (**b**) Left: Average occupancy (x-y Density) maps of the training set *top* and the test set *bottom* from the example dataset in **a**. Right: Average speed (AvgSpeed) maps of the training set *top* and the test set *bottom* from the example dataset in **a**. 2D correlations ($r$) between these maps to the error maps were computed to evaluate potential contributing factors for decoded errors. Here, where error pointed to was mostly correlated with the AvgSpeed map of the training set ($r = 0.61$, right, *top*), and where error occurred was mostly correlated with the AvgSpeed map of the test set ($r = 0.81$, right, *bottom*). (**c**) Across all datasets, where error pointed to was more correlated with the AvgSpeed map of the training set ($r = 0.56 \pm 0.02$, green; Wilcoxon signed-rank test: stats $= 134$, $p < 0.05$) relative to other factors, but overall the correlations were mild. On the other hand, where error occurred was mostly correlated with the AvgSpeed map of the test set ($r = 0.78 \pm 0.01$, blue; Wilcoxon signed-rank test: stats $= 1$, $p < 1e-5$)




**Supplementary Figure 4.9: Earlier optimal decoded states in cell somata.** (**a**) Comparison of position decoding using Clean (*top*) and Hollow ROI (*bottom*) images across time shifts. Optimal decoded states refer to the time-shift reaching least decoding error. Clean images were composed of mostly cell somata, whereas Hollow ROI images mostly contained neuropil residuals. (**b**) Zoom-in decoding comparison in □-maze (*top*) and H-maze (*bottom*). Median decoding error in each dataset was normalized by the median error decoding positions without time-shifts (i.e., 0.0 s). A ratio of 1 refers to no difference compared to decoding original positions. Optimal decoded positions have smallest ratios of errors. For □-maze, optimal decoded state of Clean images (orange, -1.0 s: 0.70 ± 0.03, -0.8 s: 0.75 ± 0.03; stats = 1, $p < 0.01$) was earlier than that of HollowROI images (pink, -1.0 s: 0.79 ± 0.04, -0.8 s: 0.76 ± 0.03; stats = 14, $p = 0.17$). Similarly, for H-maze, optimal decoded state of Clean images (-1.2 s: 0.70 ± 0.03, -0.3 s: 0.85 ± 0.04; stats = 1, $p < 0.01$) was earlier than that of Hollow ROI (-1.2 s: 0.83 ± 0.03, -0.3 s: 0.84 ± 0.03; stats = 21, $p = 0.51$). All p-values were obtained by Wilcoxon signed-rank test.




**Supplementary Figure 4.10: Optimal decoding using a 1-second input and convolutions independent of time.** (**a**) Examples of variables decoded from the same dataset, using a single-frame input (*left*) and 51 frames (*right*) when convolutions were independent of time. Orange triangles mark timepoints where we can easily see prediction differences between models. The gray line denotes true traces. Dots denote predicted traces, colored based on running speeds. (**b**) Model performance across all datasets. Top: Incorporating 11 frames (time window of 1 s) led to least median position-decoding error (7.50 $\pm$ 0.60 cm), significantly different from single frame input (10.17 $\pm$ 0.43 cm; stats = 5, $p < 1e-5$). Bottom: Incorporating 51 frames (time window of 5 s) led to least median speed-decoding error (2.42 $\pm$ 0.09 cm/s), significantly different from single frame input (3.39 $\pm$ 0.14 cm/s; stats = 1, $p < 1e-5$).




**Supplementary Figure 4.11: Encoding of behavioral topology during STOP periods in manifolds across model layers.** (**a**) Ensemble representations from the last convolutional layer and the dense layer of a trained model on a typical dataset in the T-maze (*top*), H-maze (*middle*), and □-maze (*bottom*) (same models as in **Figure 4.8a**). (**b**) Comparison between neural manifolds extracted from RUN and STOP periods (i.e., animals were either stationary or running at speeds below 5 cm/s), in the T-maze (*left*), H-maze (*middle*), and □-maze (*right*). Across all datasets, manifold similarity was significantly higher when the animals were running (Wilcoxon signed-rank test: stats $= 1$, $p < 0.05$).

## 4.4 Discussion

### Information embedded in background residuals

Our findings demonstrate that behavioral variables such as position and speed can be directly decoded from raw microendoscopic images, without a need to specifically identify putative neurons and/or deconvolve spikes. Our ConvNet-based approach outperforms a feedforward neural network that takes the denoised calcium traces as input, suggesting that neuropil/background, which is filtered in a typical calcium analysis pipeline, encodes behavioral information. One might argue that this performance difference is driven by different decoder architectures or input dimensions. However, by generating images composed of isolated background and foreground components as input for the ConvNet decoder, we showed that the residual signals encode the animal's position and speed. By occluding spatial footprints of the neurons, with mostly neuropil signals remained, we excluded the possibility of successful decoding being a byproduct of incomplete source extraction within the ROIs. Our results indicate that neuropil/background in single-photon calcium imaging in the hippocampal CA1 contains spatial information.

Neuropil signals are considered to be out-of-focus fluorescence signals from nearby cellular processes including dendritic spines and axonal segments (Gobel and Helmchen, 2007; Kerr et al., 2005; Ohki et al., 2005), and are often removed in a calcium imaging analysis pipeline (Keemink et al., 2018; Lu et al., 2018; Pachitariu et al., 2017; Pnevmatikakis et al., 2016; Zhou et al., 2018). However, studies have shown that dendritic fluorescence can exhibit calcium transients in response to synaptic inputs, back-propagating action potentials and dendritic spikes (Kleindienst et al., 2011; Larkum et al., 1999; Svoboda et al., 1999; Takahashi et al., 2012; Yuste and Denk, 1995), which are important for information processing. For example, dendritic spikes in the sensory cortex can contribute to the orientation selectivity (Euler et al., 2002; Sivyer and Williams, 2013; Smith et al., 2013) and angular tuning (Lavzin et al., 2012). In addition, dendritic signals in the CA1 can modulate the place fields and induce novel place field formations to produce feature selectivity (Bittner et al., 2015;

Sheffield and Dombeck, 2015).

Similarly, axonal calcium imaging has shown that individual axons can encode task-related variables such as touch and whisking (Petreanu et al., 2012). Together, these suggest that neuropil signals, comprising both dendritic and axonal activities, encode external stimuli and behavioral information. This hypothesis is supported by our results, consistent with previous studies that demonstrate both cell somata and neuropil patches can represent goal-directed behaviors (Allen et al., 2017) as well as orientations of the moving gratings (Lee et al., 2017). To isolate the contributions from dendrites, axons, and out-of-focus cell somata in the background signals, a calcium indicator which specifically localizes to the soma (e.g., SomaGCaMP (Shemesh et al., 2020)) or the axon (e.g., axon-GCaMP6 (Broussard et al., 2018)) will be needed for future experiments. Overall, our study challenges the idea of removing background signals from the calcium imaging data, and proposes to re-examine the analysis pipeline for extracting behavioral information from microendoscopic data.

## Asymmetric calcium dynamics and temporal bias in decoding

Our ConvNet decoder predicts behavioral variables corresponding to each frame. Given the onset delay of a calcium indicator in response to a spike (Chen et al., 2013; Dana et al., 2019), it is reasonable to expect a temporal mismatch between each frame and its corresponding animal behavior. By comparing the single-frame decoding performance at different time shifts in behavioral data, we found an optimal position-decoding lag of 1 second on average. For speed-decoding, the temporal shift varied largely depending on the dataset. However, the jGCaMP7f used in our study has a rise-time of about 56 msec (Dana et al., 2019), and therefore cannot explain the optimal temporal-shift in decoding given our 10 Hz image sampling rate. One caveat in this interpretation is that the reported characteristics of a GCaMP sensor were quantified based on *in vitro* measurements, whereas its *in vivo* kinematics might be different depending on the brain regions and neuronal types (Akerboom et al., 2012; Greenberg et al., 2018). Another factor to consider is the nonlinear relationship between the rising time of fluorescence and the number of action potentials. For

example, jGCaMP7f exhibits a wide range of times to peak fluorescence in response to presynaptic stimulation with different frequencies (Dana et al., 2019). The other possible cause for variable optimal temporal shifts in decoding is the variable contribution of different signal sources across datasets. Because dendritic, axonal, and somatic dynamics are different, each dataset can have different rising dynamics depending on the dominant signal source. Furthermore, previous studies have revealed retrospective and prospective representation of space (Ferbinteanu and Shapiro, 2003) as well as speed (Iwase et al., 2020) in the hippocampus, and thus the sampling of these representations can bias the overall spatial and speed coding. All these factors can account for the wide variance in optimal decoding lags across datasets in this study.

## Contributions from our end-to-end decoding paradigm

In addition to uncovering information embedded in imaging background, our work does not need sophisticated hyperparameter tuning that requires prior knowledge, and provides an easy-to-implement framework for decoding behavior from imaging data. Many methods have been proposed to decode position from activities of a group of single neurons in the hippocampus in the literature, but they require either spike sorting for electrophysiology recordings (Brown et al., 1998; Wilson and McNaughton, 1993; Zhang et al., 1998) or spike inference for calcium imaging data (Etter et al., 2020; Gonzalez et al., 2019; Shuman et al., 2020; Stefanini et al., 2020; Ziv et al., 2013). Although alternative decoding methods for calcium imaging have been developed to bypass spike deconvolution, such as estimating neuronal firing rates directly (Ganmor et al., 2016), approximating spiking with marked point process (Tu et al., 2020), or decoding using frequency representations (Frey et al., 2019), they all require extracting fluorescence traces from identified individual neurons. On the other hand, new statistical methods are developed to decode position with unsorted spikes or local field potentials in electrophysiology (Cao et al., 2019; Deng et al., 2015; Frey et al., 2019; Kloosterman et al., 2014). However, this idea has never been applied in single-photon calcium imaging. This study is the first demonstration to

directly decode behaviors from raw microendoscopic data without the need to demix individual neuronal signals.

Our approach benefits from the power of deep learning. With recent advances in machine learning, there have been efforts to use deep learning models such as a recurrent neural network to learn hidden features from neural data and decode animal behavior with high accuracies (Glaser et al., 2020; Tampuu et al., 2019). However, these efforts still rely on the identification of single neurons and unfortunately, fail to take full advantage of deep learning methods to extract information from noisy data. Our end-to-end decoder embeds all preprocessing steps into the network. In addition, we adapted the Grad-CAM algorithm (Selvaraju et al., 2017) to find saliency maps across layers of the decoder. We chose Grad-CAM given its reduced visual artifacts (Adebayo et al., 2018) and smoother saliency maps compared to other approaches such as Vanilla Gradients (Simonyan et al., 2013), Guided Backpropagation (Springenberg et al., 2014) and Deconvolution (Zeiler and Fergus, 2014). However, this method does have a limitation for capturing fine-grained detail. Nevertheless, our results showed that video decomposition automatically emerges in a trained decoder, and identified clusters composed of different groups of cells and neuropil patches for spatial representations in the CA1. We provide the field an alternative way to classify neural activities by visualizing the saliency maps of the decoder, without a need to run an extra clustering algorithm.

In conclusion, we have (1) demonstrated an efficient and easy-to-implement decoder for raw microendoscopic data, (2) revealed extra behavioral representations embedded in neuropil background, and (3) proposed a method to identify clusters relevant for extracting position and speed information in neural data. We believe that our decoding analysis can be extended to examine representations of other brain regions such as the entorhinal cortex, primary visual cortex, and parietal cortex, underlying different behavioral tasks.

# Chapter 5

# Simulating constraint-removal improves sparse reward navigation

## 5.1 Introduction

Imagine if you had to learn a new recipe from chopping, peeling, stirring, all the way to cooking. All these steps need to be learned in order to successfully replicate the recipe. In nature, learning arises when humans and animals explore and interact with their environments in order to gather food and other types of rewards. This is commonly described as a reinforcement learning (RL) paradigm: interactions with the environment reinforce or inhibit specific behavior depending on the cumulative reward or penalty (Solway and Botvinick, 2012; Thorndike, 1898). This paradigm is useful to solve the problem of making sequential decisions under uncertainty, which is often formulated as a Markov decision process (MDP) whose components include a state set $S$, an action set $A$, a reward set $R$, and a decision-epoch set $T$. At each time step $t \in T$, the agent at a state $s_t \in S$ executes an action $a_t \in A$. Consequently, the agent arrives at the next state $s_{t+1}$ and earns a reward $r_{t+1}$. The environment dynamics govern the state-transition probability and the reward function.

A model-free RL agent can learn policy and value functions from direct experiences. By observing states and rewards from the environment, the agent generates action based on its current policy, leading to a new state. Through iterations, the

agent can update its policy and value functions. This is in contrast to a model-based RL agent which learns to model the environment (Dean et al., 2019; Dewanto et al., 2020; Tadepalli and Ok, 1998). While model-free agents generally learn faster, model-based agents are more flexible to generalize. Often times, human and animals utilize a strategy that is a compromise between the model-free and the model-based approaches to learn a new task (Daw et al., 2011; Momennejad et al., 2017; Russek et al., 2017).

It has been shown that a simple task such as CartPole and Mountain Car can be easily learned by a simple RL model (Li et al., 2018). However, many tasks require the agent to successfully produce a long sequence of actions to achieve the goal when binary feedback (success or fail) is only given at the end. For example, a packaging robotic task requires opening a box prior to placing the object inside, and the task is not accomplished until the box is sealed. Similarly, to play a video game such as Montezuma's Revenge, the agent needs to achieve relevant subgoals such as picking up the key prior to the final goal - opening the exit door. Throughout the process, the agent also needs to accomplish subtasks such as climbing ladders and jumping over enemies. While these seem straightforward for humans, to discover a long sequence of correct actions or identify essential subgoals is difficult for an AI agent - because the agent has no motivation to favor one action over others before it obtains any reward.

As discussed in Chapter 1, learning the state-action-reward contingencies from complex environments with sparse feedback is extremely challenging (Kulkarni et al., 2016; Riedmiller et al., 2018; Salimans and Chen, 2018; Vecerik et al., 2017). Although a random exploration or a brute-force search might help in some cases, discovering a long sequence of state-action pairs becomes unlikely as the task complexity increases.

There has been several methods developed to deal with this challenge, including reward shaping (Ng et al., 1999), transfer learning from other tasks (Mehta et al., 2008; Wang et al., 2014), or learning the hierarchy with subgoals provided by human experts (Ekvall and Kragic, 2006; Kulkarni et al., 2016; Roderick et al., 2017). These methods, however, rely on prior knowledge specific to the task, and often bias the policy toward a suboptimal solution. On the other hand, work without given prior

knowledge either use auxiliary rewards defined by inputs collected from robotic sensors to favor certain actions (Riedmiller et al., 2018), or use observed demonstrations to extract the hierarchical structures and identify subtasks in an unsupervised manner (Duan et al., 2017; Kipf et al., 2019; Sermanet et al., 2016; Sohn et al., 2020). However, auxiliary rewards have bias in the way they are predefined, and a demonstration is not not always available and will largely depend on an intelligent agent who already learned the task. Therefore, learning to solve a sparse-reward task is still a challenging problem in the field.

Here we propose to address this problem through constraint relaxation during initial training. Inspired by how humans and animals use mental simulation in partially-observable environments to set their high-level planning, we hypothesize that a high-level goal trajectory is set through a violation of physical constraints, and that subgoals can be identified more easily by gradually adding back the constraints. For example, when we play Mario Kart, we often set an overall route as if pit and obstacles did not exist, and later on realize we need to either jump over the pit or find a bypass. Similarly, real-world spatial navigation also benefits from simulating options blocked by physical barriers. It is observed that hippocampal activity not only replays experienced trajectories (Ambrose et al., 2016; Foster and Wilson, 2006), but might also provide basis for exploring potential state transitions across a closed door (Dragoi and Tonegawa, 2011; Gupta et al., 2010). Taking these into account, we propose a model where the agent can relax the physical constraints through simulation in spatial navigation to accelerate initial search and help discover potential subtasks, and we hypothesize that such agent will be flexible to find new subtasks given a changed environmental configuration with the same high-level goal.

## 5.2 Related Work

Learning from direct experiences can be unstable. Previous work (Lin, 1992) suggests that storing past experiences in a memory buffer, and update the policy and value functions by sampling a subset of experienced state-action pairs from the memory

buffer can improve learning stability. These sampled experiences are called experience replays, similar to hippocampal replays (Ambrose et al., 2016; Foster and Wilson, 2006) in terms of replaying experienced state-transitions. It has been demonstrated that adding experience replays improves learning efficiency in many algorithms such as connectionist AHC-learning, connectionist Q-learning, deep Q-learning and double-deep Q-learning (Hessel et al., 2018; Lin, 1992; Mnih et al., 2015, 2016).

Instead of replaying experienced transitions at the same frequency that they were originally experienced regardless of their significance, recent studies proposed to prioritize experience replays by sampling important transitions more frequently to improve learning efficiency (Mattar and Daw, 2018; Moore and Atkeson, 1993; Schaul et al., 2015). Indeed, the hippocampal replays are known to bias toward reward and the animal's locations (Davidson et al., 2009; Foster and Wilson, 2006; Mattar and Daw, 2018).

Both experience replays and prioritized experience replays use past experiences to stabilize training. What if we could add fictitious data, by imagining what would happen had the circumstance been different? A recent study (Andrychowicz et al., 2017) proposed Hindsight Experience Replay (HER) to replay each episode with a different goal than the one that agent was initially trying to achieve. In addition to replaying the state-goal-action triplets, trajectories based on imagining a different goal are also replayed. In this case, no matter how bad the policy is, the model will always have some positive rewards to learn from. Adding HER to a Deep Deterministic Policy Gradient (DDPG) model is shown to improve performance on tasks with sparse and binary reward.

While hindsight experience replay suggests that we can replay the episode by imagining the goal is changed, our study takes a step further to explore the potential of imagining to break physical barriers/constraints in a RL agent. By using a visual cue learning task, we showed that adding this type of imaginary replays can improve learning efficiency.

## 5.3 Methods and Materials

### 5.3.1 Navigation task

#### Deterministic gridworld navigation (maze solving)

In a deterministic gridworld navigation task, the agent needs to find the shortest path from its location toward a fixed target in a fixed gridworld maze.

**Markov decision process framework** In this task, the agent acts under a Markov decision process (MDP) framework. The agent's location at each grid cell is a state $s$. At each state, the agent needs to select an action $a$ from the action list (A) consisted of up, down, left, and right. The state transition probability $P(s_{t+1} \mid s_t, a_t)$ is the transition probability that action $a_t$ in state $s_t$ at time $t$ leads to state $s_{t+1}$ at time $t+1$, which equals to 1 in our deterministic navigation task. Here we assume that after learning, the agent took homogeneous processes in which transition functions were stationary with respect to time, i.e., $T(s, a, s') = P(s_{t+1} = s' \mid a_t = a, s_t = s)$. The reward obtained at state $s_t$ followed by an action $a_t$ is deterministic, i.e., $r_t = R(s_t, a_t)$

**Learning paradigm** We used a reinforcement learning paradigm to train the agent. During the model training period, the agent received a reward upon reaching the target, and received a small punishment upon bumping into a wall or walking beyond the maze boundaries. At each episode, an agent started from a fixed initial location. An episode ended either when the agent reached the reward target, or when the number of steps were above the threshold, i.e., 150 steps. A step referred to an attempted transition moving away from the agent's current grid cell, regardless of whether the action could lead to falling outside of the maze or bumping into a wall.

**Task evaluation** If the agent learns the state transitions of the maze, it should be able to find a shortest path from any location of the maze to the target without bumping into any wall.

### Maze generation

**Perfect maze** We can describe a maze as a graph, where each grid cell is a vertex, and the path between them is an edge. A perfect maze has two properties: (1) There are no cycles in the graph; (2) There is a unique path from the starting grid cell to the end grid cell. Given that a maze can be regarded as a group of disjoint sets (one for each grid cell), a perfect maze is basically a minimum spanning tree, i.e., a subset of the edges that forms a tree that includes every vertex where the weighted sum of all edges is minimized. Therefore, we implemented a randomized version of Kruskal's algorithm (Kruskal, 1956) to generate a perfect maze.

**Algorithm 1:** Krukal's algorithm for maze generation

```
Result: Generate a perfect maze G
Create a forest as a set of edges F:=∅ ;
foreach v ∈ G.V do
    Create a set for v (grid cell);
end
foreach (u, v) ∈ G.E ordered randomly do
    if FINDS(u) ≠ FINDS(v) then
        F:=F ∪ (u, v) ;
        UNION(FINDS(u), FINDS(v));
    end
end
```

In above algorithm, FINDS finds the set for each grid cell, and UNION knocks down a wall between them. Different mazes were created using different random seeds.

**Dungeon maze** A dungeon maze is basically an imperfect maze that has walled-off rooms, and we add a few extra criteria for generating a dungeon maze: (1) Rooms are connected, i.e., all the rooms need to be reachable; (2) The size of a room is about one eighth of the size of the maze; (3) No more than 4 rooms in a maze; (4) Multiple paths can exist between two grid cells. To generate a dungeon mazes, we added rooms by randomly removing some walls in a perfect maze. Different dungeon mazes were created with different random seeds that control the number of rooms, the number and the location of walls to be removed.

#### Visual cue generalization

In a visual cue generalization task, the agent needs to follow the directions suggested by visual cues to find the shortest path toward a reward target from its location in any gridworld maze.

**Visual cue** A visual cue was presented at a grid cell to instruct the agent direction that leads to reaching the target goal. There were four types of visual cues, each suggesting the agent an action (i.e., left, right, up, or down) to take from its current location. The representation of a visual cue was described in section 5.3.2.

**Task evaluation** If the agent learns the concept of visual cues, it should be able to find a shortest path toward a target following visual cues given any novel maze.

### 5.3.2 Reinforcement learning model

#### Q-learning paradigm

We formulated navigation under a Q-learning paradigm and the agent acted under a MDP framework. At each step $t$, the agent received an observation of the world $o_t$, potentially a $r_t$, and takes an action $a_t$. As the agent received reward upon reaching the target, values were propagated to different grid cells, and the agent learned both values of states and a policy for state transition through iterative updates. The value function of a state $s_t$ under a policy $\pi$ was defined to be the expectation of the return starting from state $s_t$ according to a policy, as in Eq (5.1).

$$V^{\pi}(s_t) = \mathrm{E}_{\pi}[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_t] \tag{5.1}$$

where $\gamma$ was the discount factor. Q-function described values obtained at state $s_t$ by taking a particular action $a_t$, as shown in Eq (5.2).

$$Q^{\pi}(s_t, a_t) = r_t + \gamma \sum_{s_{t+1}} P\big(s_{t+1} \mid s_t, a_t = \pi(s_t)\big) V^{\pi}(s_{t+1}) \tag{5.2}$$

The agent learned an optimal policy $\pi^*$ which optimized accumulated rewards.

$$\pi^*(s) = \underset{a}{\text{argmax}} Q^*(s, a) \tag{5.3}$$

**Policy updates** Throughout training process, the agent updated Q-functions by correcting the prediction error generated at each interaction with the environment. With a learning rate $\alpha$, the update equation was shown in Eq (5.4).

$$Q^\pi(s_t, a_t) \leftarrow Q^\pi(s_t, a_t) + \alpha\big(r_t + \gamma \max_{a \in A} Q(s_{t+1}, a) - Q(s_t, a_t)\big) \tag{5.4}$$

**DQN algorithm** With recent advances in deep-learning, policy updates and values of states can be learned through a neural network. Here we implemented a deep reinforcement learning network (DQN) (Mnih et al., 2015) to update policy in Q-learning. Because the gridworld navigation was a much simpler task compared to playing Atari games, we used a two-layered feed-forward neural network to predict four-dimensional action values $Q(s, \cdot\,;\, \theta)$ for a given state $s$, where $\theta$ were the network parameters and $s$ was a $n$-dimensional input. In addition, the DQN algorithm used a target network to stabilize learning. The target network had a same architecture with parameters $\theta^-$, which updated its parameters by copying online network every $\tau$ steps, i.e., $\theta^-_\tau = \theta_\tau$. Traditionally, a reinforcement learning model only updated upon real-time direct experiences, the DQN algorithm added experience replay (Lin, 1992), where a subset of experienced transitions, $(s, a, r, s')\ U(D)$, were stored in a buffer and sampled uniformly to update the network. Together, the action values were updated using loss functions in Eq (5.5).

$$L_i(\theta) = \mathrm{E}_{(s,a,r,s')\ U(D)}\Big[\big(r + \gamma \max_{a' \in A} Q(s', a';\ \theta^i_i) - Q(s, a;\ \theta_i)\big)^2\Big] \tag{5.5}$$

where $\alpha$ was the learning rate, $\theta_i$ and $\theta^-_i$ were parameters of the online network and the target network at iteration $i$, respectively. In our task, we set experience replay to sample 128 experiences $(s, a, r, s')$ from a buffer that stored 200 recent experiences.

**$\epsilon$-greedy exploration procedure** To balance between exploitation and exploration, we implemented a $\epsilon$-greedy exploration policy (Sutton et al., 1998), in which the agent selected a random action $a \in A$ with probability $\epsilon$ at each step and the optimal learned action with probability $1 - \epsilon$:

$$\pi(s) = \begin{cases} \text{random action from } A(s) & \text{if } \xi \leq \epsilon \\ \text{argmax}_{a \in A(s)} Q(s, a) & \text{otherwise} \end{cases}$$

where $\xi$ is a random number drawn from a uniform distribution at each time step, and $\epsilon$ decayed from 0.1 to 0.01 over time during training.

**Softmax exploraiton procedure** Alternatively, we also implemented a softmax policy for exploration, by selecting an action probabilistically using a Boltzmann distribution:

$$\pi(a \mid s) = \frac{e^{\frac{Q(s,a)}{\tau}}}{\sum_b e^{\frac{Q(s,b)}{\tau}}}$$

where $\tau$ was a positive parameter that also decayed over time during training.

**Training and evaluation** For the simple maze solving task, we prepared 12 mazes of size 7 by 7 and 4 mazes of size 9 by 9. For the visual cue generalization task, we prepared a train set of 100 unique 7 by 7 mazes with visual cues on each grid cell, and a test set of 100 novel 7 by 7 mazes with visual cues. When the agent reached the target, it received a reward of +10. When the agent bumped into a wall or attempted to walk outside of the maze, it received a punishment of -0.1 and stayed at the original grid cell until a valid action next. Mazes in the train set were randomly ordered and sampled to train the model in each episode, and the procedure repeated with a different order after all mazes were sampled in a round. The test set was evaluated by the model every 100 iteration steps in the training. Performance was defined as the percentage of mazes in the test set that were correctly solved by the agent following visual cues. The evaluation process did not affect training, i.e., state-action-reward experiences during evaluation were not used by the model. Adam optimizer with a

learning rate at 1e-4 was used to train the model. The training stopped either when the testing performance reached 100% or when the number of iteration steps reached 30,000. The training was performed on GeForce RTX 2080 Ti using Keras.

### Features of state-space input

A given state $s$ described the observed environment, where each grid cell was represented in one-hot coding. The overall state-space input to the DQN model was $n$-dimensional, where $n$ depended on whether the agent viewed an environment in an allocentric or an egocentric map.

**One-hot coding** To represent different environmental cues as well as the agent's occupancy in a gridworld maze, we use one-hot coding to indicate the agent's location and represent walls and visual cues. For a maze solving task without visual cues, each grid cell was denoted as two-dimensional vector, where the first element was an indicator for whether the agent occupied the cell or not, and the second element was an indicator for whether it was a wall or not. For a visual cue generalization task, we used a six-dimensional vector to describe each grid cell, where the first element was an indicator for whether the agent occupied the cell or not. For the rest of elements, each was an indicator for visual cue signaling left, up, right, and down, respectively.

**Allocentric map** In an allocentric map, we assumed the agent was able to observe the whole maze. For a $m$ by $m$ maze in a maze solving task, the state-space input was a $n = 2m^2$ dimensional vector. As for a visual cue generalization task, the state-space input was a $n = 6m^2$ dimensional vector. In this study, $m = 7$ (mostly), 9, or 11.

**Egocentric map** In an egocentric map, we assumed the agent only had access to its surrounding grid cells instead of a whole maze. In a maze solving task with a $k$ by $k$ egocentric view, the state-space input was a $n = 2k^2$ dimensional vector. In a visual cue generalization task with a $k$ by $k$ egocentric view, the state-space input was a $n = 6k^2$ dimensional vector. In this study, $k = 1$ (local grid cell only) or 3.

### 5.3.3 Additional replays for constraint removal

To test if simulating constraint-removal improves the agent's learning efficiency, we added additional relays that broke physical barriers in a task. For maze navigation, the agent was able to pass through walls by getting a penalty. For each wall-passing, the agent got 0 penalty in initial episodes. The penalty for wall-passing increased by -0.001 after each successful sequence in an episode (i.e., the agent reached the reward target with a limited number of steps before an episode ended), and remained the same to previous episodes, otherwise. After the penalty for wall-passing reached -0.1, it stayed at -0.1 for all subsequent episodes.

## 5.4 Results

### 5.4.1 Wide variability in learning to solve gridworld mazes with DQN models.

To test our hypothesis that efficiency to learn a task task-learning can be improved by an agent capable of mentally simulating navigation sequences as if physical barriers were removed, we started with a deterministic maze-solving task. We generated a set of 7 by 7 mazes and 9 by 9 mazes (Figure 5.1), and implemented a DQN model that replayed previous observed experiences and a DQN model that additionally replayed imaginary experiences (Figure 5.2a). By comparing their learning efficiency in solving a maze, we found a wide variability in number of episodes required to learn state-transitions in a 7 by 7 maze (Figure 5.2b). Similarly, learning to solve 9 by 9 mazes also demonstrated a wide variability in number of steps took by both models (Figure 5.2c, left). Despite there was no significant difference in learning efficiency between two models, there was a trend that a DQN model was able to learn maze-solving faster with additional replays simulating sequences that removed physical constraints. We hypothesized that wide variability came from bias in exploitation, especially when a model acted on an $\epsilon$-greedy policy. We found that training models with a softmax policy reduced variability in learning to solve deterministic mazes (Figure 5.2c, right).

### 5.4.2 Simulating constraint-removal aids cue generalization

We proposed a model capable of mentally simulating experiences as if constraints were removed was able to aid learning with sparse rewards. To evaluate leverage brought by our proposed model in task-learning, the task should have enough complexity. For example, Dijkstra shortest path algorithm has been suggested to be a better method for solving deterministic gridworld mazes compared to reinforcment learning algorithms. Therefore, we decided to increase task complexity by asking the agent to learn visual cues from navigation and generalize to novel environments. Visual cues signaled the desired direction toward the target, just like traffic signs.

We first generated a train set of 100 perfect mazes (Figure 5.3a, left) and a test set of 100 perfect mazes (Figure 5.3a, right). Starting locations and reward targets were randomized across mazes, and the test set was completely different from the train set. If the agent learned visual cues through navigation, it should be able to generalize and easily navigate any novel maze with visual cues. Surprisingly, it was found that when a model was trained with imaginary replays passing through the walls in a perfect maze, learning actually slowed down compared to a model trained with real experiences only (Figure 5.3b). One possibility was that the agent was able to reach the target reward by random-walk within an episode. If it were the case, the benefit of simulating constraint-removal to search for the target earlier could be offset easily by the effort to correct wrong state-transitions learned from imaginary experiences. We thus examined distribution of steps a random walk model took to navigate mazes in the train set (Supplementary Figure 5.1a, b), and decided to change the step-limit to end a training episode to 40 from 150. Unfortunately, adding imaginary replays to remove physical constraints still slowed down the learning to generalize visual cues.

We suspected the above results were related to the geometric nature of a perfect maze. Specifically, obtaining state-transitions from current state to two neighboring grid cells automatically gave the agent knowledge of state-transitions to the other two neighboring grid cells, so bumping into a wall was quite informative to navigation. Therefore, we trained both models to learn visual cues from dungeon mazes instead

(Figure 5.4a, b). As expected, we found learning significantly accelerated by a model trained with imaginary replays in a dungeon maze (Figure 5.4c). Overall, these results indicated that learning a complex task could benefit from simulating sequences as if physical constraints were removed, depending on the difficulty in target searching.

### 5.4.3 Effect of state-space representations (input features)

To evaluate how representations of states of the world affected learning performance, we compared models trained with different input features. For inputs represented in an egocentric map, the agent got access to states of surrounding grid cells, and input sizes were generally smaller. For inputs represented in an allocentric map, the agent observed the whole maze regardless of its relative location, and input sizes grew in proportional to the size of the maze (see Methods and Materials). Results revealed that visual cues were learned to generalize across different mazes when state-space was represented egocentrically (Figure 5.5a, b), but not when state-space was represented allocentrically (Figure 5.5c). In addition, it is shown that a DQN model capable of simulating sequences that removed physical constraints (e.g., walls) learned faster than a normal DQN model, regardless of the size of the egocentric map.

# Figures

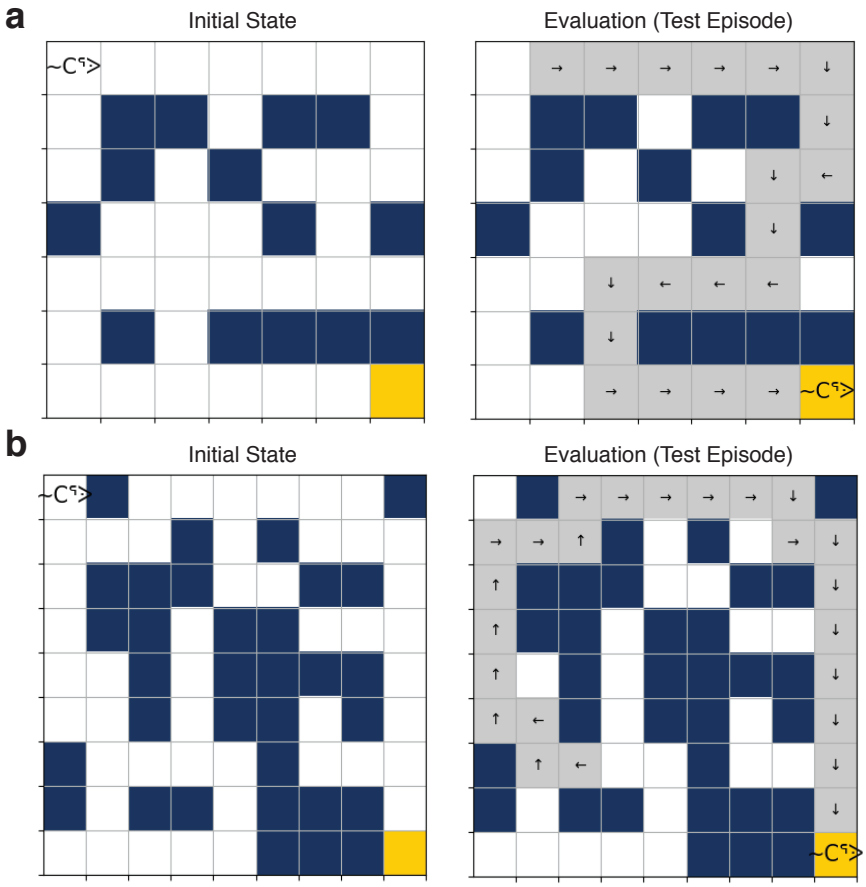



**Figure 5.1: Deterministic gridworld navigation (the maze-solving task).** (**a-b**) Two examples of gridworld mazes. The mouse symbol denoted the location of the agent. The white grid cells were available passages, whereas the dark blue grid cells were walls. The yellow grid cell represents the reward target. Left: Initial State. Right: State with a trained agent during evaluation. Light gray grid cells and arrows denoted the sequence of the agent during a test episode. (**a**) An example of a 7 by 7 maze. (**b**) An example of a 9 by 9 maze.

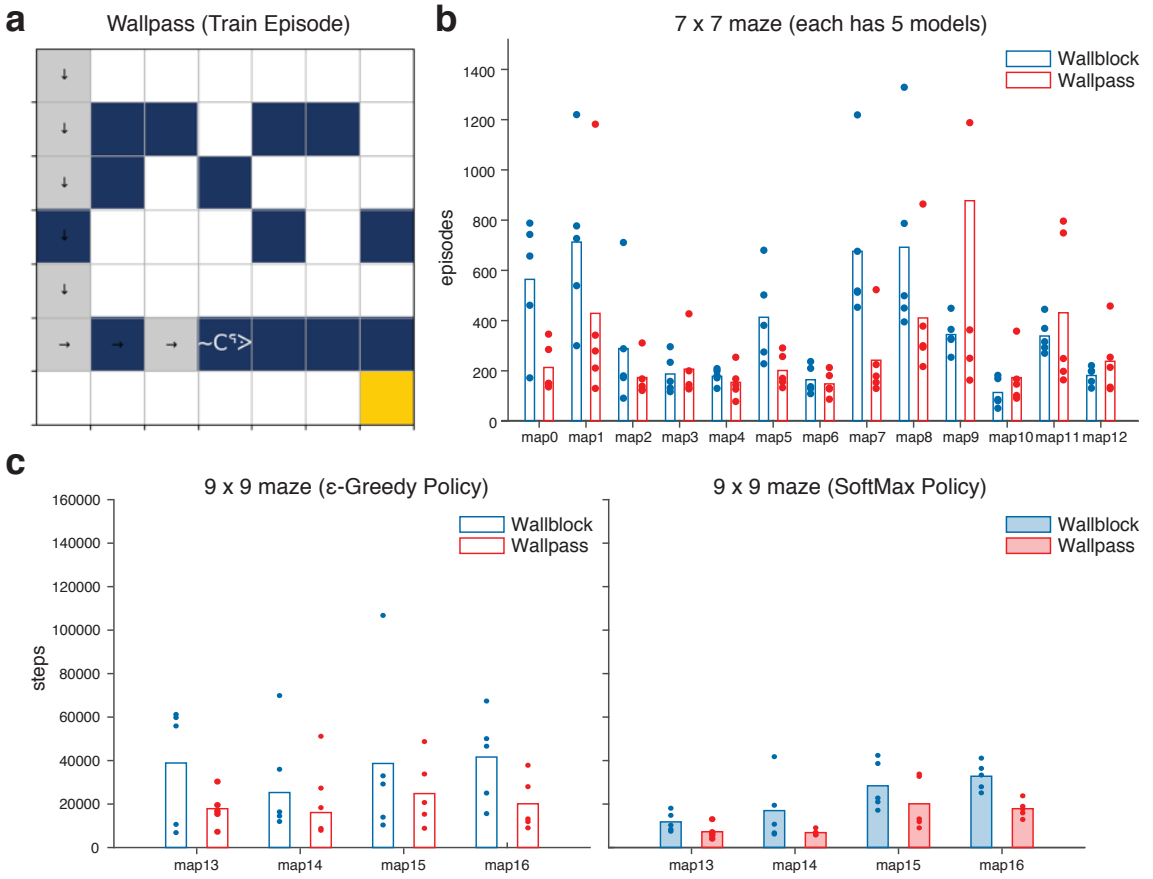



**Figure 5.2: Navigation performance with an agent simulating constraint removal.** (**a**) An example sequence in a training episode from an agent simulating constraint removal. In this example, the agent (mouse) was able to walk pass through the wall (dark blue) and received a penalty. Details see Methods and Materials. (**b**) Number of episodes required in training a normal DQN model (blue) compared with training a DQN model with additional replays that broke physical constraints in navigation (red). There were 12 7 by 7 mazes. For each maze, 5 models with different initialization seeds were trained. Each dot represented performance of an individual model, and each bar represented the mean performance among models trained to solve the same maze. (**c**) Comparison of model performance in solving 4 9 by 9 mazes using different exploration policies. Left: Models were trained with $\epsilon$-greedy exploration. Right: Models were trained with a softmax exploration policy.

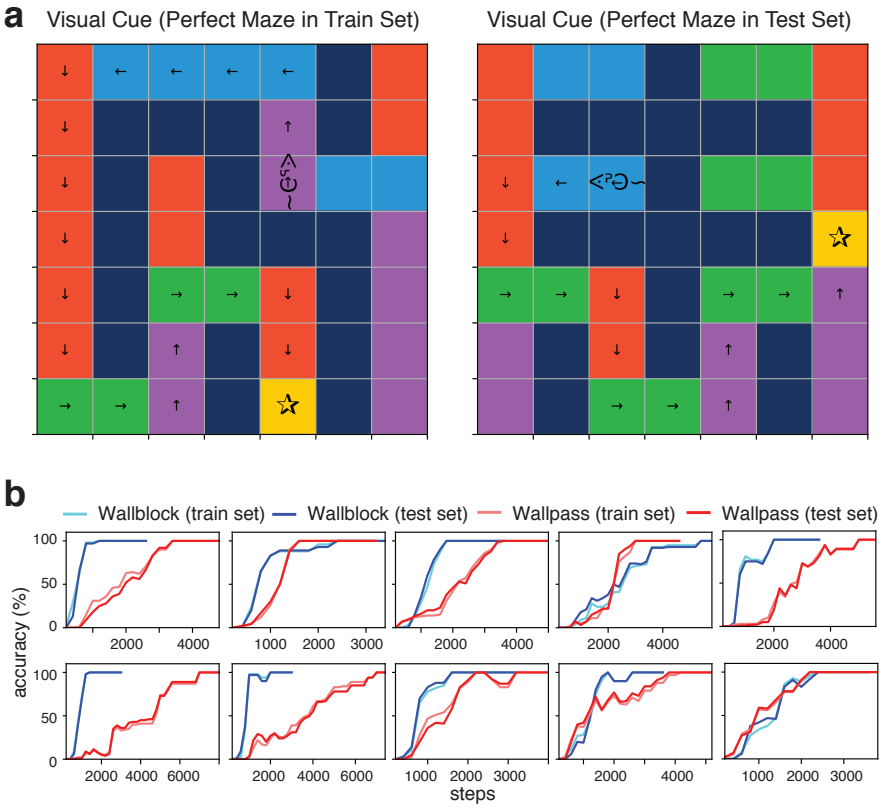



**Figure 5.3: Visual cue generalization from navigating perfect mazes.** (**a**) Left: An example sequence from the agent successfully navigated a perfect maze in the train set. Right: An example sequence from the agent successfully navigated a perfect maze in the test set. Different colors of grid cells represent visual cues signaling different directions toward the target reward (down: red; up: purple; left: blue; green: right). (**b**) Performance from 10 example models with different initialization seeds. Throughout training, the model was separately evaluated every 500 steps by navigating mazes in the train set (cyan, pink) and the test set (blue, red). Percentage of mazes that were successfully solved was used as a performance metric. Results indicated that the original DQN model (blue) learned visual cues across perfect mazes faster than the model with additional replays (red) simulating wall-pass.

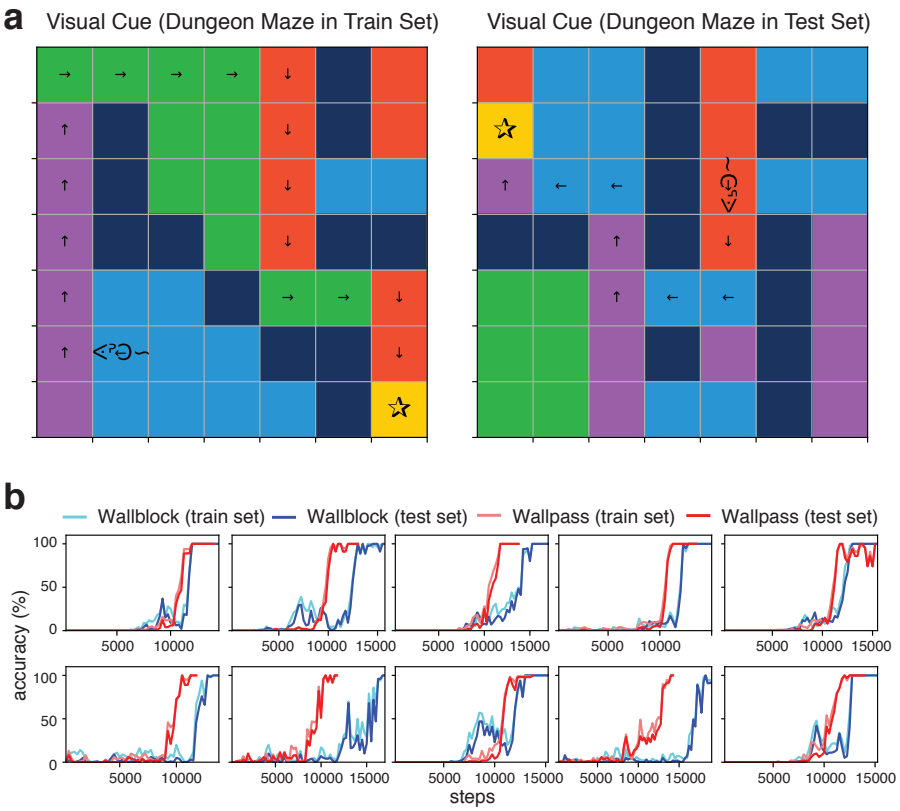



**Figure 5.4: Visual cue generalization from navigating dungeon mazes.** (**a**) Left: An example sequence from the agent successfully navigated a dungeon maze in the train set. Right: An example sequence from the agent successfully navigated a dungeon maze in the test set. Different colors of grid cells represent visual cues signaling different directions toward the target reward (down: red; up: purple; left: blue; green: right). (**b**) Performance from 10 example models with different initialization seeds. Throughout training, the model was separately evaluated every 500 steps by navigating mazes in the train set (cyan, pink) and the test set (blue, red). Percentage of mazes that were successfully solved was used as a performance metric. On the contrary to **Figure 5.3**, a model with additional replays (red) simulating wall-pass learned visual cues across dungeon mazes faster than a normal DQN model (blue).

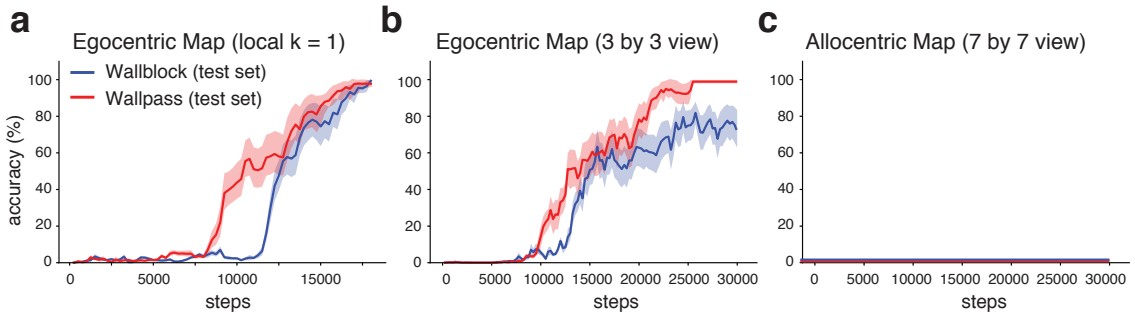
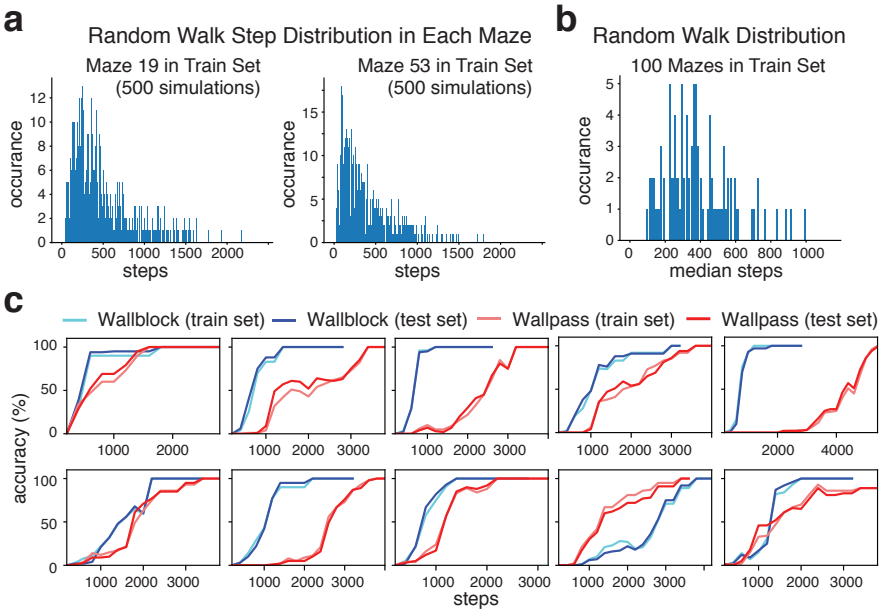



**Figure 5.5: Effect of state-space representation in learning visual cues.** (**a**) Models trained with an egocentric map of a size $k = 1$ to learn visual cue generalization. Models with additional replays that removed physical constraints (red) learned faster than a normal DQN model (blue). (**b**) Models trained with an egocentric map of size 3 by 3 to learn visual cue generalization. Models with additional replays that removed physical constraints (red) learned faster than a normal DQN model (blue), but both took longer to train than models in **a**. (**c**) Models trained with an allocentric map of the whole maze to learn visual cue generalization. Both models were unable to learn to generalize visual cues.

# Supplementary Figures




**Supplementary Figure 5.1: Training steps in a random walk model and effect of a smaller step-limit to end an episode navigating perfect mazes.** (**a**) Distribution of total steps took by a random-walk model from 500 simulations to reach the target reward in two example mazes. In this case, there was no limitation in number of steps to end an episode. (**b**) Distribution of median steps across 100 mazes in the train set solved by a random-walk model. (**c**) Performance from 10 example model trained on perfect mazes with visual cues. Similar to models in **Figure 5.3b**, but the step-limit to end a training episode was changed to 40 instead of 150.

# Chapter 6

# Discussion and Future Direction

This dissertation aimed to ask how an internal model of the world could be improved through mental simulation, and whether this process additionally utilized information which has been ignored in the field. Our results opened up new directions for future studies.

For example, Chapter 2 revealed that humans optimally integrate implicit temporal structures with kinematic information while interacting with a dynamic stimuli. It is unclear whether the coding of temporal structures are already intermingled with kinematic representations in early cortices, or there are two parallel processes in the brain. If there are two parallel processes, where in the brain are both representations integrated?

Chapter 3 demonstrated how intrinsic dynamics of a neural ensemble can be interpreted with a low-dimensional manifold within the high-dimensional state space, in both time reproduction and observational learning tasks. However, it is unclear how low-dimensional manifolds were extracted by the actual neural circuits. Future studies where we can dentify cellular identities and track their synaptic inputs will help us move from an algorithmic level of analysis to the implementational/physical level

Chapter 4 proposed an end-to-end decoding paradigm to predict navigation from raw microendoscopic calcium imaging data in the CA1. While this result was interesting, it is unfortunately difficult to directly compare ensembles identified by the

network attention maps to the identified single cells due to a limited spatial resolution in Grad-CAM. It is also unclear whether these results were specific to CA1 recordings, and will be interesting to test our paradigm in other brain regions when the animals undergo a different task.

Results in Chapter 5 were still preliminary. We proposed a RL model where the agent can relax the physical constraints in a task with sparse rewards. This agent should have the following properties: (1) faster initial search on high-level path toward the goal, (2) easier to discover potential subgoals, and (3) easier to find alternative paths when environmental configuration is slightly changed (i.e. a tunnel becomes available or a hallway gets closed, etc). Results suggested visual cue can be learned and generalized across novel mazes faster by a model trained on imaginary replays simulating to remove the physical barriers (i.e., walls) in navigation. However, learning visual cues from gridworld navigation is still a relatively easy task. The next steps is to test our model in learning subgoals in a task with hierarchical structures. Ongoing work on applying this model to play Atari games (e.g., breakouts and pong) will allow us to verify our model efficacy and capacity in learning a complex task under a sparse-reward framework.

# Bibliography

L. Acerbi, D. M. Wolpert, and S. Vijayakumar. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput Biol*, 8(11):e1002771, 2012. 48, 87

M. A. Addicott, J. M. Pearson, M. M. Sweitzer, D. L. Barack, and M. L. Platt. A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology*, 42(10):1931–1939, 2017. 24

J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim. Sanity checks for saliency maps. *Advances in neural information processing systems*, 31: 9505–9515, 2018. 181

J. E. Adriaense, J. S. Martin, M. Schiestl, C. Lamm, and T. Bugnyar. Negative emotional contagion and cognitive bias in common ravens (corvus corax). *Proceedings of the National Academy of Sciences*, 116(23):11547–11552, 2019. 70

A. Afshar, G. Santhanam, M. Y. Byron, S. I. Ryu, M. Sahani, and K. V. Shenoy. Single-trial neural correlates of arm movement preparation. *Neuron*, 71(3):555–564, 2011. 85, 89, 90

O. J. Ahmed and M. R. Mehta. Running speed alters the frequency of hippocampal gamma oscillations. *Journal of Neuroscience*, 32(21):7373–7383, 2012. 149

M. B. Ahrens and M. Sahani. Observers exploit stochastic models of sensory change to help judge the passage of time. *Current Biology*, 21(3):200–206, 2011. 47

J. Akerboom, T.-W. Chen, T. J. Wardill, L. Tian, J. S. Marvin, S. Mutlu, N. C. Calderón, F. Esposti, B. G. Borghuis, X. R. Sun, et al. Optimization of a gcamp calcium indicator for neural activity imaging. *Journal of neuroscience*, 32(40): 13819–13840, 2012. 179

D. Alais and D. Burr. The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, 14(3):257–262, 2004. 22

C. Allen and C. F. Stevens. An evaluation of causes for unreliability of synaptic transmission. *Proceedings of the National Academy of Sciences*, 91(22):10380–10383, 1994. 17

W. E. Allen, I. V. Kauvar, M. Z. Chen, E. B. Richman, S. J. Yang, K. Chan, V. Gradinaru, B. E. Deverman, L. Luo, and K. Deisseroth. Global representations of goal-directed behavior in distinct cell types of mouse neocortex. *Neuron*, 94(4):891–907, 2017. 179

S. A. Allsop, R. Wichmann, F. Mills, A. Burgos-Robles, C.-J. Chang, A. C. Felix-Ortiz, A. Vienne, A. Beyeler, E. M. Izadmehr, G. Glober, et al. Corticoamygdala transfer of socially derived information gates observational learning. *Cell*, 173(6): 1329–1342, 2018. 76

R. E. Ambrose, B. E. Pfeiffer, and D. J. Foster. Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron*, 91(5):1124–1136, 2016. 185, 186

K. C. Ames, S. I. Ryu, and K. V. Shenoy. Neural dynamics of reaching following incorrect or absent motor preparation. *Neuron*, 81(2):438–451, 2014. 84, 85, 86

M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba. Hindsight experience replay. *arXiv preprint arXiv:1707.01495*, 2017. 186

D. E. Angelaki, A. G. Shaikh, A. M. Green, and J. D. Dickman. Neurons compute internal models of the physical laws of motion. *Nature*, 430(6999):560–564, 2004. 68

E. W. Archer, U. Koster, J. W. Pillow, and J. H. Macke. Low-dimensional models of neural population activity in sensory cortical circuits. *Advances in neural information processing systems*, 27:343–351, 2014. 22

P. Atsak, M. Orre, P. Bakker, L. Cerliani, B. Roozendaal, V. Gazzola, M. Moita, and C. Keysers. Experience modulates vicarious freezing in rats: a model for empathy. *PloS one*, 6(7):e21855, 2011. 70

J.-Y. Audibert, R. Munos, and C. Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19): 1876–1902, 2009. 24

P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002. 24

B. B. Averbeck, P. E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–366, 2006. 22, 130

F. Baeyens, B. Kaes, P. Eelen, and P. Silverans. Observational evaluative conditioning of an embedded stimulus element. *European Journal of Social Psychology*, 26(1): 15–28, 1996. 69

W. Bair. Visual receptive field organization. *Current opinion in neurobiology*, 15(4): 459–464, 2005. 21

R. Barbieri, L. M. Frank, D. P. Nguyen, M. C. Quirk, V. Solo, M. A. Wilson, and E. N. Brown. Dynamic analyses of information encoding in neural ensembles. *Neural computation*, 16(2):277–307, 2004. 96

H. M. Bayer and P. W. Glimcher. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1):129–141, 2005. 23

A. K. Beery and D. Kaufer. Stress, social behavior, and resilience: insights from rodents. *Neurobiology of stress*, 1:116–127, 2015. 95

T. E. Behrens, M. W. Woolrich, M. E. Walton, and M. F. Rushworth. Learning the value of information in an uncertain world. *Nature neuroscience*, 10(9):1214–1221, 2007. 20

J. N. Betley, S. Xu, Z. F. H. Cao, R. Gong, C. J. Magnus, Y. Yu, and S. M. Sternson. Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature*, 521(7551):180–185, 2015. 128

W. Bialek. Physical limits to sensation and perception. *Annual review of biophysics and biophysical chemistry*, 16(1):455–478, 1987. 17

W. Bialek and S. Setayeshgar. Physical limits to biochemical signaling. *Proceedings of the National Academy of Sciences*, 102(29):10040–10045, 2005. 17

K. C. Bittner, C. Grienberger, S. P. Vaidya, A. D. Milstein, J. J. Macklin, J. Suh, S. Tonegawa, and J. C. Magee. Conjunctive input processing drives feature selectivity in hippocampal ca1 neurons. *Nature neuroscience*, 18(8):1133–1142, 2015. 178

R. J. Bootsma and R. R. Oudejans. Visual information about time-to-collision between two objects. *Journal of experimental psychology: human perception and performance*, 19(5):1041, 1993. 28

G. Bosco, S. Delle Monache, and F. Lacquaniti. Catching what we can't see: manual interception of occluded fly-ball trajectories. *PLoS One*, 7(11):e49381, 2012. 62

D. Bouneffouf, A. Bouzeghoub, and A. L. Gançarski. A contextual-bandit algorithm for mobile context-aware recommender system. In *International conference on neural information processing*, pages 324–331. Springer, 2012. 26

R. I. Brafman and M. Tennenholtz. R-max-a general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research*, 3 (Oct):213–231, 2002. 24

D. H. Brainard, P. Longère, P. B. Delahunt, W. T. Freeman, J. M. Kraft, and B. Xiao. Bayesian model of human color constancy. *Journal of Vision*, 6(11):10–10, 2006. 22

E. Brenner and J. B. Smeets. Sources of variability in interceptive movements. *Experimental Brain Research*, 195(1):117–133, 2009. 47, 87

E. Brenner and J. B. Smeets. Intercepting moving objects: Do eye-movements matter. *Space and time in perception and action*, pages 109–120, 2010. 64

E. Brenner and J. B. Smeets. How people achieve their amazing temporal precision in interception. *Journal of Vision*, 15(3):8–8, 2015. 28, 62

G. J. Broussard, Y. Liang, M. Fridman, E. K. Unger, G. Meng, X. Xiao, N. Ji, L. Petreanu, and L. Tian. In vivo measurement of afferent activity with axon-specific calcium imaging. *Nature neuroscience*, 21(9):1272–1280, 2018. 179

A.-M. Brouwer, E. Brenner, and J. B. Smeets. Hitting moving objects: is target speed used in guiding the hand? *Experimental Brain Research*, 143(2):198–211, 2002. 28, 62

E. N. Brown, L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18 (18):7411–7425, 1998. 81, 129, 180

R. L. Buckner. The role of the hippocampus in prediction and imagination. *Annual review of psychology*, 61:27–48, 2010. 23

C. J. Burke, P. N. Tobler, M. Baddeley, and W. Schultz. Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences*, 107(32): 14431–14436, 2010. 69

J. P. Burkett, E. Andari, Z. V. Johnson, D. C. Curry, F. B. de Waal, and L. J. Young. Oxytocin-dependent consolation behavior in rodents. *Science*, 351(6271):375–378, 2016. 95

O. Caelen and G. Bontempi. Improving the exploration strategy in bandit algorithms. In *International Conference on Learning and Intelligent Optimization*, pages 56–68. Springer, 2007. 26

D. J. Cai, D. Aharoni, T. Shuman, J. Shobe, J. Biane, W. Song, B. Wei, M. Veshkini, M. La-Vu, J. Lou, et al. A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, 534(7605):115–118, 2016. 128

W. H. Calvin and C. F. Stevens. Synaptic noise as a source of variability in the interval between action potentials. *Science*, 155(3764):842–844, 1967. 17

L. Cao, V. Varga, and Z. S. Chen. Ultrafast readout of representations from spatially distributed rodent hippocampal field potentials. *bioRxiv*, page 828467, 2019. 180

J. M. Carmena, M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov, P. G. Patil, C. S. Henriquez, and M. A. Nicolelis. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS biol*, 1(2): e42, 2003. 129

V. Cavallo and M. Laurent. Visual information and skill level in time-to-collision estimation. *Perception*, 17(5):623–632, 1988. 63

C.-J. Chang and M. Jazayeri. Integration of speed and time for estimating time to contact. *Proceedings of the National Academy of Sciences*, 115(12):E2879–E2887, 2018. 30

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357, 2002. 139

Q. Chen, J. B. Panksepp, and G. P. Lahvis. Empathy is moderated by genetic background in mice. *PloS one*, 4(2):e4387, 2009. 69, 70

T.-W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, 2013. 179

M. S. Christensen, J. Lundbye-Jensen, S. S. Geertsen, T. H. Petersen, O. B. Paulson, and J. B. Nielsen. Premotor cortex modulates somatosensory cortex during voluntary movements without proprioceptive feedback. *Nature neuroscience*, 10(4): 417–419, 2007. 23

M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012. 22, 91, 97

G. M. Cicchini, R. Arrighi, L. Cecchetti, M. Giusti, and D. C. Burr. Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, 32(3):1056–1060, 2012. 48

J. D. Cohen, S. M. McClure, and A. J. Yu. Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):933–942, 2007. 24

M. A. Conditt and F. A. Mussa-Ivaldi. Central representation of time during motor learning. *Proceedings of the National Academy of Sciences*, 96(20):11625–11630, 1999. 64

C. J. Cueva, A. Saez, E. Marcos, A. Genovesio, M. Jazayeri, R. Romo, C. D. Salzman, M. N. Shadlen, and S. Fusi. Low-dimensional dynamics for working memory and time encoding. *Proceedings of the National Academy of Sciences*, 117(37):23021–23032, 2020. 22, 91, 97

X. Cui, C. Stetson, P. R. Montague, and D. M. Eagleman. Ready… go: amplitude of the fmri signal encodes expectation of cue arrival time. *PLoS Biol*, 7(8):e1000167, 2009. 64

G. Czanner, U. T. Eden, S. Wirth, M. Yanike, W. A. Suzuki, and E. N. Brown. Analysis of between-trial and within-trial neural spiking dynamics. *Journal of neurophysiology*, 99(5):2672–2693, 2008. 81, 82, 83, 96

H. Dana, Y. Sun, B. Mohar, B. K. Hulse, A. M. Kerlin, J. P. Hasseman, G. Tsegaye, A. Tsang, A. Wong, R. Patel, et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nature methods*, 16(7): 649–657, 2019. 179, 180

T. J. Davidson, F. Kloosterman, and M. A. Wilson. Hippocampal replay of extended experience. *Neuron*, 63(4):497–507, 2009. 149, 186

N. D. Daw, Y. Niv, and P. Dayan. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8 (12):1704–1711, 2005. 20

N. D. Daw, S. J. Gershman, B. Seymour, P. Dayan, and R. J. Dolan. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215, 2011. 20, 184

C. de la Malla and J. López-Moliner. Hitting moving targets with a continuously changing temporal window. *Experimental brain research*, 233(9):2507–2515, 2015. 63, 64

F. P. de Lange, O. Jensen, and S. Dehaene. Accumulation of evidence during sequential decision making: the importance of top–down factors. *Journal of Neuroscience*, 30(2):731–738, 2010. 19, 21

S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, pages 1–47, 2019. 184

K. Del Punta, A. Puche, N. C. Adams, I. Rodriguez, and P. Mombaerts. A divergent pattern of sensory axonal projections is rendered convergent by second-order neurons in the accessory olfactory bulb. *Neuron*, 35(6):1057–1066, 2002. 21

P. R. DeLucia and G. W. Liddell. Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3):901, 1998. 45, 64

S. Deneve, P. E. Latham, and A. Pouget. Efficient computation and cue integration with noisy population codes. *Nature neuroscience*, 4(8):826–831, 2001. 22

X. Deng, D. F. Liu, K. Kay, L. M. Frank, and U. T. Eden. Clusterless decoding of position from multiunit activity using a marked point process filter. *Neural computation*, 27(7):1438–1460, 2015. 180

H. Derksen and A. Verveen. Fluctuations of resting neural membrane potential. *Science*, 151(3716):1388–1389, 1966. 17

N. S. Desai, L. C. Rutherford, and G. G. Turrigiano. Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature neuroscience*, 2(6):515–520, 1999. 22

C. A. Devine, C. Gaffney, G. M. Loughnane, S. P. Kelly, and R. G. O'connell. The role of premature evidence accumulation in making difficult perceptual decisions under temporal uncertainty. *Elife*, 8:e48526, 2019. 19, 21

V. Dewanto, G. Dunn, A. Eshragh, M. Gallagher, and F. Roosta. Average-reward model-free reinforcement learning: a systematic review and literature mapping. *arXiv preprint arXiv:2010.08920*, 2020. 184

I. C. Dezza, J. Y. Angela, A. Cleeremans, and W. Alexander. Learning the value of information and reward over time when solving exploration-exploitation problems. *Scientific reports*, 7(1):1–13, 2017. 24

G. Diaz, J. Cooper, C. Rothkopf, and M. Hayhoe. Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of vision*, 13(1):20–20, 2013. 65

J. Diedrichsen, S. E. Criscimagna-Hemminger, and R. Shadmehr. Dissociating timing and coordination as functions of the cerebellum. *Journal of Neuroscience*, 27(23): 6291–6301, 2007. 64

G. Dragoi and S. Tonegawa. Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, 469(7330):397–401, 2011. 185

Y. Duan, M. Andrychowicz, B. Stadie, O. J. Ho, J. Schneider, I. Sutskever, P. Abbeel, and W. Zaremba. One-shot imitation learning. In *Advances in neural information processing systems*, pages 1087–1098, 2017. 185

A. Dubbs, J. Guevara, and R. Yuste. moco: Fast motion correction for calcium imaging. *Frontiers in neuroinformatics*, 10:6, 2016. 128

A. Dubrowski, J. Lam, and H. Carnahan. Target velocity effects on manual interception kinematics. *Acta psychologica*, 104(1):103–118, 2000. 62

C. Dulac and A. T. Torello. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Reviews Neuroscience*, 4(7):551–562, 2003. 69

T. Ebina, Y. Masamizu, Y. R. Tanaka, A. Watakabe, R. Hirakawa, Y. Hirayama, R. Hira, S.-I. Terada, D. Koketsu, K. Hikosaka, et al. Two-photon imaging of neuronal activity in motor cortex of marmosets during upper-limb movement tasks. *Nature communications*, 9(1):1–16, 2018. 128

S. W. Egger and M. Jazayeri. A nonlinear updating algorithm captures suboptimal inference in the presence of signal-dependent noise. *Scientific reports*, 8(1):1–16, 2018. 75

S. W. Egger, E. D. Remington, C.-J. Chang, and M. Jazayeri. Internal models of sensorimotor integration regulate cortical dynamics. *Nature neuroscience*, pages 1–12, 2019. 71

H. K. Ekenel and R. Stiefelhagen. Why is facial occlusion a challenging problem? In *International Conference on Biometrics*, pages 299–308. Springer, 2009. 19

S. Ekvall and D. Kragic. Learning task models from multiple human demonstrations. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 358–363. IEEE, 2006. 184

M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002. 22, 29, 30, 40, 62, 63

G. Etter, F. Manseau, and S. Williams. A probabilistic framework for decoding behavior from in vivo calcium imaging data. *Frontiers in Neural Circuits*, 14:19, 2020. 180

T. Euler, P. B. Detwiler, and W. Denk. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature*, 418(6900):845–852, 2002. 178

D. A. Evans, A. V. Stempel, R. Vale, S. Ruehle, Y. Lefler, and T. Branco. A synaptic threshold mechanism for computing escape decisions. *Nature*, 558(7711):590–594, 2018. 128

A. A. Faisal, L. P. Selen, and D. M. Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292–303, 2008. 17, 21, 30

L. Feng and X. Wang. Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proceedings of the National Academy of Sciences*, 114(5):E840–E848, 2017. 67

J. Ferbinteanu and M. L. Shapiro. Prospective and retrospective memory coding in the hippocampus. *Neuron*, 40(6):1227–1239, 2003. 180

L. Ferrucci, S. Nougaret, and A. Genovesio. Macaque monkeys learn by observation in the ghost display condition in the object-in-place task with differential reward to the observer. *Scientific reports*, 9(1):1–9, 2019. 70

B. Fischer. Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision research*, 13(11):2113–2120, 1973. 21

B. A. Flusberg, A. Nimmerjahn, E. D. Cocker, E. A. Mukamel, R. P. Barretto, T. H. Ko, L. D. Burns, J. C. Jung, and M. J. Schnitzer. High-speed, miniaturized fluorescence microscopy in freely moving mice. *Nature methods*, 5(11):935–938, 2008. 128

J. M. Foley and R. Held. Visually directed pointing as a function of target distance, direction, and available cues. *Perception & Psychophysics*, 12(3):263–268, 1972. 17

J. Fooken, S.-H. Yeo, D. K. Pai, and M. Spering. Eye movement accuracy determines natural interception strategies. *Journal of vision*, 16(14):1–1, 2016. 62, 65

E. Formisano, D. E. Linden, F. Di Salle, L. Trojano, F. Esposito, A. T. Sack, D. Grossi, F. E. Zanella, and R. Goebel. Tracking the mind's image in the brain i: time-resolved fmri during visuospatial mental imagery. *Neuron*, 35(1):185–194, 2002. 23

M. Fortunato, M. G. Azar, B. Piot, J. Menick, I. Osband, A. Graves, V. Mnih, R. Munos, D. Hassabis, O. Pietquin, et al. Noisy networks for exploration. *arXiv preprint arXiv:1706.10295*, 2017. 26

D. J. Foster and M. A. Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683, 2006. 185, 186

M. Frey, S. Tanni, C. Perrodin, A. O'Leary, M. Nau, J. Kelly, A. Banino, C. F. Doeller, and C. Barry. Deepinsight: a general framework for interpreting wide-band neural activity. *bioRxiv*, page 871848, 2019. 130, 180

N. Fujii, H. Mushiake, and J. Tanji. Distribution of eye-and arm-movement-related neuronal activity in the sef and in the sma and pre-sma of monkeys. *Journal of Neurophysiology*, 87(4):2158–2166, 2002. 73

C. R. Gallistel and J. Gibbon. Time, rate, and conditioning. *Psychological review*, 107(2):289, 2000. 37

E. Ganmor, M. Krumin, L. F. Rossi, M. Carandini, and E. P. Simoncelli. Direct estimation of firing rates from calcium imaging data. *arXiv preprint arXiv:1601.00364*, 2016. 180

C. Geisler, D. Robbe, M. Zugaro, A. Sirota, and G. Buzsáki. Hippocampal place cell assemblies are speed-controlled oscillators. *Proceedings of the National Academy of Sciences*, 104(19):8149–8154, 2007. 149

A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, 1986. 21, 67, 129

K. K. Ghosh, L. D. Burns, E. D. Cocker, A. Nimmerjahn, Y. Ziv, A. El Gamal, and M. J. Schnitzer. Miniaturized integration of a fluorescence microscope. *Nature methods*, 8(10):871, 2011. 128

J. Gibbon, R. M. Church, W. H. Meck, et al. Scalar timing in memory. *Annals of the New York Academy of sciences*, 423(1):52–77, 1984. 36, 37, 47, 88

A. Glas, M. Hübener, T. Bonhoeffer, and P. M. Goltstein. Benchmarking miniaturized microscopy against two-photon calcium imaging using single-cell orientation tuning in mouse visual cortex. *PloS one*, 14(4):e0214954, 2019. 128

J. Gläscher, N. Daw, P. Dayan, and J. P. O'Doherty. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4):585–595, 2010. 20

J. I. Glaser, A. S. Benjamin, R. H. Chowdhury, M. G. Perich, L. E. Miller, and K. P. Kording. Machine learning for neural decoding. *Eneuro*, 7(4), 2020. 129, 181

X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010. 140, 142

E. Glowatzki and P. A. Fuchs. Transmitter release at the hair cell ribbon synapse. *Nature neuroscience*, 5(2):147–154, 2002. 21

W. Gobel and F. Helmchen. In vivo calcium imaging of neural network function. *Physiology*, 22(6):358–365, 2007. 178

Z. H. T. Góis and A. B. Tort. Characterizing speed cells in the rat hippocampus. *Cell reports*, 25(7):1872–1884, 2018. 147

M. D. Golub, M. Y. Byron, and S. M. Chase. Internal models for interpreting neural population activity during sensorimotor control. *Elife*, 4:e10015, 2015. 68

W. G. Gonzalez, H. Zhang, A. Harutyunyan, and C. Lois. Persistence of neuronal representations through time and damage in the hippocampus. *Science*, 365(6455): 821–825, 2019. 180

R. Gray and D. Regan. Accuracy of estimating time to collision using binocular and monocular information. *Vision research*, 38(4):499–512, 1998. 30, 40, 62, 63

D. S. Greenberg, D. J. Wallace, K.-M. Voit, S. Wuertenberger, U. Czubayko, A. Monsees, T. Handa, J. T. Vogelstein, R. Seifert, Y. Groemping, et al. Accurate action potential inference from a calcium sensor protein through biophysical modeling. *BioRxiv*, page 479055, 2018. 179

B. F. Grewe, J. Gründemann, L. J. Kitch, J. A. Lecoq, J. G. Parker, J. D. Marshall, M. C. Larkin, P. E. Jercog, F. Grenier, J. Z. Li, et al. Neural ensemble dynamics underlying a long-term associative memory. *Nature*, 543(7647):670–675, 2017. 128

A. S. Gupta, M. A. van der Meer, D. S. Touretzky, and A. D. Redish. Hippocampal replay is not a simple function of experience. *Neuron*, 65(5):695–705, 2010. 185

T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005. 21

D. L. Harrington, J. L. Zimbelman, S. C. Hinton, and S. M. Rao. Neural modulation of temporal encoding, maintenance, and decision processes. *Cerebral Cortex*, 20(6): 1274–1285, 2010. 64

C. M. Harris and D. M. Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394(6695):780–784, 1998. 18

M. Haruno and D. M. Wolpert. Optimal control of redundant muscles in step-tracking wrist movements. *Journal of Neurophysiology*, 94(6):4244–4255, 2005. 22

C. D. Harvey, P. Coen, and D. W. Tank. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62–68, 2012. 128

D. Hassabis, D. Kumaran, and E. A. Maguire. Using imagination to understand the neural basis of episodic memory. *Journal of neuroscience*, 27(52):14365–14374, 2007. 23

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 140

H. Hecht and G. J. Savelsbergh. Theories of time-to-contact judgment. In *Advances in psychology*, volume 135, pages 1–11. Elsevier, 2004. 28

H. Hecht, M. K. Kaiser, and M. S. Banks. Gravitational acceleration as a cue for absolute size and distance? *Perception & Psychophysics*, 58(7):1066–1075, 1996. 62

L. V. Hedges and I. Olkin. *Statistical methods for meta-analysis*. Academic press, 2014. 35

F. Helmchen, M. S. Fee, D. W. Tank, and W. Denk. A miniature head-mounted two-photon microscope: high-resolution brain imaging in freely moving animals. *Neuron*, 31(6):903–912, 2001. 128

M. Hessel, J. Modayil, H. Van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver. Rainbow: Combining improvements in deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. 186

J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy. Combining sensory information: mandatory fusion within, but not between, senses. *Science*, 298(5598): 1627–1630, 2002. 22, 29, 30, 40, 62, 63

J. M. Hillis, S. J. Watt, M. S. Landy, and M. S. Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of vision*, 4(12):1–1, 2004. 22

D. Holcman and M. Tsodyks. The emergence of up and down states in cortical networks. *PLoS Comput Biol*, 2(3):e23, 2006. 22

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982. 22

A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 140

D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574, 1959. 67

D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968. 67

D. Huber, D. A. Gutnisky, S. Peron, D. H. O'connor, J. S. Wiegert, L. Tian, T. G. Oertner, L. L. Looger, and K. Svoboda. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature*, 484(7395):473–478, 2012. 128

M. F. Huerta and J. H. Kaas. Supplementary eye field as defined by intracortical microstimulation: connections in macaques. *Journal of Comparative Neurology*, 293(2):299–330, 1990. 73

S. A. Huettel, A. W. Song, and G. McCarthy. Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *Journal of Neuroscience*, 25(13):3304–3311, 2005. 20

S. Ishii, W. Yoshida, and J. Yoshimoto. Control of exploitation–exploration meta-parameter in reinforcement learning. *Neural networks*, 15(4-6):665–687, 2002. 25

Y. Isogai, S. Si, L. Pont-Lezica, T. Tan, V. Kapoor, V. N. Murthy, and C. Dulac. Molecular organization of vomeronasal chemoreception. *Nature*, 478(7368):241–245, 2011. 69

M. Ito. Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, 9(4):304–313, 2008. 68

M. Iwase, T. Kitanishi, and K. Mizuseki. Cell type, sub-region, and layer-specific speed representation in the hippocampal–entorhinal circuit. *Scientific reports*, 10 (1):1–23, 2020. 180

R. A. Jacobs and I. Fine. Experience-dependent integration of texture and motion cues to depth. *Vision research*, 39(24):4062–4075, 1999. 29, 62

N. Jacoby and J. H. McDermott. Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27(3):359–370, 2017. 65

P. Janssen and M. N. Shadlen. A representation of the hazard rate of elapsed time in macaque area lip. *Nature neuroscience*, 8(2):234–241, 2005. 49

M. Jazayeri and M. N. Shadlen. Temporal context calibrates interval timing. *Nature neuroscience*, 13(8):1020, 2010. 35, 47

M. Jazayeri and M. N. Shadlen. A neural mechanism for sensing and reproducing a time interval. *Current Biology*, 25(20):2599–2609, 2015. 40, 48, 64

J. H. Jennings, R. L. Ung, S. L. Resendez, A. M. Stamatakis, J. G. Taylor, J. Huang, K. Veleta, P. A. Kantak, M. Aita, K. Shilling-Scrivo, et al. Visualizing hypothalamic network dynamics for appetitive and consummatory behaviors. *Cell*, 160(3):516–527, 2015. 128

D. Jeon, S. Kim, M. Chetana, D. Jo, H. E. Ruley, S.-Y. Lin, D. Rabah, J.-P. Kinet, and H.-S. Shin. Observational fear learning involves affective pain system and ca v 1.2 ca 2+ channels in acc. *Nature neuroscience*, 13(4):482, 2010. 70

S. Jewell and D. Witten. Exact spike train inference via l0 optimization. *The annals of applied statistics*, 12(4):2457, 2018. 129

K. E. Jones, A. F. d. C. Hamilton, and D. M. Wolpert. Sources of signal-dependent noise during isometric force production. *Journal of neurophysiology*, 88(3):1533–1544, 2002. 18

L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996. 24

P. Kaifosh, J. D. Zaremba, N. B. Danielson, and A. Losonczy. Sima: Python software for analysis of dynamic fluorescence imaging data. *Frontiers in neuroinformatics*, 8:80, 2014. 128

S. W. Keemink, S. C. Lowe, J. M. Pakan, E. Dylda, M. C. Van Rossum, and N. L. Rochefort. Fissa: A neuropil decontamination toolbox for calcium imaging signals. *Scientific reports*, 8(1):1–12, 2018. 178

A. Kennedy, G. Wayne, P. Kaifosh, K. Alviña, L. Abbott, and N. B. Sawtell. A temporal basis for predicting the sensory consequences of motor commands in an electric fish. *Nature neuroscience*, 17(3):416–422, 2014. 68

J. N. Kerr, D. Greenberg, and F. Helmchen. Imaging input and output of neocortical networks in vivo. *Proceedings of the National Academy of Sciences*, 102(39):14063–14068, 2005. 178

M. P. Kilgard and M. M. Merzenich. Distributed representation of spectral and temporal information in rat primary auditory cortex. *Hearing research*, 134(1-2):16–28, 1999. 21

J. M. Kilner, K. J. Friston, and C. D. Frith. Predictive coding: an account of the mirror neuron system. *Cognitive processing*, 8(3):159–166, 2007. 23

S. Kim, F. Mátyás, S. Lee, L. Acsády, and H.-S. Shin. Lateralization of observational fear learning at the cortical but not thalamic level in mice. *Proceedings of the National Academy of Sciences*, 109(38):15497–15501, 2012. 70

T. Kipf, Y. Li, H. Dai, V. Zambaldi, A. Sanchez-Gonzalez, E. Grefenstette, P. Kohli, and P. Battaglia. Compile: Compositional imitation learning and execution. In *International Conference on Machine Learning*, pages 3418–3428. PMLR, 2019. 185

S. Kira, T. Yang, and M. N. Shadlen. A neural implementation of wald's sequential probability ratio test. *Neuron*, 85(4):861–873, 2015. 19, 21

K. B. Kjelstrup, T. Solstad, V. H. Brun, T. Hafting, S. Leutgeb, M. P. Witter, E. I. Moser, and M.-B. Moser. Finite scale of spatial representation in the hippocampus. *Science*, 321(5885):140–143, 2008. 21

A. Klaus, G. J. Martins, V. B. Paixao, P. Zhou, L. Paninski, and R. M. Costa. The spatiotemporal organization of the striatum encodes action space. *Neuron*, 95(5): 1171–1180, 2017. 128, 129

T. Kleindienst, J. Winnubst, C. Roth-Alpermann, T. Bonhoeffer, and C. Lohmann. Activity-dependent clustering of functional synaptic inputs on developing hippocampal dendrites. *Neuron*, 72(6):1012–1024, 2011. 178

F. Kloosterman, S. P. Layton, Z. Chen, and M. A. Wilson. Bayesian decoding using unsorted spikes in the rat hippocampus. *Journal of neurophysiology*, 2014. 180

E. Knapska, E. Nikolaev, P. Boguszewski, G. Walasek, J. Blaszczyk, L. Kaczmarek, and T. Werka. Between-subject transfer of emotional information evokes specific pattern of amygdala activation. *Proceedings of the National Academy of Sciences*, 103(10):3858–3862, 2006. 70

E. Knapska, M. Mikosz, T. Werka, and S. Maren. Social modulation of learning in rats. *Learning & memory*, 17(1):35–42, 2010. 70

D. C. Knill and A. Pouget. The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719, 2004. 30, 40, 63

K. P. Körding and D. M. Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247, 2004. 22, 29, 30, 40, 63

K. P. Kording, S.-p. Ku, and D. M. Wolpert. Bayesian integration in force estimation. *Journal of neurophysiology*, 92(5):3161–3165, 2004. 30, 40, 63

A. S. Kozlov, T. Risler, and A. Hudspeth. Coherent motion of stereocilia assures the concerted gating of hair-cell transduction channels. *Nature neuroscience*, 10 (1):87–92, 2007. 22

E. Kropff, J. E. Carmichael, M.-B. Moser, and E. I. Moser. Speed cells in the medial entorhinal cortex. *Nature*, 523(7561):419–424, 2015. 147

F. Krueger, K. McCabe, J. Moll, N. Kriegeskorte, R. Zahn, M. Strenziok, A. Heinecke, and J. Grafman. Neural correlates of trust. *Proceedings of the National Academy of Sciences*, 104(50):20084–20089, 2007. 20

A. E. Krukowski, K. A. Pirog, B. R. Beutter, K. R. Brooks, and L. S. Stone. Human discrimination of visual direction of motion with and without smooth pursuit eye movements. *Journal of Vision*, 3(11):16–16, 2003. 63

J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1):48–50, 1956. 188

T. D. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. *Advances in neural information processing systems*, 29:3675–3683, 2016. 20, 184

G. Kumar, G. Foster, C. Cherry, and M. Krikun. Reinforcement learning based curriculum optimization for neural machine translation. *arXiv preprint arXiv:1903.00041*, 2019. 26

J. Kunimatsu and M. Tanaka. Alteration of the timing of self-initiated but not reactive saccades by electrical stimulation in the supplementary eye field. *European Journal of Neuroscience*, 36(9):3258–3268, 2012. 64

O.-S. Kwon and D. C. Knill. The brain uses adaptive internal models of scene statistics for sensorimotor estimation and planning. *Proceedings of the National Academy of Sciences*, 110(11):E1064–E1073, 2013. 28, 62

F. Lacquaniti and C. Maioli. The role of preparation in tuning anticipatory and reflex responses during catching. *Journal of Neuroscience*, 9(1):134–148, 1989. 62

M. E. Larkum, K. Kaiser, and B. Sakmann. Calcium electrogenesis in distal apical dendrites of layer 5 pyramidal cells at a critical frequency of back-propagating action potentials. *Proceedings of the National Academy of Sciences*, 96(25):14600–14604, 1999. 178

M. Lavzin, S. Rapoport, A. Polsky, L. Garion, and J. Schiller. Nonlinear dendritic processing determines angular tuning of barrel cortex neurons in vivo. *Nature*, 490 (7420):397–401, 2012. 178

C. Lee, W. H. Rohrer, and D. L. Sparks. Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, 332(6162):357–360, 1988. 21, 67

D. Lee, D. Young, P. Reddish, S. Lough, and T. Clayton. Visual timing in hitting an accelerating ball. *The Quarterly Journal of Experimental Psychology*, 35(2): 333–346, 1983. 28

D. D. Lee, P. A. Ortega, and A. A. Stocker. Dynamic belief state representations. *Current Opinion in Neurobiology*, 25:221–227, 2014. 62

S. Lee, J. F. Meyer, J. Park, and S. M. Smirnakis. Visually driven neuropil activity and information encoding in mouse primary visual cortex. *Frontiers in neural circuits*, 11:50, 2017. 179

J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:1702.03037*, 2017. 26

D. Li, Y. Yang, Y.-Z. Song, and T. Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. 184

P. Lillywhite and S. Laughlin. Transducer noise in a photoreceptor. *Nature*, 277 (5697):569–572, 1979. 17

L.-J. Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine learning*, 8(3-4):293–321, 1992. 185, 186, 190

J. F. Linden, R. C. Liu, M. Sahani, C. E. Schreiner, and M. M. Merzenich. Spectrotemporal structure of receptive fields in areas ai and aaf of mouse auditory cortex. *Journal of neurophysiology*, 90(4):2660–2675, 2003. 21

B. Lindström, J. Haaker, and A. Olsson. A common neural network differentially mediates direct and social fear learning. *NeuroImage*, 167:121–129, 2018. 69

G. Lopes, N. Bonacchi, J. Frazão, J. P. Neto, B. V. Atallah, S. Soares, L. Moreira, S. Matias, P. M. Itskov, P. A. Correia, et al. Bonsai: an event-based framework for processing and controlling data streams. *Frontiers in neuroinformatics*, 9:7, 2015. 133

J. López-Moliner, D. T. Field, and J. P. Wann. Interceptive timing: Prior knowledge matters. *Journal of Vision*, 7(13):11–11, 2007. 28

X. Lou, V. Scheuss, and R. Schneggenburger. Allosteric modulation of the presynaptic ca 2+ sensor for vesicle fusion. *Nature*, 435(7041):497–501, 2005. 17

R. J. Low, S. Lewallen, D. Aronov, R. Nevers, and D. W. Tank. Probing variability in a cognitive map using manifold inference from neural dynamics. *bioRxiv*, page 418939, 2018. 143

J. Lu, C. Li, J. Singh-Alvarado, Z. C. Zhou, F. Fröhlich, R. Mooney, and F. Wang. Min1pipe: A miniscope 1-photon-based calcium imaging signal extraction pipeline. *Cell reports*, 23(12):3673–3684, 2018. 128, 130, 178

W. J. Ma, J. M. Beck, P. E. Latham, and A. Pouget. Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432–1438, 2006. 22

V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474):78–84, 2013. 91, 97

S. Maren. Auditory fear conditioning increases cs-elicited spike firing in lateral amygdala neurons even after extensive overtraining. *European Journal of Neuroscience*, 12(11):4047–4054, 2000. 126

A. M. Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern analysis and machine intelligence*, 24(6):748–763, 2002. 19

A. H. Mason and H. Carnahan. Target viewing time and velocity effects on prehension. *Experimental Brain Research*, 127(1):83–94, 1999. 63

Y. Matsuzaka, H. Aizawa, and J. Tanji. A motor area rostral to the supplementary motor area (presupplementary motor area) in the monkey: neuronal activity during a learned motor task. *Journal of neurophysiology*, 68(3):653–662, 1992. 73

M. G. Mattar and N. D. Daw. Prioritized memory access explains planning and hippocampal replay. *Nature neuroscience*, 21(11):1609–1617, 2018. 186

S. P. McKee, G. H. Silverman, and K. Nakayama. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision research*, 26 (4):609–619, 1986. 36, 47

K. Mehlhorn, B. R. Newell, P. M. Todd, M. D. Lee, K. Morgan, V. A. Braithwaite, D. Hausmann, K. Fiedler, and C. Gonzalez. Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, 2(3):191, 2015. 24

N. Mehta, S. Natarajan, P. Tadepalli, and A. Fern. Transfer in variable-reward hierarchical reinforcement learning. *Machine Learning*, 73(3):289, 2008. 184

H. Merchant, W. Zarco, O. Pérez, L. Prado, and R. Bartolo. Measuring time with different neural chronometers during a synchronization-continuation task. *Proceedings of the National Academy of Sciences*, 108(49):19784–19789, 2011. 90

H. Merchant, O. Pérez, W. Zarco, and J. Gámez. Interval tuning in the primate medial premotor cortex as a general timing mechanism. *Journal of Neuroscience*, 33(21):9082–9096, 2013. 64

R. C. Miall and D. M. Wolpert. Forward models for physiological motor control. *Neural networks*, 9(8):1265–1279, 1996. 18, 23, 68, 89

A. Mikami, W. T. Newsome, and R. H. Wurtz. Motion selectivity in macaque visual cortex. i. mechanisms of direction and speed selectivity in extrastriate area mt. *Journal of neurophysiology*, 55(6):1308–1327, 1986. 29

S. Mineka, M. Davidson, M. Cook, and R. Keir. Observational conditioning of snake fear in rhesus monkeys. *Journal of abnormal psychology*, 93(4):355, 1984. 70

A. Mita, H. Mushiake, K. Shima, Y. Matsuzaka, and J. Tanji. Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nature neuroscience*, 12(4):502–507, 2009. 64

W. Mittmann, D. J. Wallace, U. Czubayko, J. T. Herb, A. T. Schaefer, L. L. Looger, W. Denk, and J. N. Kerr. Two-photon calcium imaging of evoked activity from l5 somatosensory neurons in vivo. *Nature neuroscience*, 14(8):1089–1093, 2011. 128

M. Miyazaki, D. Nozaki, and Y. Nakajima. Testing bayesian models of human coincidence timing. *Journal of neurophysiology*, 94(1):395–399, 2005. 48

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015. 186, 190

V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937, 2016. 26, 186

I. Momennejad, E. M. Russek, J. H. Cheong, M. M. Botvinick, N. D. Daw, and S. J. Gershman. The successor representation in human reinforcement learning. *Nature Human Behaviour*, 1(9):680–692, 2017. 184

A. W. Moore and C. G. Atkeson. Prioritized sweeping: Reinforcement learning with less data and less time. *Machine learning*, 13(1):103–130, 1993. 24, 186

L. A. Mrotek and J. F. Soechting. Target interception: hand–eye coordination and strategies. *Journal of Neuroscience*, 27(27):7297–7309, 2007. 62

L. A. Mrotek, M. Flanders, and J. F. Soechting. Interception of targets using brief directional cues. *Experimental brain research*, 156(1):94–103, 2004. 28, 62

E. A. Mukamel, A. Nimmerjahn, and M. J. Schnitzer. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron*, 63(6):747–760, 2009. 129

M. Murakami, M. I. Vicente, G. M. Costa, and Z. F. Mainen. Neural antecedents of self-initiated actions in secondary motor cortex. *Nature neuroscience*, 17(11): 1574–1582, 2014. 90

A. Y. Ng, D. Harada, and S. Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *ICML*, volume 99, pages 278–287, 1999. 184

H. Nover, C. H. Anderson, and G. C. DeAngelis. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *Journal of Neuroscience*, 25(43):10049–10060, 2005. 147

K. Ohki, S. Chung, Y. H. Ch'ng, P. Kara, and R. C. Reid. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, 433 (7026):597–603, 2005. 178

H. F. Ólafsdóttir, F. Carpenter, and C. Barry. Task demands predict a dynamic switch in the content of awake hippocampal replay. *Neuron*, 96(4):925–935, 2017. 149

A. Olsson and E. A. Phelps. Learned fear of "unseen" faces after pavlovian, observational, and instructed fear. *Psychological science*, 15(12):822–828, 2004. 69, 70

M. Pachitariu, C. Stringer, M. Dipoppa, S. Schröder, L. F. Rossi, H. Dalgleish, M. Carandini, and K. D. Harris. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *Biorxiv*, 2017. 178

S. Panzeri, S. R. Schultz, A. Treves, and E. T. Rolls. Correlations and the encoding of information in the nervous system. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1423):1001–1012, 1999. 22

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 143

G. Pellizzer and J. H. Hedges. Motor planning: effect of directional uncertainty with discrete spatial cues. *Experimental Brain Research*, 150(3):276–289, 2003. 64

A. G. Pereira, A. Cruz, S. Q. Lima, and M. A. Moita. Silence resulting from the cessation of movement signals danger. *Current biology*, 22(16):R627–R628, 2012. 70

F. Pereira, T. Mitchell, and M. Botvinick. Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1):S199–S209, 2009. 129

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990. 138

L. Petreanu, D. A. Gutnisky, D. Huber, N.-l. Xu, D. H. O'Connor, L. Tian, L. Looger, and K. Svoboda. Activity in motor–sensory projections reveals distributed coding in somatosensation. *Nature*, 489(7415):299–303, 2012. 179

B. Petreska, B. M. Yu, J. P. Cunningham, G. Santhanam, S. Ryu, K. V. Shenoy, and M. Sahani. Dynamical segmentation of single trials from population neural data. *Advances in neural information processing systems*, 24:756–764, 2011. 22, 91, 97

L. Pinto and Y. Dan. Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron*, 87(2):437–450, 2015. 128, 129

M. T. Pisansky, L. R. Hanson, I. I. Gottesman, and J. C. Gewirtz. Oxytocin enhances observational fear in mice. *Nature communications*, 8(1):1–11, 2017. 95

E. A. Pnevmatikakis. Analysis pipelines for calcium imaging data. *Current opinion in neurobiology*, 55:15–21, 2019. 128

E. A. Pnevmatikakis and A. Giovannucci. Normcorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *Journal of neuroscience methods*, 291:83–94, 2017. 129, 134

E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 89(2):285–299, 2016. 129, 178

R. Polanía, I. Krajbich, M. Grueschow, and C. C. Ruff. Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, 82(3):709–720, 2014. 19, 21

A. M. Poulos, I. Zhuravka, V. Long, C. Gannam, and M. Fanselow. Sensitization of fear learning to mild unconditional stimuli in male and female rats. *Behavioral neuroscience*, 129(1):62, 2015. 95

A. A. Prinz, D. Bucher, and E. Marder. Similar network activity from disparate circuit parameters. *Nature neuroscience*, 7(12):1345–1352, 2004. 22

G. J. Quirk, J. C. Repa, and J. E. LeDoux. Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. *Neuron*, 15(5):1029–1039, 1995. 126

R. Q. Quiroga and S. Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3): 173–185, 2009. 129

B. C. Rakitin, J. Gibbon, T. B. Penney, C. Malapani, S. C. Hinton, and W. H. Meck. Scalar expectancy theory and peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 24(1):15, 1998. 36, 37, 47, 88

J. Reidl, J. Starke, D. B. Omer, A. Grinvald, and H. Spors. Independent component analysis of high-resolution imaging data identifies distinct functional domains. *Neuroimage*, 34(1):94–108, 2007. 129

E. Remington and M. Jazayeri. Late bayesian inference in sensorimotor behavior. *bioRxiv*, page 130062, 2017. 62

E. D. Remington, D. Narain, E. A. Hosseini, and M. Jazayeri. Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron*, 98(5): 1005–1019, 2018. 85

A. Resulaj, R. Kiani, D. M. Wolpert, and M. N. Shadlen. Changes of mind in decision-making. *Nature*, 461(7261):263–266, 2009. 64

M. Rickert, O. Brock, and A. Knoll. Balancing exploration and exploitation in motion planning. In *2008 IEEE International Conference on Robotics and Automation*, pages 2812–2817. IEEE, 2008. 24

M. Riedmiller, R. Hafner, T. Lampe, M. Neunert, J. Degrave, T. Van de Wiele, V. Mnih, N. Heess, and J. T. Springenberg. Learning by playing-solving sparse reward tasks from scratch. *arXiv preprint arXiv:1802.10567*, 2018. 20, 184, 185

T. F. Roberts, E. Hisey, M. Tanaka, M. G. Kearney, G. Chattree, C. F. Yang, N. M. Shah, and R. Mooney. Identification of a motor-to-auditory pathway important for vocal learning. *Nature neuroscience*, 20(7):978, 2017. 128

M. Roderick, C. Grimm, and S. Tellex. Deep abstract q-networks. *arXiv preprint arXiv:1710.00459*, 2017. 184

H. R. Rodman and T. D. Albright. Coding of visual stimulus velocity in area mt of the macaque. *Vision research*, 27(12):2035–2048, 1987. 29

R. Romo, A. Hernández, A. Zainos, and E. Salinas. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron*, 38(4): 649–657, 2003. 22

G. Rothschild, I. Nelken, and A. Mizrahi. Functional organization and population dynamics in the mouse primary auditory cortex. *Nature neuroscience*, 13(3):353, 2010. 128

E. M. Russek, I. Momennejad, M. M. Botvinick, S. J. Gershman, and N. D. Daw. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS computational biology*, 13(9):e1005768, 2017. 184

A. A. Russo, R. Khajeh, S. R. Bittner, S. M. Perkins, J. P. Cunningham, L. Abbott, and M. M. Churchland. Neural trajectories in the supplementary motor area and motor cortex exhibit distinct geometries, compatible with different classes of computation. *Neuron*, 2020. 89, 90

P. N. Sabes. The planning and control of reaching movements. *Current opinion in neurobiology*, 10(6):740–746, 2000. 68

A. T. Sack, C. Jacobs, F. De Martino, N. Staeren, R. Goebel, and E. Formisano. Dynamic premotor-to-parietal interactions during spatial imagery. *Journal of Neuroscience*, 28(34):8417–8429, 2008. 23

I. Sajedian, H. Lee, and J. Rho. Double-deep q-learning to increase the efficiency of metasurface holograms. *Scientific reports*, 9(1):1–8, 2019. 26

T. Salimans and R. Chen. Learning montezuma's revenge from a single demonstration. *arXiv preprint arXiv:1812.03381*, 2018. 20, 184

E. Salinas and L. Abbott. Vector reconstruction from firing rates. *Journal of computational neuroscience*, 1(1-2):89–107, 1994. 129

T. D. Sanger. Probability density estimation for the interpretation of neural population codes. *Journal of neurophysiology*, 76(4):2790–2793, 1996. 129

G. Savelsbergh, H. Whiting, and R. J. Bootsma. Grasping tau. *Journal of experimental psychology: human perception and performance*, 17(2):315, 1991. 28

J. Sawinski, D. J. Wallace, D. S. Greenberg, S. Grossmann, W. Denk, and J. N. Kerr. Visually evoked activity in cortical cells imaged in freely moving animals. *Proceedings of the National Academy of Sciences*, 106(46):19557–19562, 2009. 128

D. L. Schacter, D. R. Addis, and R. L. Buckner. Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, 8(9):657–661, 2007. 23

T. Schaul, J. Quan, I. Antonoglou, and D. Silver. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*, 2015. 186

R. A. Scheidt, J. B. Dingwell, and F. A. Mussa-Ivaldi. Learning to move amid uncertainty. *Journal of neurophysiology*, 86(2):971–985, 2001. 64

J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, 2012. 138

J. Schlag and M. Schlag-Rey. Evidence for a supplementary eye field. *Journal of neurophysiology*, 57(1):179–200, 1987. 73

E. Schneidman, B. Freedman, and I. Segev. Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural computation*, 10(7):1679–1703, 1998. 17

S. H. Scott. Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, 5(7):532–545, 2004. 68

R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 143, 153, 181

P. Sermanet, K. Xu, and S. Levine. Unsupervised perceptual rewards for imitation learning. *arXiv preprint arXiv:1612.06699*, 2016. 185

R. Shadmehr and H. H. Holcomb. Neural correlates of motor memory consolidation. *Science*, 277(5327):821–825, 1997. 22

R. Shadmehr and J. W. Krakauer. A computational neuroanatomy for motor control. *Experimental brain research*, 185(3):359–381, 2008. 68, 89

R. Shadmehr and F. A. Mussa-Ivaldi. Adaptive representation of dynamics during learning of a motor task. *Journal of neuroscience*, 14(5):3208–3224, 1994. 68

L. F. Shampine. Vectorized adaptive quadrature in matlab. *Journal of Computational and Applied Mathematics*, 211(2):131–140, 2008. 38

M. E. Sheffield and D. A. Dombeck. Calcium transient prevalence across the dendritic arbour predicts place field properties. *Nature*, 517(7533):200–204, 2015. 179

O. A. Shemesh, C. Linghu, K. D. Piatkevich, D. Goodwin, O. T. Celiker, H. J. Gritton, M. F. Romano, R. Gao, C.-C. J. Yu, H.-A. Tseng, et al. Precision calcium imaging of dense neural populations via a cell-body-targeted calcium indicator. *Neuron*, 107(3):470–486, 2020. 179

B. Shook, M. Schlag-Rey, and J. Schlag. Primate supplementary eye field. ii. comparative aspects of connections with the thalamus, corpus striatum, and related forebrain nuclei. *Journal of Comparative Neurology*, 307(4):562–583, 1991. 73

T. Shuman, D. Aharoni, D. J. Cai, C. R. Lee, S. Chavlis, L. Page-Harley, L. M. Vetere, Y. Feng, C. Y. Yang, I. Mollinedo-Gajate, et al. Breakdown of spatial coding and interneuron synchronization in epileptic mice. *Nature neuroscience*, 23 (2):229–238, 2020. 180

K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013. 143, 181

B. Sivyer and S. R. Williams. Direction selectivity is computed by active dendritic integration in retinal ganglion cells. *Nature neuroscience*, 16(12):1848–1856, 2013. 178

J. B. Smeets, E. Brenner, S. Trebuchet, and D. R. Mestre. Is judging time-to-contact based on 'tau'? *Perception*, 25(5):583–590, 1996. 28

A. C. Smith and E. N. Brown. Estimating a state-space model from point process observations. *Neural computation*, 15(5):965–991, 2003. 81, 83, 96

A. C. Smith, J. D. Scalon, S. Wirth, M. Yanike, W. A. Suzuki, and E. N. Brown. State-space algorithms for estimating spike rate functions. *Computational Intelligence and Neuroscience*, 2010, 2010. 81, 84

S. L. Smith, I. T. Smith, T. Branco, and M. Häusser. Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. *Nature*, 503(7474):115–120, 2013. 178

J. F. Soechting, J. Z. Juveli, and H. M. Rao. Models for the extrapolation of target motion for manual interception. *Journal of Neurophysiology*, 102(3):1491–1502, 2009. 28, 62

S. Sohn, H. Woo, J. Choi, and H. Lee. Meta reinforcement learning with autonomous inference of subtask dependencies. *arXiv preprint arXiv:2001.00248*, 2020. 185

A. Solway and M. M. Botvinick. Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychological review*, 119(1):120, 2012. 183

M. A. Sommer and R. H. Wurtz. A pathway in primate brain for internal monitoring of movements. *Science*, 296(5572):1480–1482, 2002. 68

M. A. Sommer and R. H. Wurtz. Brain circuits for the internal monitoring of movements. *Annu. Rev. Neurosci.*, 31:317–338, 2008. 23

J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014. 181

B. J. Stankiewicz, G. E. Legge, J. S. Mansfield, and E. J. Schlicht. Lost in virtual space: Studies in human and ideal spatial navigation. *Journal of Experimental Psychology: Human Perception and Performance*, 32(3):688, 2006. 20

S. D. Stavisky, J. C. Kao, S. I. Ryu, and K. V. Shenoy. Trial-by-trial motor cortical correlates of a rapidly adapting visuomotor internal model. *Journal of Neuroscience*, 37(7):1721–1732, 2017. 68

H. W. Steenland, X.-Y. Li, and M. Zhuo. Predicting aversive events and terminating fear in the mouse anterior cingulate cortex during trace fear conditioning. *Journal of Neuroscience*, 32(3):1082–1095, 2012. 126

F. Stefanini, L. Kushnir, J. C. Jimenez, J. H. Jennings, N. I. Woods, G. D. Stuber, M. A. Kheirbek, R. Hen, and S. Fusi. A distributed neural code in the dentate gyrus and in ca1. *Neuron*, 2020. 180

A. A. Stocker and E. P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4):578–585, 2006. 22, 30, 36, 40, 47, 63

R. S. Sutton, A. G. Barto, et al. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge, 1998. 24, 25, 191

K. Svoboda, F. Helmchen, W. Denk, and D. W. Tank. Spread of dendritic excitation in layer 2/3 pyramidal neurons in rat barrel cortex in vivo. *Nature neuroscience*, 2 (1):65–73, 1999. 178

P. Tadepalli and D. Ok. Model-based average reward reinforcement learning. *Artificial intelligence*, 100(1-2):177–224, 1998. 184

N. Takahashi, K. Kitamura, N. Matsuo, M. Mayford, M. Kano, N. Matsuki, and Y. Ikegaya. Locally synchronized synaptic inputs. *Science*, 335(6066):353–356, 2012. 178

A. Tampuu, T. Matiisen, H. F. Ólafsdóttir, C. Barry, and R. Vicente. Efficient neural decoding of self-location with a deep recurrent network. *PLoS computational biology*, 15(2):e1006822, 2019. 181

Y. Tanimoto, A. Yamazoe-Umemoto, K. Fujita, Y. Kawazoe, Y. Miyanishi, S. J. Yamazaki, X. Fei, K. E. Busch, K. Gengyo-Ando, J. Nakai, et al. Calcium dynamics regulating the timing of decision-making in c. elegans. *Elife*, 6:e21629, 2017. 128

H. Tassinari, T. E. Hudson, and M. S. Landy. Combining priors and noisy visual cues in a rapid pointing task. *Journal of Neuroscience*, 26(40):10154–10163, 2006. 30, 40, 63

G. T. Taylor. Fear and affiliation in domesticated male rats. *Journal of Comparative and Physiological Psychology*, 95(5):685, 1981. 95

E. L. Thorndike. Animal intelligence: an experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4):i, 1898. 183

S. B. Thrun. Efficient exploration in reinforcement learning. 1992. 24, 25

E. Todorov. Optimality principles in sensorimotor control. *Nature neuroscience*, 7 (9):907–915, 2004. 68

M. Tokic. Adaptive $\varepsilon$-greedy exploration in reinforcement learning based on value differences. In *Annual Conference on Artificial Intelligence*, pages 203–210. Springer, 2010. 26

M. Tomasello, M. Davis-Dasilva, L. CamaK, and K. Bard. Observational learning of tool-use by young chimpanzees. *Human evolution*, 2(2):175–183, 1987. 70

R. B. Tootell, M. Silverman, S. Hamilton, E. Switkes, and R. De Valois. Functional anatomy of macaque striate cortex. v. spatial frequency. *Journal of Neuroscience*, 8(5):1610–1624, 1988. 67

V. Torre, J. Ashmore, T. Lamb, and A. Menini. Transduction and adaptation in sensory receptor cells. *Journal of Neuroscience*, 15(12):7757–7768, 1995. 17

J. Tresilian. Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction-motion and relative judgment tasks. *Perception & Psychophysics*, 57(2):231–245, 1995. 64

J. R. Tresilian. Visually timed action: time-out for 'tau'? *Trends in cognitive sciences*, 3(8):301–310, 1999. 28

Y.-w. Tseng, J. Diedrichsen, J. W. Krakauer, R. Shadmehr, and A. J. Bastian. Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of neurophysiology*, 98(1):54–62, 2007. 23

M. V. Tsodyks and H. Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the national academy of sciences*, 94(2):719–723, 1997. 17

M. Tu, R. Zhao, A. Adler, W.-B. Gan, and Z. S. Chen. Efficient position decoding methods based on fluorescence calcium imaging in the mouse hippocampus. *Neural Computation*, 32(6):1144–1167, 2020. 180

G. G. Turrigiano and S. B. Nelson. Homeostatic plasticity in the developing nervous system. *Nature reviews neuroscience*, 5(2):97–107, 2004. 22

G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. B. Nelson. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391(6670):892–896, 1998. 22

R. C. Twining, J. E. Vantrease, S. Love, M. Padival, and J. A. Rosenkranz. An intra-amygdala circuit specifically regulates social fear learning. *Nature neuroscience*, 20 (3):459–469, 2017. 69, 70

R. J. Van Beers, A. C. Sittig, and J. J. D. Van Der Gon. The precision of proprioceptive position sense. *Experimental brain research*, 122(4):367–377, 1998. 17

R. J. Van Beers, A. C. Sittig, and J. J. D. v. d. Gon. Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of neurophysiology*, 81(3):1355–1364, 1999. 22

R. J. van Beers, P. Baraduc, and D. M. Wolpert. Role of uncertainty in sensorimotor control. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1424):1137–1145, 2002. 17, 40

R. d. R. Van Steveninck and S. Laughlin. The rate of information transfer at graded-potential synapses. *Nature*, 379(6566):642–645, 1996. 21

C. Van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996. 18

M. Vecerik, T. Hester, J. Scholz, F. Wang, O. Pietquin, B. Piot, N. Heess, T. Rothörl, T. Lampe, and M. Riedmiller. Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards. *arXiv preprint arXiv:1707.08817*, 2017. 20, 184

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett,

J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. 144

J. T. Vogelstein, A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of neurophysiology*, 104(6):3691–3704, 2010. 129

J. Wang, D. Narain, E. A. Hosseini, and M. Jazayeri. Flexible timing by temporal scaling of cortical responses. *Nature neuroscience*, 21(1):102–110, 2018. 89

M. Wang, X. Liao, R. Li, S. Liang, R. Ding, J. Li, J. Zhang, W. He, K. Liu, J. Pan, et al. Single-neuron representation of learned complex sounds in the auditory cortex. *Nature communications*, 11(1):1–14, 2020. 128

X. Wang. Neural coding strategies in auditory cortex. *Hearing research*, 229(1-2): 81–93, 2007. 67

X. Wang, T.-K. Huang, and J. Schneider. Active transfer learning under model shift. In *International Conference on Machine Learning*, pages 1305–1313, 2014. 184

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 137

X.-X. Wei and A. A. Stocker. Efficient coding provides a direct link between prior and likelihood in perceptual bayesian inference. *Advances in neural information processing systems*, 25:1304–1312, 2012. 22

L. Welch. The perception of moving plaids reveals two motion-processing stages. *Nature*, 337(6209):734–736, 1989. 36, 47

J. A. White, J. T. Rubinstein, and A. R. Kay. Channel noise in neurons. *Trends in neurosciences*, 23(3):131–137, 2000. 17

R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992. 26

T. J. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723): 873–876, 2005. 22

M. A. Wilson and B. L. McNaughton. Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124):1055–1058, 1993. 129, 180

D. M. Wolpert and M. Kawato. Multiple paired forward and inverse models for motor control. *Neural networks*, 11(7-8):1317–1329, 1998. 23

D. M. Wolpert, Z. Ghahramani, and M. I. Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882, 1995. 18, 23, 68, 89

W. Wu, Y. Gao, E. Bienenstock, J. P. Donoghue, and M. J. Black. Bayesian population decoding of motor cortical activity using a kalman filter. *Neural computation*, 18(1):80–118, 2006. 129

X.-S. Wu, L. Xue, R. Mohan, K. Paradiso, K. D. Gillis, and L.-G. Wu. The origin of quantal size variation: vesicular glutamate concentration plays a significant role. *Journal of Neuroscience*, 27(11):3046–3056, 2007. 17

M. Xu, S.-y. Zhang, Y. Dan, and M.-m. Poo. Representation of interval timing by temporally scalable firing patterns in rat prefrontal cortex. *Proceedings of the National Academy of Sciences*, 111(1):480–485, 2014. 89, 90

E. Yaksi and R. W. Friedrich. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved ca 2+ imaging. *Nature methods*, 3(5):377–383, 2006. 129

T. Yoshida and K. Ohki. Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. *Nature communications*, 11(1): 1–19, 2020. 128

W. Yoshida and S. Ishii. Resolution of uncertainty in prefrontal cortex. *Neuron*, 50 (5):781–789, 2006. 20

B. M. Yu, A. Afshar, G. Santhanam, S. Ryu, K. V. Shenoy, and M. Sahani. Extracting dynamical structure embedded in neural activity. *Advances in neural information processing systems*, 18:1545–1552, 2005. 22, 91, 97

B. M. Yu, J. P. Cunningham, G. Santhanam, S. I. Ryu, K. V. Shenoy, and M. Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of neurophysiology*, 102(1):614–635, 2009. 22, 84, 91, 97

K. Yu, S. Ahrens, X. Zhang, H. Schiff, C. Ramakrishnan, L. Fenno, K. Deisseroth, F. Zhao, M.-H. Luo, L. Gong, et al. The central amygdala controls learning in the lateral amygdala. *Nature neuroscience*, 20(12):1680–1685, 2017. 128

R. Yuste and W. Denk. Dendritic spines as basic functional units of neuronal integration. *Nature*, 375(6533):682–684, 1995. 178

M. Zago, J. McIntyre, P. Senot, and F. Lacquaniti. Visuo-motor coordination and internal models for object interception. *Experimental Brain Research*, 192(4):571–604, 2009. 28, 62

M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014. 181

K. Zhang, I. Ginzburg, B. L. McNaughton, and T. J. Sejnowski. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *Journal of neurophysiology*, 79(2):1017–1044, 1998. 129, 180

P. Zhou, S. L. Resendez, J. Rodriguez-Romaguera, J. C. Jimenez, S. Q. Neufeld, A. Giovannucci, J. Friedrich, E. A. Pnevmatikakis, G. D. Stuber, R. Hen, et al. Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *Elife*, 7:e28728, 2018. 128, 129, 130, 134, 138, 178

Y. Ziv, L. D. Burns, E. D. Cocker, E. O. Hamel, K. K. Ghosh, L. J. Kitch, A. El Gamal, and M. J. Schnitzer. Long-term dynamics of ca1 hippocampal place codes. *Nature neuroscience*, 16(3):264, 2013. 128, 180

W. Zong, R. Wu, M. Li, Y. Hu, Y. Li, J. Li, H. Rong, H. Wu, Y. Xu, Y. Lu, et al. Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice. *Nature methods*, 14(7):713–719, 2017. 128