

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Kanishk Jain

Date

Multi-timescale representation of animal behavior

By

Kanishk Jain
Doctor of Philosophy

Physics

Gordon Berman, Ph.D.
Advisor

Ilya Nemenman, Ph.D.
Committee Member

Robert Liu, Ph.D.
Committee Member

Jennifer Rieser, Ph.D.
Committee Member

Eric Weeks, Ph.D.
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Multi-timescale representation of animal behavior

By

Kanishk Jain

BS MS, Indian Institute of Science Education and Research, Mohali, India, 2017

Advisor: Gordon Berman, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Physics
2024

Abstract

Multi-timescale representation of animal behavior
By Kanishk Jain

This dissertation focuses on developing computational methods to identify and characterize long-timescale dynamics in behavioral data across multiple spatiotemporal scales. Utilizing the dynamical nature of Recurrent Neural Networks (RNNs), we formulate novel approaches to learn dynamical models from large timeseries datasets of animal postures. We introduce LIDAR, a framework for training RNNs on extensive timeseries data while maintaining temporal coherence through statefulness. Applying this to human gait data, we generate quantitative gait signatures that encode individual and group-specific locomotor patterns, with potential clinical applications in diagnosis and therapy development. We extend our methodology to create multi-timescale representations of behavior using hierarchical RNN models. We apply this approach to a dataset containing simultaneous neural and behavioral recordings, allowing us to explore neural correlates underlying behavior at multiple timescales.

To address limitations in existing methods for reconstructing state-space dynamics, we propose a novel encoder-decoder RNN architecture capable of identifying long-timescale non-stationarities in time series data generated from a modified Lorenz system. Our work demonstrates the utility of RNNs in approximating dynamical systems from vast behavioral datasets, intentionally leveraging these overparameterized models to learn underlying dynamical timescales. These approaches offer promising avenues for disentangling the hierarchical organization of behavioral patterns, characterizing long-timescale physiological states, and understanding neural dynamics underlying complex behaviors.

Multi-timescale representation of animal behavior

By

Kanishk Jain

BS MS, Indian Institute of Science Education and Research, Mohali, India, 2017

Advisor: Gordon Berman, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Physics
2024

Acknowledgments

I am filled with gratitude and emotion as I reflect on the journey that has brought me to this moment. It has been long and challenging, but it has also been one of the most rewarding experiences of my life.

I would like to express my sincere gratitude to my supervisor, Dr. Gordon Berman, for his guidance, mentorship, support, and patience throughout my PhD. His expertise and knowledge in the field have been instrumental in shaping the direction and content of this research. Thank you, Gordon, for setting the benchmark for the kind of researcher I aspire to be.

Additionally, I would like to thank former and current labmates: Josuan, Itai, Preet, Ben, and the rest, for the camaraderie and sense of community that they have provided throughout this process. Their support and encouragement have kept me going.

I am deeply grateful for the support that my family has offered me throughout this process. Their love and care have been a constant source of strength and inspiration. The biggest thanks must go my Mom and Dad, who are my biggest supporters in life. Their sacrifices, guidance, and encouragement have been invaluable in helping me through the years.

My sister-in-law Zihan was always on hand to offer her wisdom, advice, and inimitable humor. And I thank my brother, Hardik, the smartest and most capable person I know, and whose judgment I trust more than anyone's. And finally, Surbhi, the love of my life. Thank you so much for being there with me along the way. I am thankful for the countless ways in which she has supported me, and for the care that she has provided throughout this process. Her kindness, understanding, and unwavering love have been critical in sustaining me through the ups and downs.

Contents

1	Introduction	1
1.1	A brief history of ethology	3
1.2	Tracking Posture	5
1.3	Postural Dynamics	10
1.4	Takens' Embedding Theorem and ensemble methods for state-space reconstruction	15
1.5	Thesis Outline	20
2	Using Recurrent Neural Networks to mimic dynamical systems	22
2.1	Introduction	22
2.2	Recurrent Neural Networks	25
2.3	LIDAR: a python package to train RNNs statefully	29
2.4	Characterizing human gait dynamics using RNNs	30
2.4.1	Introduction	30
2.4.2	Dataset	34
2.4.3	Model training and selection	36
2.4.4	Measuring gait signatures from RNNs	38
2.4.5	Gait signatures are individual and group-specific	40
2.4.6	Biomechanical interpretability of gait signatures	42
2.4.7	Using RNNs to generalize to new gait speeds	43

2.5	Building multi-timescale representation of animal behavior using RNNs	45
2.5.1	Datasets	46
2.5.2	Building representations at a single timescale	49
2.5.3	Building representations at coarsened timescales	52
2.5.4	Linking behavioral and neural representations across timescales	54
2.6	Conclusion	56
3	Unsupervised learning algorithms reveal sex biases in baseline and stress adaptive behavior	62
3.1	Introduction	62
3.2	Methods	64
3.2.1	Animals	64
3.2.2	Stress	65
3.2.3	Surgeries	65
3.2.4	Behavioral Assessments	65
3.2.5	Chemogenetic Manipulation	66
3.2.6	Posture Tracking	66
3.2.7	Quantitative identification of behaviors	67
3.2.8	Training classifiers across assays	68
3.2.9	Calculating statistical upregulation between two assays	68
3.3	Results	68
3.3.1	Sex biases in the exploration of an open field	68
3.3.2	Stress reverses baseline sex differences	69
3.3.3	Chemogenetic activation of susceptibility circuit drives similar upregulations in both sexes	73
3.4	Discussion	75
4	State-space reconstruction of dynamical systems with long timescale	

non-stationarities	82
4.1 Introduction	82
4.2 The Lorenz System	83
4.3 Modified Lorenz System	87
4.4 Reconstructing state-space dynamics using RNNs	92
4.5 Conclusion	93
5 Discussion	98
Bibliography	102

List of Figures

- 1.1 Various approaches to represent posture. a) Postural representations (such as joint angles) in fruit flies can be derived from body frame coordinates tracked using machine learning techniques. [1] b) Left panel shows tracked variations in a nematode's (*C. elegans*) centerline [2]. Right panel displays postural modes (eigenworms) identified using principal component analysis. Linear combinations of these components can reconstruct the original centerline shapes. c) Using image compression, fly images can be decomposed to identify postural modes [3]. Corresponding postural eigenvectors (right panel) concentrate data's variance in minimal directions. The original image can be reconstructed by linearly combining these modes with an overall mean, and a projection onto these modes generates a postural time series. Figure adapted from [4]. 6

1.2	Using a markerless deep learning tracking tool [5]. a) Training these tools requires extracting representative frames from the video dataset. A region of interest (ROI) can be selected for faster computation. b) Extracted frames are manually annotated with relevant postural markers on the body. When occlusions are present, annotations must be skipped. c) A deep neural network model is trained using manual annotations. Parameters that are updated during training include the backbone architecture as well as readout layers. Marker-specific readout layers generate the probability of detection for each annotated postural marker. d) A trained model can generate frame-by-frame predictions on video data, generating a postural time series. Figure taken from [5].	7
1.3	Neural network architecture for LEAP (figure borrowed from [1]). Raw images enter the system (top left), passing through a series of convolutional, max pooling, and transposed convolutional layers (center). The network outputs confidence maps matching the input image dimensions (top right). During training, these outputs are compared to ground truth maps derived from user labels (bottom left), and a mean squared error loss function quantifies the difference, which is minimized to optimize network performance (bottom right). Arrows indicate data flow and feedback loops in the training process.	8

1.4 Representing postural dynamics across species (taken from [4]). a) *C. elegans*: Histogram of projections onto the first two "eigenworms" (from Fig. 1.1b) reveals a structure parameterizable by a single phase variable ϕ . b) Fitting the dynamics of this variable to a deterministic dynamical system yields this phase map of *C. elegans* locomotion (derived from a), forward and backward movements are represented as wave trajectories, with two fixed points representing pause states. (a and b adapted from [2]). c) Time-series projections onto eigenworms can be used to identify repeated motifs (blue and red curves), and sequences of motifs can then be used for behavioral representation (adapted from [6]). d) Alternatively, worm behavior can be described as a series of discrete postures (from [7]). e) Autoregressive Hidden Markov Model (AR-HMM) trained on mouse postural data obtained from a depth-sensing camera and decomposed using ideas similar to Fig.1.1c. Here, P_t and S_t represent postural mode values and the underlying state influencing posture dynamics, respectively. Arrows indicate the direction of dependencies amongst variables in the model [8]. f) Phenotypic signatures showing behavior usage frequencies obtained using e across four mouse genotypes. g) Distinct walking gaits identified between two different genotypes using behavioral signatures from f (e-g adapted from [8]). h) Time-frequency analysis of fruit fly movement: Continuous wavelet transform amplitudes for different postural modes over time. i) 2D embedding of fruit fly movements: Probability density map where similar movements cluster together and peaks representing stereotyped behaviors. j) Behavioral map: Breakdown of i with manually labeled regions and transition probabilities (black lines). (h-j adapted from [3, 9])

- 1.5 Workflow underlying VAME [10] a) Videos were recorded of mice freely behaving in a circular arena for 25mins from a bottom-up camera. b) 4 paws, nose, and tailbase of mice were tracked using DeepLabCut [5]. c) Tracked keypoints were ego-centered and aligned along a fixed direction (facing right). d) 12-dimensional postural trajectories were used for training the subsequent model. e) The VAME model consists of bidirectional recurrent neural networks, where the encoder RNN receives a 500ms trajectory sample and embeds it into a lower dimensional variational latent space. This representation is fed into 2 RNN decoders, one reconstructing the input data and one predicting the subsequent 500ms section of the postural time series. f) Latent representations are used to train a Hidden Markov Model (HMM) for motif detection. Figure borrowed from [10]. 16
- 1.6 State-space reconstruction and ensemble dynamics (figure taken from [11]). a) From a complex dynamical system (top left), partial observations are measured and are then used to reconstruct the state space using delay-embedding (top right). To maximize short-term predictability, partitioning is performed in multiple delay-embedded spaces constructed with different delay lengths, and unpredictability is measured through the short-term entropy rate (bottom right). This allows selecting a maximally predictive delay embedded space that reflects the properties of the original dynamics (bottom left). b) The eigenspectra of transition operator $P_{ij}(\tau)$ at different τ captures multiple characteristic timescales. At appropriate τ^* , where long-timescale dynamics are approximately Markovian and the spectral gap is large, the slowest eigenvector (ϕ_2^*) reveals metastable states of the system (bottom). . . 18

1.7	Reconstructing state-space of an underdamped particle in a double-well potential (figure borrowed from [11]). Using only the position of the particle (left), delay-embedded spaces are constructed, and entropy rates are measured (center). Inset shows an increase in entropy rate with the number of partitions (N) for spaces with different delay lengths (K). $N^* = 10^3$ is picked to avoid finite-size effects, and $K^* = 7$ is chosen beyond which the entropy production rate becomes constant. Particle trajectories can be reconstructed using simulations of the Markov chain inferred in the previous step.	19
1.8	a) The characteristic timescale of the time-symmetric transition matrix P'_τ estimated from the reconstructed state-space of the underdamped double well system (left). At short transition timescales τ , transient dynamics dominate, resulting in large characteristic time scales and eigenvalues converging close to 1 (left inset). As τ increases, the longest relaxation time becomes constant at $\tau \sim 5$ seconds, equal to half the mean first passage time. The characteristic timescale for double well dynamics at different temperatures (β^{-1}) predicts the hopping rate between wells (a, right). b) For $\beta^{-1} = 0.5J$, the first two leading singular modes of the second leading eigenvector ϕ_2 splits the two wells. Figure taken from [11].	20
2.1	Example phase averaged gait kinematics. 6 joint angles were measured from participants walking at variable speeds. Here, phase-averaged kinematics are shown from one gait cycle.	35

2.2 Training RNN model on kinematics data. A. A sequence-to-sequence RNN model was trained using kinematic data, where each input to the model is a sequence of kinematic data, and the model predicts a sequence of kinematic data shifted by one time step, i.e., the model predicts kinematics data at $t+1$ given kinematics at time t . B. A simple RNN architecture with one hidden LSTM layer with 512 units and 1 dense output layer was used. The input layer and output (dense) layer had 6 units, corresponding to the dimensionality of kinematics data. C. The model was trained statefully, where each sample in each mini-batch was temporally consecutive, resulting in temporal continuity across batches. The hidden states of the model were not reset after each batch was processed.

2.3 Identifying gait-signatures from kinematic data. A) Continuous, multi-joint movement data from multiple individuals are input into a Recurrent Neural Network (RNN) model. The model is trained to predict kinematic output one time step ahead using data from all individuals. Using the trained model, high-dimensional internal states (\vec{h} and \vec{h}) from each individual are extracted from the model. Principal component analysis is then applied to reduce the data's dimensionality, forming individual gait signatures. B)(Left) Visualization of the first 3 principal components of gait signatures from 3 individuals: able-bodied (blue), high-functioning stroke survivor (red), and low-functioning stroke survivor (orange). B)(Right) 3D projections of the 6D gait signatures using multi-dimensional scaling (MDS) reveal different gait dynamics among the three groups. The size of each circle represents the individual's walking speed during the trial, with smaller circles indicating slower speeds and larger circles representing faster speeds. Figure taken from [12].

2.4 Gait signatures are individual and group-specific. 3D gait signatures for unimpaired subjects (left) and impaired subjects (right) are color-coded in two ways: i) by individual and ii) by gait phase. In the individual-based coloring (i), signatures cluster according to specific subjects within both groups. For unimpaired subjects (i, left), similar shades of blue group together, while for impaired subjects (i, right), similar shades of red form clusters. This grouping indicates that each person has a unique gait signature. When color-coded by gait phase (ii), unimpaired gait signatures (ii, left) show a consistent looped structure across the four phases of a gait cycle: leg 1 swing, leg 1 stance, leg 2 swing, and leg 2 stance. In contrast, impaired signatures (ii, right) exhibit subject-specific variations across these four phases and demonstrate greater overall variability. Figure taken from [12]. 41

2.5 Comparison of gait signatures between three distinct locomotor subgroups: able-bodied (AB), high-functioning stroke survivors (HF), and low-functioning stroke survivors (LF). 3D gait map created using multi-dimensional scaling shows the relative spatial relationships between AB subjects (in blue), HF stroke survivors (in red), and LF stroke survivors (in orange). LF stroke survivors exhibit less cohesive clustering, occupying unique regions within the map. These regions are notably distant from the centroid of the able-bodied group, which is represented by a black square marker. Figure taken from [12]. 42

- 2.6 Biomechanical analysis of gait signatures. A) The gait signatures demonstrate distinct walking dynamics among representative able-bodied (AB) individuals, low-functioning, and high-functioning stroke survivors. B) The weightings of each principal component (PC) fluctuate throughout the gait cycle. These can be contrasted with the 95% confidence interval of the AB group (shown in gray). C) When used to drive the gait model, each PC produces specific multi-joint walking coordination patterns. This enables a biomechanical interpretation of gait impairments and the effects of interventions. Figure taken from [12]. 43
- 2.7 Generalizing to new gait speeds. Joint kinematic predictions (shown in green) for intermediate walking speeds were excluded from model training. They were generated by first interpolating gait signatures (hidden-states) obtained from slow (dashed grey) and fast (dashed black) speed trials and then driving the model with the interpolated gait signatures to predict kinematics at the intermediate speed. Joint kinematics interpolated directly from kinematics data (blue) and the measured reference kinematics (solid black) are also shown. A) For a representative able-bodied (AB) participant, predictions are more accurate when interpolating gait signatures compared to interpolating gait kinematics across speeds. B) In a representative low-functioning stroke survivor, interpolated gait signatures more accurately predict non-linear changes in kinematics at intermediate speeds than interpolated gait kinematics. Simple averaging of kinematics fails in this case, where larger differences exist between the slow and fast speed paretic kinematics. The averaged kinematics (blue) follow the fast speed paretic hip kinematics, while other joint angles do not resemble waveforms from either the fast or slow speed. Figure taken from [12]. 44

2.8 Visualization of features in a multilayer convolutional neural network trained on ImageNet 2012 dataset. For layers 2-5, the nine highest activations in a randomly selected subset of feature maps across the validation dataset are shown. These activations are projected back to pixel space using the method described in [13]. Individual feature maps are highly clustered, and there is increased feature invariance in higher layers. Layers 4 and 5 show strong activation for distinct image elements such as canine eyes and noses. Figure taken from [13]. . . . 47

2.9 Behavioral datasets used here. A) Leap Fly Dataset. Postures were tracked from videos of freely behaving fruit flies (imaged at 100 Hz) obtained from [3] using LEAP [1]. 32 keypoints on the body of the fly were tracked. B) OFT Rat Dataset. Freely behaving rats in an enriched arena were imaged from above at 60 Hz, and 13 postural markers on the rat were tracked using DeepLabCut [5]. C) Motion Capture Rats. This dataset is taken from [14], 3D motion tracking of 20 postural markers was performed on rats at 300 Hz using infrared beads attached. Simultaneous neural recordings from the dorsolateral striatum (DLS) were also performed. 48

2.10 Level 1 RNN model for generating behavioral representation. (Top) A 3-layer recurrent neural network with U-Net-inspired architecture and LSTM units receives the posture of the animal at time t $\vec{X}(t)$ and predicts the posture at $t+1$ $\vec{X}(t+1)$. (Middle) After training, the entire postural time series $\vec{X}(t)$ is fed through the network, and the internal states from layer 2 $\vec{y}_1(t)$ are extracted containing both $\vec{h}(t)$ and $\vec{c}(t)$ variables of the internal states. (Bottom) $\vec{y}_1(t)$ is used to build a 2D representation following a pipeline inspired from [3]. Each dimension in $\vec{y}_1(t)$ is first decomposed into a spectrogram using wavelet transform and then concatenated together, the high-dimensional representation is then embedded in 2D (Z_1) using t-SNE [15]. 50

2.11 Power spectral density (left, normalized across layers on right) of extracted hidden state dynamics from each layer of level 1 RNN model trained Motion Capture Rat Dataset. 51

2.12 2D representations of each dataset from Fig.2.9 obtained using the level 1 RNN model(Fig.2.10. 52

2.13 Hierarchical RNNs. (Top) Hidden state dynamics $\vec{y}_1(t)$ obtained from the level 1 RNN model are subsampled by a factor of 2 ($\vec{Y}_1(t')$) and are used to statefully train a different but identical RNN model (level 2). After training, hidden states of the level 2 model are extracted $\vec{y}_2(t)$ and embedded into a 2D representation Z_2 . This process is repeated until level 6, after which there is not enough data left to train the subsequent model due to subsampling. 53

2.14 Power spectral density (normalized) of hidden state dynamics $\vec{y}_1(t)$ to $\vec{y}_6(t)$ obtained from models from each level trained on the Motion Captured Rats dataset. 54

2.15	2D representations obtained using RNN models (levels 1 to 6) for LEAP Fly Dataset (left), Open Field Rat dataset (middle), and Motion Capture Rats dataset (right). Representations show coarsening at deeper levels.	55
2.16	(Top) Using the 2D representation from level 1 (Z_1) for Motion Capture Rat dataset, Deterministic Information Bottleneck [16] (DIB) is used to minimize \mathcal{F} , identifying a minimally complex partitioning of Z_1 (for a fixed number of partitions) while preserving information between the partitions and spiking data from a neural unit i . (Bottom) Pareto front obtained during DIB optimization, optimal solutions lie in the direction indicated by the arrow indicating low complexity of partitioning while maximizing the encoded information.	57
2.17	(Continuation of Fig.2.16) Pareto curves obtained while partitioning level 1 representation from Motion Capture Rats while encoding a spike train from each neural unit. For each unit, the maximum possible information encoded (I_{max}) is measured.	58
2.18	Pareto fronts curves obtained for each unit at each level of representation (Z_1 to Z_6 from Motion Capture Rats. Unit a (solid) and b (dashed) lines represent two units with statistically significant information encoding in level 1 (in case of unit a) and level 6 (for unit b).	58
2.19	(Left) Given a behavioral representation ($p(B)$), a conditional representation is calculated for when unit a is spiking ($p(B A)$). (Right) Conditional representation of behavior when spiking occurs in unit a (top) and unit b (bottom). Unit a shows a punctate representation in level 1 (thus high information encoding). Similarly, unit b shows a clustered representation in level 6.	59

3.1	Unveiling sex-specific behaviors through unsupervised behavioral representation	70
3.2	Stress alters behavioral profiles in female mice from active to inactive behaviors. A. Mice were exposed to either a chronic (21-day) or subchronic (6d) variable stress before being tested in the OFT. B,C Time spent in the center of open field, a field-standard metric for assessing response to stress. SCVS females show less time spent in the center, while CVS females show no difference compared to the control animals. D,E. Sub-chronic stressed females show upregulation of male-typical less active behaviors like still nose sweeps, idling, and rearing in female mice, whereas control animals show upregulation of active behaviors like walking and running and controls show upregulation of female-typical active behaviors. F. Logistic regression coefficients show less active behaviors like still nose sweeps, idling, and pause and runs as more predictive of sub-chronic stress, and active behaviors like run, pause and run, walk and rear are more predictive of control condition. G,H. Chronic stressed female mice also show upregulation of less active behaviors but distinctly upregulated in walking and rearing behaviors, while controls show upregulation in active as well as inactive behaviors like running, idling, and still nose sweeps. I. Logistic regression coefficients show still and reared nose sweep behaviors to have strong predictive power for CVS females, and running, intermittent running, and still nose sweep behaviors to be predictive of control females. . . .	72

3.3 Chronic and sub-chronic stress changes male mice behaviors differentially. A. Control males show upregulation of male-typical less active behaviors like nose sweeps, idling, and grooming. B. SCVS male mice show upregulation of active behaviors like walking, intermittent running, and running. C. No SCVS-dependent differences were observed in time in the center of the open field for males. D,E. Chronically stressed male mice also show upregulation of some active behaviors like intermittent running compared to controls but retain male-typical less-active behaviors like still nose sweeps. F. CVS males spend less time in the center of OFT. G. Logistic regression coefficients show active behaviors like intermittent running and running as more predictive of SCVS males and less active behaviors like idle, nose sweeps, and rearing to be more predictive of controls. H. For CVS males, intermittent running and still nose sweeps are more predictive, whereas rearing, walk and rear, still nose sweeps, and intermittent running are more predictive of the control condition.

3.4	Chemogenetic activation of susceptibility circuit drives similar upregulations in both sexes. A. Mice were injected with a retrograding AAV-cre into the NAc and a cre-dependent excitatory DREADD into the vHIP to target NAc projecting vHIP cells. C21 injection was given 30 minutes before the OFT to activate the projections. B. Logistic regression coefficients for pairwise classification of behavioral cluster densities from control males and females. C. Behavioral upregulation in control females following CNO administration. D. Behavioral upregulation in DREADD-injected females post-CNO administration. E. Logistic regression coefficients showing sex-specific and treatment-dependent behavioral predictors in female subjects. Positive values indicate DREADD-associated behaviors, while negative values represent control-associated behaviors. F. Behavioral upregulation in control males following CNO administration. G. Behavioral upregulation in DREADD-injected males post-CNO administration. H. Logistic regression coefficients showing sex-specific and treatment-dependent behavioral predictors in male subjects.	76
4.1	Reconstructing the state=space of Lorenz attractor. A) 2D projection of the standard Lorenz system with $\sigma = 10$, $\rho = 28$ and $\delta = 8/3$, showing 2 attractor states. Only $x(t)$ is measured from the system (partial observation). B) (Inset) Entropy rate ($h(N, K)$) as a function of N (number of partitions) and K (delay length) of transition dynamics in the reconstructed state-space. $h(N, K)$ saturates after $N^* = 3162$, indicating maximum predictability. The entropy rate stops changing after $K^*=12$, indicating saturation of predictive information. Figure taken from [11].	84

4.2	a) Relaxation timescales vs. transition time (τ) of the inferred transition operator in delay embedded state-space of the Lorenz system. Transient dynamics (within a partition) contribute to the initial peak in relaxation times after which they settle to near-constant value. Beyond this regime, the relaxation times are driven by the quasi-periodicity of the Lorenz attractor. B) The first non-trivial eigenvector ϕ_2 projected onto the first 2 SVD modes of the delay-embedded space. The phase space is split into two metastable states by ϕ_2 . Figure taken from [11].	85
4.3	Modified Lorenz attractor. a) Particle in a double well potential (left) is simulated and positions ($h(t)$) are extracted (right). b) Mean escape time of the particle as a function of inverse temperature β . c) Modified Lorenz attractor equations. Only \dot{x} is modified, where parameter σ is scaled based on the position of the particle in double well potential. γ determines the strength of scaling, $\sigma = 10$, $\rho = 28$ and $\delta = 8/3$ are Lorenz attractor parameters.	87
4.4	Sample trajectories from modified Lorenz system for 4 pairs of (γ, β) values, where γ determines the extent of scaling of σ parameter and β is the inverse temperature for the particle in the double well. As the particle shifts from left to right well, sigma values modulate between σ_1 and σ_2 values (calculated at the center of the wells). Trajectories are colored red when the particle is in the left well and blue for the right well.	89
4.5	Entropy rate for the modified Lorenz system with $(\gamma, \beta) = (1, 0.40)$. Here we choose $N^*=2512$ and $K^*=6$	89
4.6	Characteristic timescales for the system with various values of β and γ with (right) and without (left) making the transition matrix reversible at a transition timescale of $\tau^*=0.10$ s.	90

4.7	First five non-trivial eigenvectors of the delay-embedded modified Lorenz system (similar to Fig.4.2B) for various values of γ and $\beta = 0.40$. . .	91
4.8	Mutual Information between the leading non-trivial eigenvector ϕ_2 and partitioned delay embedded space (KNN)(left), between the position of the particle in left or right well and KNN (center), and between the position of the particle in the well and ϕ_2	91
4.9	Encoder-Decoder RNN Model. Temporal segments of length (L) are used to drive an encoder Recurrent Neural Network model, generating a final hidden state $h_1(L)$. A decoder RNN model uses $h_1(L)$ as its initial state and is driven without any input to reconstruct the temporal segment input into the encoder RNN. Concurrent temporal segments are used during training for statefulness.	92
4.10	3D UMAP[17] projections of hidden states extracted from trained encoder model in Figure 4.9. Here 3 different models were trained for each (γ, β) pair and hidden states were extracted as each model is driven with time series from each system.	94
4.11	Level 2 model. We train another encoder-decoder RNN model (inspired from Section 2.5.3 using hidden states extracted from the previous model. We follow the same training style as before, training the model statefully. After training, hidden states $h_2(t)$ were extracted using $h_1(t)$. 95	95
4.12	3D UMAP[17] projections of hidden states extracted from level 2 encoder model from Figure 4.11. Here 3 different models were trained for each (γ, β) pair using corresponding $h_1(t)$ and hidden states $h_2(t)$ were extracted as each model is driven with $h_1(t)$	96

List of Tables

- 3.1 **Experimental design.** Table showing different experimental groups, stress conditions for each experiment, and the number of animals from each sex in each experiment. 81
- 3.2 **Logistic regression parameters.** Logistic regression parameters used to train pairwise classifiers between class 1 and class 2. All classifiers shared the following parameter values other than L1 strength: penalty=l1, dual=False, tol=0.0001, fit intercept=True, intercept scaling=1, class weight=balanced, solver=saga. 81

List of Algorithms

1	Stateful RNN Training Epoch	31
---	---------------------------------------	----

Chapter 1

Introduction

Despite the presence of conceptual frameworks to study behavior in the current literature, it remains a challenge to measure and quantify behavior across the wide range of spatio-temporal scales it spans. This difficulty often arises from a difficulty in defining behavior, which remains subjective across scientific fields [18, 19, 20]. In current literature, behavior is defined and studied across a wide range of spatiotemporal scales, from fast muscle twitches occurring over milliseconds to physiological responses such as sweating and salivating that span multiple seconds, and from complex sequences of postural actions that symbolize behaviors like courtship [21] to longer timescale internal processes such as hunger [22] and migration [4].

Advances in technology have allowed measuring some of these paradigms pragmatically and accurately. For example, we have the ability to attach electrodes to specific muscles and detect precise electrical activity associated with each muscle contraction [23] while simultaneously recording ultrasound data, allowing us to correlate electrical activity with physical movement of a muscle [24]. Similarly, thermal cameras [25] and skin conductance [26] measurements can be used to detect thermoregulatory processes like sweating. Moreover, the last two decades saw unprecedented advances in image processing, computing, and machine learning that have resulted in significant

progress in our ability to measure animal postures extremely precisely during experiments [27, 1], as well as building statistical [8, 28] or dynamical characterizations [2, 3] of stereotyped postural movements. However, many of the currently available methods focus on modeling postural dynamics, and the intricate interplay between short-term actions and long-term behavioral states still remains unknown. That is, we currently do not have the ability to quantify long-timescale states like hunger cycles, circadian rhythms, and aging. We need a theoretical framework that can bridge these disparate temporal scales while accounting for inter- and intra-individual variability, as well as other genetic, environmental, and physiological factors, all without any deterministic and empirical measures of these long-timescale states.

Overcoming this limitation can help us better understand adaptive behavioral strategies over the lifetime of an animal, the evolution and genetic predispositions of behaviors, internal metabolic processes, and the neural processing of motor control and decision-making. A coarsened understanding of behavior will also likely reveal different phenomena and inspire new theories of behavior [29, 30]. In this dissertation, I focus on developing methods to identify and measure the internal dynamics governing temporal structures and patterns in behaviors that are not directly observable, explicitly focusing on dynamics happening far beyond the typical postural timescales. I begin this chapter by briefly discussing the history of the field of ethology and its rise as a quantitative discipline. I then discuss technical advancements in the field of postural measurement and methods that build statistical and dynamic representations of postural movements. Subsequently, I will discuss the limitations and shortcomings of current behavioral mapping approaches in discovering longer timescale dynamics and discuss the merits of recently published ideas on data-based state-space reconstruction of dynamical systems.

1.1 A brief history of ethology

Ethology is the study of animal behavior in naturalistic environments. Behavior is a result of hierarchical state-dependent computations in the brain. That is, it depends on interactions between internal states, external inputs, and multiple levels of neural dynamics across various timescales. [31, 32, 33, 34]. Measuring behavior in addition to neural activity in the brain is essential to understanding the functional context of neural processing [35, 36]. However, measuring behavior is difficult, and the quest to understand and define behavior has captivated thinkers and philosophers for ages. Notably, Aristotle [37] observed and described behavioral phases in organisms such as reproduction, migration, and hibernation. His student, Theophrastus, later studied [38] physiological color changes (metachrosis) in chameleons and octopuses, noting differentiated internal states of metachrosis, particularly fear in chameleons versus foraging in octopuses [39].

In the modern age, the study of physiological phenomena driven by the environment has been brought to new light, particularly with a focus on quantitative approaches to measure animal behavior. In the late 19th and early 20th centuries, Ivan Pavlov's groundbreaking study on measuring physiological responses in dogs to external stimuli inspired the beginning of quantitative ethology [40]. Pavlov's interest resided in understanding the interplay of conditioned vs. unconditioned reflexes and sought to understand the behavioral reflexes of dogs in response to external stimuli. He studied how the salivary glands in canines responded to food, a well-known unconditioned response. In his experiments, trained dogs were able to pair novel neutral stimuli such as visual, auditory, or olfactory cues with food, turning them into conditioned stimuli that generated salivary reflexes. However, these conditioned reflexes were found to weaken if not repeated without reinforcement. He showed how lost reflexes can quickly be restored if paired unconditioned stimulus is presented to dogs. Moreover, a temporal delay of up to 2 minutes between stimuli did not affect condi-

tioning; however, there is a failure in conditioning if the delay is more than 3 minutes. The dogs also developed an ability to distinguish between similar conditioned stimuli (such as musical notes and timbre) and sensitivity to intensity and complexity stimuli (e.g. musical chords). [40].

Later in the 1930s, Konrad Lorenz and his student Nikolaas Tinbergen proposed new directions in ethology, focusing on studying a full range of animal behavior in naturalistic environments. Lorenz proposed a dichotomous separation in behavior, particularly between rigid innate behaviors that develop independently of the animal's experience and its environment and learned behaviors that are acquired during ontogenic development. Lorenz hypothesized the stereotypy in instinctive behaviors is driven by coordination centers in the central nervous systems and that environment stimuli and inner drive force animals to perform these innate behaviors. [41, 42]. He gave the example of egg retrieval behavior in Greylag geese, where geese move closer to the egg and maneuver the egg back into the nest using a stereotyped sequence of actions. When exploring learned behaviors, Lorenz focused on imprinting in young animals across species and showed that this imprinting occurred across multiple senses. Famously, he showed how Greylag goslings, when hatched in his presence, would follow and imitate him, as they would imitate the behaviors of their mother. He noted imprinting as a form of irreversible innate learning, occurring during a temporary window of time where young animals are very receptive to stimuli around them. Tinbergen progressed Lorenz's ideas forward, emphasizing on conceptual frameworks to study behavior in naturalistic environments [43]. He proposed the following four aims of ethology -

- Causation: physiological and neurological causes of behavior.
- Development: non-stationarity in behavior as a result of ontogeny (growth and development)

- Function: survival and success of the animal in the environment.
- Evolution: the evolutionary history of behavior across species.

Tinbergen supported his proposal with observations and experiments with various species of fish and birds. Notably, he presented gull chicks with a cardboard replica of adult gull heads. Some of these replicas had red spots on them (as present on real adult gulls while some did not have the red spots). He discovered that young chicks prefer pecking at adult replicas with red spots. Baby chicks were exhibiting feeding behavior [32]. He also observed and quantified eggshell removal behavior in black-headed gulls and the development of red coloring on the belly of Stickleback fish. These were a step forward in meaningfully quantifying behavior in animals. Other examples of substantial contributions to ethology in the 20th century include Karl von Frisch’s experiments with honeybees. He discovered the presence of color perception in bees and, most notably, observed the use of “waggle dance” by honeybees for social communication [44]. For their fundamental contributions to the field of ethology, Lorenz, Tinbergen, and von Frisch were jointly awarded the Nobel Prize in Physiology or Medicine in 1973.

1.2 Tracking Posture

Once the behavior of interest is identified, a quantitative framework is needed to measure that behavior. Recent progress in deep-learning techniques has allowed for the high-throughput measurement of animal posture, and current tools allow the identification of stereotypical postural movements in animals and the identification of low-dimensional sub-spaces that describe behavioral dynamics over short timescales.

The first step in modeling postural dynamics is to quantify the posture (or pose) of animals. Early tools like Ethovision [45], introduced in 1993, use simple image processing algorithms to monitor and track movements of animals in video files over

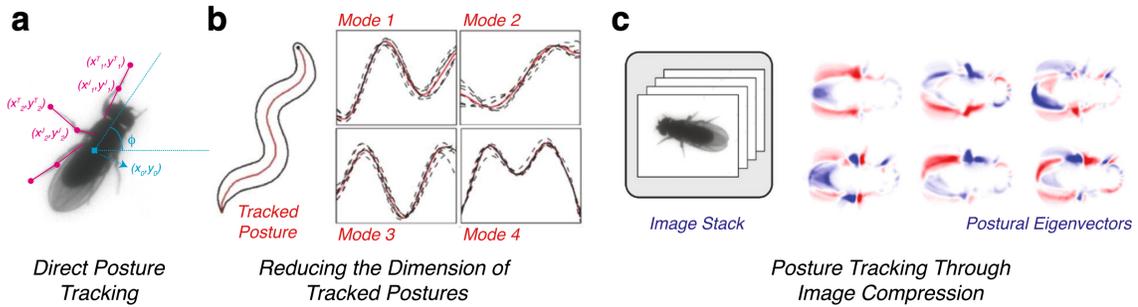


Figure 1.1: Various approaches to represent posture. a) Postural representations (such as joint angles) in fruit flies can be derived from body frame coordinates tracked using machine learning techniques. [1] b) Left panel shows tracked variations in a nematode's (*C. elegans*) centerline [2]. Right panel displays postural modes (eigenworms) identified using principal component analysis. Linear combinations of these components can reconstruct the original centerline shapes. c) Using image compression, fly images can be decomposed to identify postural modes [3]. Corresponding postural eigenvectors (right panel) concentrate data's variance in minimal directions. The original image can be reconstructed by linearly combining these modes with an overall mean, and a projection onto these modes generates a postural time series. Figure adapted from [4].

long durations, providing mobility measures such as distance traveled, 2-dimensional movement speed, and movement duration. Applications of behavioral measures from Ethovision include the analysis of motor deficits in aging monkeys [46], measuring fear-related immobility in mice [47], and quantifying opioid withdrawal in rat pups [48]. Stephens et al. [2] use image processing to go beyond kinematic descriptions and develop a dynamical description of worm posture during locomotion. They image and process freely moving worms using tracking microscopy, obtaining a single-pixel backbone of the worm. By applying a spline fit to this backbone, they measure tangential angles along the curvature to characterize the posture of worms in each frame (Fig 1.1B).

More recently, deep learning methods have significantly simplified pose estimation in video data. A major breakthrough in deep learning was AlexNet [49], a convolutional neural network with 8 layers that gained prominence after winning a state-of-the-art object detection challenge. Its success catalyzed the adoption of convolutional

DeepLabCut: markerless tracking toolbox

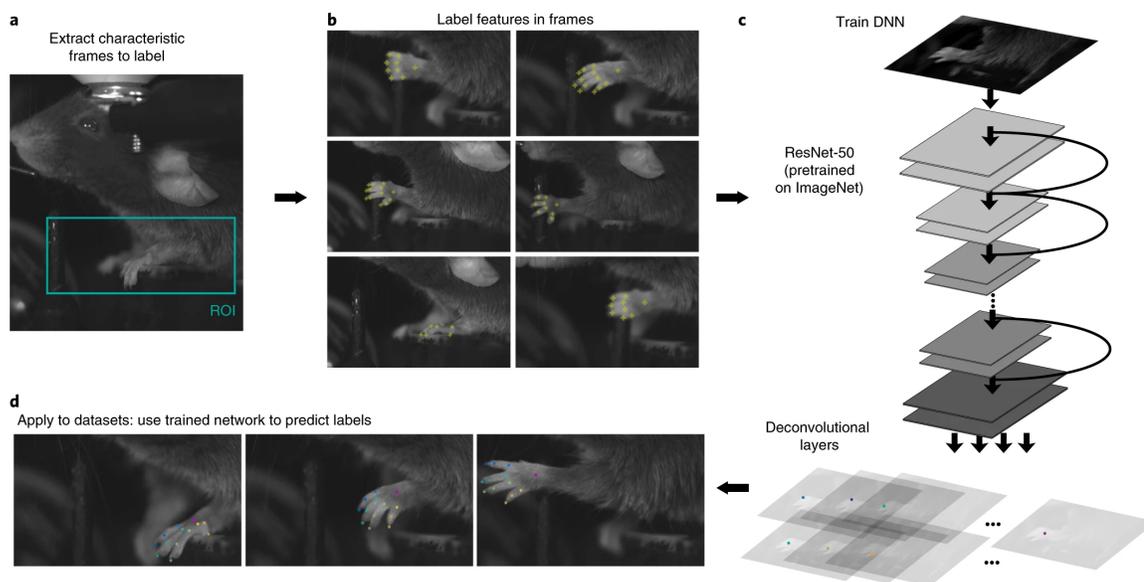


Figure 1.2: Using a markerless deep learning tracking tool [5]. a) Training these tools requires extracting representative frames from the video dataset. A region of interest (ROI) can be selected for faster computation. b) Extracted frames are manually annotated with relevant postural markers on the body. When occlusions are present, annotations must be skipped. c) A deep neural network model is trained using manual annotations. Parameters that are updated during training include the backbone architecture as well as readout layers. Marker-specific readout layers generate the probability of detection for each annotated postural marker. d) A trained model can generate frame-by-frame predictions on video data, generating a postural time series. Figure taken from [5].

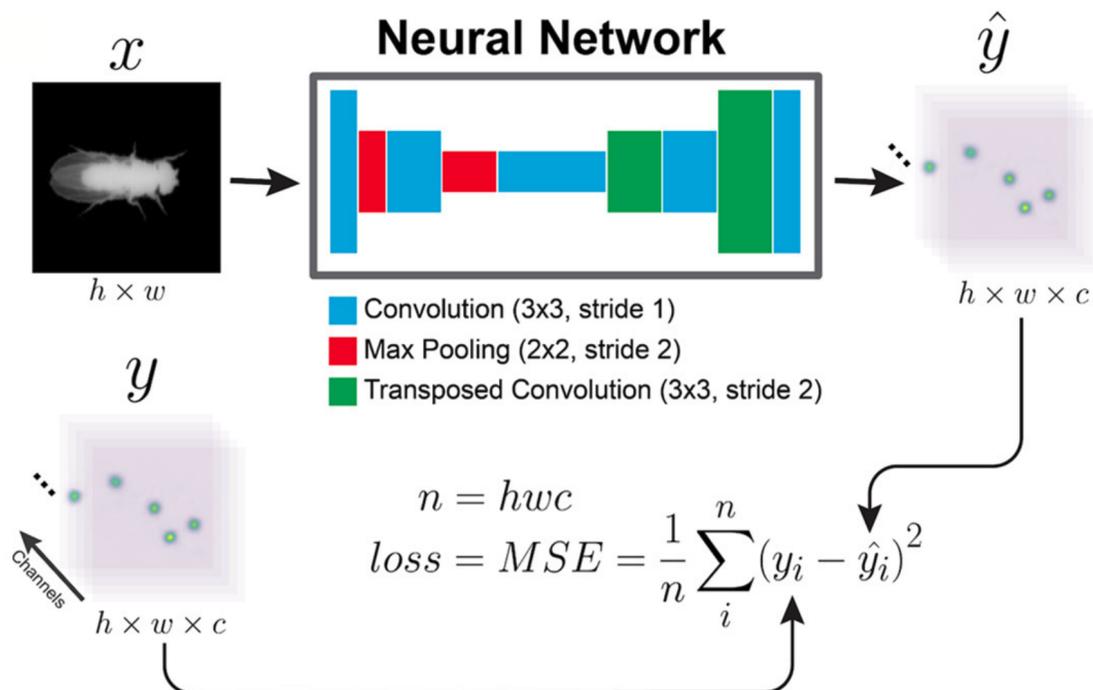


Figure 1.3: Neural network architecture for LEAP (figure borrowed from [1]). Raw images enter the system (top left), passing through a series of convolutional, max pooling, and transposed convolutional layers (center). The network outputs confidence maps matching the input image dimensions (top right). During training, these outputs are compared to ground truth maps derived from user labels (bottom left), and a mean squared error loss function quantifies the difference, which is minimized to optimize network performance (bottom right). Arrows indicate data flow and feedback loops in the training process.

layers in deep learning models, transforming the field of computer vision. Later on, enabled by hardware advancements, much deeper architectures like ResNets [50] were developed with hundreds of convolutional layers. A key innovation in ResNet was the use of skip connections that bypass one or more layers, allowing gradients to flow through the deeper architecture without vanishing. Another big idea in the era of deep learning has been transfer learning, where a foundational model trained on a large dataset (such as AlexNet) is used, and a few later layers (closer to the output) are modified and trained to adapt to a new task while retaining the architecture and weights from the foundational model for other layers. Human pose estimation models like DeeperCut [51] leveraged both ideas – using a ResNet architecture as its backbone and employing transfer learning for training later layers of the pre-trained deep learning models using annotated data of human pose from a small training set. Soon after, these ideas were adapted for scientific use by tools like DeepLabCut [5] and LEAP [1]. DeepLabCut extends the usability of DeeperCut, allowing an easy interface for training models for pose estimation on videos of single animals taken in a lab setting, as well as providing the option for other smaller architectures for training (Figure 1.2). LEAP uses stacked hourglass convolutional networks – another class of networks popular for human pose estimation (Figure 1.3).

Since the development of DeepLabCut and LEAP, many improvements and advancements have been made for human and animal pose estimation, and subsequently these new tools, including updated versions of DeepLabCut and LEAP, have emerged for tracking animal pose. More importantly, many methods extend these ideas for the problem of tracking posture in 3D, either by utilizing multiple camera angles [52, 14, 53, 54], learning the skeletal structure of animals [55] or recording pose using depth sensing cameras [8]. When multiple animals are present in a video frame (e.g., when studying social behavior), tracking the pose of each animal across time becomes challenging as there are three steps required - 1). Identifying the presence of multiple

animals in the frame, 2). Tracking poses for each identified animal 3). Maintaining identities of individuals over time (video frames). Recent tools have come close to bridging the gap between single and multi-animal posture tracking, addressing these challenges with varied approaches. However, while 1 and 2 are technically straightforward to address, 3 becomes challenging especially when physical or visual identifiers are absent or when animals cross over each other, masking each other in the camera view. Multiple deep learning tools now allow pose-estimation of multiple animals in a frame, either by detecting pose and identity using a single deep learning model [56] or employing separate multi-step processes to detect identity and estimate pose, some providing the ability to switch the order of steps [57].

Despite access to many morphological degrees of freedom, postural dynamics are highly correlated, and studies have found lower-dimensional representations that capture these stereotypes in animals [2, 3]. Hence, when tracking posture, one choice the user needs to make is which posture-relevant locations (keypoints) to track on the body of the animal. B-KinD [58] is a self-supervised method that addresses this issue, identifying *meaningful* keypoints from behavioral videos. Specifically, it forces the learned key points to have predictive information over time, and this is shown to encourage the selection of consistent and semantically meaningful key points. However, like pose detection algorithms, some limitations apply to this method as well - such as tracking errors during occlusions of the animal's body (by itself or by a cable), tracking jumps between laterally symmetric keypoints, and difficulties when animals rotate along an axis parallel to the imaged plane.

1.3 Postural Dynamics

Access to these pose estimation tools has largely discounted the cost of acquiring high throughput behavioral data over long timescales, even up to multiple weeks

[59]. This is unprecedented, and there is a need for phenomenological modeling of behavior over very long timescales. To build a representation of behavioral dynamics from posture, we must think in terms of characterizing and quantifying movements rather than postural snapshots. To that extent, a primary early goal in this direction is identifying stereotypical postural dynamics.

Current methods take varied approaches to this end. One popular direction is taking a supervised approach, where a machine learning model is trained to do frame-by-frame classification of postural data, sometimes appending additional position, locomotion, and social features to postural data. These models are usually trained using training datasets annotated by behavioral experts, and sometimes annotations are pooled from multiple annotators to reduce bias in the training set [60, 61, 28, 62].

However, there are significant limitations when using supervised approaches for behavioral classification. A primary concern is the presence of human bias when labeling training data with behavioral annotations, where the annotator can not *a priori* know the full spectrum of behavior variability. It is also possible that not all postural actions are interpretable by humans, and we may innocently anthropomorphize other actions. Moreover, some tools generate behavioral classification using data from a single frame, imposing a short behavioral timescale. Behavioral dynamics are also non-stationary, and a supervised model may not be able to capture the rich complexity and repertoire of behaviors recorded in a large dataset.

A contrary approach is using unsupervised methods for discovering a representation and repertoire of behaviors from the postural time series of an animal. We discuss some of these methods below, and while some methods predate the convenience of large-scale pose estimation tools and use image processing techniques to capture coarser measures of animal pose dynamics, these methods are also largely applicable to postural time series data with additional data-specific pre-processing.

MoSeq [8] is developed to quantify and characterize movement patterns in freely

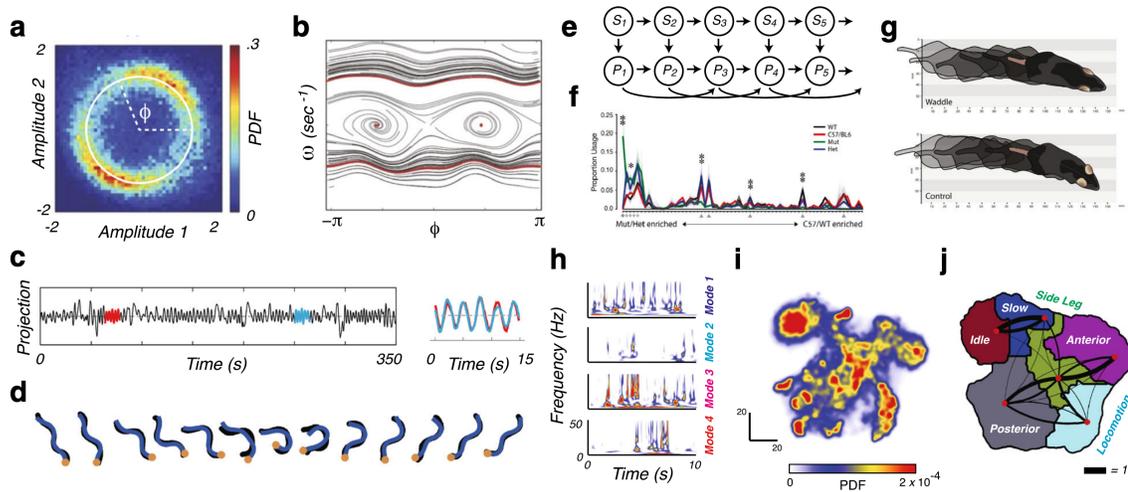


Figure 1.4: Representing postural dynamics across species (taken from [4]). a) *C. elegans*: Histogram of projections onto the first two "eigenworms" (from Fig. 1.1b) reveals a structure parameterizable by a single phase variable ϕ . b) Fitting the dynamics of this variable to a deterministic dynamical system yields this phase map of *C. elegans* locomotion (derived from a), forward and backward movements are represented as wave trajectories, with two fixed points representing pause states. (a and b adapted from [2]). c) Time-series projections onto eigenworms can be used to identify repeated motifs (blue and red curves), and sequences of motifs can then be used for behavioral representation (adapted from [6]). d) Alternatively, worm behavior can be described as a series of discrete postures (from [7]). e) Autoregressive Hidden Markov Model (AR-HMM) trained on mouse postural data obtained from a depth-sensing camera and decomposed using ideas similar to Fig.1.1c. Here, P_t and S_t represent postural mode values and the underlying state influencing posture dynamics, respectively. Arrows indicate the direction of dependencies amongst variables in the model [8]. f) Phenotypic signatures showing behavior usage frequencies obtained using e across four mouse genotypes. g) Distinct walking gaits identified between two different genotypes using behavioral signatures from f (e-g adapted from [8]). h) Time-frequency analysis of fruit fly movement: Continuous wavelet transform amplitudes for different postural modes over time. i) 2D embedding of fruit fly movements: Probability density map where similar movements cluster together and peaks representing stereotyped behaviors. j) Behavioral map: Breakdown of i with manually labeled regions and transition probabilities (black lines). (h-j adapted from [3, 9])

behaving mice, originally applied to 3D data obtained from depth-sensing cameras. Its core framework is an autoregressive hidden Markov (AR-HMM) model (Fig.1.4e-g), which is trained on postural data to identify stereotyped movement syllables. AR-HMM models behave as discrete dynamical models where each discrete state corresponds to a locally linear dynamical system with a characteristic timescale chosen from an explicitly chosen exponential distribution of timescales. Moseq successfully manages to identify about 50-100 discrete syllables of mouse behavior that persist for 300-600 milliseconds. However, several limitations apply to this approach. First, the process of discretization of behavioral motifs into finite states loses important variability at the motif scale. Also, the number of HMM states, optimized using Bayesian model selection, can dictate a false limit to the repertoire of behaviors available to the animal. Second, the persistent timescale of HMM states is required to be explicitly set by the user, limiting the timescale of each behavioral state. This requires *a priori* information about the behavioral timescale that typically spans multiple orders of magnitude [63, 9].

A complementary approach is a method known as MotionMapper [3]. Here, the authors decomposed images of freely behaving fruit flies into a lower-dimensional postural representation that captures maximal variance using radon transformation and PCA. Then, postural movements are identified using a continuous wavelet transform, obtaining power at dyadically spaced range of frequencies for each postural dimension. This results in a very high-dimensional spectral representation at each time t , which is embedded into a 2-dimensional representation using dimensionality reduction algorithms. This representation is shown to identify and cluster a variety of stereotyped behaviors such as wing movements, locomotion gaits, grooming, and idle (Fig.1.4h-j). Here, decomposition of postural data into a spectral representation allows capturing dynamics across a range of timescales, however, these timescales need to be manually selected by the user. Moreover, spectral decomposition is highly sensitive to input

postural measurements, and variations in stereotyped behaviors and measurement noise can highly affect the resulting representation.

While unsupervised discovery of behavioral motifs can address several limitations found in supervised methods, these still have some limitations of their own, either in the form of computational intractability on large datasets, or the need for hand-tuning of parameters that fundamentally limit the behavioral timescale being resolved. A favorable solution to this limitation is to build dynamical models of behavioral data, as these can learn and capture the wide range of timescales that generate behavior.

A notable approach for generating dynamical representations of behavior without explicitly setting a timescale is proposed by Stephens et al [2]. Here, they imaged freely moving worms using tracking microscopy and obtained a single-pixel backbone of the worm using image processing. They fit a spline to this backbone, measuring 100 tangential angles along the pixel backbone curvature characterizing worm pose in each frame, and found a low dimensional (4-D) subspace that captured 95% variance in worm pose during locomotion. They observe emergent oscillatory dynamics along the 2 leading eigenmodes corresponding to forward or backwards locomotion, and the phase velocity along this manifold is found to predict the physical velocity of the worm (Fig.1.4a-d). Unlike former approaches, they model the phase dynamics of the worm during locomotion as a stochastic dynamical system and find metastable behavioral states corresponding to forward crawling, backwards crawling, and pauses, with noise-induced transitions between these states.

Building such dynamical models from postural time series data is extremely useful, as they naturally represent temporal continuity and allow smooth trajectories through learned internal phase space. Dynamical models are also generative by design and can be simulated to generate new trajectories that may be experimentally unobservable. For continuous dynamical systems, models trained on data can infer low-dimensional manifold structures such as attractors and can be used to study state-dependent

transitions and bifurcations.

Recurrent Neural Networks are deep-learning models that can behave as dynamical models and are introduced in more detail in Chapter 2. VAME [10], a recent deep-learning framework, uses bidirectional Recurrent Neural Networks (RNNs) to encode snippets (500ms) of postural time series into a low-dimensional variational latent space (Fig.1.5). It then uses a Hidden Markov Model (HMM) to infer hidden states in the low dimensional latent space thus learning behavioral motifs. However, in this method, a fixed sequence length is encoded using RNNs, restricting the temporal scale that the model is able to learn to encode.

1.4 Takens' Embedding Theorem and ensemble methods for state-space reconstruction

Real-world complex systems often involve a large number of interacting components, interactions among which often lead to emergent dynamics and disparate spatio-temporal scales that we observe. However, when measuring from such systems, we may not be able to observe all dynamically relevant variables necessary to reconstruct the entire state-space of the system. For example, in the basic Lotka-Volterra model, prey and predator population dynamics follow a pair of first-order nonlinear differential equations

$$\frac{dx}{dt} = \alpha x - \beta xy \quad (1.1)$$

$$\frac{dy}{dt} = \delta xy - \gamma y \quad (1.2)$$

where (x) is the prey population and (y) is the predator population. If we attempt to reconstruct these dynamics from only the predator population, we will be unable to reconstruct the 2-dimensional phase space, losing important information about the

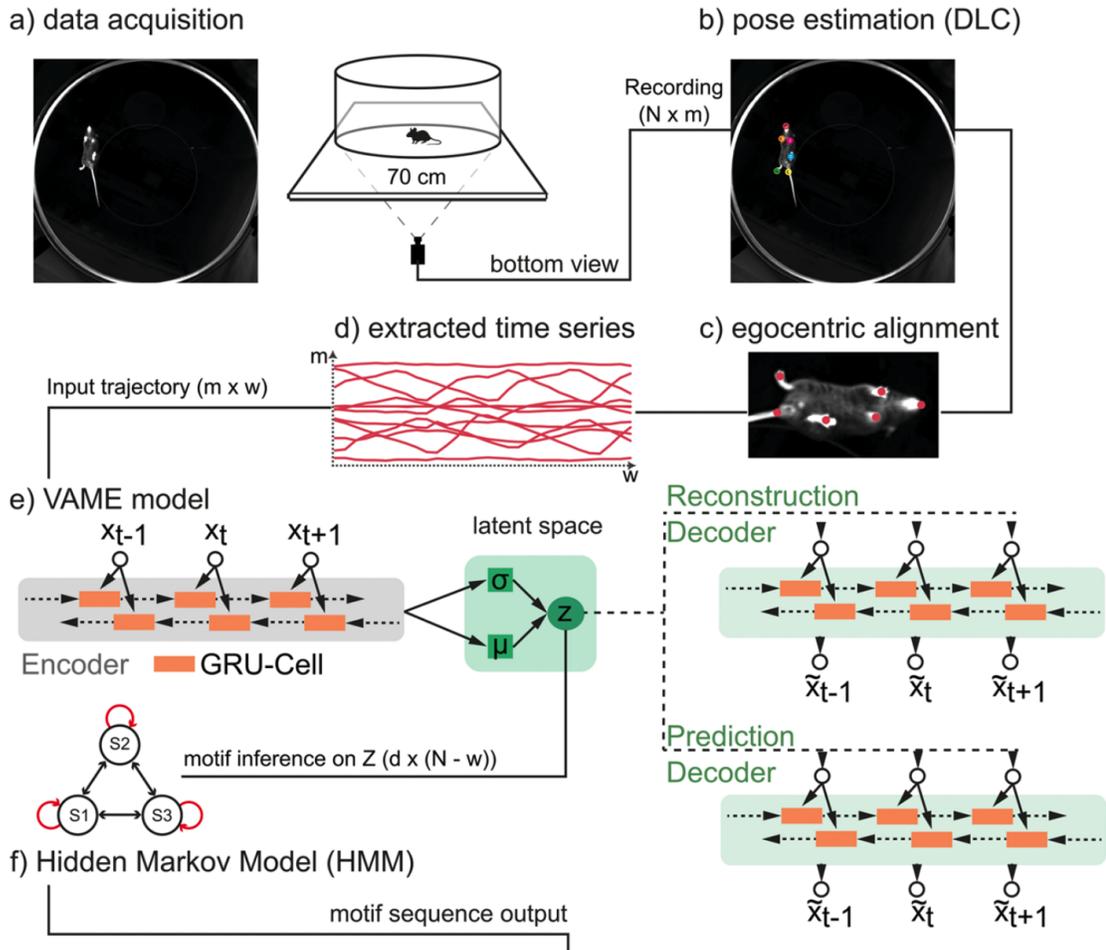


Figure 1.5: Workflow underlying VAME [10] a) Videos were recorded of mice freely behaving in a circular arena for 25mins from a bottom-up camera. b) 4 paws, nose, and tailbase of mice were tracked using DeepLabCut [5]. c) Tracked keypoints were ego-centered and aligned along a fixed direction (facing right). d) 12-dimensional postural trajectories were used for training the subsequent model. e) The VAME model consists of bidirectional recurrent neural networks, where the encoder RNN receives a 500ms trajectory sample and embeds it into a lower dimensional variational latent space. This representation is fed into 2 RNN decoders, one reconstructing the input data and one predicting the subsequent 500ms section of the postural time series. f) Latent representations are used to train a Hidden Markov Model (HMM) for motif detection. Figure borrowed from [10].

trajectories and steady states of this system.

Takens' embedding theorem, proposed by Floris Takens in 1981 [64], provides a robust framework that can reconstruct the topological structure of phase spaces without observing all state variables of the system. Consider the case where only a single variable $x(t)$ is observed from the system. Takens' embedding theorem states we first create a delay vector $\hat{x}(t)$ from $x(t)$ such that

$$\hat{x}(t) = x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (K - 1)\tau), \quad (1.3)$$

where τ is a chosen time delay and d is the embedding dimension. Takens proved that if K is sufficiently large (specifically more than twice the dimension of the original system), the delay vectors $\hat{x}(t)$ form an embedding of the original phase space that is topologically equivalent to the original phase space, preserving the dynamical properties of the original system.

Recently, these ideas were extended to discover long-timescale structures from partially observed dynamics by Costa et al in a series of publications [30, 11, 65]. Specifically, their goal is to reconstruct the dynamics of a system of differential equations $\dot{\bar{x}} = f(\bar{x})$ where \bar{x} is the unobserved state of the system. The partially observed time-series data is then denoted by $\bar{y} = M(\bar{x}(t))$, which will have non-Markovian dynamics due to the underlying long-timescale dynamical structure that remains unobserved. To reconstruct the state space from \bar{y} , authors use delay embedding with varying delay lengths K , obtaining candidate state-space reconstructions (Fig.1.6a top).

To search for a maximally Markovian state-space from data, they analyze ensemble dynamics in this reconstructed space. Specifically, high-dimensional space spanned by \bar{y} is partitioned into N Voronoi cells using clustering algorithms, and a transition matrix P_τ at timescale τ is then estimated from the transitions observed between

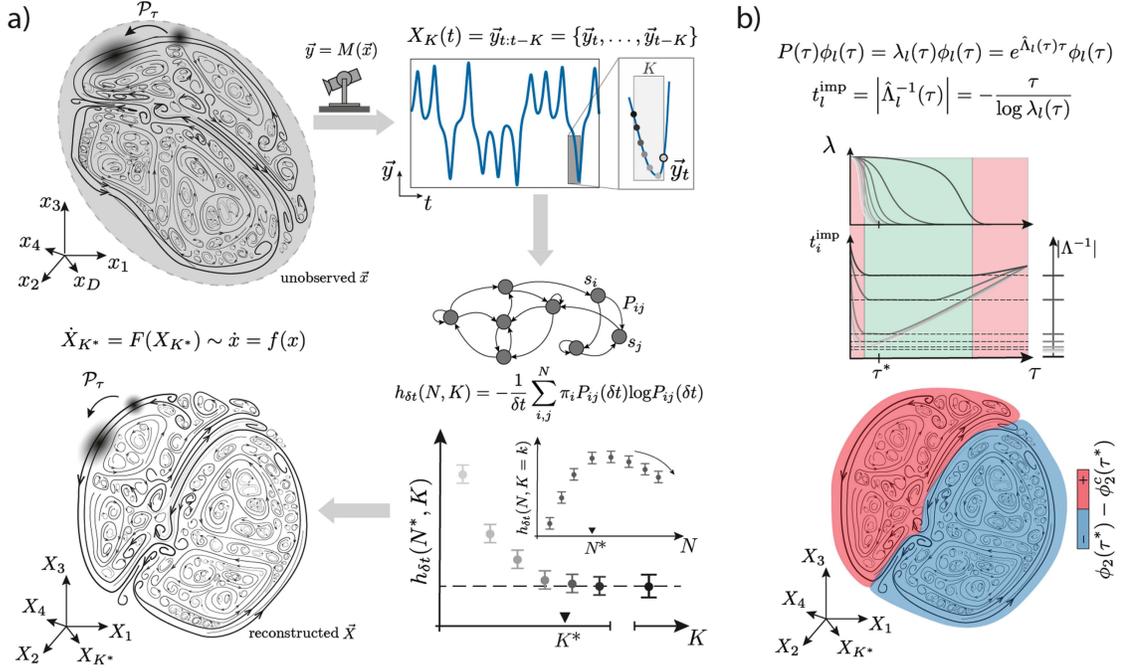


Figure 1.6: State-space reconstruction and ensemble dynamics (figure taken from [11]). a) From a complex dynamical system (top left), partial observations are measured and are then used to reconstruct the state space using delay-embedding (top right). To maximize short-term predictability, partitioning is performed in multiple delay-embedded spaces constructed with different delay lengths, and unpredictability is measured through the short-term entropy rate (bottom right). This allows selecting a maximally predictive delay embedded space that reflects the properties of the original dynamics (bottom left). b) The eigenspectra of transition operator $P_{ij}(\tau)$ at different τ captures multiple characteristic timescales. At appropriate τ^* , where long-timescale dynamics are approximately Markovian and the spectral gap is large, the slowest eigenvector (ϕ_2^*) reveals metastable states of the system (bottom).

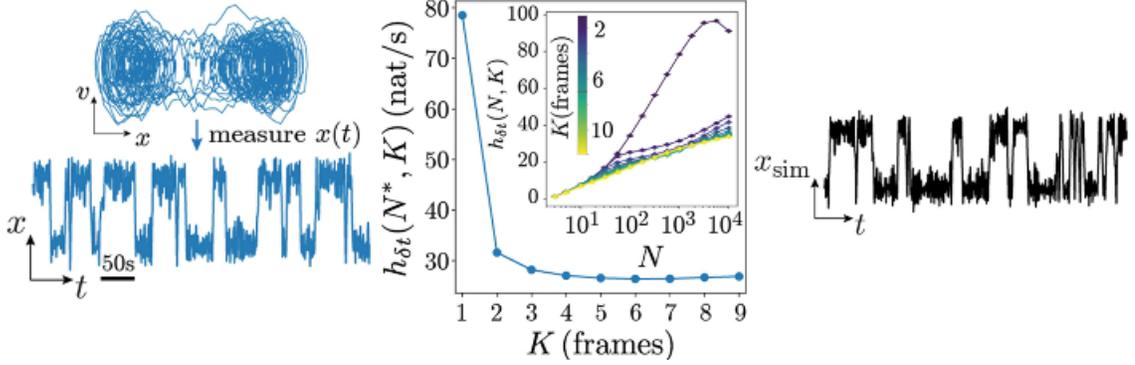


Figure 1.7: Reconstructing state-space of an underdamped particle in a double-well potential (figure borrowed from [11]). Using only the position of the particle (left), delay-embedded spaces are constructed, and entropy rates are measured (center). Inset shows an increase in entropy rate with the number of partitions (N) for spaces with different delay lengths (K). $N^* = 10^3$ is picked to avoid finite-size effects, and $K^* = 7$ is chosen beyond which the entropy production rate becomes constant. Particle trajectories can be reconstructed using simulations of the Markov chain inferred in the previous step.

these cells. This creates a dynamical representation of the reconstructed structure, however, these dynamics may not be strictly Markovian and close to the true state-space description of the system (Fig.1.6a bottom). To test this, authors vary K and N , measuring the Kolmogorov-Sinai (KS) entropy $h_{P_{N,\tau}}(K) = -\sum_{ij} \pi_i P_{ij} \log P_{ij}$ where π is the stationary distribution of P_τ . KS entropy is a fundamental measure of compressibility of a dynamical system, and measures the degree of unpredictability in the dynamics.

Once a maximally predictive state-space is reconstructed, spectral properties of a time-symmetric transition operator $P'_\tau = (P_\tau + P_\tau^\dagger)/2$ are analyzed. Specifically, the largest non-trivial eigenvalue λ_2 of P'_τ determines the longest relaxation timescales of the inferred dynamics (Fig.1.7), and the corresponding eigenvectors ϕ_2 determines the directions of slow relaxation in the state space (Fig.1.8).

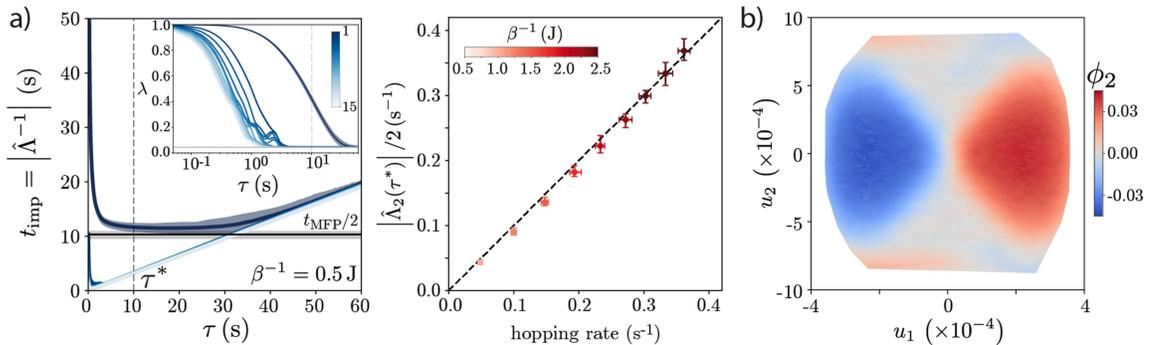


Figure 1.8: a) The characteristic timescale of the time-symmetric transition matrix P'_τ estimated from the reconstructed state-space of the underdamped double well system (left). At short transition timescales τ , transient dynamics dominate, resulting in large characteristic time scales and eigenvalues converging close to 1 (left inset). As τ increases, the longest relaxation time becomes constant at $\tau \sim 5$ seconds, equal to half the mean first passage time. The characteristic timescale for double well dynamics at different temperatures (β^{-1}) predicts the hopping rate between wells (a, right). b) For $\beta^{-1} = 0.5J$, the first two leading singular modes of the second leading eigenvector ϕ_2 splits the two wells. Figure taken from [11].

1.5 Thesis Outline

The objective of this dissertation is to formulate and develop tools and methodologies to identify long-timescale structures within behavioral dynamics. I have developed computational approaches that utilize the large time series data available from the postural tracking of animals. Specifically, I focus on utilizing the dynamical nature of Recurrent Neural Networks (RNNs), and employ them not just for predictive learning of postural dynamics, but also utilizing their internal representations to learn the dynamics encoded in temporal data.

In Chapter 2, I introduce RNNs and their ability to model dynamical systems. I introduce a novel tool based on RNNs designed to learn dynamical models from data. I apply this tool to gait data from humans and show its ability to generate gait signatures that quantify properties of gait cycles in human subjects. Subsequently, I develop novel RNN architectures for creating dynamical representations of behavior at multiple timescales and discuss its potential applications.

In Chapter 3, I highlight my contribution to extending an existing method to generate behavioral representations of freely behaving mice. I use these representations across a cohort of mice with and without exposure to stress and identify sex-specific behavioral signatures in mice exposed to acute and chronic stress.

In Chapter 4, I investigate the limits of a recently published method where authors reconstruct state space dynamics and identify long-timescale structure from time series data measured at very short timescales. I apply this method to time series data generated from a modified Lorenz system with a very long-timescale latent non-stationarity and discuss its shortcomings in identifying the slow latent dynamics. I then propose a novel encoder-decoder RNN architecture that is able to identify this long-timescale non-stationarity from data.

In Chapter 5, I summarize the contributions of this thesis, highlighting various computational methods developed here to identify long-timescale structures from large datasets. Moreover, the strengths and limitations of each approach are discussed, and potential solutions are proposed to address these limitations. Finally, potential applications of these approaches are presented, which outline promising avenues for future research.

Chapter 2

Using Recurrent Neural Networks to mimic dynamical systems¹

2.1 Introduction

Dynamical systems have a time-dependent behavior that is governed by underlying deterministic laws or equations. When observing physical or biological phenomena with disparate timescales where analytical solutions aren't possible, it is often desirable to have the ability to infer governing equations using experimental data and use these equations to build a fundamental understanding of the system. A traditional theme to discover governing equations from data involves defining a repertoire of candidate functions that can capture the non-linearities underlying experimental data, and use symbolic regression to find data-appropriate coefficients. This remains a highly active area of research, and recent work in this direction addresses several limitations of this approach such as being computationally intractable for systems with large degrees of freedom, overfitting from regression, incomplete library of basis

¹Portions of this chapter appear in T.S. Winner, M.C. Rosenberg, K. Jain, T.M. Kesar, L.H. Ting, and G.J. Berman, "Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics," PLOS Computational Biology 19(10), e1011556 (2023).

functions, noisy datasets, and complex systems. Notably, the SINDy algorithm [66] uses sparse regression techniques to select the most important nonlinear terms that govern a system’s dynamics from a large library of candidate functions. This allows it to balance model complexity with descriptive ability, producing parsimonious models that avoid overfitting. However, adapting these methods to data from a system with disparate timescales still remains a challenge because of: a) noisy and/or incomplete measurements, b) large experimental datasets, and c) the complexity of the underlying phenomena. When underlying dynamics are nonlinear, non-stationary, chaotic, and embedded in higher-dimensional spaces, using traditional approaches to building phenomenological models from such data becomes difficult.

Recurrent Neural Networks (RNNs), which will be described in detail in the next section, are designed to behave as overparametrized dynamical systems and can learn to mimic complex dynamical systems when trained on vast amounts of data. The recurrent links in RNNs allow them to have a dynamic memory [67], allowing RNNs to integrate information over time and synthesize internal representations that incorporate task-based memory. A single-layer Recurrent Neural Network (also known as the Elman network) can be mathematically described as follows [67]:

$$\vec{h}_t = \sigma_h(W_{hh}\vec{h}_{t-1} + W_{xh}\vec{x}_t + \vec{b}_h) \quad (2.1)$$

$$\vec{y}_t = \sigma_o(W_{hy}\vec{h}_t + \vec{b}_y) \quad (2.2)$$

where

\vec{x}_t : input to the network at time t

\vec{h}_t : state of RNN units at time t

\vec{y}_t : network output at time t

σ_h, σ_o : nonlinearities or activation functions

and the learnable parameters are

W_{hh}, W_{hy}, W_{xh} : the recurrence, input and output kernels (weights)

\vec{b}_h, \vec{b}_y : biases

Elman RNNs have been theoretically proven to be highly efficient at modeling a wide range of discrete and continuous non-linear dynamical systems [68, 69, 70], even when training data is sparse [71]. Their continuous, multivariate internal states’ information capacity is shown to grow linearly with the network’s size [72]. RNNs are able to recapitulate chaotic, periodic, or fixed point attractor-like dynamics, even in the absence of any input to the network [73], and can provide a robust universal approximation of any open dynamical system [74].

However, current implementations train RNNs over a finite timescale (i.e., training over a short sequence of multi-dimensional inputs at a time), utilizing mini-batch training where network parameters are updated as each batch is processed and gradients are calculated using Back Propagation Through Time (BPTT). Each batch contains a number of training samples (equal to the user-defined batch size), and each sample contains a short sequence of training data. As discussed in the following sections, this training timescale (also known as ‘timesteps’ in implementations) needs to be sufficiently small to avoid instabilities when calculating gradients after each batch is processed. Moreover, during training, after each batch in the training set has been trained on (e.g., over one epoch of training), the internal state of the RNN layers can be reset to their predefined initialization scheme (commonly set to be initialized as $\vec{0}$) or can be passed over to the next batch. If we choose to reset states after each batch, information accumulated over the last batch is not handed over to the next batch, and any long-timescale training signal is lost. If states obtained after processing each batch are utilized as the initial hidden state for the next batch, we can force continuity in the internal dynamics of the RNN over long timescales. This process is called *statefulness* and can allow the RNNs to learn long-timescale structure in data.

In this chapter, we introduce a new framework to train RNNs on large biological time series datasets. Our new approach allows training RNNs on longer timescales by allowing persistent internal states to propagate across batches of training data and allowing RNN hidden states to evolve continuously through each epoch of training. Subsequently, we showcase the application of this framework to train RNN models on human gait data, leading to individual-specific gait signatures in healthy and stroke survivor patients. We also propose a method to train novel RNN architectures on behavioral data, using this method to identify behavioral representations at multiple timescales.

2.2 Recurrent Neural Networks

Recurrent Neural Networks (RNNs), first proposed in 1986 [75], are artificial neural networks constructed using recurrent neural units (or cells). These recurrent units are interconnected to each other, unlike traditional feedforward networks, forming directed cycles [67]. That is, during training and inference, each forward pass through the network updates the *state* of the recurrent units, and this update process depends on the current input and the previous state of the recurrent units. This updates the internal state of each recurrent unit following this equation 2.3

$$\vec{h}(t) = f(\vec{x}(t), \vec{h}(t-1)) \quad (2.3)$$

where $\vec{x}(t)$ is the input to the network, \vec{h}_i is the state of unit i in the current layer, and \vec{h} is the state of all units in the current layer, thus emulating the classical form of a dynamical system.

As ideas on using recurrence in neural networks were being developed in the late 20th century, parallel efforts were made to invent methods to train such networks. A popular algorithm to train multi-layer neural networks is the Back-propagation

algorithm [75], grounded in the principles of gradient descent and the chain rule of derivatives. It was later modified to adapt to recurrent connections within the network, and this modified approach is called Back-Propagation Through Time (BPTT) [76]. BPTT unrolls a traditional recurrent network over multiple time steps (also referred to as truncated BPTT if done over τ time steps), effectively creating a multi-layer feedforward network. It then applies standard backpropagation to this unrolled network and is able to propagate errors through the temporal dimension in the data. Additional methods like Real-Time Recurrent Learning [77] (RTRL) also exist. However, they were never adapted in practice because of quadratic and cubic complexities in space (memory) and time respectively [78]. Back-Propagation Through Time has been practically tractable and is the primary algorithm used to train Recurrent Neural Networks today.

Some early applications of RNNs have shown complex linguistic structures represented by the internal state dynamics of the RNNs [79]. They surpassed Hidden Markov Models in identifying continuous handwritten text trajectories and were shown to store a long temporal context length of 120 time steps [72]. Later on, they were also shown to significantly outperform standard n-gram techniques when performing speech recognition tasks [80].

RNNs, however, are famously prone to instability in gradients during learning, where correction gradients flowing through the network could diminish or explode exponentially as the network is unrolled over time when training [81, 82]. If we train an RNN over k steps, the error (ϵ) with respect to network parameters (θ) can be written as

$$\frac{\delta\epsilon_t}{\delta\theta} = \sum_{1 \leq k \leq t} \left(\frac{\delta\epsilon_t}{\delta x_t} \frac{\delta x_t}{\delta x_k} \frac{\delta^+ x_k}{\delta\theta} \right), \quad (2.4)$$

where

$$\frac{\delta x_t}{\delta x_k} = \prod_{k < i \leq t} \frac{\delta x_i}{\delta x_{i-1}}, \quad (2.5)$$

which is the propagation of error in time from step t back to step k . When $k \ll t$, it refers to the error contributions from longer terms (far back in history).

Now, for all k that exist such that $\frac{\delta x_{k+1}}{\delta x_k} \leq \eta < 1$, the temporal error term can be shown to be

$$\left| \frac{\delta \epsilon_t}{\delta x_t} \left(\prod_{i=k}^{t-1} \frac{\delta x_{i+1}}{\delta x_i} \right) \right| \leq \eta^{t-1} \left| \frac{\delta \epsilon_t}{\delta x_t} \right|. \quad (2.6)$$

If $(t - k)$ is large and $\eta < 1$, it is apparent that long-term contributions will vanish (go to 0) exponentially fast. Exploding gradients can be seen by inverting this argument and assuming $\eta \gg 1$.

One alternative to deal with this gradient instability is the use of gating within recurrent units, as proposed in the Long Short-Term Memory (LSTM) [83] or the Gated Recurrent Units [84] (GRUs). LSTMs introduce forget, input, and output gating mechanisms, as shown in the equations below

$$\vec{f}_t = \sigma(W_f[\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_f) \quad (2.7)$$

$$\vec{i}_t = \sigma(W_i[\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_i) \quad (2.8)$$

$$\vec{\tilde{C}}_t = \tanh(W_c[\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_c) \quad (2.9)$$

$$\vec{C}_t = \vec{f}_t * \vec{C}_{t-1} + \vec{i}_t * \vec{\tilde{C}}_t \quad (2.10)$$

$$\vec{o}_t = \sigma(W_o[\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_o) \quad (2.11)$$

$$\vec{h}_t = \vec{o}_t * \tanh(\vec{C}_t). \quad (2.12)$$

Here, \vec{f}_t , \vec{i}_t , and \vec{o}_t are the forget, input, and output gates, \vec{C}_t is the cell state, and \vec{h}_t is the hidden state, and σ is the sigmoid activation function. Note that each unit now maintains a cell (\vec{C}_t) and a hidden state (\vec{h}_t). W_f, W_i, W_c, W_o are weight matrices learned during training. $\vec{\tilde{C}}_t$ represents the candidate cell state containing new information that can be added to the actual cell state \vec{C}_t . This addition of

new information to the cell state is controlled by the forget and input gates through Equation 2.10.

GRUs combine the forget and input gates found in LSTMs into a single update gate, obtaining a single state for each cell. Their equations are as follows:

$$\vec{z}_t = \sigma(W_x[\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_z) \quad (2.13)$$

$$\vec{r}_t = \sigma(W_r[\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_r) \quad (2.14)$$

$$\vec{\tilde{h}}_t = \tanh(W_g[\vec{r}_t * \vec{h}_{t-1}, \vec{x}_t] + \vec{b}_h) \quad (2.15)$$

$$\vec{h}_t = (1 - \vec{z}_t) * \vec{h}_{t-1} + \vec{z}_t * \vec{\tilde{h}}_t. \quad (2.16)$$

These gating mechanisms in LSTMs and GRUs help recurrent networks learn long-timescale dependencies in data in two ways: 1) they control the flow of information through time as the cell (in the case of LSTMs) and state updates are controlled, and 2) gradients flowing through time in the network are regularized, i.e., the partial derivative for cell updates is $\frac{dC_t}{dC_{t-1}} = f_t$, and the gradients cannot explode or vanish as long as f_t is close to 1 (the same can be extended for GRU cells). Hence, gating allows for better control of timescales and dimensionality in network dynamics [85, 86], and the update gate in GRUs and forget gate in LSTMs are particularly important for accumulating slow modes and maintaining long-term dependencies [87]. However, the nonlinearities in gates force saturation of gradients at the extrema, hindering their ability to learn new information, and can be avoided by reparameterizing the gate equations and careful initialization [88]. Additionally, novel gated architectures like the STAR cell have been developed to allow for deeper and more efficient RNNs while mitigating gradient issues [89].

The introduction of gating allowed RNNs to be adapted for numerous applications in science and technology, notably for music generation [90], speech recognition [91,

92], machine translation [93], acoustic modeling [94], image captioning [95], sentiment analysis [96], video analysis [97], and bioinformatics [98].

2.3 LIDAR: a python package to train RNNs statefully

LIDAR is a Python package that implements a specialized routine to train Recurrent Neural Networks on large, high-dimensional time series datasets obtained from biological systems. The goal of LIDAR is to build RNN models that can learn to mimic the multi-scale dynamical structure inherent in the underlying data, capturing both long-timescale non-stationarities and short-term dynamics. LIDAR uses Long Short-Term Memory (LSTM) units to avoid vanishing gradients and support the network’s capacity to learn long-timescale structures. Crucially, it employs statefulness during training, which allows the model to have persistent states across training batches. It also includes easy functionality for extracting internal state variables from the trained model, as well as the ability to drive the network forward autonomously in generative mode to observe the model dynamics to perturbed inputs. In this section, we explain the mechanism to train RNNs statefully as implemented in LIDAR.

As discussed in the previous section, RNNs are sequential models and are well-suited for training on temporal data. Each term in the input data is sequentially processed, and the output from the model depends both on the input as well as the hidden state vector carried forward from the previous time step. In stateful RNNs, the hidden state is not reset after processing each batch, and requires temporal continuity of input data across batches. Thus, to perform stateful training, we need to meticulously structure the training data to ensure the continuity of the hidden state is maintained appropriately.

We can illustrate this concept by first assuming that we have a single d -dimensional

time series data set $X = (x_1, x_2, \dots, x_n)$, where each $x_i \in \mathbb{R}^d$ and n is the total number of time steps. To curate a training set with a batch size of 1 and sequence length τ , we first partition X into $\lfloor n/\tau \rfloor$ subsequences, each of length τ . Let S_i denote the i -th subsequence:

$$S_i = (x_{(i-1)\tau+1}, x_{(i-1)\tau+2}, \dots, x_{i\tau}) \text{ for } i = 1, 2, \dots, \lfloor n/\tau \rfloor.$$

With a batch size of 1, each subsequence S_i constitutes a single batch for training, and the network can be trained statefully. The input tensor for each batch thus has dimensions $(1, \tau, d)$. Please note that during training, shuffling should not be performed as is typical when not training statefully, to maintain the temporal order of the sequences.

In the case where we have multiple datasets, $X^{(1)}, X^{(2)}, \dots, X^{(k)}$, where each $X^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots, x_{n_j}^{(j)})$ and $x_i^{(j)} \in \mathbb{R}^d$, and n_j represents the length of the j -th time series which may vary across the set but are close to each other, we partition each time series $X^{(j)}$ into $\lfloor n_j/\tau \rfloor$ subsequences of length τ . Let $S_i^{(j)}$ denote the i -th subsequence of the j -th time series: $S_i^{(j)} = (x_{(i-1)\tau+1}^{(j)}, x_{(i-1)\tau+2}^{(j)}, \dots, x_{i\tau}^{(j)})$ for $i = 1, 2, \dots, \lfloor n_j/\tau \rfloor$.

With k time series, we construct batches of size k , where each batch contains one subsequence from each time series: $B_i = (S_i^{(1)}, S_i^{(2)}, \dots, S_i^{(k)})$ for $i = 1, 2, \dots, \min_j \lfloor n_j/\tau \rfloor$.

The input tensor for each batch has dimensions (k, τ, d) . Before a training epoch begins, k hidden states, one for each time series, are initialized $h_0^{(j)} \in \mathbb{R}^h$, all set to zero initially. The stateful RNN training process then proceeds following Algorithm 1.

2.4 Characterizing human gait dynamics using RNNs

2.4.1 Introduction

Locomotion is ubiquitous and often involves repeating a pattern of limb movements, resulting in behaviors like walking, running, crawling, flying, and swimming. This

Algorithm 1: Stateful RNN Training Epoch

Input: N : number of batches
Input: k : number of time series
Input: τ : sequence length
Input: $B_{i=1}^N$: batches where $B_i = (S_i^{(1)}, S_i^{(2)}, \dots, S_i^{(k)})$
Input: $f(\cdot, \theta)$: RNN cell function with parameters θ
Input: $g(\cdot, \phi)$: output function with parameters ϕ

```

1 for  $i \leftarrow 1$  to  $N$  do
2   if  $i = 1$  then
3     |  $h_0^{(j)} \leftarrow 0$  for  $j = 1$  to  $k$ ;
4   else
5     | Use  $h_{\tau}^{(j)}$  from previous batch as initial state;
6   end
7   for  $t \leftarrow 1$  to  $\tau$  do
8     | for  $j \leftarrow 1$  to  $k$  do
9       | |  $h_t^{(j)} \leftarrow f(x_t^{(j)}, h_{t-1}^{(j)}; \theta)$ ;
10      | |  $y_t^{(j)} \leftarrow g(h_t^{(j)}; \phi)$ ;
11      | end
12    end
13    Compute loss  $L(\{y_{1:\tau}^{(j)}\}, \{\hat{y}_{1:\tau}^{(j)}\})$ ;
14    Update parameters  $\theta$  and  $\phi$  via backpropagation;
15    Retain  $h_{\tau}^{(j)}$  for each  $j$  for the next batch;
16 end

```

precise movement of limbs, however, requires a complex interaction between our nervous system, muscles, bones, and the environment. Even though these behaviors appear highly rigid and stereotyped, studies have shown that there are significant variations within and across individuals in these behaviors that can stem from neural [99, 100, 101, 102] and biomechanical perturbations [103, 104, 105, 106], environmental challenges [107, 108], psychological state [109, 110], social status [111, 112], presence of injuries [113, 114, 115] and diseases [102, 116, 117, 118, 119, 120, 121]. For example, lower socioeconomic status was found to aggravate gait speed decline with aging [111]. Another study used CO₂ inhalation to induce anxiety in healthy volunteers and observed them walking through an aperture slightly larger than their shoulder width. They found volunteers with induced anxiety decreased their gait speed, but also observed superfluous yaw rotation as they passed through the aperture, indicating increased caution and alteration of gait [109].

Deficits in crucial neural and biomechanical mechanisms underlying gait often lead to gait patterns that are qualitatively and subjectively apparent and easy to detect using simple kinematic variables such as gait speed and joint angles. However, what would be desirable is the ability to quantitatively and precisely characterize each gait cycle with or without such neuromotor deficits. Analyzing just the raw kinematic parameters to characterize the nuances of each gait cycle is difficult, as these result from complex neuromechanical dynamics that remain unobservable [122, 123, 124]. One approach to this is building musculoskeletal models that learn optimal control strategies to build predictive models of gait in humans [125, 126, 127]. While these approaches are able to simulate unimpaired gait, as well as gait affected by biomechanical or neural perturbations, they do not accurately predict gait dynamics resulting from neurological injuries [126] or more subtle perturbations [128, 127]. An alternate approach is building models based on the neural circuitry that generates movement [129, 130]. However, these models are typically trained across all individuals and

are thus limited in their ability to capture the unique gait characteristics of each individual [123, 131]. A holistic approach would be to discover individual-specific parameters using a large model based on both neural and biomechanical dynamics, however, these parameters may not be unique as in a high-dimensional model different set of parameters could lead to the same dynamical output.

These observations beg the question of whether building an individual-specific model is even required. Do we all just walk the same walk? We know that the latter is not the case, however, as numerous studies point to evidence that humans can identify gender [132], body size [133], sexual orientation [134], emotion [135], individual differences in dancing [136], perceived affective states [137] and underlying intention [138] based on gait. There is also evidence that humans use gait to determine attractiveness [139], derive diagnoses [140, 141], and to make treatment plans [142, 143].

The ability to measure high-throughput kinematics data has allowed researchers to create supervised machine learning models [144, 145], such as prediction of individual or group-level differences in orthosis design needs in a cohort of children with cerebral palsy [144]. However, these models do not provide individual-specific signatures of gait as they do not model individual-specific inter-limb coordination dynamics.

Here, we utilize LIDAR to establish a novel framework to characterize gait dynamics across multiple individuals within a low-dimensional latent space, identifying individual and group-level variations in the gait dynamics. Kinematic data is collected from a diverse cohort including both healthy and neurologically impaired individuals walking at varying gait speeds across trials. We train Recurrent Neural Networks using this data, building a generative model that captures the intricate dynamical structure underlying observed kinematic data. We extract the internal dynamics of the model after training and find a stride-averaged representation of gait dynamics across individuals. We denote this representation as gait signatures and utilize it

as a quantitative descriptor of gait cycles for individuals. We find that gait signatures elucidate differences across individuals, groups, walking speeds, and degree of neurological impairment. To test the generalizability of this model, predictions are generated from the model at novel gait speeds, and find that the model performs better compared to direct kinematic interpolation in healthy individuals.

Lower-dimensional projections of gait signatures are found to be biomechanically interpretable and relate to inter- and intra-limb coordination patterns. Manipulating these projections results in changes to model predicted gait dynamics, creating a framework to infer relationships between perturbed dynamical components and their kinematic manifestations. When this approach is applied to a different dataset of healthy individuals with widely different walking speeds from exceptionally slow (0.3m/s) to exceptionally fast (as transitioning from walk to run), gait signatures are found to remain individual-specific across all walking speeds and are found to contain predictable linear changes across the entire speed range.

2.4.2 Dataset

Continuous gait kinematics were recorded from 12 individuals - 5 able-bodied (age = 24 ± 4 years; 4 female) and 7 stroke survivors (age = 56 ± 12 years; 2 females; 48 ± 25 months post-stroke; > 6 months post-stroke). Post-stroke participants experienced a cortical or subcortical ischemic stroke and were able to walk on a treadmill for one minute without an orthotic device without any signs of hemi-neglect or orthopedic conditions. All participants walked for 15 seconds, each on the treadmill at 6 different speeds that were gradually and evenly increased from the participant's self-selected speed to the fastest safe and comfortable speed. Participants rested for 1-2 minutes between consecutive gait trials. Gait speeds ranged from 0.1 to 0.8 m/s for stroke-surviving participants. The fastest safe speed was determined as the speed that participants on the treadmill could safely and comfortably maintain for 30 seconds.

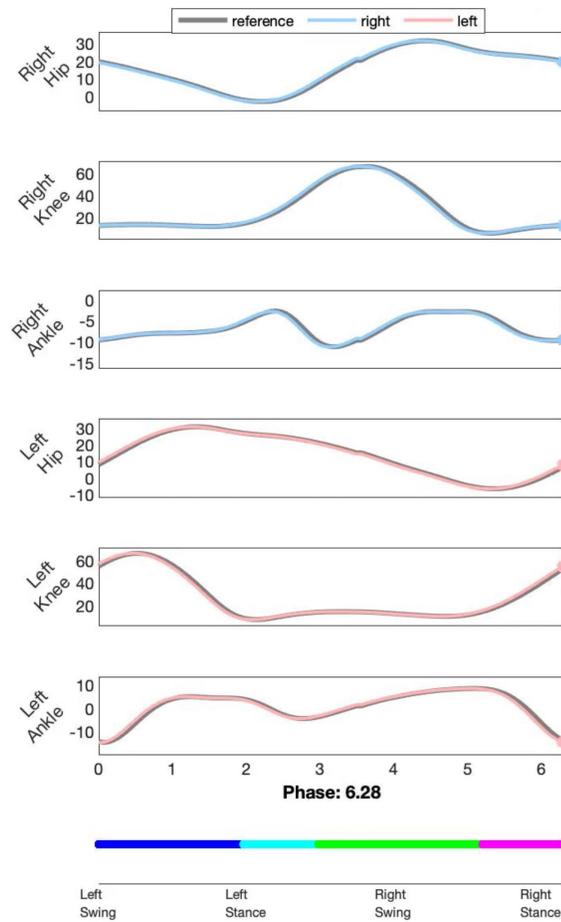


Figure 2.1: Example phase averaged gait kinematics. 6 joint angles were measured from participants walking at variable speeds. Here, phase-averaged kinematics are shown from one gait cycle.

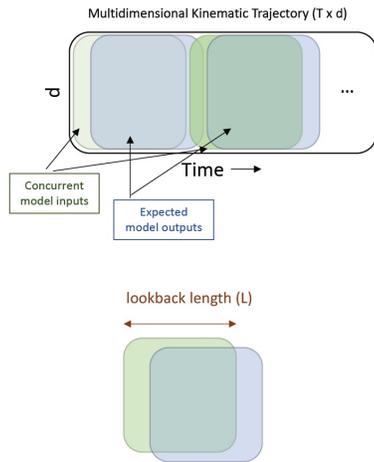
Sagittal-plane kinematic data was recorded using motion capture from bilateral hip, knee, and ankle joints, resulting in a 6-dimensional time series from 72 distinct trials across all participants. Each time series was sampled at 100 Hz, resulting in 1500 time points per trial.

2.4.3 Model training and selection

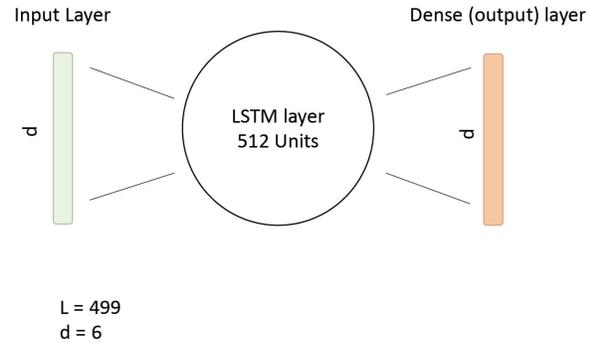
We trained a sequence-to-sequence RNN model that uses input data (0 to $\tau-1$) to predict kinematics one time step in the future (1 to $\tau + 1$). The model was trained statefully with a batch size equal to the total number of trials in the dataset (72) and training time steps equal to 499. With 1500 total time points in each trial, we were able to create 3 mini-batches (2 training batches and 1 validation batch) with 499 time points in each sample. For example, the first mini-batch consisted of time points 0 to 499 from all trials, and the corresponding 1-shifted output consisted of time points 1 to 500 from all trials. The networks were trained using mean-squared-error as the loss function and ADAM optimizer algorithm for 5000 iterations or until training and validation errors converged ($< 0.75^\circ$).

Single-layer RNN architectures were evaluated with varying the number of LSTM units and training time steps. Specifically, models with the following LSTM units : [128,256,512,1024] and training time steps : [249,499,749] were trained, and the training and validation loss associated with each pair of hyperparameters was obtained. Following are the training and validation losses for each [units-time steps] pair of hyperparameters - 512–499 ($MSE_{train} = 0.010 \pm 0.001 \text{ deg}^2$; $MSE_{val} = 0.018 \pm 0.000 \text{ deg}^2$), 256–749 ($MSE_{train} = 0.010 \pm 0.002 \text{ deg}^2$; $MSE_{val} = 0.015 \pm 0.001 \text{ deg}^2$), 256–499 ($MSE_{train} = 0.010 \pm 0.001 \text{ deg}^2$; $MSE_{val} = 0.017 \pm 0.000 \text{ deg}^2$). The training loss was not different between hyperparameter pairs ($p > 0.235$). The validation loss differed between all three models ($p < 0.001$), with the 256–749 model having the lowest validation loss. However, if the differences in validation loss of less

A) Training Dataset



B) Model Architecture



C) Stateful Training

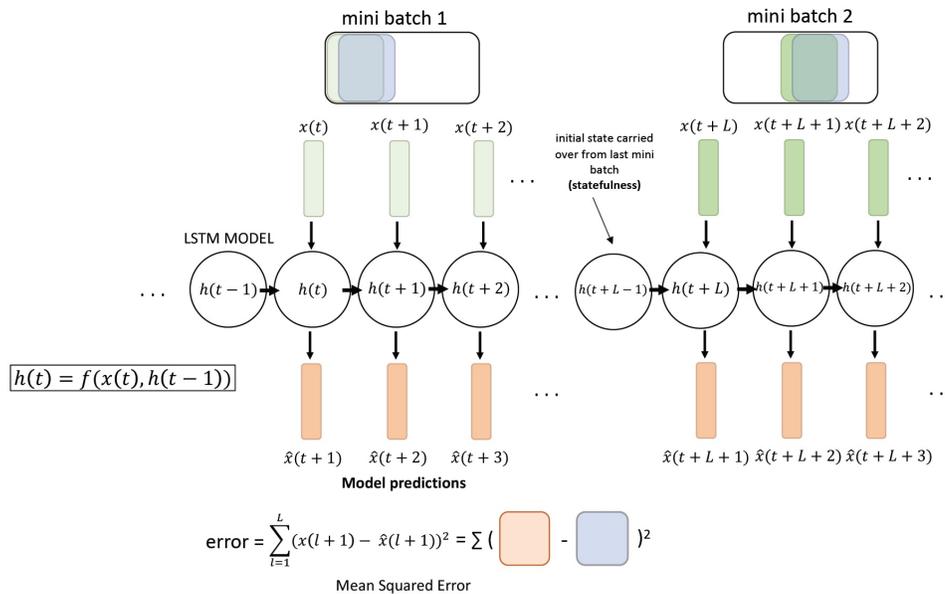


Figure 2.2: Training RNN model on kinematics data. A. A sequence-to-sequence RNN model was trained using kinematic data, where each input to the model is a sequence of kinematic data, and the model predicts a sequence of kinematic data shifted by one time step, i.e., the model predicts kinematics data at $t+1$ given kinematics at time t . B. A simple RNN architecture with one hidden LSTM layer with 512 units and 1 dense output layer was used. The input layer and output (dense) layer had 6 units, corresponding to the dimensionality of kinematics data. C. The model was trained statefully, where each sample in each mini-batch was temporally consecutive, resulting in temporal continuity across batches. The hidden states of the model were not reset after each batch was processed.

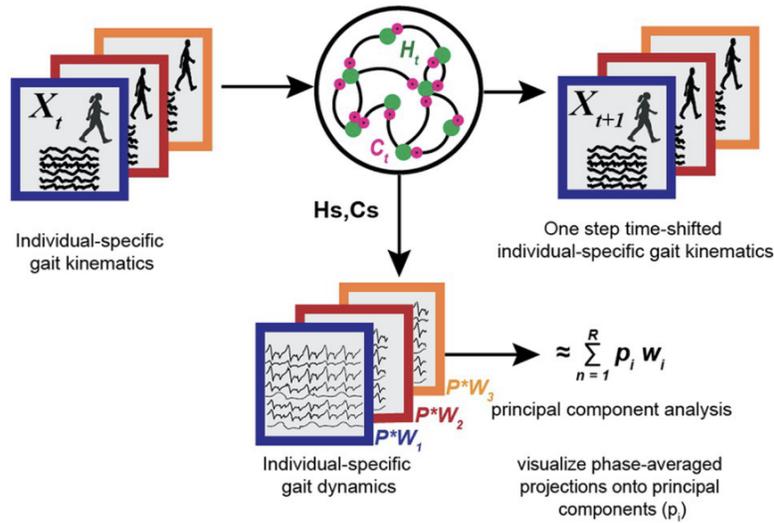
than 0.003 deg^2 corresponded to meaningful differences in performance was unclear.

2.4.4 Measuring gait signatures from RNNs

Once the RNN model was trained, the hidden and cell states (h and c variables) from the LSTM units were extracted from the model by feeding the kinematics of each trial through the model. As shown in equations 2.7 to 2.12, LSTM units use 2 variables to characterize their internal state; a cell state (c_t) and a hidden state (h_t). With 512 units in the LSTM layer, this results in a 1500×1024 dimensional time series for each trial, concatenating h and c state variables (2.3A). To reduce the dimensionality of this representation, Principal Components Analysis (PCA) was performed to linearly project high-dimensional dynamics along directions with maximal variance. The first 6 PCA components account for 72% variance of the high-dimensional dataset and are denoted as *gait signatures*. When the first 3 dominant principal components are plotted in 3D (2.3B) for 3 representative individuals (able-bodied, high-functioning, and low-functioning stroke survivor), differences are observed in the dynamical trajectories between able-bodied and stroke survivor participants. To determine whether some structure exists amongst the three different subject groups, the time series of 6-dimensional gait signatures from each participant was projected into 3-dimensions map using Multidimensional scaling, an unsupervised dimensionality reduction algorithm that preserves relative distances between low-dimensional projections and the high-dimensional gait signature space, to visualize relative distances between all gait signatures (2.3B, right). Multi-dimensional Scaling (MDS) projections from participants belonging to each group cluster together, highlighting the potential ability of gait signatures to identify participant groups and other clinically relevant features.

Since participants performed each trial at different speeds but with well-defined periodic gaits, we can use phase averaging to align kinematics. This process was

A) Data-driven approach: train sequence to sequence RNN to extract dynamics



B) 3D visualizations to examine individual differences in gait dynamics

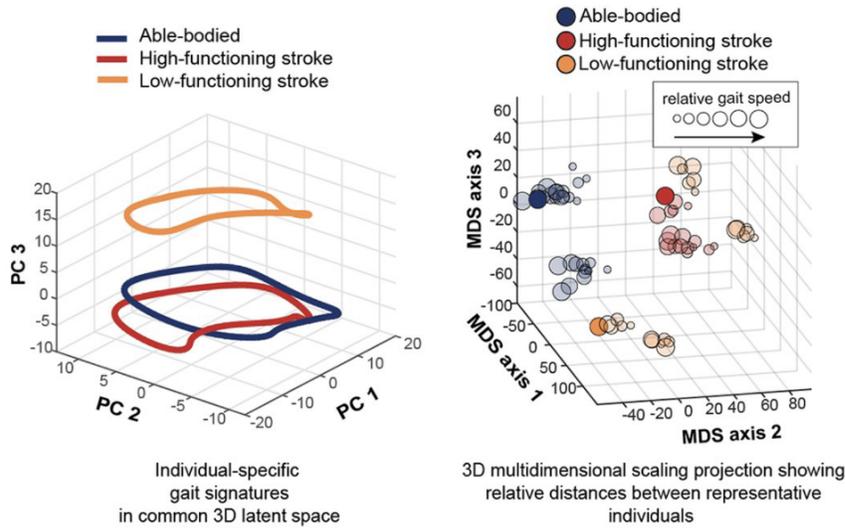


Figure 2.3: Identifying gait-signatures from kinematic data. A) Continuous, multi-joint movement data from multiple individuals are input into a Recurrent Neural Network (RNN) model. The model is trained to predict kinematic output one time step ahead using data from all individuals. Using the trained model, high-dimensional internal states (\vec{h} and \vec{c}) from each individual are extracted from the model. Principal component analysis is then applied to reduce the data's dimensionality, forming individual gait signatures. B)(Left) Visualization of the first 3 principal components of gait signatures from 3 individuals: able-bodied (blue), high-functioning stroke survivor (red), and low-functioning stroke survivor (orange). B)(Right) 3D projections of the 6D gait signatures using multi-dimensional scaling (MDS) reveal different gait dynamics among the three groups. The size of each circle represents the individual's walking speed during the trial, with smaller circles indicating slower speeds and larger circles representing faster speeds. Figure taken from [12].

performed using the Phaser algorithm² [146] on gait signatures, translating each gait signature to a phase from 0 to 2π , resulting in a speed-independent measure of each gait cycle.

2.4.5 Gait signatures are individual and group-specific

To what degree do gait signatures encode individual-specific information? To answer this question, we plot the first three principal components of the phase-averaged gait signatures from all trials from all individuals (Fig2.4i). Gait signatures of individuals' six speed trials within both healthy and impaired groups are tightly grouped together. Gait signatures within the unimpaired group are highly stereotyped and live on a low-dimension structure across individuals, whereas those from the impaired group exhibit highly variable gait signatures. Moreover, we see that gait signatures from both groups are aligned over the four gait phases (leg 1 swing, leg 1 stance, leg 2 swing, leg 2 stance). However, there is more variation in the impaired group as expected. To quantify the similarity between gait signatures, we reduce gait signatures into a 3-dimensional representation using MDS, as before. The unimpaired group forms a tight cluster in this lower-dimensional representation (Fig2.4Bi) away from the impaired individuals highlighting distinct traits of gait signatures from the two groups. The representation of impaired individuals in this space is very sparse, highlighting heterogeneity in gait deficits in the stroke cohort. We find that within an individual, walking speed imparts a relatively minuscule change in gait signatures compared to the change in gait signatures between individuals(Fig2.4Bii).

Moreover, to quantify the degree of separability between unimpaired and impaired gaits, we calculate the Euclidean distance between the centroid of all unimpaired gait signatures and each of the impaired gait signatures in the 3-dimensional MDS space. When plotting this distance against gait speed (Fig2.4Bii orange inset), it is clear that

²<https://github.com/sheim/phaser>

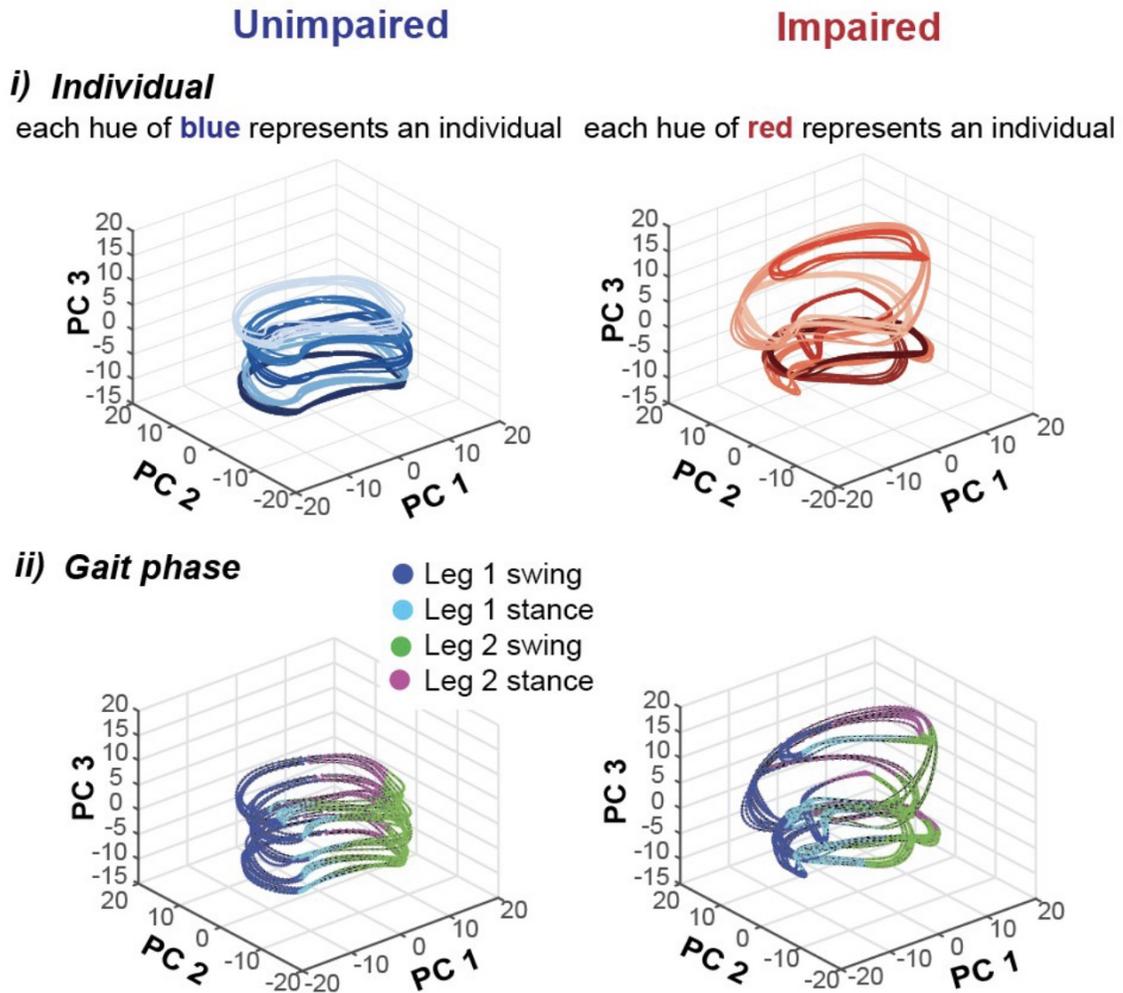


Figure 2.4: Gait signatures are individual and group-specific. 3D gait signatures for unimpaired subjects (left) and impaired subjects (right) are color-coded in two ways: i) by individual and ii) by gait phase. In the individual-based coloring (i), signatures cluster according to specific subjects within both groups. For unimpaired subjects (i, left), similar shades of blue group together, while for impaired subjects (i, right), similar shades of red form clusters. This grouping indicates that each person has a unique gait signature. When color-coded by gait phase (ii), unimpaired gait signatures (ii, left) show a consistent looped structure across the four phases of a gait cycle: leg 1 swing, leg 1 stance, leg 2 swing, and leg 2 stance. In contrast, impaired signatures (ii, right) exhibit subject-specific variations across these four phases and demonstrate greater overall variability. Figure taken from [12].

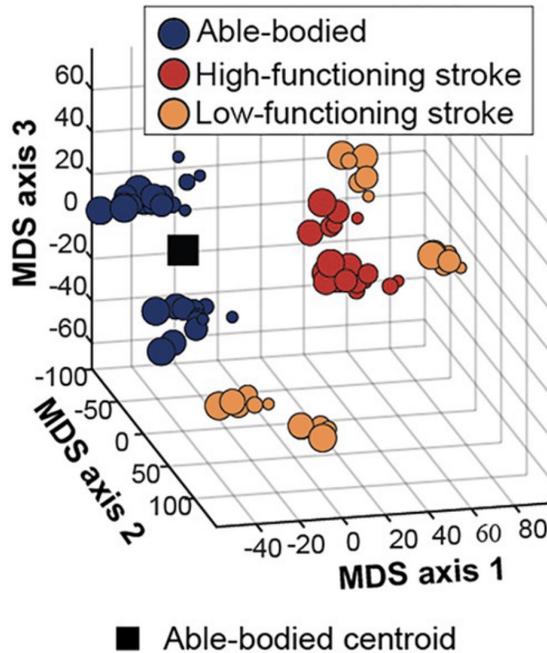


Figure 2.5: Comparison of gait signatures between three distinct locomotor subgroups: able-bodied (AB), high-functioning stroke survivors (HF), and low-functioning stroke survivors (LF). 3D gait map created using multi-dimensional scaling shows the relative spatial relationships between AB subjects (in blue), HF stroke survivors (in red), and LF stroke survivors (in orange). LF stroke survivors exhibit less cohesive clustering, occupying unique regions within the map. These regions are notably distant from the centroid of the able-bodied group, which is represented by a black square marker. Figure taken from [12].

low-functioning stroke survivors are further away from the unimpaired individuals than high-functioning stroke survivors. Gait signatures in the MDS space also show separation between high and low-functioning stroke survivors (Fig2.5).

2.4.6 Biomechanical interpretability of gait signatures

When projected onto the first six principal components, gait signatures exhibit a periodic pattern that fluctuates over the gait cycle (Fig. 2.6B). Using these projections, we see that high and low-functioning stroke survivors have very similar dynamics to able-bodied participants over the first 2 principal components (within 95% confidence interval, shaded gray). However low-functioning survivors differ significantly over the

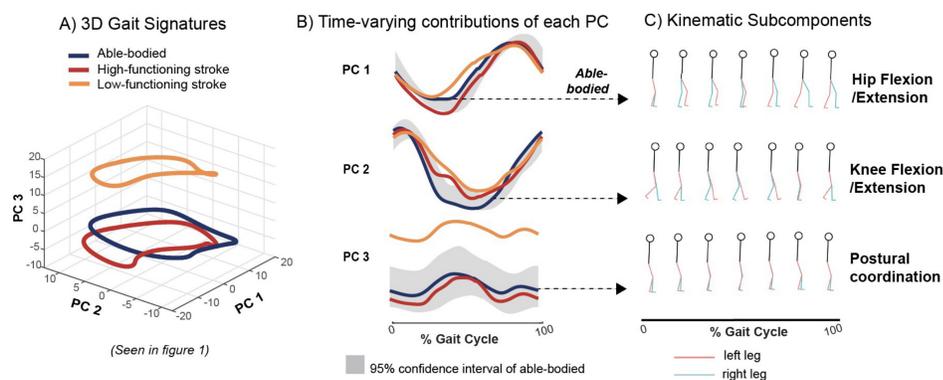


Figure 2.6: Biomechanical analysis of gait signatures. A) The gait signatures demonstrate distinct walking dynamics among representative able-bodied (AB) individuals, low-functioning, and high-functioning stroke survivors. B) The weightings of each principal component (PC) fluctuate throughout the gait cycle. These can be contrasted with the 95% confidence interval of the AB group (shown in gray). C) When used to drive the gait model, each PC produces specific multi-joint walking coordination patterns. This enables a biomechanical interpretation of gait impairments and the effects of interventions. Figure taken from [12].

third principle component. To derive biomechanical interpretability of these principal components, we re-project each principal component into the high-dimensional hidden-state space of the model and use these hidden states to predict joint angles using our trained model. We visualize the corresponding kinematics through movie representations³ and find the first three components represent hip flexion/extension, knee flexion/extension, and high-level postural coordination, respectively.

2.4.7 Using RNNs to generalize to new gait speeds

To test the generalizability of the RNN model, we train a new model with data from only the 2 slowest and 2 fastest gait speed trials from each participant and use a speed-weighted average of gait signatures at these speeds as a predictive measure of gait signatures at the average speed of the 2 slowest and 2 fastest trials. When driving the trained model with interpolated gait signatures, the generated joint kinematic predictions in unimpaired (Fig. 2.7A) and low-functioning stroke survivors (Fig.

³<https://doi.org/10.1371/journal.pcbi.1011556.s001>

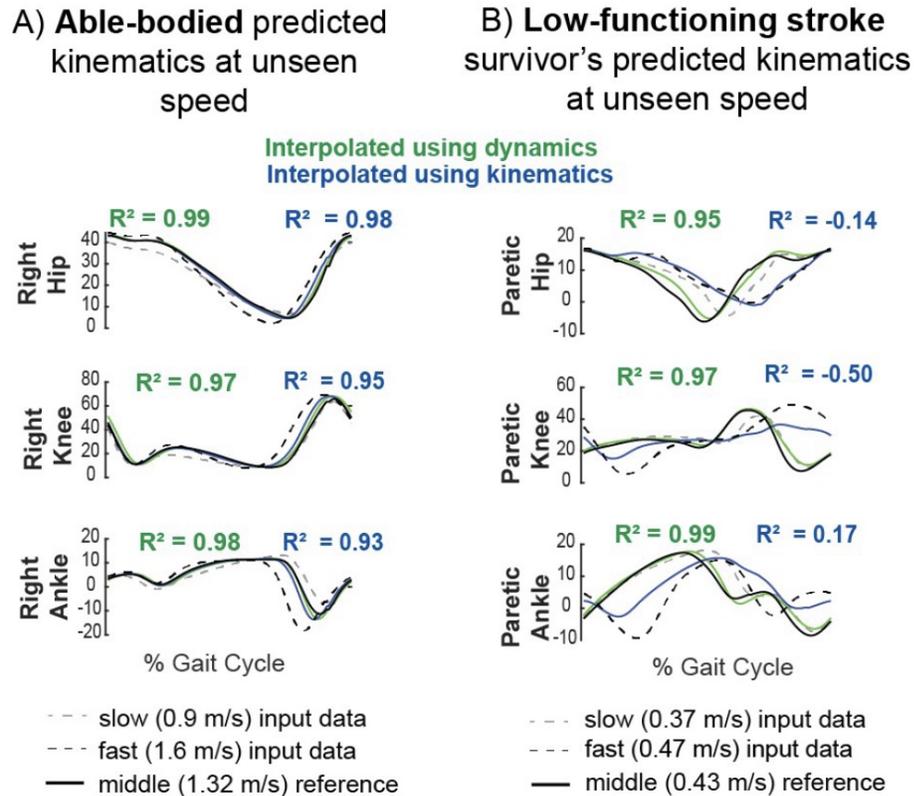


Figure 2.7: Generalizing to new gait speeds. Joint kinematic predictions (shown in green) for intermediate walking speeds were excluded from model training. They were generated by first interpolating gait signatures (hidden-states) obtained from slow (dashed grey) and fast (dashed black) speed trials and then driving the model with the interpolated gait signatures to predict kinematics at the intermediate speed. Joint kinematics interpolated directly from kinematics data (blue) and the measured reference kinematics (solid black) are also shown. A) For a representative able-bodied (AB) participant, predictions are more accurate when interpolating gait signatures compared to interpolating gait kinematics across speeds. B) In a representative low-functioning stroke survivor, interpolated gait signatures more accurately predict non-linear changes in kinematics at intermediate speeds than interpolated gait kinematics. Simple averaging of kinematics fails in this case, where larger differences exist between the slow and fast speed paretic kinematics. The averaged kinematics (blue) follow the fast speed paretic hip kinematics, while other joint angles do not resemble waveforms from either the fast or slow speed. Figure taken from [12].

2.7A) show similarities with experimental data acquired at the average gait speed from the same participants. We also predict joint kinematics at the average gait speed by directly interpolating the phase-averaged kinematics data (joint angles) at slower and faster speeds. We find that kinematics predicted from the model perform considerably better, highlighting the generalizability of the model.

2.5 Building multi-timescale representation of animal behavior using RNNs

Recurrent Neural Networks are powerful deep-learning tools that build internal representations that can reflect dynamical structures underlying data. To mimic the dynamical structure underlying postural time series data, we can use RNNs as an over-parameterized dynamical model and use the representations learned by the model to understand the dynamical structure. In this section, we begin with developing a novel multi-layer Recurrent Neural Network model to build representations of behavioral time series data using multiple distinct datasets of flies and rats. We show these representations are meaningful and capture stereotyped behaviors performed by these animals.

We first draw inspiration for the model architecture from the work of Zeiler et al [13], where they visualize the representations learned by each layer in a deep learning model. They train a multi-layer convolutional network with 8 layers that takes a 224x224 pixel image as input and classifies the image. They train this network on ImageNet 2012 training set, a standard dataset with 1.3 million images spanning 1000 different classes such as animals, household objects, vehicles, food items, etc. To understand the convolution structures learned by the network, the learned convolution filters are reconstructed and projected back to the input image pixels to identify spatial structures learned by the convolution in the original image. Fig2.8 presents

this visualization of convolutional filters from layers 1 to 5 of the trained network. They find these features are not random or uninterpretable, but rather show compositionality and increased invariance in deeper layers of the network. The convolution filters discriminate specific structures in the input image, hierarchically identifying features across layers. For example, filters in layer 2 responds to linear features like corners and edges in the image, layer 3 captures more complex features such as textures (mesh patterns, text). Layer 4 identifies compositionally complex class-specific structures such as dog faces, and bird legs, and layer 5 represents entire objects with significant postural variations.

Further in this section, we discuss a novel method to build coarsened representations of behavior at timescales longer than postural dynamics by training Recurrent Neural Network models recursively, using the hidden state dynamics from each network to train the subsequent model. We find that each subsequent network builds an internal representation of the data at a coarser timescale across all datasets. To showcase potential applications of this method, we apply the deterministic information bottleneck [16, 147] formalism to the obtained representations from a dataset containing simultaneous neural activity recorded from the dorsolateral striatum, a brain region relevant to motor learning and control. We ask if there is any predictive information about neural dynamics in the behavioral representations at different timescales, and find different timescales encode predictive information for distinct neural units.

2.5.1 Datasets

Fly Dataset

LEAP Fly Dataset: This dataset was original taken from Berman et al. [3]. Individual flies were imaged in a clear plastic dome 100mm in diameter and with gently sloping sides forcing flies to avoid jumping, flying, and adhering upside-down

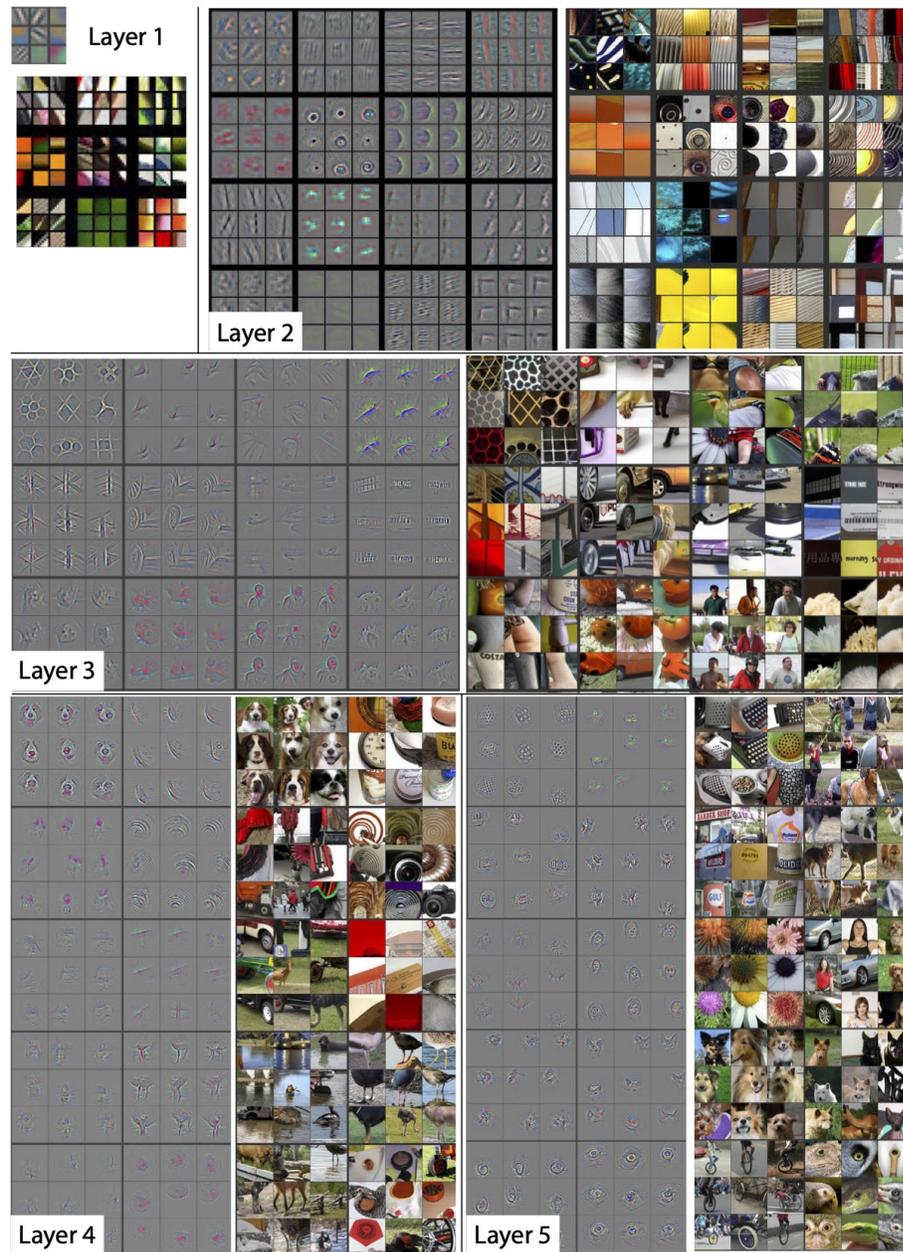


Figure 2.8: Visualization of features in a multilayer convolutional neural network trained on ImageNet 2012 dataset. For layers 2-5, the nine highest activations in a randomly selected subset of feature maps across the validation dataset are shown. These activations are projected back to pixel space using the method described in [13]. Individual feature maps are highly clustered, and there is increased feature invariance in higher layers. Layers 4 and 5 show strong activation for distinct image elements such as canine eyes and noses. Figure taken from [13].

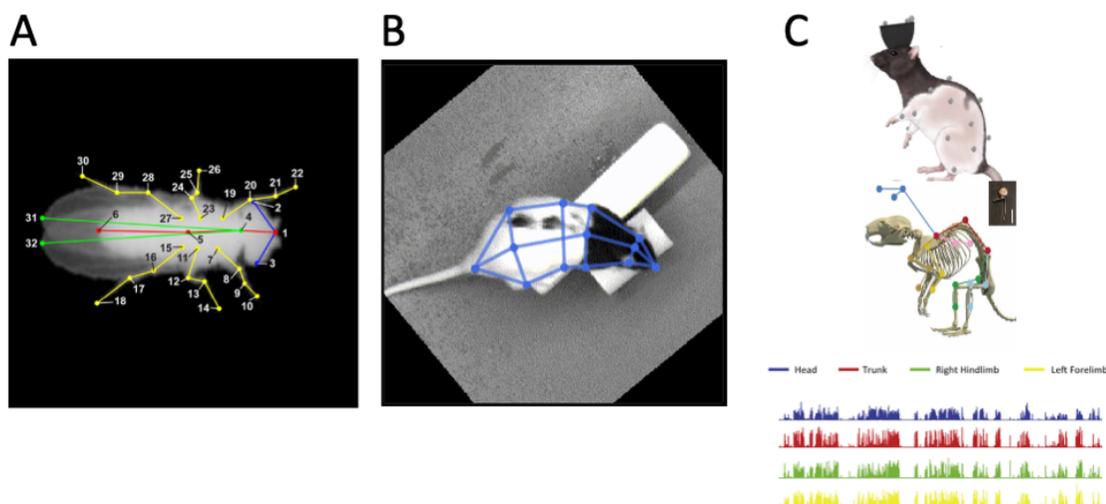


Figure 2.9: Behavioral datasets used here. A) Leap Fly Dataset. Postures were tracked from videos of freely behaving fruit flies (imaged at 100 Hz) obtained from [3] using LEAP [1]. 32 keypoints on the body of the fly were tracked. B) OFT Rat Dataset. Freely behaving rats in an enriched arena were imaged from above at 60 Hz, and 13 postural markers on the rat were tracked using DeepLabCut [5]. C) Motion Capture Rats. This dataset is taken from [14], 3D motion tracking of 20 postural markers was performed on rats at 300 Hz using infrared beads attached. Simultaneous neural recordings from the dorsolateral striatum (DLS) were also performed.

in this arena. Flies were imaged from above with a camera, and online tracking was performed to move the arena to keep flies centered in the camera view. A total of 59 male flies and 51 female flies were imaged at 100 Hz for 1 hour, generating frames 1088 x 1088 pixels in size. The flies were isolated in each frame, and the frame was cropped and rotated to keep flies rotationally and translationally invariant. This resulted in images that were 200 x 200 pixels in size (Fig. 2.9A). Pereira et al. [1] develop an open-source deep-learning tool for animal pose estimations and apply it to this dataset, training their model using manually annotated frames from the fly dataset. 32 locations on the body of the fly were manually annotated (Fig. 2.9A). The trained model generates a postural time series tracking 32 locations in 2 dimensions, resulting in a 64-dimensional time series dataset.

Open-field Rat Dataset

OFT Rat Dataset: This dataset was acquired in Dr. Aman Saleem’s lab at University College London by Drs. Tomaso Muzzu and Elena Menichini. The dataset consists of individual rats exploring a 1m x 1m open arena for 3 hours imaged from above for 3 hours using an IR camera. 13 locations on the rat’s body were tracked using DeepLabCut [148], and the tracked postures were rotationally and translationally aligned (Fig2.9B). The arena is enriched with toys, food, and water to encourage interactions, and, as a result, more active behaviors. The resulting video is 1280 x 1024 pixels in resolution and is recorded at 60 Hz.

Motion captured Rat Dataset

Motion Capture Rats: This dataset was introduced in Marshall et al. [14], where a novel approach is developed for 3-dimensional tracking of rat posture using retroreflective markers. 20 such markers were fused with transdermal body piercings that were attached to each rat. This allowed 3-D triangulation and tracking using 12 cameras, providing robustness against occlusions that typically occur in 2-D posture tracking setups(Fig2.9C). 5 rats were continuously tracked for a week at 300hz with sub-millimeter precision. The dataset also contains simultaneous electrophysiological recordings from the dorsolateral striatum (DLS) region in the brain.

2.5.2 Building representations at a single timescale

Each dataset described above contains a multi-dimensional postural time series obtained from multiple animals. We create a model with 3 LSTM layers: layer 1 receives the multidimensional postural input, it is then connected directly to layers 2 and 3, and layer 2 is also connected to layer 3. A dense layer after the third LSTM layer projects the hidden states from layer 3 back into the postural space (Fig. 2.10 top). This U-shaped architecture with a skip connection between layers 1 and 3 helps preserve fine-grained temporal information encoded in layer 1 (as layer 1 is driven by inputs), allowing layers 2 and 3 to build representations of longer-timescale dynamics.

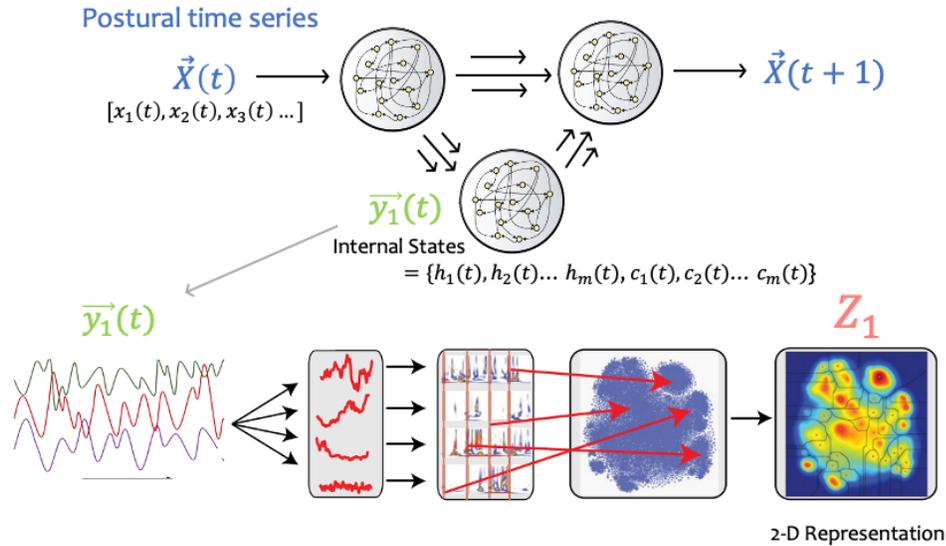


Figure 2.10: Level 1 RNN model for generating behavioral representation. (Top) A 3-layer recurrent neural network with U-Net-inspired architecture and LSTM units receives the posture of the animal at time t $\vec{X}(t)$ and predicts the posture at $t+1$ $\vec{X}(t+1)$. (Middle) After training, the entire postural time series $\vec{X}(t)$ is fed through the network, and the internal states from layer 2 $\vec{y}_1(t)$ are extracted containing both $\vec{h}(t)$ and $\vec{c}(t)$ variables of the internal states. (Bottom) $\vec{y}_1(t)$ is used to build a 2D representation following a pipeline inspired from [3]. Each dimension in $\vec{y}_1(t)$ is first decomposed into a spectrogram using wavelet transform and then concatenated together, the high-dimensional representation is then embedded in 2D (Z_1) using t-SNE [15].

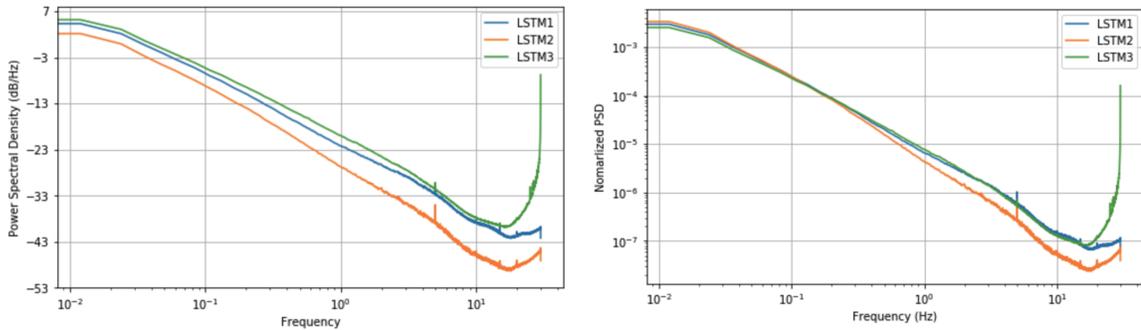


Figure 2.11: Power spectral density (left, normalized across layers on right) of extracted hidden state dynamics from each layer of level 1 RNN model trained Motion Capture Rat Dataset.

We create a stateful training dataset using our approach explained before, and train the model statefully, predicting postural tracking 1 time step in the future.

Once the model is trained, the network is driven statefully by using the postural time series, and hidden layer activities (both \vec{h} and \vec{c} variables) from all layers are extracted and saved. Fig. 2.11 shows the power spectral density of the concatenated hidden state activities from each layer when trained on the *Motion capture Rats* dataset. We observe that layer 2 activities encode slower dynamics, as evident from low normalized power at higher frequencies (Fig. 2.11 right). To find the longest-timescale structure possible, we pick the 2nd LSTM layer to build our behavioral representations.

Using a previously established pipeline [3], a 2-D representation of the 2nd LSTM layer hidden layer activities ($\vec{y}_1(t)$) is created (Fig. 2.10 bottom). Specifically, the extracted hidden state activities are transformed into a spectrogram using a Morlet wavelet transform, decomposing each hidden state variable into 25 dyadically spaced frequencies between 1 Hz and the Nyquist frequency (half of the sampling rate of the dataset). This results in a $25 \times d$ dimensional time series, where d is the dimensionality of the hidden states. To reduce the dimensionality of this data while preserving the local structure, we use t-distributed stochastic neighbor embedding (t-SNE) technique [15], a dimensionality reduction algorithm, to embed this data into

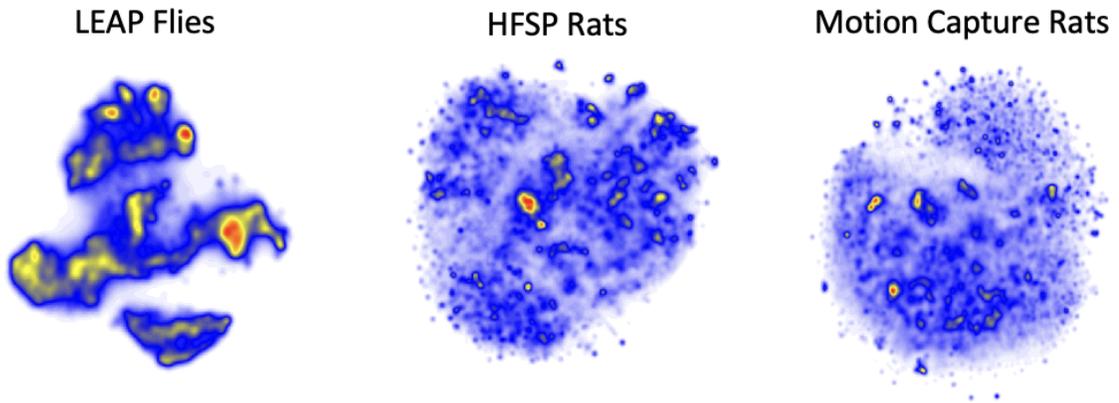


Figure 2.12: 2D representations of each dataset from Fig.2.9 obtained using the level 1 RNN model(Fig.2.10).

2-dimensions (Z_1). t-SNE works by minimizing the difference between the probability distribution of pairwise distances between points in the high-dimensional space and the low-dimensional embedding of these points. The 2-D representations obtained for each dataset are shown in Fig. 2.12.

2.5.3 Building representations at coarsened timescales

To build coarser representations, we use the hidden state activities obtained from layer 2 from the previous model $\vec{y}_1(t)$. RNN hidden states integrate information over time, and to identify long-timescale structure from data, we downsample $\vec{y}_1(t)$ by a factor of 2 to get $\vec{Y}_1(t')$. We create a stateful training set using $\vec{Y}_1(t')$ to train a new RNN model (calling it Level 2 model) with an identical architecture as before. After training, we obtain hidden state activities from the 2nd LSTM layer of this network and denote it as $(\vec{y}_2(t'))$. We perform this process recursively, training 6 models (Level 1 to Level 6) in total, extracting hidden states from each model after training and obtaining $\{\vec{y}_1(t), \vec{y}_2(t), \vec{y}_3(t) \dots \vec{y}_6(t)\}$. To check for timescales encoded in each level, we look at the distribution of power across different frequencies in each signal (Fig. 2.14). We observe that earlier levels are richer in power at higher frequencies and

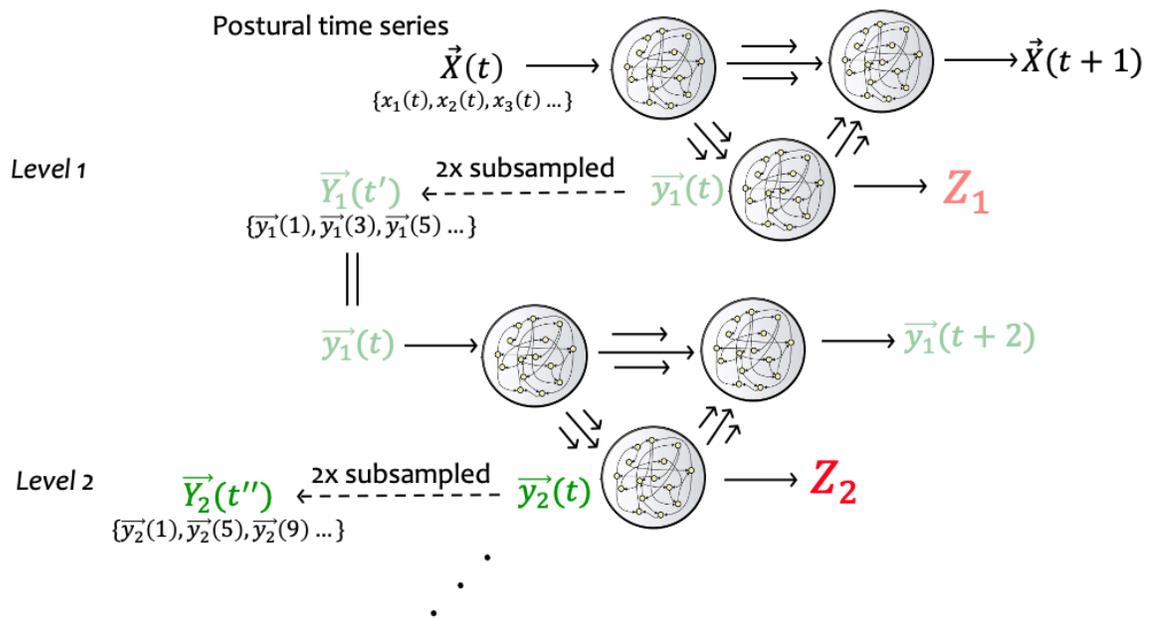


Figure 2.13: Hierarchical RNNs. (Top) Hidden state dynamics $\vec{y}_1(t)$ obtained from the level 1 RNN model are subsampled by a factor of 2 ($\vec{Y}_1(t')$) and are used to statefully train a different but identical RNN model (level 2). After training, hidden states of the level 2 model are extracted $\vec{y}_2(t)$ and embedded into a 2D representation Z_2 . This process is repeated until level 6, after which there is not enough data left to train the subsequent model due to subsampling.

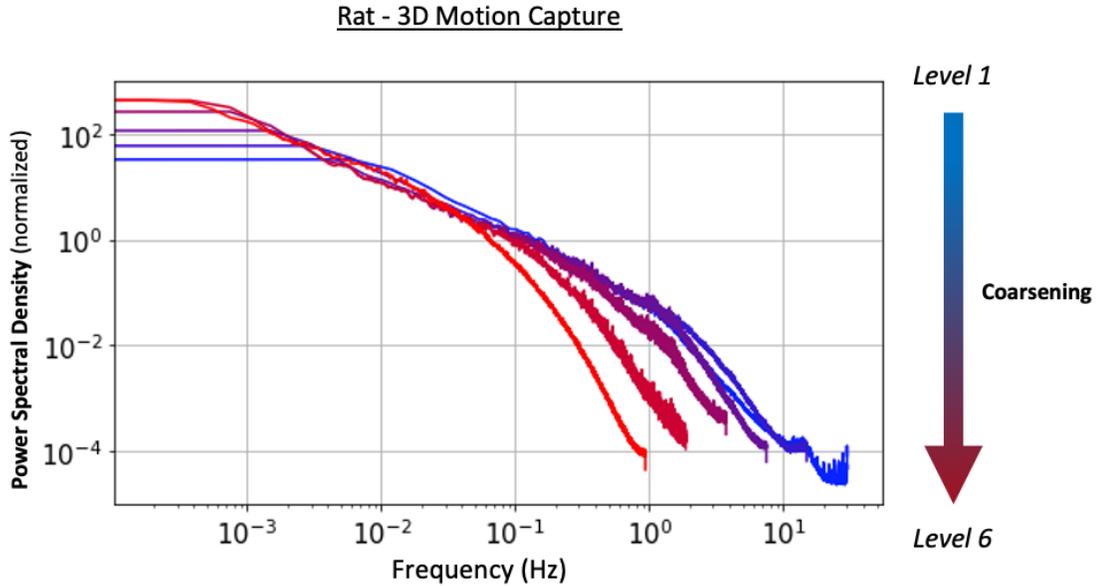


Figure 2.14: Power spectral density (normalized) of hidden state dynamics $\vec{y}_1(t)$ to $\vec{y}_6(t)$ obtained from models from each level trained on the Motion Captured Rats dataset.

latter levels are rich in power at lower frequencies. There is a gradual shift in this power, denoting a coarsening of temporal representation at each level.

Similar to the level 1 model, we reduce the hidden state activities from each level to a 2-dimensional representation obtaining $\{Z_1, Z_2, Z_3 \dots Z_6\}$. We apply this process to all 3 datasets, obtaining the representations shown in Figure 2.15.

2.5.4 Linking behavioral and neural representations across timescales

An important aspect of building computational tools for creating behavioral representations at disparate timescales is the ability to explain neural correlates that underlie behavior. This is crucial to generate a mechanistic understanding of neural processes in the brain and could help propose new experimental designs [149].

We explore such neural encoding of behavior using the *Motion Capture Rats*

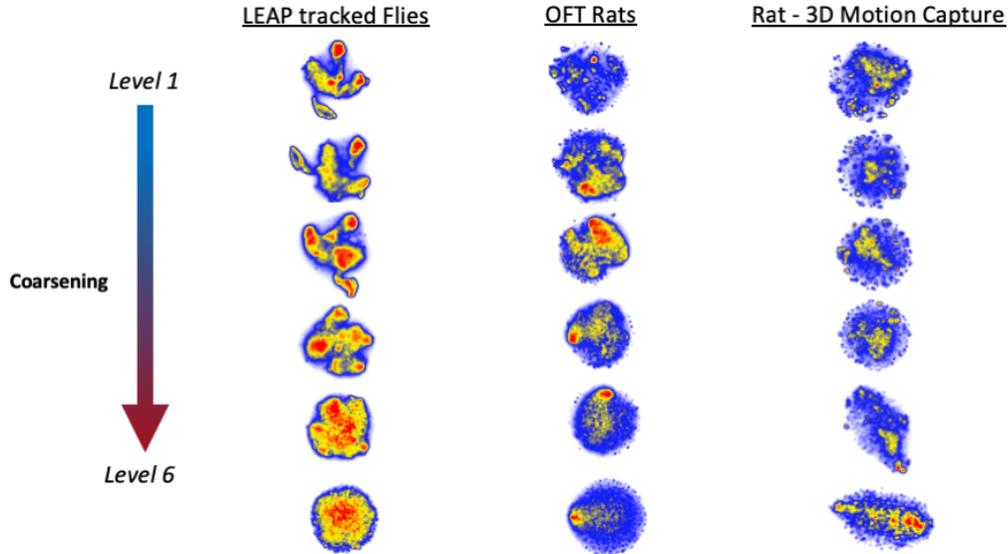


Figure 2.15: 2D representations obtained using RNN models (levels 1 to 6) for LEAP Fly Dataset (left), Open Field Rat dataset (middle), and Motion Capture Rats dataset (right). Representations show coarsening at deeper levels.

dataset, which contains simultaneous neural and behavioral recording. For these animals, neural activity recorded from the dorsolateral striatum was spike-sorted, yielding 60 neural units for each animal. Using the representations $\{Z_1, Z_2, Z_3 \dots Z_6\}$ obtained from this dataset, we try to infer if we can optimally decode behavioral repertoires in each representation using these units and if different units encode behavioral repertoires differently at each coarsening level.

To find this encoding, we use the deterministic information bottleneck formalism [16, 147]. Here, we try to partition the space Z into a clustered representation T in a way that minimizes the function:

$$\min_T \mathcal{F} = H(T) - \beta I(T; spikes), \quad (2.17)$$

where $H(T)$ corresponds to the entropy of the partitioned space, β is the Lagrange multiplier, and $I(T; Spikes)$ is the mutual information preserved between the partitioned space and the neural spikes from a unit. As we minimize \mathcal{F} for varying

values of β and the number of partitions in T , we find solutions that try to minimize the complexity (and thus the entropy) of T while preserving the predictability between the T and neural spikes. This optimization results in the Pareto front shown in Figure 2.16. Here, ideal solutions lie in the direction denoted by the large arrow where solutions with maximum information and minimum complexity exist. We obtain such Pareto fronts for each unit (Fig. 2.17) and measure the maximum possible information (I_{max}) encoded by each neural unit at each level (Fig. 2.18) of behavioral representation without constraining the complexity of partitioning. We observed a lot of variance in information encoded by each unit across all 6 levels, denoting different encoding capacities of each unit across coarsening levels of behavioral representation. To illustrate, we identify units a and b (Fig. 2.18 solid and dashed black lines) which have statistically high I_{max} in levels 1 and 6 respectively, and check the conditional representation of behavior at each level when these units were spiking (Fig. 2.19). Unit 'a' has a punctate representation in level 1, which spreads out into a sparse representation in level 6, whereas unit 'b' shows a sparse representation in level 1, which subsequently becomes relatively peaked in level 6. This method provides a new framework to get insights into how these neural units encode behaviors at different timescales.

2.6 Conclusion

At the beginning of this chapter, I introduced RNNs and motivated their utility in modeling dynamical systems. However, RNNs are shown to be prone to gradient instabilities when training on large time series datasets, hindering their ability to learn long-timescale structures from data. Here, we introduce LIDAR, a Python package and training formalism that uses a careful curation of training data from large time series data sets to allow for stateful propagation of hidden states of the

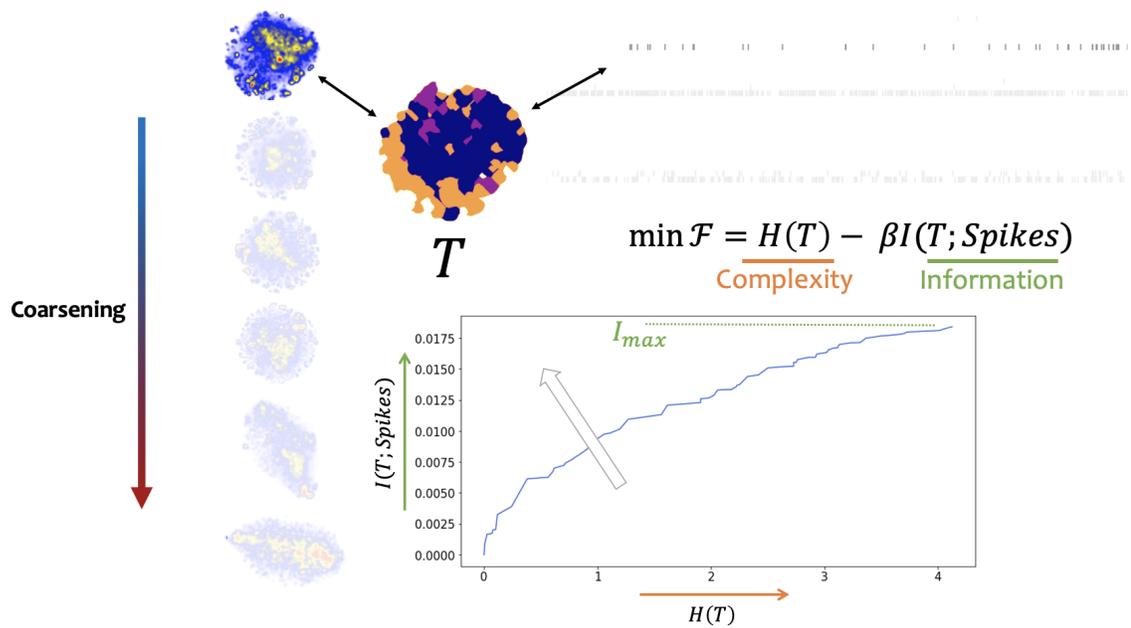


Figure 2.16: (Top) Using the 2D representation from level 1 (Z_1) for Motion Capture Rat dataset, Deterministic Information Bottleneck [16] (DIB) is used to minimize \mathcal{F} , identifying a minimally complex partitioning of Z_1 (for a fixed number of partitions) while preserving information between the partitions and spiking data from a neural unit i . (Bottom) Pareto front obtained during DIB optimization, optimal solutions lie in the direction indicated by the arrow indicating low complexity of partitioning while maximizing the encoded information.

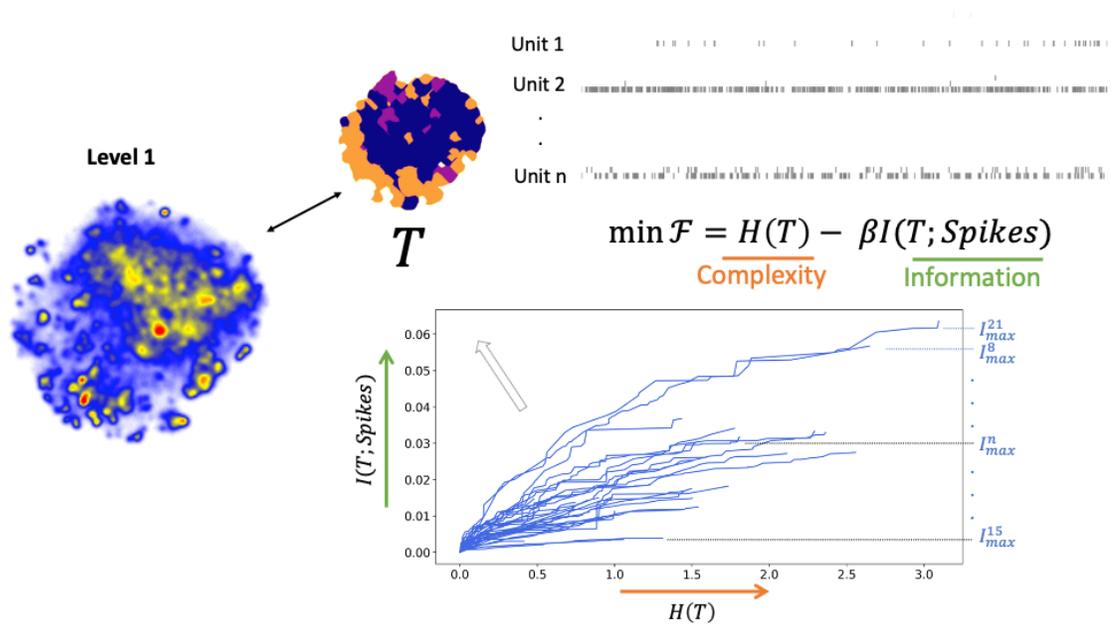


Figure 2.17: (Continuation of Fig.2.16) Pareto curves obtained while partitioning level 1 representation from Motion Capture Rats while encoding a spike train from each neural unit. For each unit, the maximum possible information encoded (I_{max}) is measured.

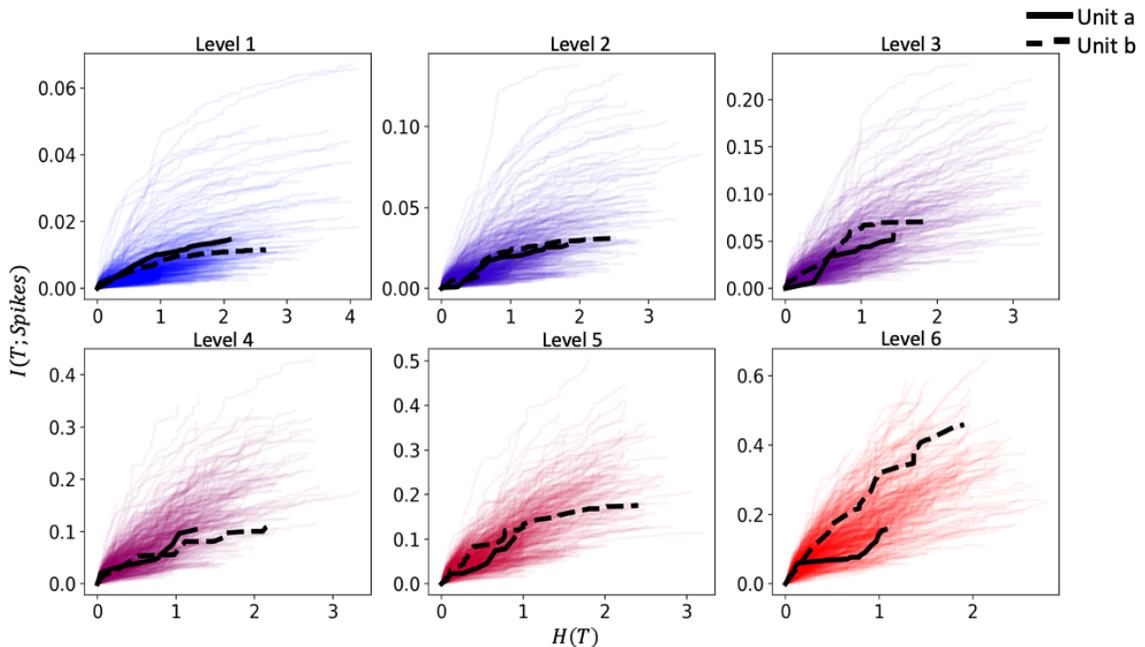


Figure 2.18: Pareto fronts curves obtained for each unit at each level of representation (Z_1 to Z_6 from Motion Capture Rats. Unit a (solid) and b (dashed) lines represent two units with statistically significant information encoding in level 1 (in case of unit a) and level 6 (for unit b).

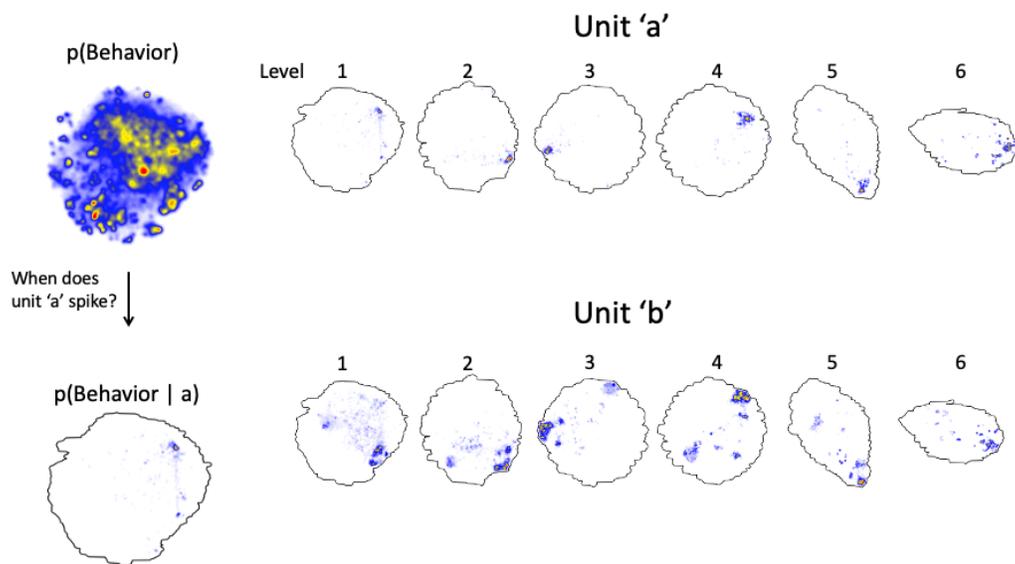


Figure 2.19: (Left) Given a behavioral representation ($p(B)$), a conditional representation is calculated for when unit a is spiking ($p(B|A)$). (Right) Conditional representation of behavior when spiking occurs in unit a (top) and unit b (bottom). Unit *a* shows a punctate representation in level 1 (thus high information encoding). Similarly, unit *b* shows a clustered representation in level 6.

RNN during training and inference.

We use this framework to train RNN models on gait data from humans, finding that emergent dynamics from these simple RNN models characterize and quantify gait signatures in able-bodied and stroke survivor participants. These gait signatures encode individual as well as group-specific information, capturing variance in gaits within and across groups that allows comparison and prediction of locomotor patterns without requiring physiological mechanistic models. These signatures also provide biomechanical interpretability and are shown to generalize to new dynamical regimes. Overall, our gait signatures provide a robust framework that can be applied to diagnose disease, develop targeted therapies, and characterize the neuromechanical mechanisms underlying locomotion.

A primary limitation here is the linear projection of the internal dynamics of the RNNs for generating gait signatures. While these projections still capture important representations that were previously unavailable, the neurophysiological mechanisms that generate gait are highly non-linear, and our low-dimensional linear projection could limit a precise characterization of gait signatures. A potential future direction here would be to constrain these dynamics to a lower-dimensional space within the model, either by regularizing the internal dynamics of the RNN or by introducing a bottleneck layer within the model.

In the second portion of this chapter, I introduced a new approach to build representations of animal behavior at multiple timescales using RNNs. Here, I use a hierarchical modeling method, where internal states from each model are used to train the subsequent model, allowing information to accumulate in these states over longer and longer timescales. I apply this approach to three distinct datasets and obtain representations that show temporal coarsening for each dataset. Using simultaneous neural recording from one of the datasets, I showcase important potential applications for building multi-timescale representations of behavior.

Building these representations, however, requires sub-sampling of hidden states from each level to train the next model. This quickly depletes the length of the data set available at each level, highly limiting the ability to train RNN models at deeper levels without overfitting. This limitation can be potentially mitigated by acquiring longer behavioral datasets (over multiple days to weeks), an approach that has been recently adopted by some research labs [59].

Chapter 3

Unsupervised learning algorithms reveal sex biases in baseline and stress adaptive behavior¹

3.1 Introduction

Stress-related disorders such as depression and anxiety continue to increase in prevalence [150, 151, 152, 153], resulting in a pressing need for more targeted treatments supported by mechanistic understanding. Pre-clinical animal models probing behaviors relevant to psychiatric symptoms are an essential foundation for mechanistic studies that could potentially generate insight into the underlying mechanisms of these disorders [154, 155]. An inherent limitation of much existing research, however, is the exclusive focus on males [156, 157, 158, 159] in most of these studies. Mandates from NIH and other funding agencies to include females in research prompted a surge in interest in studying both sexes [160, 159] and a modest increase in publications studying

¹Data presented in this chapter was collected by Jessie Muir in Bagot Lab (McGill University), and the following analyses were performed by KJ. This work will appear in K. Jain*, J. Muir*, G. Berman, and R. Bagot, "Unsupervised learning algorithms reveal sex biases in baseline and stress adaptive behavior" (to be submitted to Digital Psychiatry & Neuroscience)

female animals. However, existing animal models and behavioral metrics were developed in males and have not been validated for use with females. The approach of simply including females rests on an untested (and almost certainly erroneous) assumption that the behavioral repertoire of females is similar to males. This assumption can lead to questionable conclusions that behavior diverging from the male-defined norm indicates a diverging emotional state when it could simply reflect baseline divergences in behavioral repertoires [158]. In behavioral paradigms where sex differences have been investigated, there is ample evidence of female-specific behavioral profiles. For example, females show increased generalization in fear conditioning paradigms, as well as increased active escape behaviors [161, 162, 163, 164, 165]. To study the neural mechanisms of behavior in females as well as males, we must begin with behavioral models that capture variability in both sexes [159].

Many sex differences in stress adaptation have been documented, most notably that females are more susceptible to chronic stress than males, developing depressive and anxiety-like behaviors at times when males remain similar to the controls [166, 167, 168, 169]. These have been associated with distinct functional, structural, and molecular changes throughout the brain [170, 171, 172, 173, 167, 174, 175]. Specifically, differences in excitability in the ventral hippocampal inputs to the nucleus accumbens (vHIP-NAc), a pathway whose activity has been associated with vulnerability to stress in both sexes [169, 176, 177], has been linked to this increased vulnerability in females [168]. However, assessments of these behavioral adaptations have been benchmarked in males, and very little has been done to interrogate how females adapt to stress and how these sex differences in neuronal properties may underlie qualitatively different reactions to stress.

While sex differences in commonly used tests, such as the open field or light-dark box have been reported, a comprehensive mapping of male and female behavioral repertoires and their adaptation to stress is lacking. Emerging methods employing

machine learning algorithms to generate high-resolution behavioral maps allow for the high-throughput analysis of moment-to-moment behaviors [3, 8]. These approaches offer unprecedented insight into the rich behavioral repertoire of model organisms and have high sensitivity in revealing behavioral modulation across varying conditions [3, 8]. Here we leverage a data-driven approach to understand sex biases in behavior in a simple, widely used behavioral test of habituation and anxiety-like behavior, the open field test. We systematically interrogate how a range of stress manipulations used to model stress-induced psychopathology shape behavior in male and female mice. We identify distinct behavioral signatures enriched in females and in males, with females displaying a characteristic active behavioral profile. We find that stress upregulates opposite-sex signatures, such that stress increases male-biased behaviors in females and female-biased behaviors in males. Finally, we show that increasing activity in a neural circuit that is known to increase susceptibility to chronic stress, the vHIP-NAc drives similar upregulations in both sexes, driving upregulations in both female and male-biased behaviors. This work illustrates an important point about sex differences in behavior, showing that, even in simple behavioral tasks, males and females may adopt different strategies. As such, studying stress through an exclusively male- (or female-) centered lens risks incorrect assumptions.

3.2 Methods

3.2.1 Animals

7-week-old male and female C57BL/6J mice from Jackson Laboratories (Bar Harbor, Maine) were maintained on a 12-h light-dark cycle (lights on at 7:00AM) at 22-25°C with ad libitum access to food and water. Mice were group-housed in groups of 5 same-sex cage mates and habituated to the colony room for 1 week before the start of experiments. All experimental manipulations occurred during the light cycle, and

the testing order was counterbalanced. All experiments were conducted in accordance with the guidelines of McGill University's Comparative Medicine and Animal Resources Center and approved by the McGill Animal Care Committee.

3.2.2 Stress

Variable stress was performed as previously described in LaPlant et al [178]. Briefly, animals were subjected to either 28 days (chronic variable stress; CVS) or 6 days (subchronic variable stress; SCVS) of variable stress with one of three stressors presented in the same order across days. Stressors were: 100 random mild foot shocks (0.45mA/1s) for 1hr (administered in same-sex groups of 10); 1hr tail suspension; 1hr restraint inside a 50-mL falcon tube, with holes for air circulation inside the home cage.

3.2.3 Surgeries

Stereotaxic surgery was performed under ketamine (100 mg/kg)/xylazine (10 mg/kg) anesthesia. To achieve projection-specific DREADD expression in ventral hippocampal (vHIP) nucleus accumbens (NAc)-projecting cells, 0.5 μ l of AAV-pkg-Cre and 0.5 μ l of was infused bilaterally in the NAc (A/P: +1.3, M/L: +0.60, D/V: -4.9) and pAAV-hSyn-DIO-hM3D(Gq)-mCherry or pAAV-hSyn-DIO-mCherry in vHIP (A/P: -3.40, M/L: -3.95, D/V: -4.17 @ 12 deg angle) at a rate of 0.1 μ l per min and allowed to diffuse for 10 min before withdrawing the needle.

3.2.4 Behavioral Assessments

Open Field Test (OFT). Mice were placed in the center of an open arena (44 x 44cm; white matte acrylic) under red light. A standalone camera (FLIR Blackfly) and FLIR spinnaker software running through Bonsai recorded behavior during one 20-minute

session. A field-standard metric of anxiety-like behavior was assessed using time in the center of the open field in the first 5 minutes, calculated by video tracking software (Ethovision XT 13, Noldus). Animals were also tracked throughout the entirety of the 20-minute video using automated tracking software (DeepLabCut [5]) (see below).

3.2.5 Chemogenetic Manipulation

To inhibit the vHip-NAc pathway, male and female mice with AAV-DIO-hM4Di-mCherry or control AAV-DIO-mCherry in vHip were intraperitoneally injected with Compound 21 (3mg/kg) 30 minutes before the start of the OFT.

3.2.6 Posture Tracking

170 open-field video datasets across various experimental paradigms (Table 1) between 15 and 20 minutes in length were recorded at 30 frames per second. We used DeepLabCut (DLC) [5] (v2.1.4) to track 14 postural marks on the mice. We trained the DLC model iteratively, extracting outliers from the test set using heuristics (joint distances, lateral swapping) and retraining the model with manually labeled outlier frames. The tracked posture was transformed into an egocentric frame using one tracked point on the spine. We then rotationally aligned the postures across time to enforce rotational invariance. To accommodate the postural variances generated from morphological variations between individuals, rotationally and translationally aligned postural tracking was used to calculate 18 body angles to represent the posture of the mice. These angles were selected to maximize the representation of various postural dynamics of the animals.

3.2.7 Quantitative identification of behaviors

Autoencoder Model

To reduce the dimensionality of the postural representation an Autoencoder model was trained using the joint angle data. The joint angles were first independently centered using the median values and then scaled based on their respective interquartile ranges. The training and validation datasets were created from the scaled data. A hyperparameter grid search was performed across different architectures, layer parameters, and regularizations. The best-performing model was selected.

Behavioral Representation

Using output from the bottleneck layer of the autoencoder, the 16-dimensional postural representations were used to create a 2-D behavioral density map, following the approach in Berman et al [3]. A continuous wavelet transform was applied to the autoencoder representation using Morlet wavelets, decomposing them into 25 dyadically spaced frequencies between 0.5 Hz and 15 Hz. These wavelet amplitudes were obtained for each of the 16 dimensions, resulting in a 400-dimensional time series. This high dimensional data was then embedded into 2 dimensions using UMAP [17], a dimensionality reduction algorithm that preserves connections in local and global neighborhoods when embedding high-dimensional data into lower-dimensional embeddings. We subsequently use the PhenoGraph [179] method was used to cluster the resulting 2-dimensional data into 149 clusters. The Phenograph method first identifies clusters of nearest-neighbor networks by connecting each point to its closest neighbors, and then uses the Louvain community detection algorithm to prune these clusters, keeping only densely connected communities of points in a cluster. The method does this process iteratively to identify clusters that form densely connected communities. After clustering the data, each cluster was then manually assigned a

behavioral name by observing video segments extracted from each of the clusters.

3.2.8 Training classifiers across assays

Logistic Regression (LR) classifier models were trained on experimental assay pairs (Table 3.1). For each pair, an LR classifier was trained to generate a binary class prediction using the probability distribution across 149 Phenograph clusters from individual mice. L1 regularization was used with different regularization strengths along with other model parameters for each classifier (Table 3.2).

3.2.9 Calculating statistical upregulation between two assays

Using the set of cluster assignments from two assays, which we shall refer to as ϕ_1 and ϕ_2 , hierarchical bootstrapping [180] (random sampling with replacement both within and between individuals) was performed 1000 times to produce a distribution over $p(c|\phi_1)$ and $p(c|\phi_2)$ for each of the 149 potential cluster assignments, c . Following the procedure described in [181], we assumed that these two distributions are independent of one another and performed numerical integration to find the probability that $p(c|\phi_1) > p(c|\phi_2)$. Small probabilities imply that c is upregulated in ϕ_2 compared to ϕ_1 , and large probabilities imply the reverse. Here, we used Bonferroni corrections to account for multiple hypotheses.

3.3 Results

3.3.1 Sex biases in the exploration of an open field

To characterize male and female behavioral repertoires, we placed mice in an open field test (OFT), a standard test for probing anxiety-like behavior, and recorded for 20 minutes while the animal explored the arena. The field standard metric for quantify-

ing anxiety-like behavior in an OFT, time in center, did not reveal sex differences (Fig. 3.1D). Simple metrics incompletely sample the rich behavioral repertoire. We thus used a modified behavioral mapping approach [3] to examine the structure of mouse behavior while exploring an open field. Using DeepLabCut, we tracked 14 postural markers across 170 open-field video datasets and used tracked points to calculate 18 body angles, accounting for varying body sizes. We trained an Autoencoder model to reduce the dimensionality of normalized body angle data, generating a 16-dimensional postural representation that we used to create a 2-D behavioral density map, referred to as the behavioral space. Males and females across all experiments were embedded into a unified behavioral space to allow for comparisons between sexes and different experimental conditions. The map clustered into 149 clusters (Fig. 3.1B), each of which was manually annotated, revealing nine broad behavioral clusters (Fig. 3.1C).

Comparing baseline behavioral profiles for males and females reveals clusters of behaviors that are upregulated in females (Fig. 3.1E) and others that are upregulated in males (Figure 3.1F). Females engage in more active behaviors including running, intermittent running, and walking and rearing, while males exhibit elevated idling, still and reared nose sweeps, grooming, and rearing. We then trained logistic regression models to generate a binary class prediction using the behavioral probability distributions across these 149 clusters for each individual mouse. We find that idling and still nose sweeps were the best predictors of maleness while running, pausing and running, and walking and rearing were the best predictors of femaleness in these baseline data (Figure 3G), revealing novel sex-biased signatures of behavior.

3.3.2 Stress reverses baseline sex differences

We applied our behavioral space mapping technique to examine stress-induced behavioral change in males and females. Mice were exposed to either a subchronic or

³<https://github.com/bermanlabemory/motionmapperpy>

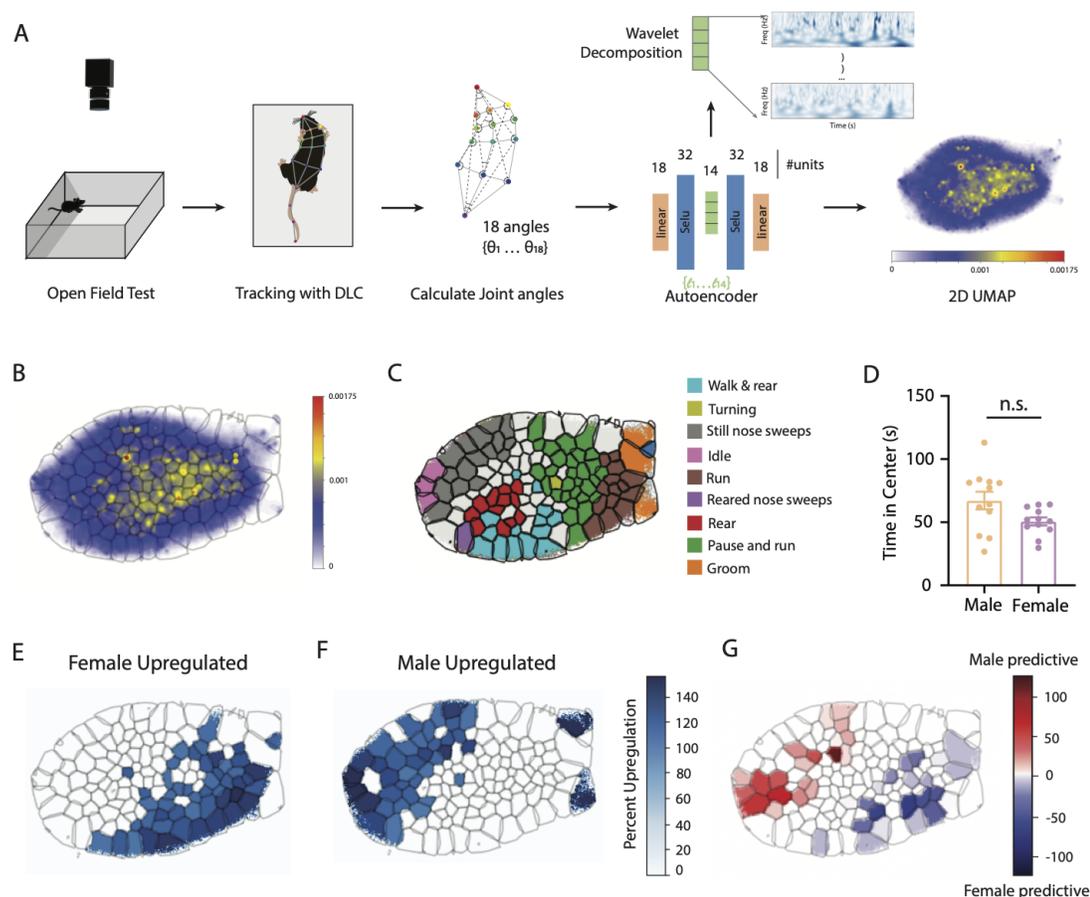


Figure 3.1: Unveiling sex-specific behaviors through unsupervised behavioral representation. A. Data from mice across all sessions was tracked using DeepLabCut and 18 joint angles were calculated from head-to-rear scaled postures of freely behaving mice. These angles were reduced to 14 dimensions via an autoencoder, and each dimension was further transformed into a spectrogram, and finally embedded into 2 dimensions UMAP using MotionMapper³, obtaining a behavioral map. B. 2-D behavioral density of all animals clustered into 149 regions using the Phenograph clustering method [179]. C. Multiple videos from each Phenograph cluster were extracted, and behaviors were manually annotated. D. No sex differences were observed in time in the center of the open field. E,F. Relative upregulation of behavioral clusters in female and male mice obtained from bootstrapping (see Methods). Female mice show active behavior upregulation (such as locomotion and rearing). Male mice exhibit less active behaviors such as idle and grooming) G. Logistic regression coefficients for pairwise classification of behavioral cluster densities from control males and females.

chronic variable stress protocol (SCVS or CVS) with daily exposure to one of three stressors (footshock, tail suspension, or restraint) for 6 or 21 days respectively (Fig. 3.2A) and were then tested in the OFT.

In females, relative to stressed animals, controls show upregulation in the frequency of female-typical behaviors, including pausing and running, walking and rearing, and running (Fig. 3.2B) and increased the frequency of male-typical behaviors such as still nose sweeps, idling, and rearing, while still exhibiting pausing and running (Figure 3.2C). Logistic regression identified still nose sweeps, idling and pausing, and running as the primary predictors of SCVS, whereas active behaviors such as run, pause and run, and walk and rear are the primary predictors of the control condition (Figure 3.2F). Exposure to an extended CVS also upregulated male-typical behaviors: still nose sweeps, idling, and rearing (Figs. 3.2D,E). However, patterns of upregulation differ from that following SCVS, notably with upregulations in walking and rearing, while, in comparison, controls show upregulations in running and intermittent running (female-typical behaviors) but also idling and still nose sweeps. Logistic regression defined the CVS condition mainly through still nose sweeps, reared nose sweeps, and the control condition as running, intermittent running, and still nose sweeps. Overall, in female mice, subchronic and chronic stress induces a more male-like behavioral profile, upregulating idle and slow behaviors compared to control animals, where active behaviors are upregulated, although CVS females still display upregulations in some female-typical behaviors.

We observed a similar phenomenon in males with stress upregulating female-typical active behaviors. As before, control males show upregulations in less active behaviors such as idling, still nose sweeps, reared nose sweeps, and grooming (Fig. 3.3A), with idling and still nose sweeps being the main predictors of the no stress condition (Fig. 3.3E). Following a 6d SCVS, males show elevated running, intermittent running, and walking and rearing (Fig. 3.3B), with running, walking and running and

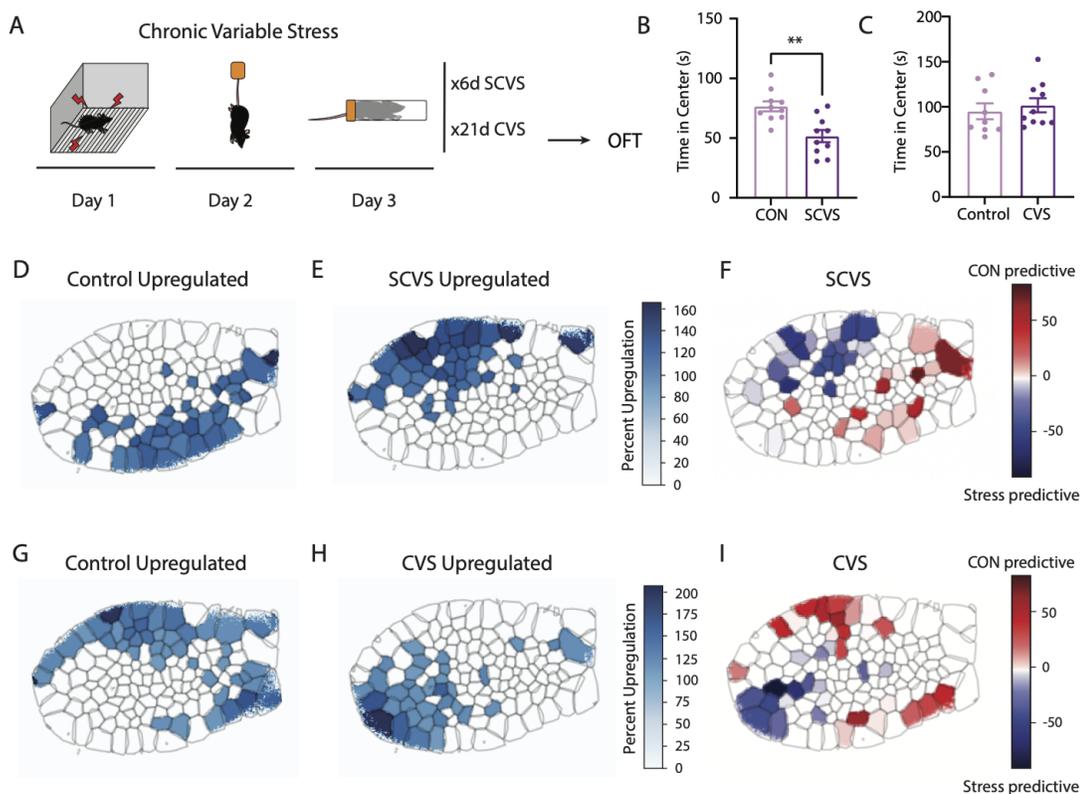


Figure 3.2: Stress alters behavioral profiles in female mice from active to inactive behaviors. A. Mice were exposed to either a chronic (21-day) or subchronic (6d) variable stress before being tested in the OFT. B,C Time spent in the center of open field, a field-standard metric for assessing response to stress. SCVS females show less time spent in the center, while CVS females show no difference compared to the control animals. D,E. Sub-chronic stressed females show upregulation of male-typical less active behaviors like still nose sweeps, idling, and rearing in female mice, whereas control animals show upregulation of active behaviors like walking and running and controls show upregulation of female-typical active behaviors. F. Logistic regression coefficients show less active behaviors like still nose sweeps, idling, and pause and runs as more predictive of sub-chronic stress, and active behaviors like run, pause and run, walk and rear are more predictive of control condition. G,H. Chronic stressed female mice also show upregulation of less active behaviors but distinctly upregulated in walking and rearing behaviors, while controls show upregulation in active as well as inactive behaviors like running, idling, and still nose sweeps. I. Logistic regression coefficients show still and reared nose sweep behaviors to have strong predictive power for CVS females, and running, intermittent running, and still nose sweep behaviors to be predictive of control females.

intermittent running being the main predictors of the stress condition (Fig. 3.3E). A 21d CVS also induces upregulation in running and intermittent running (Fig. 3.3C) compared to control while still maintaining male-typical behaviors such as still nose sweeps and rearing. Logistic regression defines CVS through intermittent running and still nose sweeps and the control condition through rearing, walking, and rearing, still nose sweeps, and intermittent running (Fig. 3.3D).

Overall, while we observe that stress modulates the frequency of distinct behavioral motifs in males and females, interpreting these changes in relation to baseline differences in behavioral repertoire reveals that, in both sexes, stress reverses sex biases in behavior. However, CVS animals still maintain sex-typical behaviors, indicating perhaps that stress slowly attenuates sex biases in behavior.

3.3.3 Chemogenetic activation of susceptibility circuit drives similar upregulations in both sexes

Our algorithm revealed sex differences in baseline exploratory behavior which reverse as mice are exposed to chronic stress, with males and females converging on similar behavioral patterns as stress continues. Stress is also accompanied by dysregulations in neural circuits which are thought to underlie these behavioral adaptations. The ventral hippocampus (vHip) projection to the nucleus accumbens (NAc) modulates stress susceptibility in both males and females, and androgen signaling suppresses vHip-NAc excitability, mediating male resilience to SCVS [176]. We hypothesized that increasing activity in this pathway would at least partially recapitulate the effects of stress on the expression of sex-biased behavioral repertoires. To test this idea, we used an intersectional viral strategy for pathway-specific inhibition, injecting retrograding AAV-cre into NAc and AAV-DIO-hM3Dq-mCherry, or AAV-DIO-mCherry into vHip (Fig. 3.4A). We then injected the DREADD actuator C21, 30 minutes prior to behavioral testing. Comparing male and female mCherry controls revealed similar

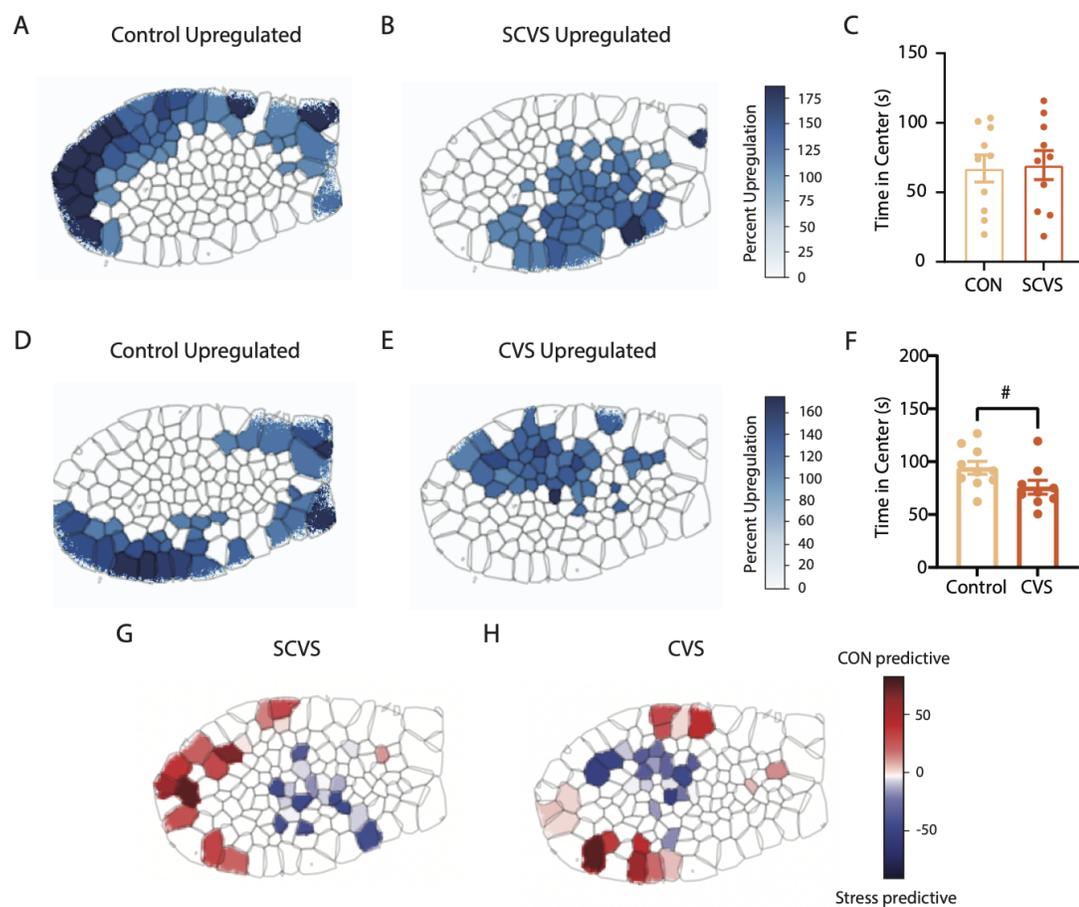


Figure 3.3: Chronic and sub-chronic stress changes male mice behaviors differentially. A. Control males show upregulation of male-typical less active behaviors like nose sweeps, idling, and grooming. B. SCVS male mice show upregulation of active behaviors like walking, intermittent running, and running. C. No SCVS-dependent differences were observed in time in the center of the open field for males. D,E. Chronically stressed male mice also show upregulation of some active behaviors like intermittent running compared to controls but retain male-typical less-active behaviors like still nose sweeps. F. CVS males spend less time in the center of OFT. G. Logistic regression coefficients show active behaviors like intermittent running and running as more predictive of SCVS males and less active behaviors like idle, nose sweeps, and rearing to be more predictive of controls. H. For CVS males, intermittent running and still nose sweeps are more predictive, whereas rearing, walk and rear, still nose sweeps, and intermittent running are more predictive of the control condition.

sex biases as in non-stressed controls (Fig. 3.4E-G), with still nose sweeps and grooming predicting maleness and running and pausing and running predicting femaleness (Fig. 3.4B). Examining the effect of vHip-NAc activation surprisingly revealed similar effects in males and females. DREADD-injected females show upregulations in pause and run, still nose sweeps and running, and are identified primarily by still nose sweeps and running using a logistic regression model (Figs. 3.4D,E). Controls, however, show scattered upregulations in several behaviors but are primarily defined by grooming and pause and run (Figs. 3.4C,E). Similarly, DREADD injected males show upregulations in still nose sweeps, idling, and running, but are identified primarily by running and still nose sweeps (Figs. 3.4F,H), while controls show elevations in rearing, walking and rearing, pause and run, and still nose sweeps, but are mainly defined by grooming (Figs. 3.4G,H). Overall, controls maintain typical sex differences in their behavioral space. In both sexes, DREADD activation of vHip-NAc upregulates the same pattern of behaviors across both sexes, and these behaviors map onto sex-typical male and female behaviors, suggesting that this manipulation partially reverses sex-biases in behavior.

3.4 Discussion

Despite the increasing inclusion of female animals in basic research, much of this work still defaults to a male-centric lens. Commonly used behavioral models have not been systematically validated in both sexes. This is despite increasing evidence of sex differences in behavior in paradigms wherein both sexes have been rigorously examined. The lack of an unbiased account of female behavior across different environments presents a major limitation to behavioral neuroscience. Here, we used an unsupervised learning algorithm to quantify sex differences in a mouse's behavioral repertoire at baseline and following chronic stress. We identify sex-specific behavioral

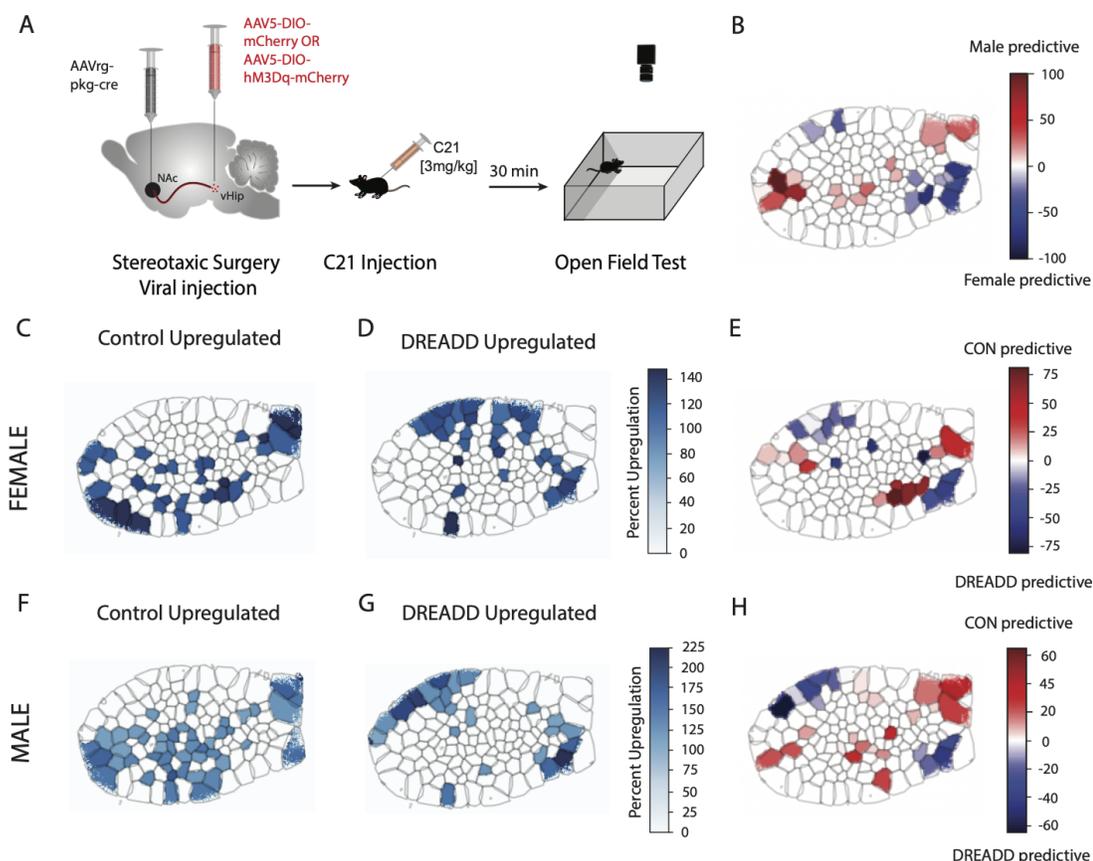


Figure 3.4: Chemogenetic activation of susceptibility circuit drives similar upregulations in both sexes. A. Mice were injected with a retrograding AAV-cre into the NAc and a cre-dependent excitatory DREADD into the vHIP to target NAc projecting vHIP cells. C21 injection was given 30 minutes before the OFT to activate the projections. B. Logistic regression coefficients for pairwise classification of behavioral cluster densities from control males and females. C. Behavioral upregulation in control females following CNO administration. D. Behavioral upregulation in DREADD-injected females post-CNO administration. E. Logistic regression coefficients showing sex-specific and treatment-dependent behavioral predictors in female subjects. Positive values indicate DREADD-associated behaviors, while negative values represent control-associated behaviors. F. Behavioral upregulation in control males following CNO administration. G. Behavioral upregulation in DREADD-injected males post-CNO administration. H. Logistic regression coefficients showing sex-specific and treatment-dependent behavioral predictors in male subjects.

motifs that emerge as animals explore an open field. Intriguingly, these differences are attenuated when animals are exposed to subchronic stress, with males displaying a more female-like profile and females a more male-like profile. Longer stress exposure also leads to males and females displaying similar behavioral profiles. Further, manipulating the vHIP-NAc, a susceptibility pathway, partially recapitulates the pattern of attenuated sex differences. These results identify unique behavioral signatures in males and females, hinting at sex-specific behavioral strategies that may drive differences in stress vulnerability.

Using an OFT to examine exploratory behavior in a novel context, we identify sex-specific behavioral motifs, whereby females exhibit more active behaviors such as running, intermittent running, and walking and rearing, while males exhibit elevated idling, still and reared nose sweeps, grooming, and rearing. Previous literature has shown that activity levels are higher in female rodents compared to males as measured in home cage activity and voluntary wheel running [182]. This effect may be mediated by estrogen signaling, as ovariectomized females show reduced activity that is rescued through estrogen supplementation [183]. Increased estradiol increases DA levels in the striatum and, as a consequence, drives elevated activity levels. This effect has been explained in terms of evolutionary pressure, with female rodents needing to be more active to seek out food sources for their offspring [184]. We confirm that females display a more active behavioral profile, but we also identify specific behavioral motifs that define the two sexes, potentially revealing sex-specific strategies in exploring a novel environment. Indeed, sex differences in behavioral strategies are well documented, specifically in threat learning with males typically being slower to discriminate threat and safety conditions as well as displaying riskier behaviors in the face of threat, while females tend to display active avoidance and typically will avoid risk in situations of threat.

In addition to baseline sex differences, we also identified changes in male and fe-

male behavioral repertoires following a 6d SCVS. Chronic stress increases anxiety- and depressive-like behavior across a series of tests. Previously, females were shown to be more vulnerable than males to chronic stress, exhibiting anxiety and depressive-like behavior following a 6d-sub-chronic stress, whereas behavioral differences were not observed in males until 21 days of repeated stress exposure. Indeed, using traditional metrics for anxiety and depressive-like behavior, we observe the same phenomenon. However, using semi-automated pose tracking combined with unsupervised learning algorithms to quantify an animal's adaptation to stress in a data-driven manner, we identify behavioral adaptation in both sexes following a 6d SCVS. We see females exhibiting a more male-typical profile, with increases in still nose sweeps, pausing and running, and idling, while males adopt a more female-typical profile with upregulations in more active behaviors such as running, pausing and running, and walking and rearing. This finding points to a potential limitation of the standard approach of evaluating the effects of stress manipulations using a limited set of metrics that have only been validated in males. There is ample evidence to suggest that under similar conditions, females and males exhibit divergent behavior. Fear conditioning presents a prominent example; males typically show freezing behavior when presented with a cue paired with foot-shock. Experiments assessing freezing as a metric of learning in females concluded that females were slower to learn cue-shock associations. However, work by Gruene et al[165] revealed that female rats are learning similarly to male rats but they express this learning through distinct behaviors. A careful analysis of behavioral responses showed that female rats are more likely to exhibit darting behavior to shock-associated stimuli. Failure to account for sex differences in behavioral expression has led to erroneously characterizing females as poor learners when, in fact, they simply express this learning through active, rather than passive responses. Similarly, our analyses demonstrate that male and female mice exhibit distinct behaviors while exploring an open-field test that should be taken into account when interpreting the

effects of experimental manipulations.

Here, through highly resolved behavioral analysis, we identified the effects of experimental manipulations that are missed in analyses relying on the field-standard time in center metric. For example, we observed that, while SCVS reduced time in center in females, there was no effect in males, which is consistent with the established interpretation that male mice are insensitive to shorter periods of stress and that longer duration stress exposure is necessary to produce effects in males. Our analyses show that SCVS already induces considerable behavioral adaptation in male mice, showing that both sexes are adapting their behavioral repertoire to SCVS. Similarly, manipulations of the vHIP-NAc pathway have been shown to have no impact on depressive or anxiety-like behavior in stress-naive mice. However, we found that, in both males and females, excitation of this pathway drives similar changes, including upregulations in still nose sweeps and pausing and running, examples of female and male-typical behaviors. Individual differences in vHip-NAc activity measured during an OFT are predictive of adaptation to future stress, with higher vHip-NAc activity associating with increased stress vulnerability in both sexes. Our analysis suggests that this increased excitability may induce some of the behavioral profiles of susceptibility, without inducing the full behavioral adaptation of chronic stress exposure. Females show upregulations in still nose sweeps, which is a male-typical behavior, while still maintaining elevated pausing and running, a female-typical behavior; the same holds true in males as well. Similar to the effects of stress, increasing the vHip-NAc activity appears to reverse baseline sex differences although without inducing the extent of the shift that is associated with stress. This finding again illustrates the potential of highly resolved behavioral analysis for capturing subtle behavioral changes that are missed with conventional metrics.

While established metrics can offer ease of interpretability and simplicity for high-throughput experiments, they necessarily obscure the intricacies of an animal's be-

havior. Here, we identify fundamental changes in behavioral motifs in circumstances where the conventional metric finds no effect, indicating that time in the center does not fully describe the effects of stress on behavior in an open field. For this reason, more highly resolved behavioral analyses are becoming increasingly common in the field, with new supervised and unsupervised algorithms allowing for more thorough and precise investigation of behavior. This presents both opportunities and challenges for researchers. Comprehensively profiling behavioral repertoires can support stronger conclusions when determining if the manipulation has any effect or if measuring the extent to which two groups differ. For example, a similar approach has been used to quantify individual variability in female behavior across the estrous cycle and between male and female animals. The method revealed that variability between female mice is greater than variability within female mice attributable to the estrous cycle, and that, in general, inter-individual variability in males exceeds that in females. This finding provides important evidence to counter a contentious claim in the field that cycling females are inherently more variable, an argument that has long been used to justify the exclusive use of male animals. However, these approaches also raise new questions. Here we reveal novel stress-induced modulation in male mice after SCVS, suggesting they are not entirely insensitive to shorter stress. This finding suggests that more work is needed to fully characterize sex differences in behavioral adaptation. For example, using highly resolved behavioral analysis to probe the effects of stress sampling across increasingly chronic stress in male and female animals may reveal differences in the trajectories of stress adaptation to inform a more accurate understanding of sex differences in susceptibility. Further, there is a question of the interpretability of behavioral motifs, requiring thoughtful consideration of the relevance of specific patterns of behaviors within the experimental context and potentially further experimentation. Overall, our model provides a thorough investigation of behavior in both male and female mice, revealing sex differences in baseline

behaviors and exploration strategies, as well as surprising similarities in responses to subchronic stress. This work makes an important point about not only the inclusion of, but also the proper investigation of, females in behavioral studies.

Table 3.1: **Experimental design.** Table showing different experimental groups, stress conditions for each experiment, and the number of animals from each sex in each experiment.

Experiment	Conditions	Males	Females
Control	CON	12	12
DREADD	DR	9	9
	CON	6	10
SCVS	SCVS	10	10
	CON	10	10
CVS	CON	10	10
	CVS	10	10

Table 3.2: **Logistic regression parameters.** Logistic regression parameters used to train pairwise classifiers between class 1 and class 2. All classifiers shared the following parameter values other than L1 strength: penalty=l1, dual=False, tol=0.0001, fit intercept=True, intercept scaling=1, class weight=balanced, solver=saga.

Class 1	Class 2	L1 Penalty Strength	Classifier Accuracy
CON CON M	CON CON F	300	1.00
CVS CON M	CVS CVS M	300	0.90
CVS CON F	CVS CVS F	300	1.00
SCVS CON M	SCVS SCVS M	300	0.95
SCVS CON F	SCVS SCVS F	300	1.00
CSDS CON M	CSDS CON F	300	1.00
DREADD CON M	DREADD CON F	600	0.94
CVS CON M	CVS CON F	300	1.00
SCVS CON M	SCVS CON F	300	0.95
DREAD CON M	DREAD DR M	600	0.93
DREAD CON F	DREAD DR F	600	0.89
CVS CVS M	CVS CVS F	400	0.95
SCVS SCVS M	SCVS SCVS F	600	0.95
DREADD DR M	DREADD DR F	600	0.89

Chapter 4

State-space reconstruction of dynamical systems with long timescale non-stationarities

4.1 Introduction

In section 1.4, we discussed a new approach[11] where authors use delay embedding to reconstruct the state-space of the dynamical system underlying partially observed data. Specifically, the authors assume there is a dynamical system $\dot{\vec{x}}(t) = f(\vec{x}(t))$ whose phase variable remains observable, but we can only measure some noisy non-linear projections of these variables, $\vec{y} = M(\vec{x}(t))$. They then use an ensemble dynamics-inspired approach to find long-timescale dynamics within the system. While the authors show that this method is effective at finding these dynamics for a variety of systems, none of these systems possess non-stationarities in their dynamics with large gaps in timescales. In this Chapter, I will investigate how well this approach is able to perform on systems of this type – which are common across fields like animal behavior and neuroscience – finding that if the gap between system timescales be-

comes too large, the method no longer is able to find the longest timescale structures. In addition, I will demonstrate a novel method that is able to capture these longest timescale dynamics through the use of recurrent neural networks.

4.2 The Lorenz System

The initial study demonstrating the method[11] uses the Lorenz system as an example of the types of complicated dynamics that it can uncover. The Lorenz system exhibits 3-dimensional dynamics that are governed by the following differential equations:

$$\frac{dx}{dt} = \sigma(y - x) \quad (4.1)$$

$$\frac{dy}{dt} = x(\rho - z) - y \quad (4.2)$$

$$\frac{dz}{dt} = xy - \delta z \quad (4.3)$$

If $\sigma = 10$, $\rho = 28$, and $\delta = 8/3$, the system exhibits chaotic behavior and results in two unstable attractor states (Fig.4.1A). Here, they assume partial observations of this system, generating simulated trajectories of $\bar{x}(t)$ at $\delta t = 0.01s$ for $T = 5 \times 10^5s$. To reconstruct the state-space of the system, the authors used delay embedding at varying delays (K) and partitioned the delay-embedded space into N clusters to obtain a discrete representation of the dynamics that are maximally Markovian. Thus, they found K and N by maximizing the short-term predictability of the partitioned representation, which can be quantified by the entropy production rate [185] as follows:

$$h_{\delta t}(N, K) = -\frac{1}{\delta t} \sum_{i,j}^N \pi_i P_{ij}(\delta t) \log P_{ij}(\delta t), \quad (4.4)$$

where $P_{ij}(\delta t) = p(s_j(t + \delta t) | s_i(t))$ is a row-stochastic Markov chain and π is the stationary distribution of P . Increasing number of partitions, N , highlights finer struc-

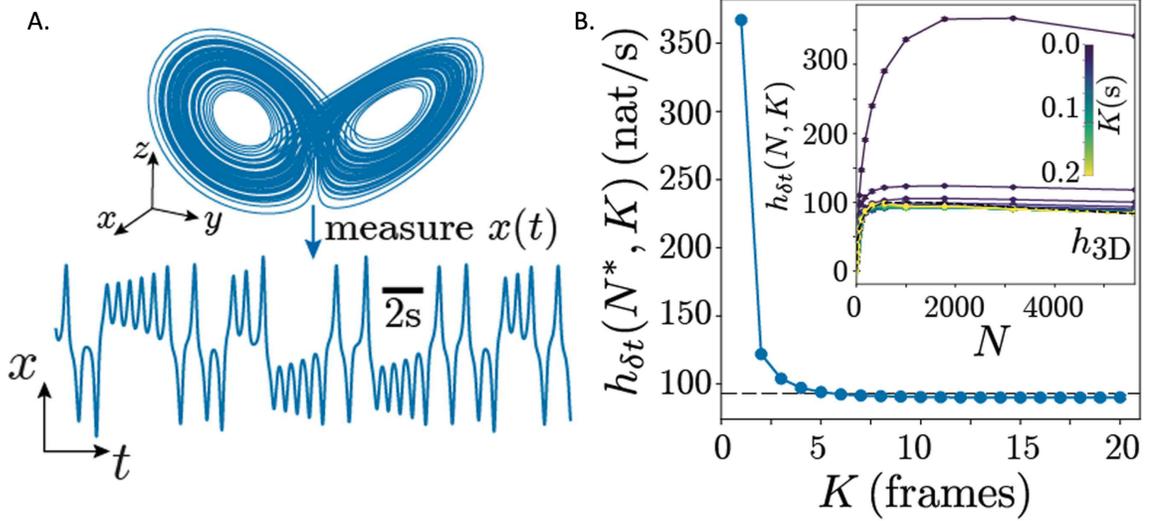


Figure 4.1: Reconstructing the state-space of Lorenz attractor. A) 2D projection of the standard Lorenz system with $\sigma = 10$, $\rho = 28$ and $\delta = 8/3$, showing 2 attractor states. Only $x(t)$ is measured from the system (partial observation). B) (Inset) Entropy rate ($h(N, K)$) as a function of N (number of partitions) and K (delay length) of transition dynamics in the reconstructed state-space. $h(N, K)$ saturates after $N^* = 3162$, indicating maximum predictability. The entropy rate stops changing after $K^*=12$, indicating saturation of predictive information. Figure taken from [11].

tures in the reconstructed space, and for stochastic processes where information is not bounded by the scale of the system, the entropy rate, $h_{\delta t}(N, K)$, is expected to increase monotonically. However, for chaotic systems, the fractal structure imparts deterministic dynamics, and $h_{\delta t}(N, K)$ will saturate after a certain N . In practice, however, finite size effects affect the estimation of $P_{ij}(\delta t)$, which results in an underestimation of the entropy rate when N is sufficiently large. Thus, to find a maximally Markovian structure in this system, N^* is selected where the rate of entropy production stops increasing (Fig. 4.1B inset). Similarly, K is gradually increased until reaching a point where $h_{\delta t}(N^*, K)$ stops changing, indicating a saturation of predictive information or memory in the system (Fig. 4.1B inset). The authors select $N^* = 3162$ and $K^* = 12$ for the Lorenz system.

After choosing N^* and K^* , the transition matrix P_{ij}^{τ} is estimated by counting the number of transitions between the partitions in the discretized state-space. Then, the

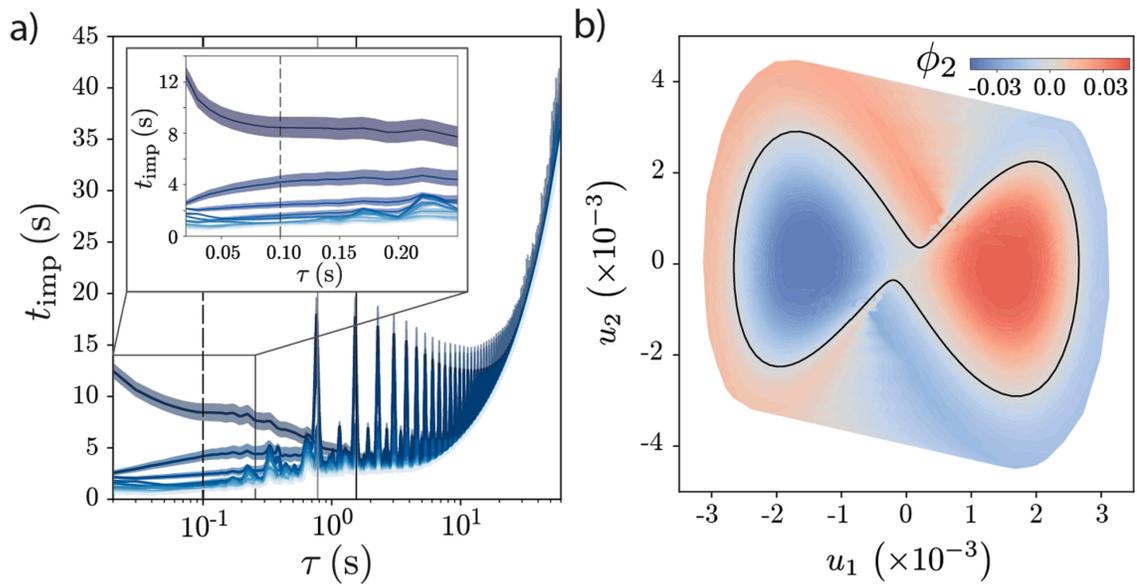


Figure 4.2: a) Relaxation timescales vs. transition time (τ) of the inferred transition operator in delay embedded state-space of the Lorenz system. Transient dynamics (within a partition) contribute to the initial peak in relaxation times after which they settle to near-constant value. Beyond this regime, the relaxation times are driven by the quasi-periodicity of the Lorenz attractor. B) The first non-trivial eigenvector ϕ_2 projected onto the first 2 SVD modes of the delay-embedded space. The phase space is split into two metastable states by ϕ_2 . Figure taken from [11].

authors choose to time-symmetrize the transition matrix $P_{ij}^{\tau^*}(r) = \frac{P_{ij}^{\tau^*} + P_{ij}^{\tau^*\dagger}}{2}$ where $P_{ij}^{\tau^*\dagger}$ is the dual operator of $P_{ij}^{\tau^*}$. The eigenspectrum of $P_{ij}^{\tau^*}(r)$ then reveals the directions of long-lived dynamics as well as the relaxation times associated with each direction. The chosen discretization of state-space maximizes predictability while ignoring finer-scale structure within each partition. Because of this, the eigenspectrum of $P_{ij}^{\tau^*}(r)$ at different values of τ reveals a spectral gap after faster dynamics are settled (Fig.4.2A), and slower dynamics emerge, and relaxation times become constant to show the underlying Markovian nature of the dynamics (Fig.4.2A inset). In this way, τ^* is chosen to capture the important long-term dynamics of the reconstructed state-space. The implied relaxation times corresponding to each eigenvalue of $P_{ij}^{\tau^*}(r)$ are then as follows:

$$t_i^{imp}(\tau^*) = \frac{-\tau^*}{\log \lambda_i(\tau)} \quad (4.5)$$

The eigenvector ϕ_2 corresponding to the longest timescale $t_2^{imp}(\tau^* = 0.1)$ 8sec is 3162-dimensional, equal to the number of partitions chosen in the delay-embedded spaces. To visualize this eigenvector, trajectories in the reconstructed state-space are decomposed using SVD (singular value decomposition). The partition centers in the delay-embedded space are then mapped onto the obtained singular vectors, and a contour plot of ϕ_2 onto the first two singular vectors is shown in (Fig.4.2A). This plot shows ϕ_2 dividing the state-space into its almost invariant sets, partially split along the shortest unstable periodic orbit. It is important to note that the unstable periodic orbit for this Lorenz system is 1.3 seconds, and the identified long-timescale dynamics are predictable for 3.04 Lyapunov times.

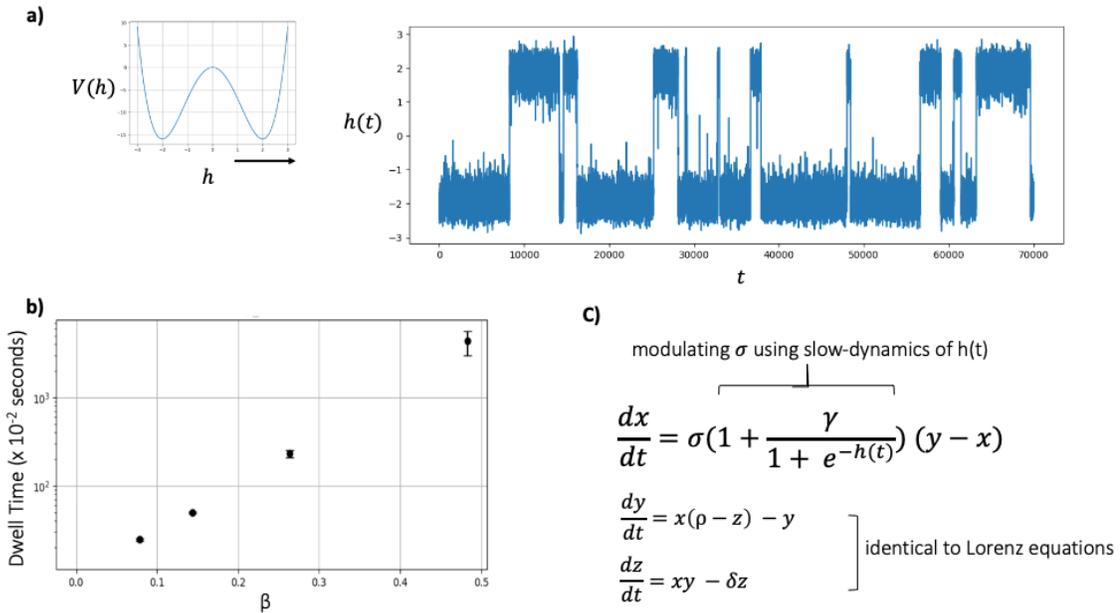


Figure 4.3: Modified Lorenz attractor. a) Particle in a double well potential (left) is simulated and positions ($h(t)$) are extracted (right). b) Mean escape time of the particle as a function of inverse temperature β . c) Modified Lorenz attractor equations. Only \vec{x} is modified, where parameter σ is scaled based on the position of the particle in double well potential. γ determines the strength of scaling, $\sigma = 10$, $\rho = 28$ and $\delta = 8/3$ are Lorenz attractor parameters.

4.3 Modified Lorenz System

The Lorenz system, however, has only a single long timescale – the timescale of switching between the two unstable periodic orbits. To determine whether the ensemble approach described above works on a system with an additional, potentially much longer, timescale, we apply this approach to a modified Lorenz system that is driven by a function, $h(t)$:

$$\frac{dx}{dt} = \sigma \left(1 + \frac{\gamma}{1 + e^{-h(t)}} \right) (y - x) \quad (4.6)$$

$$\frac{dy}{dt} = x(\rho - z) - y \quad (4.7)$$

$$\frac{dz}{dt} = xy - \delta z. \quad (4.8)$$

Here, $h(t)$ is generated by simulating a particle in a 1-dimensional double-well potential at thermal equilibrium $V(h) = ah^4 - bh^2$ (Fig. 4.3a), and $\beta = 1/(k_B T)$ defines the inverse temperature of the system. The trajectory of the particle hopping between the two wells is shown in Figure 4.3 a (right). As the position $h(t)$ of the particle moves between the left and right well, it drives scaling of the σ parameter, and the strength of this scaling depends on γ . This allows us to modulate the σ parameter of the Lorenz system. We set the inverse temperature β in a range where the mean dwell time, $\langle\tau^w\rangle$, of the particle in a well changes linearly between 3.8s and 31s, which is much longer than the timescale of the switching between periodic orbits in the Lorenz attractor (~ 1.4 s). This setup allows us to introduce a non-stationarity in the Lorenz system without greatly altering the underlying state-space. Figure 4.4) shows trajectories of the modified Lorenz system for various values of β and γ . In this figure, the color of the trajectory is blue when the particle is in the left well and red when the particle is in the right well.

We apply the ensemble method for state-space reconstruction to the modified Lorenz system, selecting K^* and N^* as described in the previous section. As an example, the entropy production rate for the system with $\gamma = 1$ and $\beta = 0.40$ is shown in Fig. 4.5. We note that the entropy rate does not become constant in this case, as we do not encounter finite-size effects because of the longer simulation times we chose to perform.

Following the previously stated analyses, we construct the transition matrix and calculate the eigenvectors to find metastable states for each system. We visualize these eigenvectors for $\beta = 0.40$ ($\langle\tau^w\rangle \approx 14$ s) in Fig. 4.7. Qualitatively, we observe that the second leading eigenvector ϕ_2 captures metastable states of the Lorenz attractor for values of $\gamma \leq 1.00$, and for $\gamma > 1$, coupling with the double well dynamics modulates σ by a larger factor, and the metastable states of Lorenz attractor are not detected by ϕ_2 . The corresponding characteristic timescales for these plots are shown in Fig.

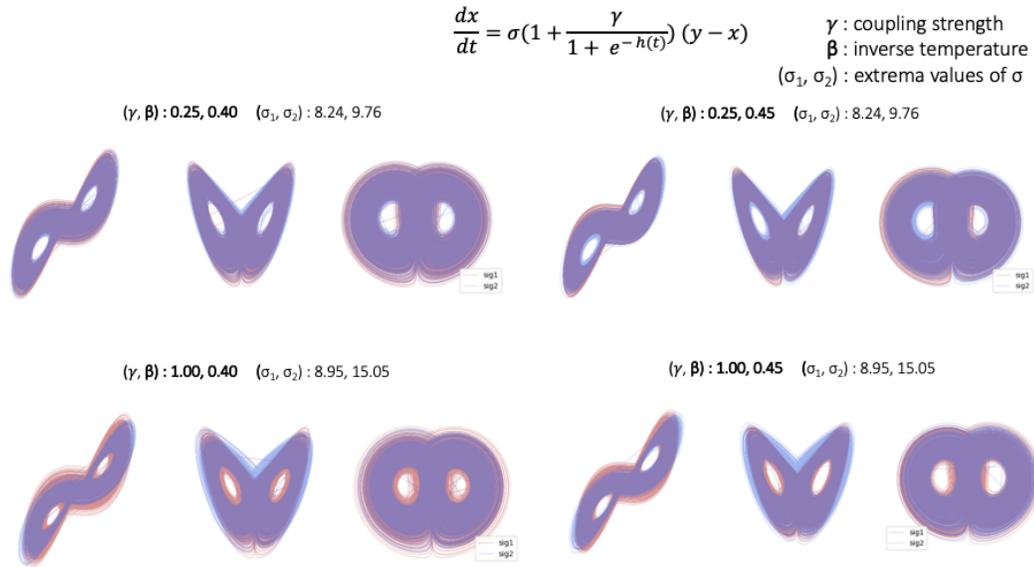


Figure 4.4: Sample trajectories from modified Lorenz system for 4 pairs of (γ, β) values, where γ determines the extent of scaling of σ parameter and β is the inverse temperature for the particle in the double well. As the particle shifts from left to right well, sigma values modulate between σ_1 and σ_2 values (calculated at the center of the wells). Trajectories are colored red when the particle is in the left well and blue for the right well.

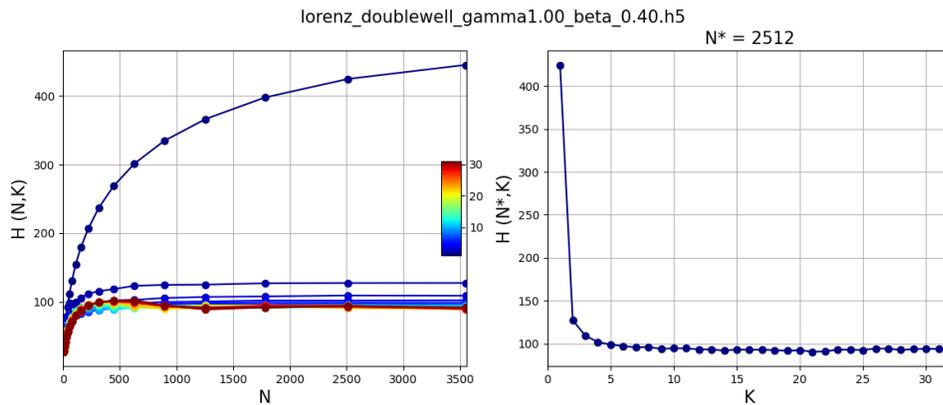


Figure 4.5: Entropy rate for the modified Lorenz system with $(\gamma, \beta) = (1, 0.40)$. Here we choose $N^* = 2512$ and $K^* = 6$.

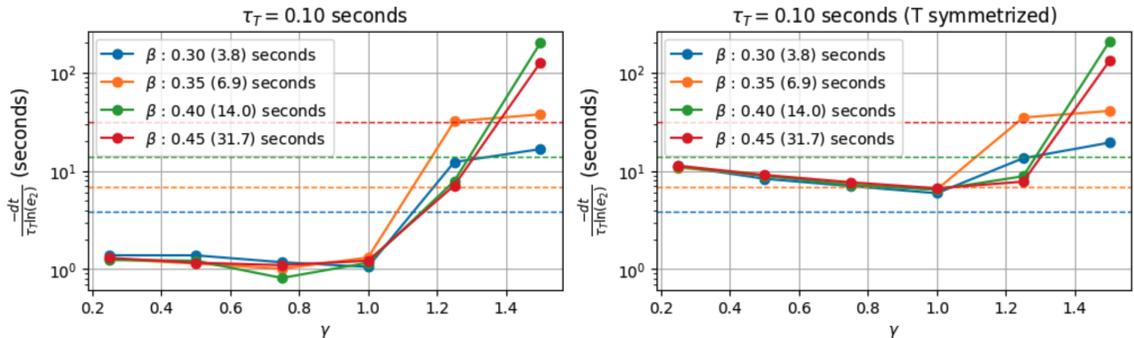


Figure 4.6: Characteristic timescales for the system with various values of β and γ with (right) and without (left) making the transition matrix reversible at a transition timescale of $\tau^*=0.10$ s.

4.6.

To measure whether ϕ_2 encodes information about the position of the particle in the wells, I calculate the Mutual Information between the position of the driving particle (left or right well) and ϕ_2 (Fig.4.8 right). The mutual information is very low compared to the maximal value of 1 bit for all pairs of values of β and γ . We measure a low MI value between the partitioned delay-embedded space and ϕ_2 (Fig.4.8 left), as well as between the delay-embedded space and the position in the well.

Here, the Lorenz system is modified in a way where the dynamics are modulated at timescales at the same order of magnitude as the dynamics of the Lorenz system for $\beta = (0.30, 0.35)$; Fig.4.3b). These modulations occur at much longer timescales $\beta = (0.30, 0.35)$. We attempt to reconstruct the state-spaces of these modified systems by finding maximally predicted delay-embedded spaces. However, we find the ensemble-based delay embedding approach fails to identify these timescales, even when non-stationarities are present at timescales equivalent to the timescale of the Lorenz system. This is evident from the low predictive power of the double well dynamics by the leading non-trivial eigenvector. In the next section, we address this limitation, proposing a new framework to reconstruct state-space dynamics using recurrent neural networks.

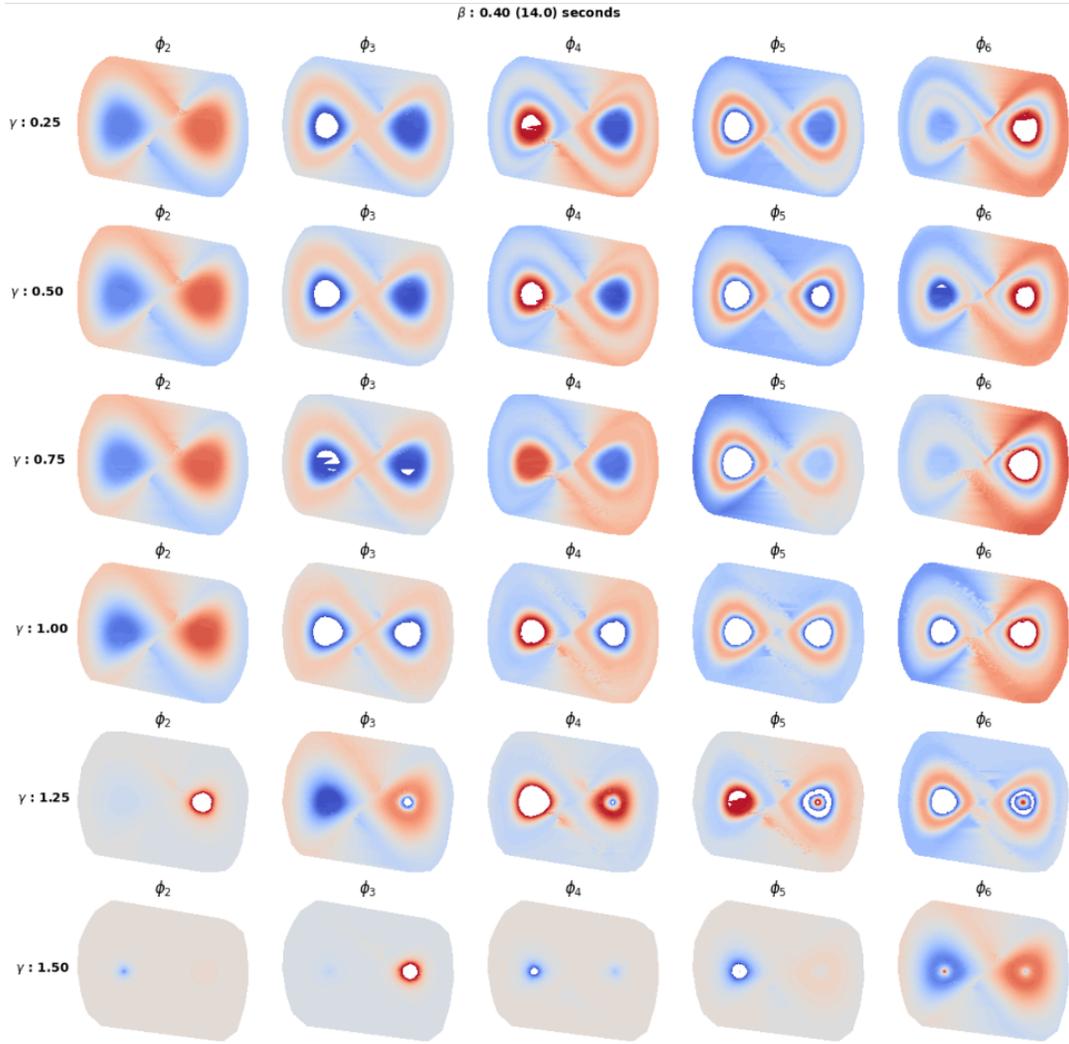


Figure 4.7: First five non-trivial eigenvectors of the delay-embedded modified Lorenz system (similar to Fig.4.2B) for various values of γ and $\beta = 0.40$.

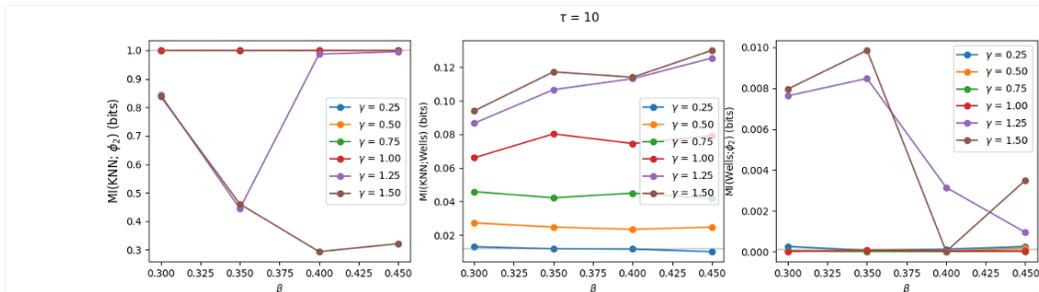


Figure 4.8: Mutual Information between the leading non-trivial eigenvector ϕ_2 and partitioned delay embedded space (KNN)(left), between the position of the particle in left or right well and KNN (center), and between the position of the particle in the well and ϕ_2 .

Encoder-Decoder RNNs

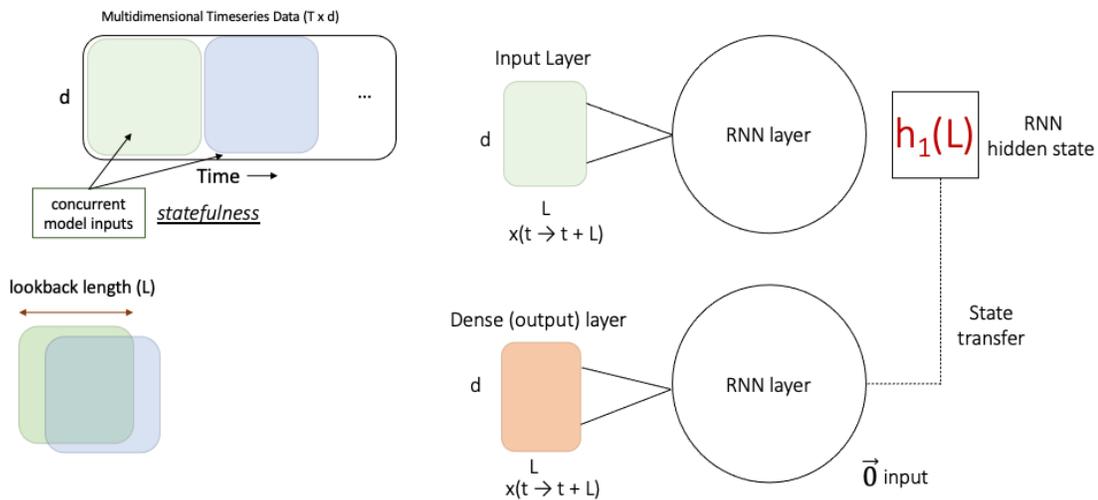


Figure 4.9: Encoder-Decoder RNN Model. Temporal segments of length (L) are used to drive an encoder Recurrent Neural Network model, generating a final hidden state $h_1(L)$. A decoder RNN model uses $h_1(L)$ as its initial state and is driven without any input to reconstruct the temporal segment input into the encoder RNN. Concurrent temporal segments are used during training for statefulness.

4.4 Reconstructing state-space dynamics using RNNs

To identify the hidden long-timescale driving underlying dynamical system, a novel approach using Encoder-Decoder Recurrent Neural Networks is introduced in this section. Here, a section of partially observed dynamics (of length T_{RNN}) is used to drive an encoder RNN model, while the hidden state representation of the network evolves statefully (see section 2.3), encoding information from the input in this representation. The final hidden state from the encoder RNN is then transferred to a decoder RNN and is utilized as the initial state for the decoder RNN as the decoder RNN runs forward autonomously (with $\vec{0}$ input) for T_{RNN} time steps. These time steps are projected back into reconstructing the input to the encoder RNN. The model is trained statefully, utilizing concurrent sections of time series data during training. Separate models are trained for each pair of β and γ pair.

After the model is trained, trajectories are fed through the encoder model, and the hidden states computed by the model are extracted. To visualize the structure of these hidden states, UMAP (Uniform Manifold Approximation and Projection)[17] is used to project these into a 3-dimensional space. These projections are shown in Fig. 4.10 for three different values of γ for $\beta = 0.40$, and colored to reflect the position of the driving particle in left (red) and right (well). A separation in the hidden state-space correlated with the identity of the wells is observed for $\gamma \geq 1$.

We train another encoder-decoder RNN model to encourage encoding of longer timescales in the hidden states (Fig. 4.11), calling it the level 2 model.

The representations obtained by the level 2 model are shown in Fig. 4.12. Here, a larger separation in the encoded representation is observed, indicating more predictive information about the driving system.

4.5 Conclusion

This chapter explores the limitations of the ensemble dynamics-inspired approach for state-space reconstruction when applied to systems with underlying non-stationarities at large time scales. This approach fails to identify dynamical timescales introduced by the double-well potential underlying the modulated dynamics of a Lorenz system. The ensemble approach, while adept at capturing the metastable states of the Lorenz attractor for modest values of the coupling parameter ($\gamma \leq 1.00$), fails to detect the longer timescale dynamics introduced by the double-well potential. This limitation is evident from the low mutual information between the leading non-trivial eigenvector ϕ_2 and the position of the driving particle, even for cases where the timescales of the non-stationarities equal those of the Lorenz system's intrinsic dynamics.

In subsequent sections, a novel framework utilizing Encoder-Decoder Recurrent Neural Networks is implemented to address these limitations. The hidden state dy-

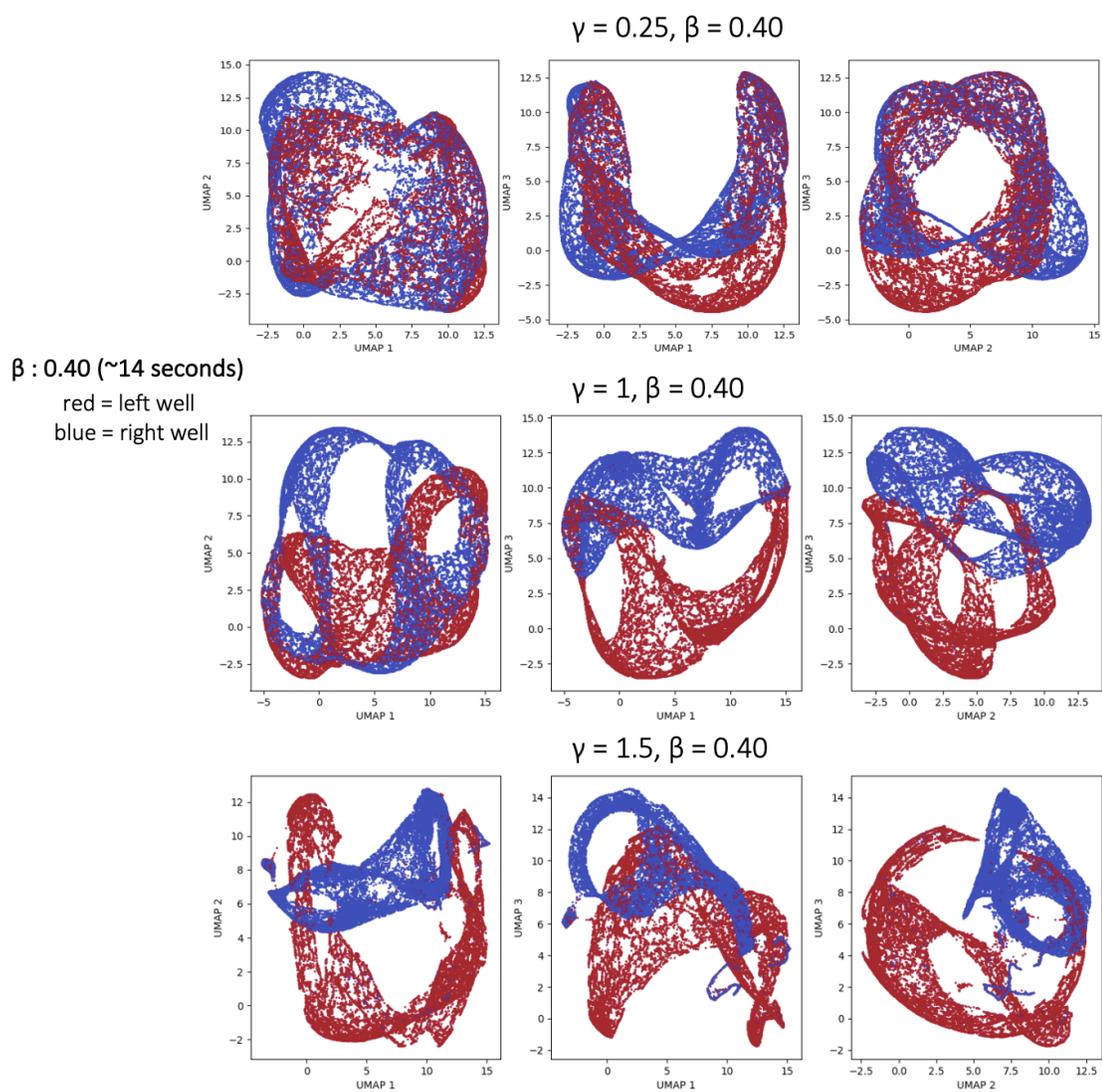


Figure 4.10: 3D UMAP[17] projections of hidden states extracted from trained encoder model in Figure 4.9. Here 3 different models were trained for each (γ, β) pair and hidden states were extracted as each model is driven with time series from each system.

LEVEL 2 Model

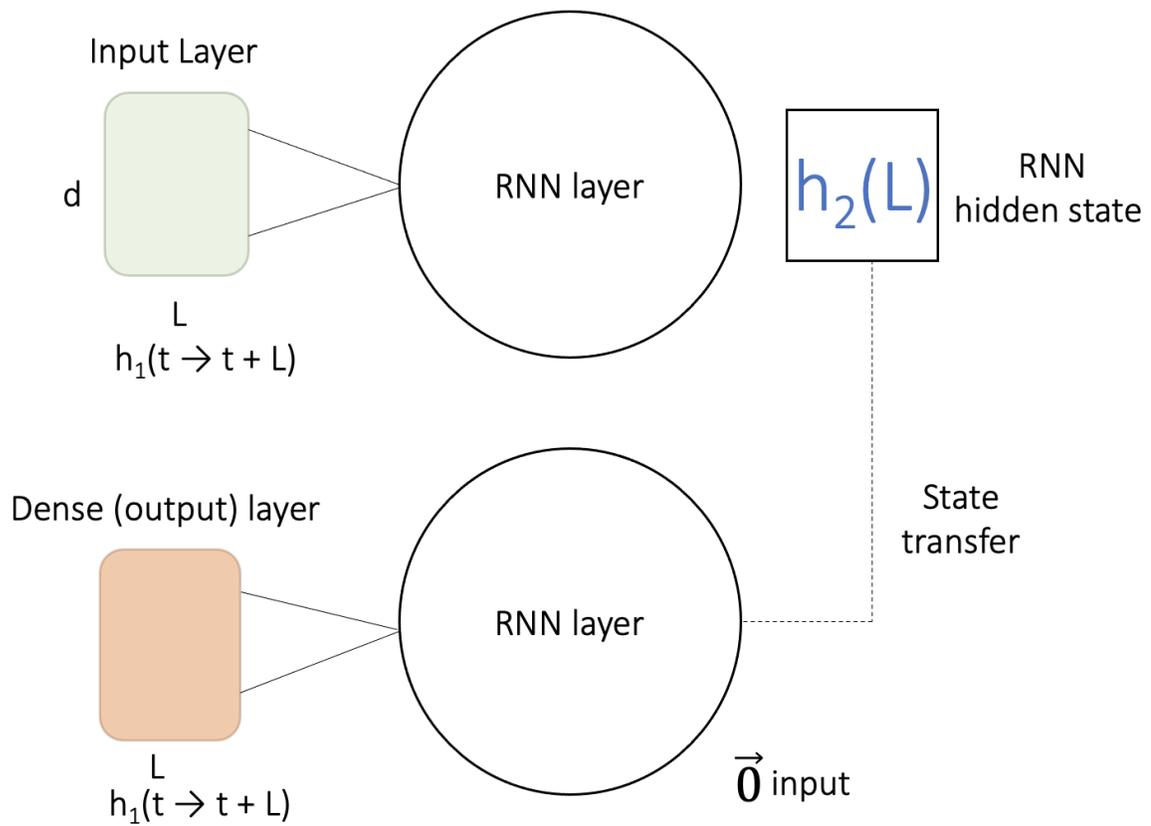


Figure 4.11: Level 2 model. We train another encoder-decoder RNN model (inspired from Section 2.5.3 using hidden states extracted from the previous model. We follow the same training style as before, training the model statefully. After training, hidden states $h_2(t)$ were extracted using $h_1(t)$.

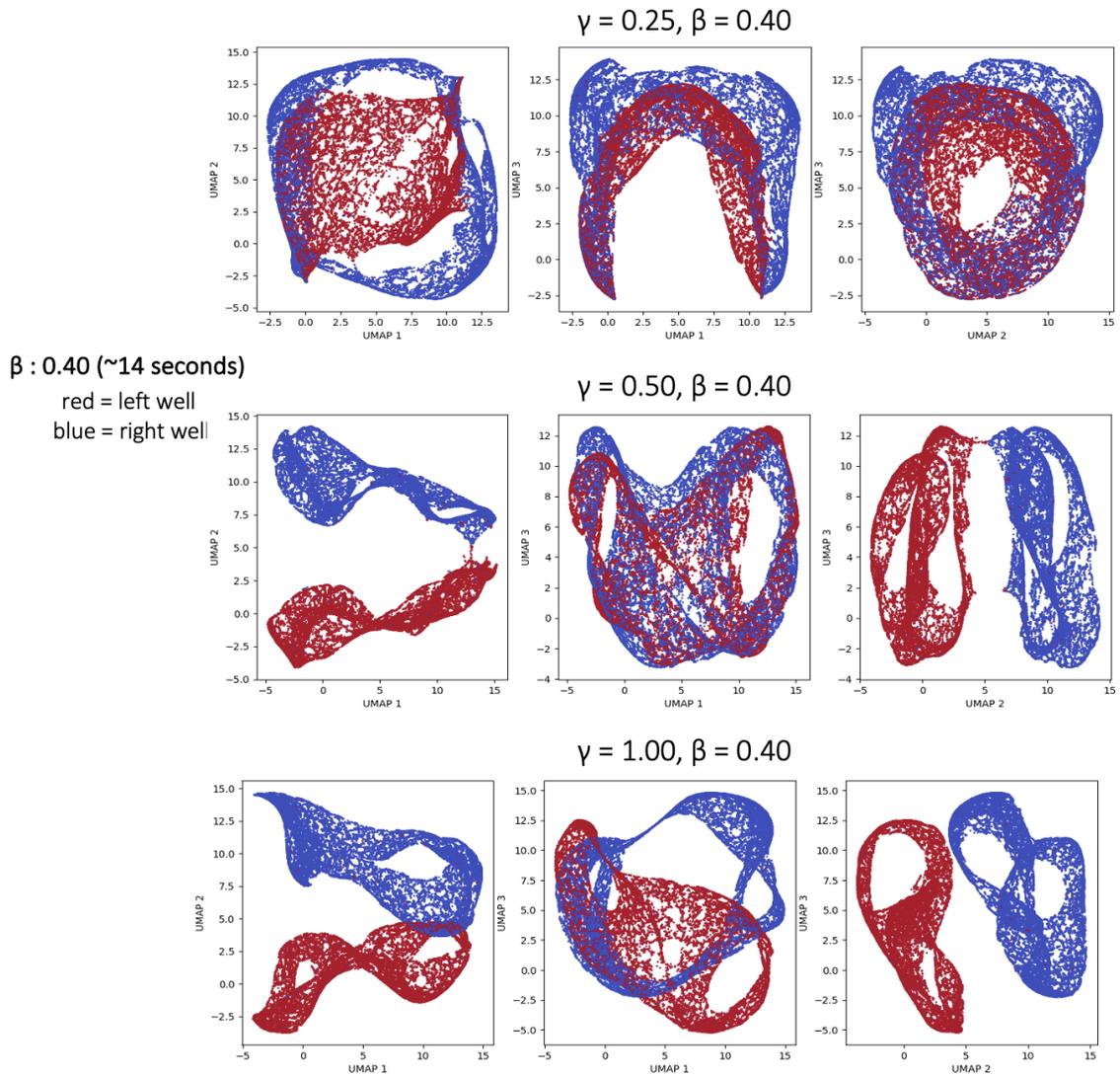


Figure 4.12: 3D UMAP[17] projections of hidden states extracted from level 2 encoder model from Figure 4.11. Here 3 different models were trained for each (γ, β) pair using corresponding $h_1(t)$ and hidden states $h_2(t)$ were extracted as each model is driven with $h_1(t)$.

namics of the trained encoder model show superior capability in identifying and encoding the hidden long-timescale dynamics, as evident from the physical separation emerging in hidden state-space when the particle is in different wells. Training a level 2 model hierarchically is shown to amplify this separation, indicating an improvement in capturing the latent non-stationary timescales.

Looking ahead, these frameworks can be applied to behavioral time series data, where non-stationarities are ubiquitous and often occur across multiple timescales. This approach could potentially disentangle the hierarchical organization of behavioral patterns and could help characterize long-timescale physiological states and understand neural dynamics underlying complex behaviors and the evolution of behavior over long timescales such as aging.

Chapter 5

Discussion

In this thesis, I focus on developing computational methods for identifying and characterizing long-timescale dynamics in behavioral data. In particular, I use Recurrent Neural Networks as a tool to approximate dynamical systems from vast amounts of data, using these overparameterized models to learn about the dynamical timescales that underlie large behavioral datasets. I focus on training RNNs while maintaining temporal coherence through statefulness to closely resemble a dynamical system. This allows RNNs to capture and propagate information over long timescales and generate interpretable hidden states, as found in this thesis.

While I focus specifically on using RNNs in this thesis, several deep learning models have been developed to emulate the behavior of dynamical systems. A notable class of models is Neural Ordinary Differential Equations (NeuralODEs), which parameterize the derivative of the hidden state using a feedforward neural network and learn *continuous* time dynamics in the latent space. When compared to RNNs, these architectures are computationally efficient, more interpretable, and do not encounter gradient instabilities of the kind seen in RNNs. However, the (current) Achilles heel for NeuralODEs is the difficulty in training them, as operating in continuous time requires the numerical integration of a vector field using ODE solvers, making these

models harder to optimize because of numerical instabilities. Their continuous time design also creates a more convoluted loss landscape, leading to highly sensitive hyperparameters and gradient instabilities. However, optimizing NeuralODEs remains a highly active area of research, and potential future solutions that address these limitations could accelerate the use of Neural ODEs for multi-timescale modeling of dynamics.

Where does this work go from here? In Chapter 2, I highlighted potential clinical applications of gait signatures for diagnosis and therapy development. Here, we decomposed RNN dynamics linearly using PCA and embedded these into a low-dimensional description using multi-dimensional scaling. However, the neurophysiological mechanisms underlying gait are highly non-linear, and to use gait signatures for clinical applications, it would be desirable to have more precise measures of characterizing gait signatures. This can potentially be achieved by regularizing the internal state dynamics of the RNN network to constrain learned dynamics onto simpler manifolds. Another approach could be to use an encoder-decoder architecture, as seen in Chapter 4, which can help constrain the representation of information flowing through the network. Since gait data is highly periodic, it would also be intriguing to extend these approaches to simple RNN units (instead of LSTM units) to make the model more interpretable. This requires careful training of the RNN model to avoid gradient instabilities, one approach being incrementally changing the RNN timescale during training.

I also propose hierarchical RNN models in Chapter 2 for creating multi-timescale representations of behavioral data. A primary limitation of that approach is that we exhaust our datasets because of sub-sampling at each level. Training deep learning models from scratch requires a large amount of data, and a small dataset at higher levels of training greatly limits our ability to identify non-stationarities at long timescales. However, this limitation is possible to address by taking longer and longer

behavioral measurements, and research labs have already extended continuous behavioral measurements to multiple weeks [59]. Another goal here is to go beyond markerless tracking data as a measure of the behavior of the animal at short timescales, perhaps adding additional behavioral degrees of freedom like vocalizations.

Another area of research where these ideas can be extended is in understanding long-timescale dynamics in the social behavior of animals. Social behavior is complex and requires careful consideration of interaction measures one chooses to measure (e.g., the distance between animals, relative orientation, etc.) to build social behavioral representations using current methods[54]. Deep-learning tools like RNNs are readily adaptable to complex datasets and can potentially infer joint representations of posture and interactions with just postural markers. It remains a really exciting avenue of research to extend these approaches for social behavior analysis.

In Chapter 3, we extend a current method to quantify behavioral repertoires[3] to study stress-induced changes in mice behavior. Our unsupervised approach is data-driven and provides unbiased quantification of behavioral adaptations to stress in male and female mice. This highly resolved behavioral analysis revealed effects that are overlooked by field-standard coarse metrics such as time spent in the center of a field. Notably, we find novel behavioral changes in male mice exposed to sub-chronic variable stress, contradicting previous conclusions of short-term stress resilience in male mice. We also find sex-specific differences in baseline behaviors, as well as surprising similarities in behavioral changes in response to subchronic stress. For future efforts, our approach can be extended to understand the behavioral trajectory of adaptations to stress, leading to a more accurate account of sex-specific differences in stress susceptibility.

In Chapter 4, I discuss a novel method to identify non-stationarities acting at disparate timescales in a dynamical system. A major limitation of the delay-embedding approach is that the reconstructed state-space is assumed to partition into disjoint

non-overlapping sets, which is ill-posed for dynamical systems with a fractal structure (such as the Lorenz system). However, if projected to high enough dimensions, such fractal measures can be smoothed out if the system has a rich dynamical structure consisting of many chaotic systems with fractal attractors [186]. While the encoder-decoder RNN model proposed in Chapter 4 captures non-stationarities encoded into the Lorenz system at very long timescales, it remains a possible direction to explore if simply projecting the delay-embedded space to a higher dimension is able to identify the underlying non-stationarity. It is possible that we are essentially performing this projection using the encoder-decoder architecture, however this connection remains unexplored.

Bibliography

- [1] Talmo D. Pereira, Diego E. Aldarondo, Lindsay Willmore, Mikhail Kislin, Samuel S.-H. Wang, Mala Murthy, and Joshua W. Shaevitz. Fast animal pose estimation using deep neural networks. *Nature Methods*, 16(1):117–125, January 2019. Publisher: Nature Publishing Group.
- [2] Greg J. Stephens, Bethany Johnson-Kerner, William Bialek, and William S. Ryu. Dimensionality and Dynamics in the Behavior of *C. elegans*. *PLOS Computational Biology*, 4(4):e1000028, April 2008.
- [3] Gordon J. Berman, Daniel M. Choi, William Bialek, and Joshua W. Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672, October 2014.
- [4] Gordon J. Berman. Measuring behavior across scales. *BMC Biology*, 16(1):23, February 2018.
- [5] Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281, September 2018.
- [6] André E. X. Brown, Eviatar I. Yemini, Laura J. Grundy, Tadas Jucikas, and William R. Schafer. A dictionary of behavioral motifs reveals clusters of

- genes affecting *Caenorhabditis elegans* locomotion. *Proceedings of the National Academy of Sciences*, 110(2):791–796, January 2013. Publisher: Proceedings of the National Academy of Sciences.
- [7] Alex Gomez-Marin, Greg J. Stephens, and André E. X. Brown. Hierarchical compression of *Caenorhabditis elegans* locomotion reveals phenotypic differences in the organization of behaviour. *Journal of the Royal Society, Interface*, 13(121):20160466, August 2016.
- [8] Alexander B. Wiltschko, Matthew J. Johnson, Giuliano Iurilli, Ralph E. Peterson, Jesse M. Katon, Stan L. Pashkovski, Victoria E. Abraira, Ryan P. Adams, and Sandeep Robert Datta. Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, 88(6):1121–1135, December 2015.
- [9] Gordon J. Berman, William Bialek, and Joshua W. Shaevitz. Predictability and hierarchy in *Drosophila* behavior. *Proceedings of the National Academy of Sciences*, 113(42):11943–11948, October 2016.
- [10] Kevin Luxem, Petra Mocellin, Falko Fuhrmann, Johannes Kürsch, Stephanie R. Miller, Jorge J. Palop, Stefan Remy, and Pavol Bauer. Identifying behavioral structure from deep variational embeddings of animal motion. *Communications Biology*, 5(1):1–15, November 2022. Publisher: Nature Publishing Group.
- [11] Antonio C. Costa, Tosif Ahamed, David Jordan, and Greg J. Stephens. Maximally predictive states: From partial observations to long timescales. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(2):023136, February 2023.
- [12] Taniel S. Winner, Michael C. Rosenberg, Kanishk Jain, Trisha M. Kesar, Lena H. Ting, and Gordon J. Berman. Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics. *PLOS Com-*

- putational Biology*, 19(10):e1011556, October 2023. Publisher: Public Library of Science.
- [13] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 818–833, Cham, 2014. Springer International Publishing.
- [14] Jesse D. Marshall, Diego E. Aldarondo, Timothy W. Dunn, William L. Wang, Gordon J. Berman, and Bence P. Ölveczky. Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire. *Neuron*, 109(3):420–437.e8, February 2021.
- [15] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.
- [16] DJ Strouse and David J. Schwab. The deterministic information bottleneck. *Neural Comput.*, 29(6):1611–1630, June 2017.
- [17] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, September 2020. arXiv:1802.03426.
- [18] Adam J. Calhoun and Ahmed El Hady. Everyone knows what behavior is but they just don’t agree on it. *iScience*, 26(11):108210, November 2023.
- [19] William M. Baum. What Counts as Behavior? The Molar Multiscale View. *The Behavior Analyst*, 36(2):283–293, 2013.
- [20] Daniel A. Levitis, William Z. Lidicker, and Glenn Freund. Behavioural biologists don’t agree on what constitutes behaviour. *Animal behaviour*, 78(1):103–110, July 2009.

- [21] J. C. Hall. The mating of a fly. *Science (New York, N.Y.)*, 264(5166):1702–1714, June 1994.
- [22] Verónica María Corrales-Carvajal, Aldo A Faisal, and Carlos Ribeiro. Internal states drive nutrient homeostasis by modulating exploration-exploitation trade-off. *eLife*, 5:e19920, October 2016. Publisher: eLife Sciences Publications, Ltd.
- [23] Bryce Chung, Muneeb Zia, Kyle A Thomas, Jonathan A Michaels, Amanda Jacob, Andrea Pack, Matthew J Williams, Kailash Nagapudi, Lay Heng Teng, Eduardo Arrambide, Logan Ouellette, Nicole Oey, Rhuna Gibbs, Philip Anschutz, Jiaao Lu, Yu Wu, Mehrdad Kashefi, Tomomichi Oya, Rhonda Kersten, Alice C Mosberger, Sean O’Connell, Runming Wang, Hugo Marques, Ana Rita Mendes, Constanze Lenschow, Gayathri Kondakath, Jeong Jun Kim, William Olson, Kiara N Quinn, Pierce Perkins, Graziana Gatto, Ayesha Thanawalla, Susan Coltman, Taegyo Kim, Trevor Smith, Ben Binder-Markey, Martin Zaback, Christopher K Thompson, Simon Giszter, Abigail Person, Martyn Goulding, Eiman Azim, Nitish Thakor, Daniel O’Connor, Barry Trimmer, Susana Q Lima, Megan R Carey, Chethan Pandarinath, Rui M Costa, J Andrew Pruszynski, Muhannad Bakir, and Samuel J Sober. Myomatrix arrays for high-definition muscle recording. *eLife*, 12:RP88551, December 2023. Publisher: eLife Sciences Publications, Ltd.
- [24] Emma Lubel, Bruno Grandi Sgambato, Deren Y. Barsakcioglu, Jaime Ibáñez, Meng-Xing Tang, and Dario Farina. Kinematics of individual muscle units in natural contractions measured in vivo using ultrafast ultrasound. *Journal of Neural Engineering*, 19(5):056005, September 2022. Publisher: IOP Publishing.
- [25] Boris G Vainer. FPA-based infrared thermography as applied to the study of cutaneous perspiration and stimulated vascular response in humans. *Physics in Medicine and Biology*, 50(23):R63–R94, December 2005.

- [26] Marieke van Dooren, J. J. G. (Gert-Jan) de Vries, and Joris H. Janssen. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, 106(2):298–304, May 2012.
- [27] Alexander Mathis, Alexander Mathis, Pranav Mamidanna, Pranav Mami-danna, Taiga Abe, Taiga Abe, Kevin M. Cury, Kevin M. Cury, Venkatesh Murthy, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, Mackenzie Weygandt Mathis, Matthias Bethge, and Matthias Bethge. Markerless tracking of user-defined features with deep learning. *arXiv: Computer Vision and Pattern Recognition*, 2018.
- [28] Alexander I. Hsu and Eric A. Yttri. B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nature Communi-cations*, 12(1):5188, August 2021. Publisher: Nature Publishing Group.
- [29] André E. X. Brown and Benjamin de Bivort. Ethology as a physical science. *Nature Physics*, 14(7):653–657, July 2018. Publisher: Nature Publishing Group.
- [30] Antonio C. Costa, Tosif Ahamed, David Jordan, and Greg J. Stephens. Maximally predictive ensemble dynamics from data, October 2021. Pages: 2021.05.26.445816 Section: New Results.
- [31] Konrad Z. Lorenz. The comparative method in studying innate behavior pat-terns. In *Physiological mechanisms in animal behavior. (Society’s Symposium IV.)*, pages 221–268. Academic Press, Oxford, England, 1950.
- [32] N. Tinbergen and A. C. Perdeck. On the stimulus situation releasing the begging response in the newly hatched Herring Gull chick (*Larus argentatus argentatus* Pont.). *Behaviour*, 3:1–39, 1950. Place: United Kingdom Publisher: Brill Academic Publishers.

- [33] Adam J. Calhoun, Jonathan W. Pillow, and Mala Murthy. Unsupervised identification of the internal states that shape natural behavior. *Nature neuroscience*, 22(12):2040–2049, December 2019.
- [34] Drew N. Robson and Jennifer M. Li. A dynamical systems view of neuroethology: Uncovering stateful computation in natural behaviors. *Current Opinion in Neurobiology*, 73:102517, April 2022.
- [35] Michael H. Dickinson, Claire T. Farley, Robert J. Full, M. A. R. Koehl, Rodger Kram, and Steven Lehman. How Animals Move: An Integrative View. *Science*, 288(5463):100–106, April 2000. Publisher: American Association for the Advancement of Science.
- [36] Mark A Frye and Michael H Dickinson. Closing the loop between neurobiology and flight behavior in *Drosophila*. *Current Opinion in Neurobiology*, 14(6):729–736, December 2004.
- [37] Aristotle Edited, translated with Notes by D. M. Balme, With a Report on Recent Work, and an updated Bibliography by Allan Gotthelf. *De Partibus Animalium I and De Generatione Animalium I: with Passages from II. 1-3*. Clarendon Aristotle Series. Oxford University Press, Oxford, New York, revised edition, revised edition edition, December 1992.
- [38] Andrea Falcon. *The Architecture of the Science of Living Beings: Aristotle and Theophrastus on Animals and Plants*. Cambridge University Press, Cambridge, 2024.
- [39] Maria Michela Sassi. *Colour psychology in the Graeco-Roman world*. Fondation Hardt, 2020. Google-Books-ID: cxPnzQEACAAJ.
- [40] Ivan P. Pavlov. *The Scientific Investigation of the Psychological Faculties or Pro-*

- cesses in the Higher Animals. *Science*, 24(620):613–619, November 1906. Publisher: American Association for the Advancement of Science.
- [41] Ingo Brigandt. “The Instinct Concept of the Early Konrad Lorenz”. *Journal of the History of Biology*, 38(3):571–608, November 2005.
- [42] Daniel S. Lehrman. A Critique of Konrad Lorenz’s Theory of Instinctive Behavior. *The Quarterly Review of Biology*, 28(4):337–363, 1953. Publisher: University of Chicago Press.
- [43] N. Tinbergen. On aims and methods of Ethology. *Zeitschrift für Tierpsychologie*, 20(4):410–433, 1963. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0310.1963.tb01161.x>.
- [44] Karl von Frisch. *The dance language and orientation of bees*. The dance language and orientation of bees. Harvard University Press, Cambridge, MA, US, 1967.
- [45] L. P. J. J. Noldus, Lucas P. J. J. Noldus, Andrew Spink, Andrew Spink, Ruud A. J. Tegelenbosch, and Ruud A. J. Tegelenbosch. EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods Instruments & Computers*, 2001.
- [46] Ashley Walton, Amy Branham, Don M. Gash, and Richard Grondin. Automated video analysis of age-related motor deficits in monkeys using EthoVision. *Neurobiology of Aging*, 27(10):1477–1483, October 2006.
- [47] Jon Pham, Sara M. Cabrera, Carles Sanchis-Segura, and Marcelo A. Wood. Automated scoring of fear related behavior using EthoVision software. *Journal of neuroscience methods*, 178(2):323–326, April 2009.

- [48] Hanna J. Kulbeth, Saki Fukuda, and Lisa K. Brents. Automated quantification of opioid withdrawal in neonatal rat pups using Ethovision® XT software. *Neurotoxicology and Teratology*, 84:106959, March 2021.
- [49] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [50] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015. arXiv:1512.03385 [cs].
- [51] Eldar Insafutdinov, Leonid Pishchulin, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 34–50, Cham, 2016. Springer International Publishing.
- [52] Pierre Karashchuk, Pierre Karashchuk, Katie L. Rupp, Katie L. Rupp, Patsy S. Dickinson, Eryn S. Dickinson, Eryn S. Dickinson, Sarah Walling-Bell, Elischa Sanders, Elischa Sanders, Eiman Azim, Eiman Azim, Bingni W. Brunton, Bingni W. Brunton, John C. Tuthill, and John C. Tuthill. Anipose: a toolkit for robust markerless 3D pose estimation. *bioRxiv*, 2020.
- [53] Timothy W. Dunn, Jesse D. Marshall, Kyle S. Severson, Diego E. Aldarondo, David G. C. Hildebrand, Selmaan N. Chettih, William L. Wang, Amanda J. Gellis, David E. Carlson, Dmitriy Aronov, Winrich A. Freiwald, Fan Wang, and Bence P. Ölveczky. Geometric deep learning enables 3D kinematic profiling across species and environments. *Nature Methods*, 18(5):564–573, May 2021. Publisher: Nature Publishing Group.
- [54] Ugnė Klibaite, Tianqing Li, Diego Aldarondo, Jumana Akoad, Bence Ölveczky,

and Timothy W. Dunn. Mapping the landscape of social behavior, September 2024. Pages: 2024.09.27.615451 Section: New Results.

- [55] Arne Monsees, Kay-Michael Voit, Damian J. Wallace, Juergen Sawinski, Edyta Charyasz, Klaus Scheffler, Jakob H. Macke, and Jason N. D. Kerr. Estimation of skeletal kinematics in freely moving rodents. *Nature Methods*, 19(11):1500–1509, November 2022. Publisher: Nature Publishing Group.
- [56] Jessy Lauer, Mu Zhou, Shaokai Ye, William Menegas, Steffen Schneider, Tanmay Nath, Mohammed Mostafizur Rahman, Valentina Di Santo, Daniel Soberanes, Guoping Feng, Venkatesh N. Murthy, George Lauder, Catherine Dulac, Mackenzie Weygandt Mathis, and Alexander Mathis. Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods*, 19(4):496–504, April 2022. Publisher: Nature Publishing Group.
- [57] Talmo D. Pereira, Nathaniel Tabris, Arie Matsliah, David M. Turner, Junyu Li, Shruthi Ravindranath, Eleni S. Papadoyannis, Edna Normand, David S. Deutsch, Z. Yan Wang, Grace C. McKenzie-Smith, Catalin C. Mitelut, Marielisa Diez Castro, John D’Uva, Mikhail Kislin, Dan H. Sanes, Sarah D. Kocher, Samuel S.-H. Wang, Annegret L. Falkner, Joshua W. Shaevitz, and Mala Murthy. SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods*, 19(4):486–495, April 2022. Publisher: Nature Publishing Group.
- [58] Jennifer J. Sun, Serim Ryou, Roni Goldshmid, Brandon Weissbourd, John Dabiri, David J. Anderson, Ann Kennedy, Yisong Yue, and Pietro Perona. Self-Supervised Keypoint Discovery in Behavioral Videos, April 2022. arXiv:2112.05121 [cs].
- [59] Grace C. McKenzie-Smith, Scott W. Wolf, Julien F. Ayroles, and Joshua W.

- Shaevitz. Capturing continuous, long timescale behavioral changes in *Drosophila melanogaster* postural data. *ArXiv*, page arXiv:2309.04044v1, September 2023.
- [60] Nastacia L. Goodwin, Jia J. Choong, Sophia Hwang, Kayla Pitts, Liana Bloom, Aasiya Islam, Yizhe Y. Zhang, Eric R. Szelenyi, Xiaoyu Tong, Emily L. Newman, Klaus Miczek, Hayden R. Wright, Ryan J. McLaughlin, Zane C. Norville, Neir Eshel, Mitra Heshmati, Simon R. O. Nilsson, and Sam A. Golden. Simple Behavioral Analysis (SimBA) as a platform for explainable machine learning in behavioral neuroscience. *Nature Neuroscience*, 27(7):1411–1424, July 2024. Publisher: Nature Publishing Group.
- [61] Cristina Segalin, Jalani Williams, Tomomi Karigo, May Hui, Moriel Zelikowsky, Jennifer J Sun, Pietro Perona, David J Anderson, and Ann Kennedy. The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice. *eLife*, 10:e63720, November 2021. Publisher: eLife Sciences Publications, Ltd.
- [62] Mayank Kabra, Alice A. Robie, Marta Rivera-Alba, Steven Branson, and Kristin Branson. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, 10(1):64–67, January 2013. Publisher: Nature Publishing Group.
- [63] Greg J. Stephens, Matthew Bueno de Mesquita, William S. Ryu, and William Bialek. Emergence of long timescales and stereotyped behaviors in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, 108(18):7286–7289, May 2011. Publisher: Proceedings of the National Academy of Sciences.
- [64] Floris Takens. Detecting strange attractors in turbulence. In David Rand and

- Lai-Sang Young, editors, *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics, pages 366–381. Springer Berlin Heidelberg, 1981.
- [65] Antonio C. Costa, Tosif Ahamed, David Jordan, and Greg J. Stephens. A Markovian dynamics for *Caenorhabditis elegans* behavior across scales. *Proceedings of the National Academy of Sciences*, 121(32):e2318805121, August 2024. Publisher: Proceedings of the National Academy of Sciences.
- [66] S. Brunton, J. Proctor, and J. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113:3932 – 3937, 2015.
- [67] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, April 1990.
- [68] D.R. Seidl and R.D. Lorenz. A structure by which a recurrent neural network can approximate a nonlinear dynamic system. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume ii, pages 709–714 vol.2, July 1991.
- [69] K. Doya. Universality of Fully-Connected Recurrent Neural Networks. *IEEE Transactions on Neural Networks*, 1993.
- [70] K. Narendra and K. Parthasarathy. Neural Networks In Dynamical Systems. *Other Conferences*, 1990.
- [71] Coryn A L Bailer-Jones, David J C MacKay, and Philip J Withers. A recurrent neural network for modelling dynamical systems. *Network: Computation in Neural Systems*, 9(4):531–547, January 1998.
- [72] Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber. A Novel Connectionist System for Uncon-

- strained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):855–868, May 2009. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [73] Claus Metzner and Patrick Krauss. Dynamics and Information Import in Recurrent Neural Networks. *Frontiers in Computational Neuroscience*, 16, April 2022. Publisher: Frontiers.
- [74] Anton Maximilian Schäfer and Hans Georg Zimmermann. Recurrent Neural Networks Are Universal Approximators. In Stefanos D. Kollias, Andreas Stafylopatis, Włodzisław Duch, and Erkki Oja, editors, *Artificial Neural Networks – ICANN 2006*, pages 632–640, Berlin, Heidelberg, 2006. Springer.
- [75] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986. Publisher: Nature Publishing Group.
- [76] Backpropagation through time: what it does and how to do it | IEEE Journals & Magazine | IEEE Xplore.
- [77] Ronald J. Williams and David Zipser. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, 1(2):270–280, June 1989.
- [78] Kazuki Irie, Anand Gopalakrishnan, and Jürgen Schmidhuber. Exploring the Promise and Limits of Real-Time Recurrent Learning, May 2023.
- [79] Jeffrey L. Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7(2):195–225, September 1991.
- [80] T. Mikolov, Martin Karafiát, Lukas Burget, J. Cernocký, and S. Khudanpur.

- Recurrent neural network based language model. *Proceedings of Interspeech*, 2, January 2010.
- [81] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, March 1994. Conference Name: IEEE Transactions on Neural Networks.
- [82] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning*, pages 1310–1318. PMLR, May 2013. ISSN: 1938-7228.
- [83] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. Place: US Publisher: MIT Press.
- [84] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches, October 2014. arXiv:1409.1259 [cs, stat].
- [85] Kamesh Krishnamurthy, Tankut Can, and David J. Schwab. Theory of Gating in Recurrent Neural Networks. *Physical Review X*, 12(1):011011, January 2022. Publisher: American Physical Society.
- [86] Tankut Can, Kamesh Krishnamurthy, and David J. Schwab. Gating creates slow modes and controls phase-space complexity in GRUs and LSTMs. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, pages 476–511. PMLR, August 2020. ISSN: 2640-3498.
- [87] Tankut Can and Kamesh Krishnamurthy. Emergence of robust memory manifolds, September 2024. arXiv:2109.03879.

- [88] Albert Gu, Caglar Gulcehre, Tom Le Paine, Matt Hoffman, and Razvan Pascanu. Improving the Gating Mechanism of Recurrent Neural Networks, June 2020. arXiv:1910.09890 [cs].
- [89] Mehmet Ozgur Turkoglu, Stefano D’Aronco, Jan Dirk Wegner, and Konrad Schindler. Gating Revisited: Deep Multi-layer RNNs That Can Be Trained, March 2021. arXiv:1911.11033 [cs].
- [90] Douglas Eck and Juergen Schmidhuber. A First Look at Music Composition using LSTM Recurrent Neural Networks. Technical Report, Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, February 2002.
- [91] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech Recognition with Deep Recurrent Neural Networks, March 2013. arXiv:1303.5778.
- [92] Wei-Ning Hsu, Yu Zhang, and James Glass. A prioritized grid long short-term memory RNN for speech recognition. In *2016 IEEE Spoken Language Technology Workshop (SLT)*, pages 467–473, December 2016.
- [93] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to Sequence Learning with Neural Networks. *arXiv:1409.3215 [cs]*, December 2014.
- [94] Haşim Sak, Andrew Senior, and Françoise Beaufays. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition, February 2014. arXiv:1402.1128.
- [95] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3156–3164, Boston, MA, USA, June 2015. IEEE.

- [96] Duyu Tang. Sentiment-Specific Representation Learning for Document-Level Sentiment Analysis. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15*, pages 447–452, New York, NY, USA, February 2015. Association for Computing Machinery.
- [97] Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, and Trevor Darrell. Long-term Recurrent Convolutional Networks for Visual Recognition and Description, May 2016. arXiv:1411.4389.
- [98] Vanessa Isabell Jurtz, Alexander Rosenberg Johansen, Morten Nielsen, Jose Juan Almagro Armenteros, Henrik Nielsen, Casper Kaae Sønderby, Ole Winther, and Søren Kaae Sønderby. An introduction to deep learning on biological sequence data: examples and solutions. *Bioinformatics (Oxford, England)*, 33(22):3685–3690, November 2017.
- [99] R. F. Frisk, J. Lorentzen, and J. B. Nielsen. Contribution of corticospinal drive to ankle plantar flexor muscle activation during gait in adults with cerebral palsy. *Exp Brain Res*, 237(6):1457–1467, June 2019.
- [100] Y. P. Ivanenko, R. Grasso, M. Zago, M. Molinari, G. Scivoletto, V. Castellano, and others. Temporal components of the motor patterns expressed by the human spinal cord reflect foot kinematics. *J Neurophysiol*, 90(5):3555–3565, November 2003.
- [101] S. Krogh, P. Aagaard, A. B. Jønsson, K. Figlewski, and H. Kasch. Effects of repetitive transcranial magnetic stimulation on recovery in lower limb muscle strength and gait function following spinal cord injury: a randomized controlled trial. *Spinal Cord*, 60(2):135–141, February 2022.
- [102] D. R. Young, C. L. Banks, T. E. McGuirk, and C. Patten. Evidence for shared

- neural information between muscle synergies and corticospinal efficacy. *Sci Rep*, 12(1):8953, May 2022.
- [103] K. Genthe, C. Schenck, S. Eicholtz, L. Zajac-Cox, S. Wolf, and T. M. Kesar. Effects of real-time gait biofeedback on paretic propulsion and gait biomechanics in individuals post-stroke. *Top Stroke Rehabil*, 25(3):186–193, April 2018.
- [104] Y. T. Pan, I. Kang, J. Joh, P. Kim, K. R. Herrin, T. M. Kesar, and others. Effects of Bilateral Assistance for Hemiparetic Gait Post-Stroke Using a Powered Hip Exoskeleton. *Ann Biomed Eng*, August 2022.
- [105] D. S. Reisman, A. J. Bastian, and S. M. Morton. Neurophysiologic and rehabilitation insights from the split-belt and other locomotor adaptation paradigms. *Phys Ther*, 90(2):187–195, February 2010.
- [106] S. Wang, Y. C. Pai, and T. Bhatt. Neuromuscular mechanisms of motor adaptation to repeated gait-slip perturbations in older adults. *Sci Rep*, 12(1):19851, November 2022.
- [107] J. D’souza, D. M. Natarajan, and D. S. Kumaran. Does the Environment Cause Changes in Hemiparetic Lower Limb Muscle Activity and Gait Velocity During Walking in Stroke Survivors? *J Stroke Cerebrovasc Dis*, 29(10):105174, October 2020.
- [108] R. J. Larsen, R. M. Queen, and D. Schmitt. Adaptive locomotion: Foot strike pattern and limb mechanical stiffness while running over an obstacle. *J Biomech*, 143:111283, October 2022.
- [109] Angela S. Attwood, Casimir J. H. Ludwig, Ian S. Penton-Voak, Jade Poh, Alex S. F. Kwong, and Marcus R. Munafò. Effects of state anxiety on gait: a 7.5% carbon dioxide challenge study. *Psychological Research*, 85(6):2444–2452, September 2021.

- [110] E. Elkjær, M. B. Mikkelsen, J. Michalak, D. S. Mennin, and M. S. O’Toole. Motor alterations in depression and anxiety disorders: A systematic review and meta-analysis. *J Affect Disord*, 317:373–387, November 2022.
- [111] Andrew Steptoe and Paola Zaninotto. Lower socioeconomic status and the acceleration of aging: An outcome-wide analysis. *Proceedings of the National Academy of Sciences*, 117(26):14911–14917, June 2020. Publisher: Proceedings of the National Academy of Sciences.
- [112] P. Zaninotto, A. Sacker, and J. Head. Relationship between wealth and age trajectories of walking speed among older adults: evidence from the English Longitudinal Study of Ageing. *J Gerontol A Biol Sci Med Sci*, 68(12):1525–1531, December 2013.
- [113] S. A. Garcia, A. K. Johnson, S. R. Brown, E. P. Washabaugh, C. Krishnan, and R. M. Palmieri-Smith. Dynamic knee stiffness during walking is increased in individuals with anterior cruciate ligament reconstruction. *J Biomech*, 146:111400, November 2022.
- [114] J. Jang and E. A. Wikstrom. Ankle joint contact force profiles differ between those with and without chronic ankle instability during walking. *Gait Posture*, 100:1–7, November 2022.
- [115] C. E. Milner, E. Foch, J. M. Gonzales, and D. Petersen. Biomechanics associated with tibial stress fracture in runners: A systematic review and meta-analysis. *J Sport Health Sci*, pages S2095–2546(22)00116–8, December 2022.
- [116] T. Ijmker and C. J. C. Lamoth. Gait and cognition: the relationship between gait stability and variability with executive function in persons with and without dementia. *Gait Posture*, 35(1):126–130, January 2012.

- [117] I. Jonkers, S. Delp, and C. Patten. Capacity to increase walking speed is limited by impaired hip and ankle power generation in lower functioning persons post-stroke. *Gait Posture*, 29(1):129–137, January 2009.
- [118] R. Moura Coelho, H. Hirai, J. Martins, and H. I. Krebs. Biomarkers for rhythmic and discrete dynamic primitives in locomotion. *Sci Rep*, 12(1):20165, November 2022.
- [119] L. A. Prosser, H. L. Atkinson, J. M. Alfano, S. K. Kessler, and R. B. Ichord. Walking speed influences spatiotemporal but not symmetry measures of gait in children and adolescents with hemiplegia. *Gait Posture*, 98:233–236, October 2022.
- [120] Y. Russo, S. Stuart, C. Silva-Batista, B. Brumbach, G. Vannozzi, and M. Mancini. Does visual cueing improve gait initiation in people with Parkinson’s disease? *Hum Mov Sci*, 84:102970, August 2022.
- [121] E. Troisi Lopez, P. Sorrentino, M. Liparoti, R. Minino, A. Polverino, A. Romano, and others. The kinectome: A comprehensive kinematic map of human motion in health and disease. *Ann N Y Acad Sci*, 1516(1):247–261, October 2022.
- [122] T. A. Correa, R. Baker, H. K. Graham, and M. G. Pandy. Accuracy of generic musculoskeletal models in predicting the functional roles of muscles in human gait. *J Biomech*, 44(11):2096–2105, July 2011.
- [123] H. Geyer and H. Herr. A Muscle-Reflex Model That Encodes Principles of Legged Mechanics Produces Human Walking Dynamics and Muscle Activities. *IEEE Trans Neural Syst Rehabil Eng*, 18(3):263–273, June 2010.
- [124] A. D. Kuo and J. M. Donelan. Dynamic Principles of Gait and Their Clinical Implications. *Physical Therapy*, 90(2):157–174, February 2010.

- [125] A. Falisse, G. Serrancolí, C. L. Dembia, J. Gillis, I. Jonkers, and F. De Groot. Rapid predictive simulations with complex musculoskeletal models suggest that diverse healthy and pathological human gaits can emerge from similar control strategies. *Journal of The Royal Society Interface*, 16(157):20190402, August 2019.
- [126] A. J. Meyer, I. Eskinazi, J. N. Jackson, A. V. Rao, C. Patten, and B. J. Fregly. Muscle Synergies Facilitate Computational Prediction of Subject-Specific Walking Motions. *Front Bioeng Biotechnol*, 4:77, 2016.
- [127] L. Pitto, H. Kainz, A. Falisse, M. Wesseling, S. Van Rossom, H. Hoang, and others. SimCP: A Simulation Platform to Predict Gait Performance Following Orthopedic Intervention in Children With Cerebral Palsy. *Front Neurobot*, 13:54, 2019.
- [128] F. De Groot and A. Falisse. Perspective on musculoskeletal modelling and predictive simulations of human movement to assess the neuromechanics of gait. *Proceedings of the Royal Society B: Biological Sciences*, 288(1946):20202432, March 2021.
- [129] F. Dzeladini, J. van den Kieboom, and A. Ijspeert. The contribution of a central pattern generator in a reflex-based neuromuscular model. *Front Hum Neurosci*, 8:371, 2014.
- [130] G. Taga, Y. Yamaguchi, and H. Shimizu. Self-organized control of bipedal locomotion by neural oscillators in unpredictable environment. *Biol Cybern*, 65(3):147–159, 1991.
- [131] E. Angelidis, E. Buchholz, J. Arreguit, A. Rougé, T. Stewart, A. von Arnim, and others. A spiking central pattern generator for the control of a simulated

- lamprey robot running on SpiNNaker and Loihi neuromorphic boards. *Neuro-morph Comput Eng*, 1(1):014005, September 2021.
- [132] F. E. Pollick, J. W. Kay, K. Heim, and R. Stringer. Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6):1247, 2005.
- [133] N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5):2, September 2002.
- [134] K. L. Johnson, S. Gill, V. Reichman, and L. G. Tassinary. Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology*, 93(3):321, 2007.
- [135] M. M. Gross, E. A. Crane, and B. L. Fredrickson. Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Human Movement Science*, 31(1):202–221, February 2012.
- [136] D. D. Brown, G. Wijffels, and R. G. J. Meulenbroek. Individual Differences in Sequential Movement Coordination in Hip-Hop Dance: Capturing Joint Articulation in Practicing the Wave. *Front Psychol*, 12:731901, 2021.
- [137] R. Edey, D. Yon, J. Cook, I. Dumontheil, and C. Press. Our own action kinematics predict the perceived affective states of others. *J Exp Psychol Hum Percept Perform*, 43(7):1263–1268, July 2017.
- [138] C. Becchio, V. Manera, L. Sartori, A. Cavallo, and U. Castiello. Grasping intentions: from thought experiments to empirical evidence. *Front Hum Neurosci*, 6:117, 2012.
- [139] N. Neave, K. McCarty, J. Freynik, N. Caplan, J. Hönekopp, and B. Fink. Male dance moves that catch a woman’s eye. *Biol Lett*, 7(2):221–224, April 2011.

- [140] A. Habersack, S. F. Fischerauer, T. Kraus, H. P. Holzer, and M. Svehlik. Kinematic and Kinetic Gait Parameters Can Distinguish between Idiopathic and Neurologic Toe-Walking. *Int J Environ Res Public Health*, 19(2):804, January 2022.
- [141] J. Heinik, P. Werner, T. Dekel, I. Gurevitz, and S. Rosenblum. Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: an exploratory study. *Int Psychogeriatr*, 22(3):479–488, May 2010.
- [142] S. Ferrante, N. Chia Bejarano, E. Ambrosini, A. Nardone, A. M. Turcato, M. Monticone, and others. A Personalized Multi-Channel FES Controller Based on Muscle Synergies to Support Gait Rehabilitation after Stroke. *Frontiers in Neuroscience*, 10, 2016.
- [143] T. Mikolajczyk, I. Ciobanu, D. I. Badea, A. Iliescu, S. Pizzamiglio, T. Schauer, and others. Advanced technology for gait rehabilitation: An overview. *Advances in Mechanical Engineering*, 10(7):1687814018783627, July 2018.
- [144] A. J. Ries, T. F. Novacheck, and M. H. Schwartz. A data driven model for optimal orthosis selection in children with cerebral palsy. *Gait Posture*, 40(4):539–544, September 2014.
- [145] E. Halilaj, A. Rajagopal, M. Fiterau, J. L. Hicks, T. J. Hastie, and S. L. Delp. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J Biomech*, 81:1–11, November 2018.
- [146] Shai Revzen and John M. Guckenheimer. Estimating the phase of synchronized oscillators. *Physical Review E*, 78(5):051907, November 2008. Publisher: American Physical Society.

- [147] Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. *arXiv:physics/0004057*, April 2000. arXiv: physics/0004057.
- [148] Tanmay Nath, Tanmay Nath, Alexander Mathis, Alexander Mathis, An Chi Chen, An Chi Chen, Amir Patel, Amir Patel, Matthias Bethge, Matthias Bethge, and Mackenzie Weygandt Mathis. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 2019.
- [149] John W. Krakauer, Asif A. Ghazanfar, Alex Gomez-Marin, Malcolm A. MacIver, and David Poeppel. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, 93(3):480–490, February 2017.
- [150] Goodwin Rd, Dierker Lc, Wu M, Galea S, Hoven Cw, and Weinberger Ah. Trends in U.S. Depression Prevalence From 2015 to 2020: The Widening Treatment Gap. *American journal of preventive medicine*, 63(5), November 2022. Publisher: Am J Prev Med.
- [151] Damian F. Santomauro, Ana M. Mantilla Herrera, Jamileh Shadid, Peng Zheng, Charlie Ashbaugh, David M. Pigott, Cristiana Abbafati, Christopher Adolph, Joanne O. Amlag, Aleksandr Y. Aravkin, Bree L. Bang-Jensen, Gregory J. Bertolacci, Sabina S. Bloom, Rachel Castellano, Emma Castro, Suman Chakrabarti, Jhilik Chattopadhyay, Rebecca M. Cogen, James K. Collins, Xiaochen Dai, William James Dangel, Carolyn Dapper, Amanda Deen, Megan Erickson, Samuel B. Ewald, Abraham D. Flaxman, Joseph Jon Frostad, Nancy Fullman, John R. Giles, Ababi Zergaw Giref, Gaorui Guo, Jiawei He, Monika Helak, Erin N. Hulland, Bulat Idrisov, Akiaja Lindstrom, Emily Linebarger, Paulo A. Lotufo, Rafael Lozano, Beatrice Magistro, Deborah Carvalho Malta, Johan C. Månsson, Fatima Marinho, Ali H. Mokdad, Lorenzo Monasta, Paulami Naik, Shuhei Nomura, James Kevin O’Halloran, Samuel M. Ostroff, Maja Pasovic, Louise Penberthy, Robert C. Reiner Jr, Grace Reinke, Antonio Luiz P.

- Ribeiro, Aleksei Sholokhov, Reed J. D. Sorensen, Elena Varavikova, Anh Truc Vo, Rebecca Walcott, Stefanie Watson, Charles Shey Wiysonge, Bethany Zigler, Simon I. Hay, Theo Vos, Christopher J. L. Murray, Harvey A. Whiteford, and Alize J. Ferrari. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, 398(10312):1700–1712, November 2021. Publisher: Elsevier.
- [152] Jaimie Gradus. Prevalence and prognosis of stress disorders: a review of the epidemiologic literature. *Clinical Epidemiology*, Volume 9:251–260, May 2017.
- [153] M. Kramer. The rising pandemic of mental disorders and associated chronic diseases and disabilities. *Acta Psychiatrica Scandinavica*, 62(S285):382–397, September 1980.
- [154] Richard McCarty. Stress, behavior and experimental hypertension. *Neuroscience & Biobehavioral Reviews*, 7(4):493–502, December 1983.
- [155] Alline C. Campos, Manoela V. Fogaca, Daniele C. Aguiar, and Francisco S. Guimaraes. Animal models of anxiety disorders and stress. *Revista Brasileira de Psiquiatria*, 35(suppl 2):S101–S111, 2013.
- [156] Kelimer Lebron-Milad and Mohammed R. Milad. Sex differences, gonadal hormones and the fear extinction network: implications for anxiety disorders. *Biology of Mood & Anxiety Disorders*, 2:3, February 2012.
- [157] Beery Ak and Zucker I. Sex bias in neuroscience and biomedical research. *Neuroscience and biobehavioral reviews*, 35(3), January 2011. Publisher: Neurosci Biobehav Rev.
- [158] Rebecca M. Shansky. Are hormones a "female problem" for animal research? *Science (New York, N.Y.)*, 364(6443):825–826, May 2019.

- [159] Lopez J and Bagot Rc. Defining Valid Chronic Stress Models for Depression With Female Rodents. *Biological psychiatry*, 90(4), August 2021. Publisher: Biol Psychiatry.
- [160] Leah R. Miller, Cheryl Marks, Jill B. Becker, Patricia D. Hurn, Wei-Jung Chen, Teresa Woodruff, Margaret M. McCarthy, Farida Sohrabji, Londa Schiebinger, Cora Lee Wetherington, Susan Makris, Arthur P. Arnold, Gillian Einstein, Virginia M. Miller, Kathryn Sandberg, Susan Maier, Terri L. Cornelison, and Janine A. Clayton. Considering sex as a biological variable in preclinical research. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 31(1):29–34, January 2017.
- [161] Sayaka Yokota, Yusuke Suzuki, Keigo Hamami, Akiko Harada, and Shoji Komai. Sex differences in avoidance behavior after perceiving potential risk in mice. *Behavioral and Brain Functions*, 13(1):9, December 2017.
- [162] M.A. Morgan and D.W. Pfaff. Effects of Estrogen on Activity and Fear-Related Behaviors in Mice. *Hormones and Behavior*, 40(4):472–482, December 2001.
- [163] José Colom-Lapetina, Anna J. Li, Tatiana C. Pelegrina-Perez, and Rebecca M. Shansky. Behavioral Diversity Across Classic Rodent Models Is Sex-Dependent. *Frontiers in Behavioral Neuroscience*, 13:45, March 2019.
- [164] Chandrashekhar D. Borkar, Mariia Dorofeikova, Quan-Son Eric Le, Rithvik Vutukuri, Catherine Vo, Daniel Hereford, Alexis Resendez, Samhita Basavanhalli, Natalia Sifnugel, and Jonathan P. Fadok. Sex differences in behavioral responses during a conditioned flight paradigm. *Behavioural Brain Research*, 389:112623, July 2020.
- [165] Tina M. Gruene, Katelyn Flick, Alexis Stefano, Stephen D. Shea, and Re-

- becca M. Shansky. Sexually divergent expression of active and passive conditioned fear responses in rats. *eLife*, 4:e11352, November 2015.
- [166] Haley L. Goodwill, Gabriela Manzano-Nieves, Meghan Gallo, Hye-In Lee, Esther Oyerinde, Thomas Serre, and Kevin G. Bath. Early life stress leads to sex differences in development of depressive-like outcomes in a mouse model. *Neuropsychopharmacology*, 44(4):711–720, March 2019. Publisher: Nature Publishing Group.
- [167] Hodes Ge, Pfau Ml, Purushothaman I, Ahn Hf, Golden Sa, Christoffel Dj, Magida J, Brancato A, Takahashi A, Flanigan Me, Ménard C, Aleyasin H, Koo Jw, Lorsch Zs, Feng J, Heshmati M, Wang M, Turecki G, Neve R, Zhang B, Shen L, Nestler Ej, and Russo Sj. Sex Differences in Nucleus Accumbens Transcriptome Profiles Associated with Susceptibility versus Resilience to Subchronic Variable Stress. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(50), December 2015. Publisher: J Neurosci.
- [168] Kayla R. Nygaard, Susan E. Maloney, Raylynn G. Swift, Katherine B. McCullough, Rachael E. Wagner, Stuart B. Fass, Krassimira Garbett, Karoly Mirnics, Jeremy Veenstra-VanderWeele, and Joseph D. Dougherty. Extensive characterization of a Williams Syndrome murine model shows Gtf2ird1-mediated rescue of select sensorimotor tasks, but no effect on enhanced social behavior. *bioRxiv*, page 2023.01.18.523029, January 2023.
- [169] Muir J, Tse Yc, Iyer Es, Biris J, Cvetkovska V, Lopez J, and Bagot Rc. Ventral Hippocampal Afferents to Nucleus Accumbens Encode Both Latent Vulnerability and Stress-Induced Susceptibility. *Biological psychiatry*, 88(11), December 2020. Publisher: Biol Psychiatry.
- [170] Yinfeng Dong, Xuyang Wang, Yan Zhou, Qiaomu Zheng, Zheng Chen, Hua

- Zhang, Zhiling Sun, Guihua Xu, and Gang Hu. Hypothalamus-pituitary-adrenal axis imbalance and inflammation contribute to sex differences in separation- and restraint-induced depression. *Hormones and Behavior*, 122:104741, June 2020.
- [171] Ryan Shepard, Chloe E. Page, and Laurence Coutellier. Sensitivity of the prefrontal GABAergic system to chronic stress in male and female mice: Relevance for sex differences in stress-related disorders. *Neuroscience*, 332:1–12, September 2016.
- [172] Bhanu Chandra Karisetty, Pranav Chintamani Joshi, Arvind Kumar, and Sumana Chakravarty. Sex differences in the effect of chronic mild stress on mouse prefrontal cortical BDNF levels: A role of major ovarian hormones. *Neuroscience*, 356:89–101, July 2017.
- [173] Labonté B, Engmann O, Purushothaman I, Menard C, Wang J, Tan C, Scarpa Jr, Moy G, Loh Ye, Cahill M, Lorsch Zs, Hamilton Pj, Calipari Es, Hodes Ge, Issler O, Kronman H, Pfau M, Obradovic Alj, Dong Y, Neve Rl, Russo S, Kazarskis A, Tamminga C, Mechawar N, Turecki G, Zhang B, Shen L, and Nestler Ej. Sex-specific transcriptional signatures in human depression. *Nature medicine*, 23(9), September 2017. Publisher: Nat Med.
- [174] Thibault P. Bittar, Mari Carmen Pelaez, Jose Cesar Hernandez Silva, Francis Quessy, Andrée-Anne Lavigne, Daphnée Morency, Léa-Jeanne Blanchette, Eric Arsenault, Yoan Cherasse, Josée Seigneur, Igor Timofeev, Chantelle F. Sephton, Christophe D. Proulx, and Benoit Labonté. Chronic Stress Induces Sex-Specific Functional and Morphological Alterations in Corticoaccumbal and Corticotegmental Pathways. *Biological Psychiatry*, 90(3):194–205, August 2021.
- [175] Anna Brancato, Dana Bregman, H. Francisca Ahn, Madeline L. Pfau, Caroline Menard, Carla Cannizzaro, Scott J. Russo, and Georgia E. Hodes. Sub-chronic

- variable stress induces sex-specific effects on glutamatergic synapses in the nucleus accumbens. *Neuroscience*, 350:180–189, May 2017.
- [176] Bagot Rc, Parise Em, Peña Cj, Zhang Hx, Maze I, Chaudhury D, Persaud B, Cacho R, Bolaños-Guzmán Ca, Cheer Jf, Deisseroth K, Han Mh, and Nestler Ej. Ventral hippocampal afferents to the nucleus accumbens regulate susceptibility to depression. *Nature communications*, 6, May 2015. Publisher: Nat Commun.
- [177] Bagot Rc, Cates Hm, Purushothaman I, Lorsch Zs, Walker Dm, Wang J, Huang X, Schlüter Om, Maze I, Peña Cj, Heller Ea, Issler O, Wang M, Song Wm, Stein Jl, Liu X, Doyle Ma, Scobie Kn, Sun Hs, Neve Rl, Geschwind D, Dong Y, Shen L, Zhang B, and Nestler Ej. Circuit-wide Transcriptional Profiling Reveals Brain Region-Specific Gene Networks Regulating Depression Susceptibility. *Neuron*, 90(5), June 2016. Publisher: Neuron.
- [178] Quincey LaPlant, Sumana Chakravarty, Vincent Vialou, Shibani Mukherjee, Ja Wook Koo, Geetha Kalahasti, Kathryn R. Bradbury, Shameeke V. Taylor, Ian Maze, Arvind Kumar, Ami Graham, Shari G. Birnbaum, Vaishnav Krishnan, Hoang-Trang Truong, Rachael L. Neve, Eric J. Nestler, and Scott J. Russo. Role of nuclear factor kappaB in ovarian hormone-mediated stress hypersensitivity in female mice. *Biological Psychiatry*, 65(10):874–880, May 2009.
- [179] Jacob H. Levine, Erin F. Simonds, Sean C. Bendall, Kara L. Davis, El-ad D. Amir, Michelle Tadmor, Oren Litvin, Harris Fienberg, Astraea Jager, Eli Zunder, Rachel Finck, Amanda L. Gedman, Ina Radtke, James R. Downing, Dana Pe'er, and Garry P. Nolan. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell*, 162(1):184–197, July 2015.

- [180] Varun Saravanan, Gordon J. Berman, and Samuel J. Sober. Application of the hierarchical bootstrap to multi-level data in neuroscience. *Neurons, Behavior, Data Analysis, and Theory*, 3(5):<https://nbdt.scholasticahq.com/article/13927-application-of-the-hierarchical-bootstrap-to-multi-level-data-in-neuroscience>, 2020.
- [181] Jacqueline M Tabler, Maggie M Rigney, Gordon J Berman, Swetha Gopalakrishnan, Eglantine Heude, Hadeel Adel Al-lami, Basil Z Yannakoudakis, Rebecca D Fitch, Christopher Carter, Steven Vokes, Karen J Liu, Shahragim Tajbakhsh, SE Roian Egnor, and John B Wallingford. Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. *eLife*, 6:e19153, February 2017.
- [182] Annika Reuser, Kristin Wenzel, Stephan B. Felix, Marcus Dörr, Martin Bahls, and Stephanie Könemann. Simultaneous assessment of spontaneous cage activity and voluntary wheel running in group-housed mice. *Scientific Reports*, 12(1):4444, 2022.
- [183] Miklós Sárvári, Imre Kalló, Erik Hrabovszky, Norbert Solymosi, and Zsolt Liposits. Ovariectomy and Subsequent Treatment with Estrogen Receptor Agonists Tune the Innate Immune System of the Hippocampus in Middle-Aged Female Rats. *PLoS ONE*, 9(2):e88540, 2014.
- [184] Rui-Wu Wang, Ya-Qiang Wang, Jun-Zhou He, and Yao-Tang Li. Resource Elasticity of Offspring Survival and the Optimal Evolution of Sex Ratios. *PLoS ONE*, 8(1):e53904, 2013.
- [185] Dirk Aeyels. Generic Observability of Differentiable Systems. *SIAM Journal on Control and Optimization*, 19(5):595–603, September 1981. Publisher: Society for Industrial and Applied Mathematics.

- [186] Yuri Kifer. *Ergodic Theory of Random Transformations*. Birkhäuser, Boston, MA, 1986.