UNIVERSITY OF CALIFORNIA

Los Angeles

Neural network dynamics of temporal processing

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Neuroscience

by

Nicholas Hardy

2018

ABSTRACT OF THE DISSERTATION

Neural network dynamics of temporal processing

by

Nicholas Hardy

Doctor of Philosophy in Neuroscience

University of California, Los Angeles, 2018

Professor Dean Buonomano, Chair

Time is centrally involved in most tasks the brain performs. However, the neurobiological mechanisms of timing remain a mystery. Signatures of temporal processing related to sensory and motor behavior have been observed in several brain regions and behavioral contexts. This activity is often complex, representing time in the activity of large populations of neurons. A major question is whether this observed activity is generated by a specialized clock in the brain or whether it is arises locally via the emergent dynamics of neural networks. State dependent theories of timing argue for the latter: neural activity evolving over time produces a trajectory of network states that can encode temporal information.

In this work, I examine the role of network dynamics in encoding temporal information. Combining mathematical models, *in vitro* neural recordings, and human psychophysics, the studies presented here describe potential network level mechanisms for timing in the brain. Chapter 2 presents research examining sequential activity observed in brain regions characterized by recurrent connectivity. This study describes a theoretical mechanism that

recurrent neural networks may use to autonomously produce sequential activity and encode temporal information. Next, Chapter 3 examines the mechanisms of producing the same complex movement at a variety of speeds, a fundamental feature of motor timing. This study combines theoretical and psychophysical studies to predict and test a novel feature of motor timing: temporal accuracy improves with speed, termed the Weber-speed effect. Finally, Chapter 4 examines how cortical neural networks encode temporal information. Using organotypic slice cultures, this study demonstrates that the cortex processes temporal input patterns in a state dependent manner, supporting theoretical predictions. Taken together, the results of this work strongly support state dependent theories of timing, providing insight into the neural basis temporal processing.

The dissertation of Nicholas Hardy is approved.

Carlos Portera-Cailliau

Sotiris Masmanidis

Dario L. Ringach

Dean Buonomano, Committee Chair

University of California, Los Angeles

2018

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my adviser Dr. Dean Buonomano for the time and effort he invested in training me. Without his guidance this work would not have been possible. I greatly admire the rigor and creativity he applies to all of his work, and hope that I have absorbed even a small portion of his scientific skill. I would also like to thank Dr. Vishwa Goudar for his help and guidance in conducting simulations and data analysis. I am excited to observe his career as he progresses through the field. In addition, I would like to thank Dr. Helen Motanis for her guidance and input in designing the research I describe in this dissertation, Juan Romero-Sosa for his help conducting psychophysics experiments. Also, the assistance of other members of the lab, especially Alisa Martinez and Jason Lee, was essential for maintaining experimental tissue and keeping the lab functioning. I also thank Dr. Carlos Portera-Cailliau for his guidance and for allowing me to attend his lab's weekly meeting, in which I learned a great deal about 2-photon calcium imaging. Lastly, I thank the other members of my committee, Dr. Sotiris Masmanidis and Dr. Dario Ringach, for the help and guidance in shaping my thesis research.

Chapter 2 is a version of: Hardy, N.F., and Buonomano, D.V. (2017). Encoding Time in Feedforward Trajectories of a Recurrent Neural Network Model. Neural Computation 1–19. https://doi.org/10.1162/neco_a_01041.

For this work, N.F.H. and D.V.B. designed the experiments and wrote the manuscript. N.F.H. conducted the simulations and analyzed the data.

Chapter 3 is from a manuscript under review. It can be viewed on the preprint server bioRxiv as: Hardy, N. F., Goudar, V., Romero-Sosa, J. L. & Buonomano, D. A Model of Temporal Scaling Correctly Predicts that Weber's Law is Speed-dependent. bioRxiv 159590 (2017). http://www.biorxiv.org/content/early/2017/07/05/159590.

For this work, D.V.B. conceived of the approach. N.F.H. performed the simulations and data analysis for the model. J.L.S., N.F.H. and D.V.B. designed the psychophysics experiments. J.L.S. conducted the psychophysics experiments and data collection. J.L.S. and D.V.B. performed data analysis for the psychophysics experiments. V.G. and N.F.H designed and performed the mechanisms analyses. N.F.H., V.G. and D.V.B. wrote the paper.

This research was supported by NIH grants MH60163 and T32 NS058280, and NSF grant IIS-1420897.

VITA

EDUCATION

University of Maryland, College Park (UMCP) May 2009
B.S. Psychology, Honors Program

NYU Study Abroad Program, Shanghai, China June 2008

RESEARCH PROJECTS AND EXPERIENCE

Graduate Student Researcher June 2013 - Present
Lab of Professor Dean Buonomano, UCLA
- Researching the computational properties of recurrent neural networks (RNNs) and their role in time perception using mathematical simulations, electrophysiology, and two-photon microscopy.

Training Temporally Invariant Recurrent Neural Networks 2014 - Present
- Developed novel deep learning algorithms to train RNNs to decode temporally invariant information
  - https://github.com/nhardy01/RNN

Intern, Johns Hopkins University Applied Physics Lab April – July 2017
Intelligent Systems Center
- Developed applications for neuromorphic computing and brain-computer interface technologies.

Rotation Student January – April 2013
Lab of Dr. Itzhak Fried, UCLA
- Studied the mechanisms of learning and memory in human epilepsy patients using local field potentials and single unit recordings.

Research Fellow, National Institute of Mental Health August 2009 – June 2013
Lab of Dr. Daniel Weinberger and Dr. Keri Martinowich. NIH Bethesda,
- Examined the genetic and neural network mechanisms underlying Schizophrenia and depression.

HONORS AND AWARDS

Neural Microcircuits Training Grant ($50k) October 2014-2016
- Competitively awarded NIH funded grant to study the computational properties of neural circuits

UCLA Brain Research Institute Travel Award 2014, 2015, 2016

Post Baccalaureate Intramural Research Training Fellowship August 2009-2011
- National Institutes of Health Bethesda, Maryland

UMCP President's Scholarship for academic merit ($20k) 2005-2009

PUBLICATIONS

- **Hardy, N.F**., Goudar, V., Romero-Sosa, J.L., and Buonomano, D. *A Model of Temporal Scaling Correctly Predicts that Weber's Law is Speed-dependent*. BioRxiv 159590. (2017, under review).
    - https://www.biorxiv.org/content/early/2017/07/05/159590
- **Hardy, N.F.**, and Buonomano, D.V. *Encoding Time in Feedforward Trajectories of a Recurrent Neural Network Model*. Neural Computation 1–19 (2017).
- **Hardy NF**, Buonomano DV. *Neurocomputational models of interval and pattern timing.* Current Opinion in Behavior Sciences (2016).
- Hill, J.L., **Hardy, N.F.**, Jimenez, D.V., Maynard, K.R., Kardian, A.S., Pollock, C.J., Schloesser, R.J., and Martinowich, K. *Loss of promoter IV-driven BDNF expression impacts oscillatory activity during sleep, sensory information processing and fear regulation.* Translational Psychiatry 6, e873 (2016).
- Schloesser RJ, Orvoen S, Jimenez DV, **Hardy NF**, Maynard KR, Sukumar M, Manji HK, Gardier AM, David DJ, Martinowich K. *Antidepressant-like effects of electroconvulsive seizures require adult neurogenesis in a neuroendocrine model of depression.* Brain Stimulation (2015).
- Shtrahman M, Aharoni DB, **Hardy NF**, Buonomano DV, Arisaka K, Otis TS. *Multifocal Fluorescence Microscope for Fast Optical Recordings of Neuronal Action Potentials.* Biophysical Journal (2015).
- Schloesser RJ, Jimenez DV, **Hardy NF**, Paredes D, Catlow BJ, Manji HK, McKay RD, Martinowich K. *Atrophy of pyramidal neurons and increased stress-induced glutamate levels in CA3 following chronic suppression of adult neurogenesis.* Brain Structure and Function (2014).

See Google Scholar profile for updated publications.

PRESENTATIONS

Invited talks

- Hardy N, Jimenez D, Martinowich K, Schloesser R. *Neurogenesis Dependent Modulations of Neuroplasticity in Cornu Ammonis Regions of the Adult Hippocampus.* Session 025 Oral Session Mood Disorders. 2010 Society of Biological Psychiatry meeting planner, New Orleans, LA, 2010.

Poster Presentations

- Hardy NF, Romero JL, Buonomano DV. *Understanding the mechanisms of motor temporal scaling using a recurrent neural network model*. Society for Neuroscience, 2016.
- Hardy NF, Romero JL, Buonomano DV. *Temporal scaling in functionally feedforward recurrent neural network models.* Society for Neuroscience, 2015.
- Hardy N, Buonomano DV. *Novel regimes in dynamic attractor networks account for timing and temporal warping.* Washington, DC: Society for Neuroscience, 2014.
- Hardy N, Martinowish K, Jimenez D, Schloesser R, Weinberger DR, Lu B. *Contribution of activity-dependent BDNF signaling to cortical slow-wave activity. Program No. 45.08. 2011 Neuroscience Meeting Planner.* Washington, DC: Society for Neuroscience, 2011.
- Hardy N, Martinowich K, Jimenez D, Schloesser R. *Suppression of Adult Neurogenesis Causes Neuroplastic Changes in the CA3 Sub-region of the Hippocampus.* Program No. 31.12. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010.

Chapter 1: Introduction

The ability to process temporal information is necessary for normal behavior. For example, activities such as tying your shoes or handwriting require precise muscle coordination, while listening to music involves processing the spectral structure of sound over time. However, unlike other sensory modalities (vision, hearing, etc), we have no sensory organ that detects the passage of time. Therefore, much of the research on temporal processing attempts to elucidate how the brain generates internal representations of temporal information.

We process time using multiple mechanisms, depending on the time scale involved. The most familiar mechanism, the circadian clock, is one of the slowest. It tracks the passage of hours during a day to regulate appetite, the sleep-wake cycle, and other bodily functions. Circadian rhythms are generated via transcription-translation autoregulatory feedback loops and light-sensitive inputs to the suprachiasmatic nucleus. On the other extreme, sound localization relies on computations about nine orders of magnitude faster than circadian rhythms. A sound wave will reach your right and left ears at different times depending in part on your head's width and its orientation relative to the sound's source. Your ears then generate differential responses in the left and right auditory nerves, which converge at the medial superior olive (MSO). To calculate the horizontal position of the source of a sound, the MSO detects microsecond differences in the arrival of action potentials from the two auditory nerves. Between these two extremes, the brain must encode time at the timescale of milliseconds and seconds to generate movement, perceive speech and motion, and consciously estimate the passage of time. The research I present here examines timing at this scale.

When thinking about research on temporal processing, it can be useful to parse a particular study into one of two categories: sensory or motor timing. Studies of sensory timing examine how the brain encodes temporal information about sensory inputs. Sound, for example, is intrinsically

dynamic, requiring the auditory system to extract its temporal information to detect even simple musical notes. On the other hand, movement requires the brain to generate time-varying signals that can coordinate the activation of different muscle groups. Though these categories provide a useful heuristic, it is important to note that they are often not independent: studies of motor timing use sensory cues to deliver task information, and studies of sensory timing often require a motor response. Another type of study, interval timing, bears mentioning. The goal of interval timing experiments is to examine the subject's ability to estimate the passage of time. In many studies of interval timing, subjects are asked to replicate a cued interval by pressing a button at the appropriate time. Using this protocol, researchers can directly examine a subject's sense of time and its underlying neural mechanisms. For this reason, interval timing is one of the most common methods of studying temporal processing.

At the time of this writing there is no consensus for how the brain tracks the passage of milliseconds and seconds. One reason for this is that, as mentioned above, there is no sensory organ that directly detects time at this scale, making it exceeding hard to manipulate. Thus, while vision researchers can modify the input to the visual system, research on temporal processing must infer how subjects track time by asking them to perform time dependent tasks. Another difficulty arises from the temporal dependency of nearly everything we do. Even when looking at a picture, your eyes will generally make saccades to different parts of the image, fixating on a single point for relatively short periods. Consequently, it is difficult to disentangle temporal processing from the myriad of other computations that the brain performs. Fascinatingly, however, the ubiquity of temporal processing in the brain has inspired theories that use the dynamics inherent to neural networks to tell time. Importantly, as these theories are built up from basic principles, they do not impose particular functions on neural networks: information is encoded via the emergent properties of the system (Buzsáki and Llinás, 2017).

In this chapter, I review relevant background on temporal processing and neural network dynamics in the brain, laying the groundwork for the original research presented in later chapters. In general, I attempt to provide insight into how the dynamics of neural networks can encode the temporal information necessary for sensory and motor processing, and account for certain important aspects of timing behavior. To do so, I discuss both theoretical and experimental studies of neural function, building on the idea that neural computations, temporal and otherwise, are an emergent property of neural networks.

1.1: Sensory timing

The role of time in sensory perception is most prominent in the auditory system. To understand spoken words, the brain must recognize each syllable at a rate of one syllable every 200-400 milliseconds (ms) and store them in the proper order. Errors in processing the temporal context of speech can also alter semantic content. For example, in the song “Lucy In The Sky With Diamonds” John Lennon sings the lyric “the girl with kaleidoscope eyes” at a reduced cadence, often creating the humorously misperceived phrase “the girl with colitis goes by.” At an even finer time scale, differences in voice-onset time (the time between air release and vocal cord activation) changes a “ba” syllable to “pa” (Mauk and Buonomano, 2004).

An additional layer of complexity emerges when you consider the interaction of multiple sensory systems. Speech processing commonly involves integrating visual and auditory information, as demonstrated in the McGurk illusion for example. In this illusion, utterances such as “ba” and “pa” can be confused when a normally-sighted subject cannot observe the speaker’s lip movements (Macdonald and McGurk, 1978). We also expect a certain temporal relationship between vision and sound, which makes watching movies with desynced sound and video tracks unsettling. In a study of this phenomenon, researchers presented subjects with a computer-generated image of moving lips and a sound that preceded, followed, or coincided with the

initiation of lip movement. The subjects were asked to indicate whether a test stimulus differed from a synchronized control (McGrath and Summerfield, 1985). Their results indicate that humans are more sensitive to visual rather than auditory delays. In other words, when sound preceded lip movement, subjects were able to correctly detect changes in the stimulus at smaller time differences (μ = 78.5 ms) than when movement preceded the sound (μ = 137.8 ms). This implies that the temporal relationship between vision and sound is learned through experience. In daily life, we often encounter longer sound delays when we are further from the sound's source. However, the converse should never occur: a sound cannot reach your ears before your retina detects the related image (assuming you're looking in the correct direction). Discovering how the brain learns the temporal relationship between stimuli is a fundamental issue in temporal processing and learning and memory overall.

1.2: Motor timing

To generate fine movement, the brain activates agonist muscles to initiate motion and antagonist muscles to halt it. This requires temporally precise coordination of the activation of each muscle, with a resolution of tens of milliseconds (Mauk and Buonomano, 2004). The timing becomes increasingly complex depending on the kinematics of a particular motion. In the simplest case, e.g. movements using only a single joint, the kinematics are relatively straightforward. However, when more joints are used the complexity of the movement increases significantly, requiring increasingly complex control of muscle contraction (Scott, 2004).

The mammalian brain region responsible for controlling fine movement is the motor cortex (MC) (Kakei et al., 1999; Scott, 2003). MC is organized topographically, in that the areas of MC responsible for controlling each muscle group form a map of the body (a homunculus) on the surface of the cortex. Therefore, producing a particular movement requires the brain to generate specific spatiotemporal patterns of neural activity in MC. Studies of motor control have confirmed

this, showing that the kinematics of movement are reflected in the network dynamics of motor cortex (Shenoy et al., 2013; Sussillo et al., 2015). In addition, motor control often requires the brain to make quick adjustments to a motor pattern on the fly. To do so, the brain continuously integrates time-varying feedback from sensory systems to optimize a motor trajectory. Several studies have examined this by asking subjects to use one hand to reach for a target. When experimenters perturbed the motion by externally applying either transient or continuous forces to the arm, subjects were able to correct for the perturbation and maintain smooth motion to reach the target point (Scott, 2004). These corrections require alterations in the spatiotemporal structure of MC network dynamics, illustrating the complexity and robustness of the motor timing system.

The motor system also learns to anticipate certain events and generate appropriately timed conditioned responses. In classic studies, experimenters conditioned rabbits to associate a tone with a delayed puff of air to the eye. During conditioning, the tone followed the air puff at a fixed interval ranging from 100 milliseconds to 3 seconds. After repeated pairings, the subjects acquired a condition response to the tone: they blinked in anticipation of the air puff, at latencies that reflected the trained pairing interval. A large body of work has shown that this form of motor learning is mediated by the cerebellum (Mauk and Buonomano, 2004). This aspect of motor timing is discussed in more detail below, in the section on neural signatures of temporal processing in the cerebellum.

1.3: Temporal processing in disease

Temporal processing deficits are implicated in a number of neurological disorders, underscoring the significance of timing in brain function. Aphasic developmental disorders have long been believed to be related to deficits in temporal processing (Tallal and Piercy, 1973). Specifically, patients with language-based learning impairments are deficient in distinguishing syllables when played at high speeds. These deficits are related to altered timing of auditory

processing in the cortex (Abrams et al., 2009) and are ameliorated by training (Merzenich et al., 1996). Interval timing, i.e. the ability to process stimulus duration, is associated with other developmental disorders, including Autism Spectrum Disorders and Attention-Deficit/Hyperactivity Disorder (Falter and Noreika, 2011). Neurodegenerative disorders including Parkinson's Disease and Huntington's Disease are associated with deficits in controlling rhythmic motor behavior (Freeman et al., 1996; Nakamura et al., 1978) and discriminating temporal intervals (Artieda et al., 1992; Beste et al., 2007). Schizophrenic patients and subjects at high risk for Schizophrenia also show interval timing deficits at a range of durations and in multiple sensory domains (Carroll et al., 2009; Densen, 1977; Penney et al., 2005). Though the causal relationship between temporal processing deficits and these disorders is unknown, understanding the neurobiological mechanisms of timing will provide insight into the disruptions related to all of these diseases and further the development of novel treatments.

1.4: Neural signatures of temporal processing

1.4.A: Birdsong:

Birdsong has a rich temporal structure at the scale of tens to hundreds of milliseconds. The smallest units of a song's structure are notes, which are grouped together to form syllables, then phrases, and finally songs (Doupe and Kuhl, 1999). The structure of a song is often highly consistent, requiring temporally precise motor control.

The avian brain regions responsible for learning and generating songs have been identified, and shown to exhibit temporally selective neural activity (Doupe and Kuhl, 1999). For example, some neurons in area HVC of the songbird brain are selective to the order of a song's syllables. When two sequential syllables (A-B) from a bird's own song are played back to it, these neurons will fire, but not when the order is reversed (B-A), nor when either syllable is played alone (Lewicki and Arthur, 1996).

In addition, neurons in the same brain region control the precise timing required for generating song. During singing, the network dynamics of the HVC exhibit a sequential structure, with neurons bursting in a stereotyped series (Long et al., 2010). This pattern of activity is believed to be controlled by “synaptic chains,” or synfire chains, in which groups of neurons generate bursts of activity in downstream neurons via feedforward connections (**Fig. 1.1**).

1.4.B: Cerebellum:

Though it is well known for its role in motor control, the cerebellum also plays an important role in temporal processing. For example, patients with lesions of the lateral cerebellum show impaired rhythmic movement resulting from errors in timing rather than simply poor motor control (Ivry et al., 1988; Mauk and Buonomano, 2004). In addition, patients with cerebellar damage are worse at discriminating different intervals than healthy controls (Ivry and Keele, 1989).

Because its microcircuitry is well known, it is relatively tractable to understand the neural network mechanisms of temporal processing in the cerebellum. Probably the best understood example is learning timed conditioned responses, such as eyeblink conditioning mentioned above in the Motor timing section. In this paradigm, animals are conditioned to blink in response to a brief tone (CS) which predicts the timing of an air puff (US) directed at the eye. Learning the temporal relationship between the CS and US is dependent on the cerebellar cortex: when experimenters lesioned the cerebellar cortex of trained animals, the timing of the conditioned response was abolished (Perrett et al., 1993). During conditioning, information about the CS is conveyed via mossy fibers from the pons, while US information is conveyed by the climbing fibers from the inferior olive. Theoretical studies have predicted that sets of granule cells in the cerebellar cortex are activated at different times following the CS, encoding temporal information which can be used to control learning (Medina and Mauk, 1999, 2000).

1.4.C: Basal ganglia

Multiple lines of evidence suggest that the structures of the basal ganglia are involved in temporal processing. In humans, patients with Parkinson's Disease (PD) are often studied to measure the effects of basal ganglia disruptions on timing behavior and perception. PD results from the degeneration of dopaminergic neurons in the substantia nigra that project to the striatum (among other brain regions), making it a useful model for basal ganglia dysfunction. Though PD patients do not show deficits in implicit timing (e.g. predicting trajectories of moving objects), they are impaired when explicitly discriminating intervals. The patient's impairment is correlated with the severity of the PD symptoms, and attenuated by L-DOPA treatment (Artieda et al., 1992; Coull et al., 2011). Surprisingly, as PD is primarily viewed as motor disease, motor timing in PD is not universally impaired, though evidence points to a subpopulation of PD patients with motor timing deficits (Merchant et al., 2008).

Using rodents and non-human primates, researchers have been able to measure neural activity in the basal ganglia associated with temporal processing. Many of these studies have examined cortico-thalamic-basal ganglia loops, i.e. networks formed between cortical regions and the striatum involved in time-dependent tasks. Their findings describe population dynamics that can be used to reliably infer task-relevant temporal information across trials (Bakhurin et al., 2017; Jin et al., 2009; Mello et al., 2015). These findings indicate that temporal coding in the basal ganglia may emerge via integrating the input of multiple cortical regions (Bakhurin et al., 2017; Merchant et al., 2013). The striatum has also been implicated in generating scalable population codes, in which the neural activity changes to reflect the target duration (e.g. the duration of firing increases with longer durations), an important but complex feature of temporal processing (Mello et al., 2015). Such temporally scaled activity is again proposed to be mediated by cortical activity shaping synaptic plasticity in the striatum (Murray and Escola, 2017). The mechanisms underlying temporal scaling in the cortex are addressed by original research discussed later in this work.

1.4.D: Hippocampus:

Determining the role of the hippocampus in temporal processing has been especially difficult because of its prominent role in spatial memory. To study spatial processing in the hippocampus, researchers often train rodents to navigate environments with specific spatial cues which the animal can use to generate a map of the environment. As the animal navigates the environment, it therefore receives continuously changing spatial information, which confounds attempts to determine whether the hippocampus generates an internal representation of time. To overcome this problem, one of the first studies of temporal processing in the hippocampus used a behavioral task with separate spatial and temporal components (Pastalkova et al., 2008). In this study, the experimenters trained rats to alternate between navigating the left and right arms of a figure-eight maze. Between maze runs, rats were also trained to run steadily in the same direction on a running wheel for 10-20 seconds. Thus, during the delay period each subject had to remember the direction of the previous run in order to choose the correct arm of the maze. Recording from neurons in area CA1 of the hippocampus, Pastalkova and colleagues found cells which activated preferentially at different times of the delay epoch, creating a continuous representation of time (**Fig. 1.2**) Extending these results, MacDonald et al. (2011) showed that hippocampal activity can also represent intervals even when making non-spatial decisions. Interestingly, the authors found that the activity of some neurons which fired selectively at a certain location (i.e. place cells) was also modulated by time. Taken together, studies of time-dependent processing show that the hippocampus integrates both spatial and internally generated temporal information to form associations between events, going beyond its traditionally understood role in spatial memory (Eichenbaum, 2017).

1.4.E: Cortex:

There is mounting evidence showing that the cortex processes temporal information in a distributed manner. An early study of human stroke patients demonstrated a role for the right, but not left, hemisphere prefrontal–inferior parietal network in interval discrimination, despite attention deficits in both groups (Harrington et al., 1998). Another study using PET and a visual interval discrimination task by Maquet et al. (1996) argued that the temporal and spatial components of a visual stimulus are processed in the same areas of visual cortex. Even more converging evidence using fMRI has shown a primary role for the supplementary motor area in interval timing (Merchant et al., 2013). Thus, several lines of evidence indicate that temporal processing can be performed by multiple cortical areas, depending on the task parameters.

Many studies of timing ask the subject to estimate the passage of time by pressing a button. Thus, motor areas are obvious candidates for examining the neural activity underlying temporal processing. Using the synchronization-continuation task (SCT), Hugo Marchant's groups has examined whether medial premotor cortex (MPC) activity represents temporal information. The SCT is uses a compelling design in which subjects are first cued to press a button at a certain frequency and then continue the rhythm autonomously a certain number of times. Thus, subjects must remember temporal information at both local (the inter-tap interval) and global (the number of produced taps) scales. The results of several SCT studies using multiple tapping frequencies have demonstrated that MPC activity dynamically represents both the local and global structure of the movement via the sequential activation of neural ensembles (Crowe et al., 2014; Merchant and Averbeck, 2017).

Separate studies have found evidence for temporal processing in higher order cortical areas. In the "Ready, Set, Go" task (RSG), subjects must first measure a temporal interval demarcated by two brief sensory cues and then reproduce the cued interval by making a saccade, thus

requiring the integration of sensory and motor timing. Using the RSG, researchers found evidence that activity in the lateral intraparietal cortex (LIP), which is believed to play a role in sensory-motor integration, represents temporal information about both the measured and produced interval (Jazayeri and Shadlen, 2015). Further, using a version of the RSG in which subjects must scale the duration of produced interval, the same researchers found activity in medial frontal cortex that itself scales with the target interval (Wang et al., 2017). This important study was the first to show evidence for “temporal scaling” of recorded neural activity, an important question in the timing field eluded to above.

Even primary sensory cortical areas have exhibited evidence of temporal processing, challenging the canonical idea that such areas only compute lower-order sensory information, such as the orientation of lines. In rodents, Shuler and Bear (2006) demonstrated that neurons in primary visual cortex develop responses that predict the timing of reward. The same researchers further showed that learning such timed responses is likely mediated by cholinergic inputs to the cortex, generating ramping firing rates that reflect the stimulus-reward pairing interval (Chubykin et al., 2013; Gavornik et al., 2009; Namboodiri et al., 2015).

It is essential to understand the role of the cortex in producing behavioral timing, and the work described above provides strong evidence that it is crucially involved in such tasks. However, those studies do not answer another important question regarding the origins of temporal computation in the brain: is the cortex able to perform these computations autonomously, or does it require input from other brain regions? To examine this issue, researchers have used *in vitro* studies of cortical slices, allowing them examine temporal computations inherent to cortical networks. In line with theoretical studies about recurrent neural networks, these studies have revealed that organotypic slice cultures can be trained to produce activity that represents a trained stimulus interval (Buonomano and Maass, 2009; Goel and Buonomano, 2016; Johnson et al.,

2010). This activity was specific to the trained stimulus and could encode multiple trained intervals. Moreover, the researchers demonstrated that timed cortical activity was generated by dynamic shifts in the balance of excitation and inhibition, supporting theoretical predictions about the origins of timing (Buonomano, 2000; Goel and Buonomano, 2016; Hardy and Buonomano, 2017). Together, studies of the cortex provide strong evidence that it is not only crucially involved in timing, temporal computations are an emergent property of recurrent cortical circuits.

1.5: Theories of sub-second temporal processing

Temporal processing is a complex phenomenon, making simplified computational models especially useful in uncovering its underlying neural mechanisms. Several classes of models proposing various mechanisms exist, which, though they implement distinct phenomenological examples of temporal processing, complement each other to provide a framework for understanding timing in the brain. All of these models should be assessed on their ability to account for observed features of timing behavior and neural activity. The most prominent such feature is the scalar property, or Weber's law, which states the variability in time estimation increases linearly with elapsed time (Church et al., 1994; Gibbon, 1977). However, Weber's law has been difficult to implement in many biologically inspired timing models (Hass and Durstewitz, 2014). Other criteria, such as generalization of time estimation across sensory modalities, are assessed less often but provide insight into a specific modeling approach and timing task. No model will incorporate every aspect of timing behavior. However, all of them should be able to either provide an intuitive understanding of temporal processing, generate testable predictions, or both. This section discusses a few prominent classes of biologically inspired timing models, that is, models that are implemented with some form of neural network.

1.5.A: Delay lines:

One of the simplest ways time might be represented in neural networks is using a delay line, in which activity spreads between neural ensembles at some relatively constant rate. Simple implementations use feedforward synfire chains, which are classically composed of distinct ensembles of neurons that sequentially propagate activity from one ensemble to the next following some triggering event (Abeles, 1991; Diesmann et al., 1999; Haß et al., 2008). Similar to a singly linked list in computer science, this type of network can only propagate activity in one direction. Thus, it is possible to determine the elapsed time since the initial trigger event by observing which ensemble is currently active.

It is easy to see how a synfire chain (or any implementation of a delay line) could be used to detect a 100 ms interval by forming an "and" gate (which activates an output only during coincident activation of two input neurons) between an ensemble active at 100 ms and some sensory input. Importantly, these simple networks can be used to control much more complex activity, such as movement. For example, in songbirds synfire chains are believed to underlie the complex forms of timing necessary for song generation, as briefly discussed above. This timing appears to be generated by sequential bursting of a subset of neurons in the songbird sensorimotor nucleus HVC. These HVC neurons fire at specific moments during a song, providing the timing necessary for the structure of each syllable and the sequence of syllables within a song (Hahnloser et al., 2002; Long and Fee, 2008). Using *in vivo* recordings and spiking neural network simulations, Long et al. (2010) provided evidence that these firing patterns are consistent with a feedforward synfire chain network architecture (**Fig. 1.1**).

Activity reminiscent of feedforward networks has been observed in multiple brain areas characterized by recurrent connectivity, including the cortex (Harvey et al., 2012; Song et al., 2005). These observations have in turn necessitated to the development of new theories of

sequential activity. While synfire chains can in principle incorporate recurrent connections, in practice they are typically implemented within purely feedforward architectures. Consequently, it is highly unlikely that cortical networks are actually feedforward synfire chains. However, it is unclear how a recurrent neural network may implement a delay line. A network's structure can also impact its capacity to encode temporal information. Specifically: how many trajectories of a given duration can one feedforward synfire network encode? In one sense the capacity is low. For example, in a purely feed-forward network, if we assume that each neuron only fires once and must participate in every pattern, then the capacity is essentially one trajectory (which is not the case in a recurrent network given these same assumption (Liu and Buonomano, 2009)). However, if we assume that different subpopulations of neurons within a pool fire during different trajectories, then the capacity increases significantly (Herrmann et al., 1995).

Several theoretical studies have examined the mechanisms underlying experimentally observed sequential activity by implementing delay lines in networks with recurrent activity (including a study discussed later in this work). In one the authors develop a "controller" network which could account for sequential activity observed in hippocampal area CA1 (Pastalkova et al., 2008). The controller network was designed to have excitatory and inhibitory connections structured to resemble a Laplacian-of-gaussian (LOG), such that peaks of excitation are bounded by troughs of inhibition (Itskov et al., 2011). This network was able to produce "moving bumps" of activity (i.e. activations of overlapping neural ensembles) that propagated in distinct trajectories depending on which ensemble was initially activated. Thus, the network could uniquely encode elapsed time from multiple inputs. The authors proposed that this architecture could be implemented by hippocampal area CA3, which is characterized by recurrent connectivity and projects onto CA1. In another study examining the origins of experimentally observed sequential activity in the cortex, the authors trained a recurrent neural network model (RNN) to precisely reproduce the observed cortical activity. In this case, the authors did not define a specific synaptic

architecture *a priori* but used a supervised learning rule to adjust the weights of an initially random network. After training, the RNN contained LOG connectivity patterns which generated the desired activity (Rajan et al., 2016). However, because that study provided time-dependent input to the network it did not encode time *per se*. An original study discussed later in this work addresses this issue.

1.5.B: State dependent computations

One of the first neurobiologically-based models of timing and temporal processing proposed that networks of neurons are intrinsically able to tell time as a result of dynamic changes in the state of neural networks (Buonomano and Mauk, 1994; Buonomano and Merzenich, 1995; Mauk and Donegan, 1997). At the sensory level, the hypothesis is that the discrimination of temporal intervals arises from the interaction between the internal-state of a network and incoming stimuli. In this sensory mode, the recurrent weights of these networks are generally fairly weak—that is, not capable of sustaining self-perpetuating activity. Thus, much of the temporal information emerges from neural and synaptic properties that are naturally time-varying (the so-called hidden states—e.g. short-term synaptic plasticity). Such models have been shown to effectively discriminate not only simple intervals, but complex temporal patterns as well (Buonomano, 2000; Karmarkar and Buonomano, 2007; Lee and Buonomano, 2012; Maass et al., 2002). The hypothesis is that each sensory event interacts with the current state of the network, forming a pattern of network states that naturally encodes each event in the context of the recent stimulus history—much as the ripples generated by each raindrop falling in a pond will interact with the ripples created by previous raindrops. Experimental studies have supported this hypothesis by demonstrating that cortical networks contain information about not only the current stimulus, but also the interval and order of recent events (Buonomano et al., 1997; Dranias et al., 2015; Kilgard and Merzenich, 2002; Nikolić et al., 2009; Zhou et al., 2010).

The same general framework has also been applied to timing in the motor domain (Buonomano and Laje, 2010; Laje and Buonomano, 2013; Mauk and Donegan, 1997; Medina and Mauk, 2000). In contrast to sensory timing, motor timing relies on the active production of a response at the appropriate interval after a start cue. Therefore, in the motor regime, the recurrent connections need to be relatively strong, i.e. capable of self-perpetuating activity. In state-dependent models of motor behavior, time is encoded in the dynamically changing patterns of active neurons, forming a *population clock* (Buonomano and Karmarkar, 2002). The activity in the network traces out a trajectory in neural state space, in which each point in time corresponds to a unique population of active neurons. These patterns can be sparse: a few neurons activated at any point in time and each neuron activated at only one point, or "dense": with many neurons activated at a time, and each neuron potentially active at different points in the same trajectory. Experimental studies have reported numerous examples of either sparse functional feed-forward patterns of activity (Hahnloser et al., 2002; Long et al., 2010; MacDonald et al., 2013; Mello et al., 2015), or complex high-entropy patterns (Crowe et al., 2010, 2014; Jin et al., 2009; Kim et al., 2013) of activity that encode time. A recent experimental and computational study also provided support for the notion that time is represented in high-dimensional trajectories (Carnevale et al., 2015). In this work, recordings from over 100 neurons in the premotor cortex revealed a neural trajectory that evolved over a period of seconds during a task in which monkeys expected a reward between 1.5 and 3.5 seconds after the start cue. Analysis suggested that the reward window was represented in a trajectory segment, and that temporal expectation was intrinsically represented because this segment was the closest to a boundary that, if crossed, triggered a motor response.

1.6: Conclusions and goals of the dissertation

In this chapter I have outlined a selection of key concepts in the neuroscience of temporal processing. It is hopefully clear from this discussion that timing is involved in many behaviors, and that there are a variety of potential mechanisms underlying temporal processing. This behavioral

and mechanistic variety has made research on timing challenging, and the field is still searching for unifying laws that describe how the brain tracks time. However, the timing field seems to be converging on a consensus that there likely is not a singular source of timing, but that it is generated in a distributed manner by circuits throughout the brain (Hardy and Buonomano, 2016; Hass and Durstewitz, 2016; Merchant et al., 2013).

The central thesis of this dissertation is that the cortex processes temporal information using state dependent computations (Buonomano and Maass, 2009). Arising from the study of dynamical systems, the state dependent theory of temporal processing argues that timing is an intrinsic property of neural networks: the state of network activity evolving over time can be used to encode temporal information. State dependent temporal processing is an appealing framework because it uses basic principles to build neural networks that can encode time, rather than designing specific timers to meet the demands of a task. In essence, the theory argues that temporal computations are an emergent property of neural network dynamics. In addition, state dependent processing can act as a general-purpose timer, accounting for many experimentally observed neural signatures of timing. In fact, many other theoretical models of timing can be described as specific cases of the state dependent framework (Hass and Durstewitz, 2016).

The original research I present in the following chapters examines the state dependent timing hypothesis by studying the role of neural network dynamics in temporal processing. Chapter 2 presents research examining sequential (or delay line) activity observed in brain regions characterized by recurrent connectivity. This study presents a theoretical mechanism that recurrent neural networks may use to autonomously produce sequential activity, termed “functionally feedforward”, and encode temporal information. Next, Chapter 3 examines the mechanisms of producing the same complex movement at a variety of speeds, a fundamental feature of motor timing. This study combines theoretical and psychophysical studies to predict

and test a novel feature motor timing: temporal accuracy improves with speed, termed the Weber-speed effect. Finally, Chapter 4 examines how cortical neural networks encode temporal information. Using organotypic slice cultures, this study demonstrates that the cortex processes temporal input patterns in a state dependent manner, supporting theoretical predictions. Overall, this dissertation presents novel findings on the mechanisms of temporal processing in the brain, advancing the knowledge of the timing field and neuroscience in general.

1.7: Figures

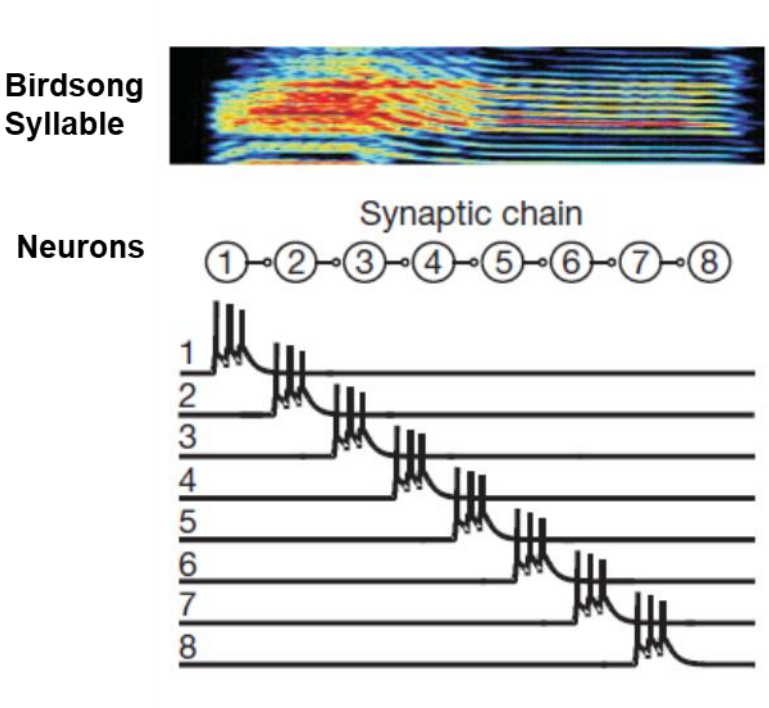


Figure 1.1. Timing with synfire chains.

Neurons in the songbird HVC nucleus burst sparsely at specific points in a song's syllable. Long and Fee (2010) suggest that the temporal structure both within a syllable and throughout a song may be generated by a synfire chain-like network. Here the synfire chain is schematized by bursting neurons connected in a feedforward architecture. Modified from (Long et al., 2010).

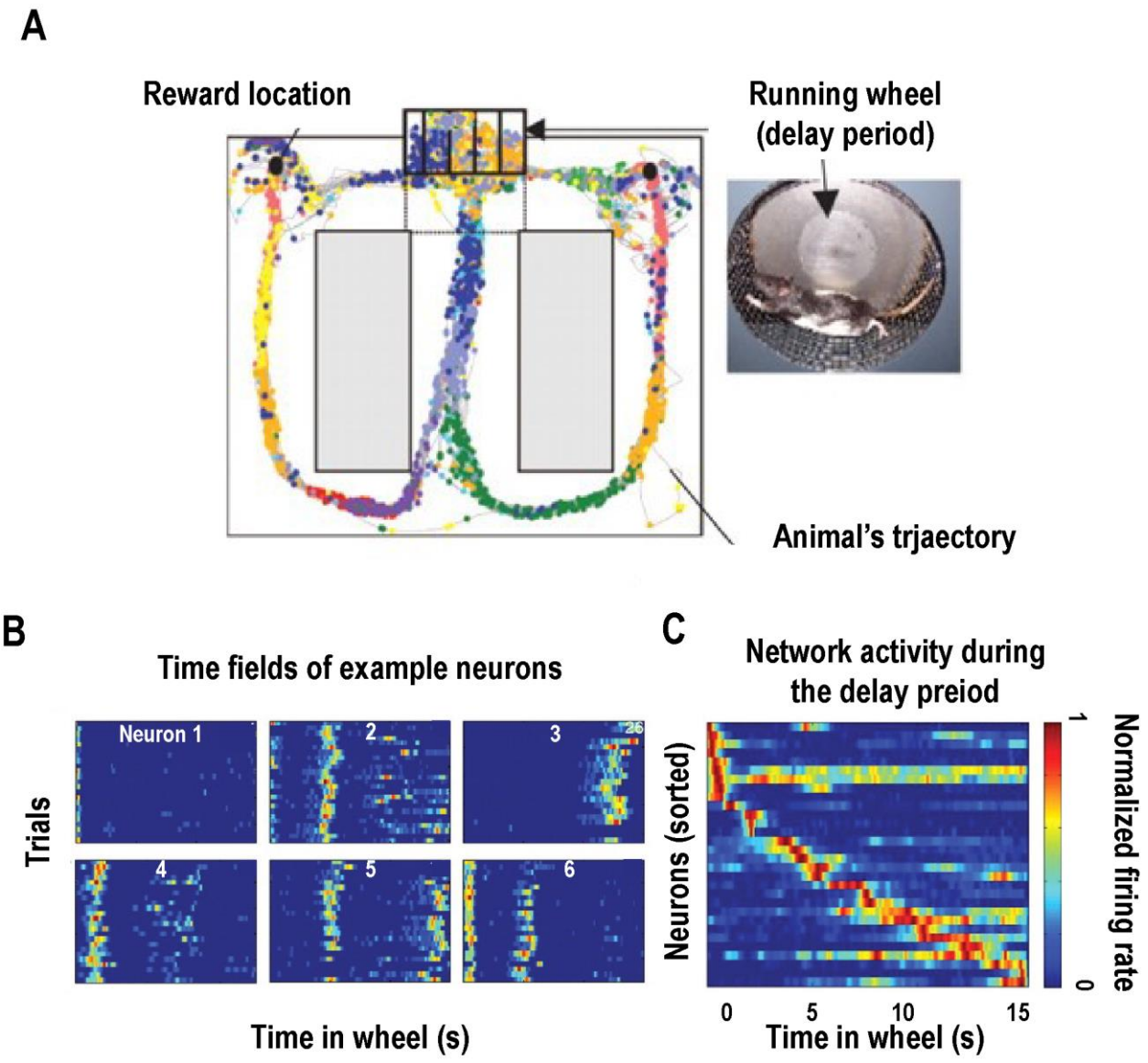


Figure 1.2. Sequential firing of hippocampal neurons during a behavioral delay period

A) Rats were trained to alternate between the left and right arms of a figure-8 maze to receive a reward (black dots). Between trials, rats had to run for several seconds on a running wheel, requiring them to remember the correct arm to navigate when they exited the running wheel. **B**) Normalized firing rates of six simultaneously recorded neurons during wheel running over trials. The neurons show remarkably precise tuning for time during the delay period, reaching their maximum firing rate after consistent amounts of time in the running wheel. **C**) Normalized firing rate of 30 simultaneously recorded neurons during wheel running, ordered by the latency of their peak firing rate. When sorted in this way, the activity of the neurons tiles the delay period. Adapted from (Pastalkova et al., 2008).

1.8: References cited

Abeles, M. (1991). Corticonics: Neural Circuits of the Cerebral Cortex (Cambridge University Press).

Artieda, J., Pastor, M.A., Lacruz, F., and Obeso, J.A. (1992). Temporal discrimination is abnormal in Parkinson's disease. Brain J. Neurol. *115 Pt 1*, 199–210.

Bakhurin, K.I., Goudar, V., Shobe, J.L., Claar, L.D., Buonomano, D.V., and Masmanidis, S.C. (2017). Differential Encoding of Time by Prefrontal and Striatal Network Dynamics. J. Neurosci. *37*, 854–870.

Beste, C., Saft, C., Andrich, J., Müller, T., Gold, R., and Falkenstein, M. (2007). Time processing in Huntington's disease: a group-control study. PloS One *2*, e1263.

Buonomano, D.V. (2000). Decoding temporal information: A model based on short-term synaptic plasticity. J. Neurosci. Off. J. Soc. Neurosci. *20*, 1129–1141.

Buonomano, D.V., and Karmarkar, U.R. (2002). Book Review: How Do We Tell Time? The Neuroscientist *8*, 42–51.

Buonomano, D.V., and Laje, R. (2010). Population clocks: motor timing with neural dynamics. Trends Cogn. Sci. *14*, 520–527.

Buonomano, D.V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. Nat. Rev. Neurosci. *10*, 113–125.

Buonomano, D.V., and Mauk, M.D. (1994). Neural Network Model of the Cerebellum: Temporal Discrimination and the Timing of Motor Responses. Neural Comput. *6*, 38–55.

Buonomano, D.V., and Merzenich, M.M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. Science *267*, 1028–1030.

Buonomano, D.V., Hickmott, P.W., and Merzenich, M.M. (1997). Context-sensitive synaptic plasticity and temporal-to-spatial transformations in hippocampal slices. Proc. Natl. Acad. Sci. *94*, 10403–10408.

Buzsáki, G., and Llinás, R. (2017). Space and time in the brain. Science *358*, 482–485.

Carnevale, F., de Lafuente, V., Romo, R., Barak, O., and Parga, N. (2015). Dynamic Control of Response Criterion in Premotor Cortex during Perceptual Detection under Temporal Uncertainty. Neuron *86*, 1067–1077.

Carroll, C.A., O'Donnell, B.F., Shekhar, A., and Hetrick, W.P. (2009). Timing dysfunctions in schizophrenia span from millisecond to several-second durations. Brain Cogn. *70*, 181–190.

Chubykin, A.A., Roach, E.B., Bear, M.F., and Shuler, M.G.H. (2013). A cholinergic mechanism for reward timing within primary visual cortex. Neuron *77*, 723–735.

Church, R.M., Meck, W.H., and Gibbon, J. (1994). Application of scalar timing theory to individual trials. J. Exp. Psychol. Anim. Behav. Process. *20*, 135–155.

Coull, J.T., Cheng, R.-K., and Meck, W.H. (2011). Neuroanatomical and Neurochemical Substrates of Timing. Neuropsychopharmacology *36*, 3–25.

Crowe, D.A., Averbeck, B.B., and Chafee, M.V. (2010). Rapid Sequences of Population Activity Patterns Dynamically Encode Task-Critical Spatial Information in Parietal Cortex. J. Neurosci. *30*, 11640–11653.

Crowe, D.A., Zarco, W., Bartolo, R., and Merchant, H. (2014). Dynamic Representation of the Temporal and Sequential Structure of Rhythmic Movements in the Primate Medial Premotor Cortex. J. Neurosci. *34*, 11972–11983.

Densen, M.E. (1977). Time perception and schizophrenia. Percept. Mot. Skills *44*, 436–438.

Diesmann, M., Gewaltig, M.-O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. Nature *402*, 529–533.

Doupe, A.J., and Kuhl, P.K. (1999). BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms. Annu. Rev. Neurosci. *22*, 567–631.

Dranias, M.R., Westover, M.B., Cash, S., and VanDongen, A.M.J. (2015). Stimulus information stored in lasting active and hidden network states is destroyed by network bursts. Front. Integr. Neurosci. *9*.

Eichenbaum, H. (2017). On the Integration of Space, Time, and Memory. Neuron *95*, 1007–1018.

Falter, C.M., and Noreika, V. (2011). Interval Timing Deficits and Abnormal Cognitive Development. Front. Integr. Neurosci. *5*.

Freeman, J.S., Cody, F.W., O'Boyle, D.J., Craufurd, D., Neary, D., and Snowden, J.S. (1996). Abnormalities of motor timing in Huntington's disease. Parkinsonism Relat. Disord. *2*, 81–93.

Gavornik, J.P., Shuler, M.G.H., Loewenstein, Y., Bear, M.F., and Shouval, H.Z. (2009). Learning reward timing in cortex through reward dependent expression of synaptic plasticity. Proc. Natl. Acad. Sci. *106*, 6826–6831.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. Psychol. Rev. *84*, 279–325.

Goel, A., and Buonomano, D.V. (2016). Temporal Interval Learning in Cortical Cultures Is Encoded in Intrinsic Network Dynamics. Neuron *91*, 320–327.

Hahnloser, R.H.R., Kozhevnikov, A.A., and Fee, M.S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature *419*, 65–70.

Hardy, N.F., and Buonomano, D.V. (2016). Neurocomputational models of interval and pattern timing. Curr. Opin. Behav. Sci. *8*, 250–257.

Hardy, N.F., and Buonomano, D.V. (2017). Encoding Time in Feedforward Trajectories of a Recurrent Neural Network Model. Neural Comput. 1–19.

Harrington, D.L., Haaland, K.Y., and Knight, R.T. (1998). Cortical Networks Underlying Mechanisms of Time Perception. J. Neurosci. *18*, 1085–1095.

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

Hass, J., and Durstewitz, D. (2014). Neurocomputational models of time perception. Adv. Exp. Med. Biol. *829*, 49–71.

Hass, J., and Durstewitz, D. (2016). Time at the center, or time at the side? Assessing current models of time perception. Curr. Opin. Behav. Sci. *8*, 238–244.

Haß, J., Blaschke, S., Rammsayer, T., and Herrmann, J.M. (2008). A neurocomputational model for optimal temporal processing. J. Comput. Neurosci. *25*, 449–464.

Herrmann, M., Hertz, J.A., and Prügel-Bennett, A. (1995). Analysis of synfire chains. Netw. Comput. Neural Syst. *6*, 403–414.

Itskov, V., Curto, C., Pastalkova, E., and Buzsáki, G. (2011). Cell Assembly Sequences Arising from Spike Threshold Adaptation Keep Track of Time in the Hippocampus. J. Neurosci. *31*, 2828–2834.

Ivry, R.B., and Keele, S.W. (1989). Timing Functions of The Cerebellum. J. Cogn. Neurosci. *1*, 136–152.

Ivry, R.B., Keele, S.W., and Diener, H.C. (1988). Dissociation of the lateral and medial cerebellum in movement timing and movement execution. Exp Brain Res *73*, 167–180.

Jazayeri, M., and Shadlen, M.N. (2015). A Neural Mechanism for Sensing and Reproducing a Time Interval. Curr. Biol. *25*, 2599–2609.

Jin, D.Z., Fujii, N., and Graybiel, A.M. (2009). Neural representation of time in cortico-basal ganglia circuits. Proc. Natl. Acad. Sci. *106*, 19156–19161.

Johnson, H.A., Goel, A., and Buonomano, D.V. (2010). Neural dynamics of in vitro cortical networks reflects experienced temporal patterns. Nat. Neurosci. *13*, 917–919.

Kakei, S., Hoffman, D.S., and Strick, P.L. (1999). Muscle and Movement Representations in the Primary Motor Cortex. Science *285*, 2136–2139.

Karmarkar, U.R., and Buonomano, D.V. (2007). Timing in the absence of clocks: encoding time in neural network states. Neuron *53*, 427–438.

Kilgard, M.P., and Merzenich, M.M. (2002). Order-sensitive plasticity in adult primary auditory cortex. Proc. Natl. Acad. Sci. *99*, 3205–3209.

Kim, J., Ghim, J.-W., Lee, J.H., and Jung, M.W. (2013). Neural Correlates of Interval Timing in Rodent Prefrontal Cortex. J. Neurosci. *33*, 13834–13847.

Laje, R., and Buonomano, D.V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. Nat. Neurosci. *16*, 925–933.

Lee, T.P., and Buonomano, D.V. (2012). Unsupervised Formation of Vocalization-Sensitive Neurons: A Cortical Model Based on Short-Term and Homeostatic Plasticity. Neural Comput. *24*, 2579–2603.

Lewicki, M.S., and Arthur, B.J. (1996). Hierarchical Organization of Auditory Temporal Context Sensitivity. J. Neurosci. *16*, 6987–6998.

Liu, J.K., and Buonomano, D.V. (2009). Embedding Multiple Trajectories in Simulated Recurrent Neural Networks in a Self-Organizing Manner. J. Neurosci. *29*, 13172–13181.

Long, M.A., and Fee, M.S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature *456*, 189–194.

Long, M.A., Jin, D.Z., and Fee, M.S. (2010). Support for a synaptic chain model of neuronal sequence generation. Nature *468*, 394–399.

Maass, W., Natschläger, T., and Markram, H. (2002). Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. Neural Comput. *14*, 2531–2560.

Macdonald, J., and McGurk, H. (1978). Visual influences on speech perception processes. Percept. Psychophys. *24*, 253–257.

MacDonald, C.J., Lepage, K.Q., Eden, U.T., and Eichenbaum, H. (2011). Hippocampal “time cells” bridge the gap in memory for discontiguous events. Neuron *71*, 737–749.

MacDonald, C.J., Carrow, S., Place, R., and Eichenbaum, H. (2013). Distinct Hippocampal Time Cell Sequences Represent Odor Memories in Immobilized Rats. J. Neurosci. *33*, 14607–14616.

Maquet, P., Lejeune, H., Pouthas, V., Bonnet, M., Casini, L., Macar, F., Timsit-Berthier, M., Vidal, F., Ferrara, A., Degueldre, C., et al. (1996). Brain Activation Induced by Estimation of Duration: A PET Study. NeuroImage *3*, 119–126.

Mauk, M.D., and Buonomano, D.V. (2004). The Neural Basis of Temporal Processing. Annu. Rev. Neurosci. *27*, 307–340.

Mauk, M.D., and Donegan, N.H. (1997). A model of Pavlovian eyelid conditioning based on the synaptic organization of the cerebellum. Learn. Mem. *4*, 130–158.

McGrath, M., and Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. J. Acoust. Soc. Am. *77*, 678–685.

Medina, J.F., and Mauk, M.D. (1999). Simulations of Cerebellar Motor Learning: Computational Analysis of Plasticity at the Mossy Fiber to Deep Nucleus Synapse. J. Neurosci. *19*, 7140–7151.

Medina, J.F., and Mauk, M.D. (2000). Computer simulation of cerebellar information processing. Nat. Neurosci. *3*, 1205–1211.

Mello, G.B.M., Soares, S., and Paton, J.J. (2015). A Scalable Population Code for Time in the Striatum. Curr. Biol. *25*, 1113–1122.

Merchant, H., and Averbeck, B.B. (2017). The Computational and Neural Basis of Rhythmic Timing in Medial Premotor Cortex. J. Neurosci. *37*, 4552–4564.

Merchant, H., Luciana, M., Hooper, C., Majestic, S., and Tuite, P. (2008). Interval timing and Parkinson's disease: heterogeneity in temporal performance. Exp. Brain Res. *184*, 233–248.

Merchant, H., Harrington, D.L., and Meck, W.H. (2013). Neural Basis of the Perception and Estimation of Time. Annu. Rev. Neurosci. *36*, 313–336.

Merzenich, M.M., Jenkins, W.M., Johnston, P., Schreiner, C., Miller, S.L., and Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. Science *271*, 77–81.

Murray, J.M., and Escola, G.S. (2017). Learning multiple variable-speed sequences in striatum via cortical tutoring. ELife *6*, e26084.

Nakamura, R., Nagasaki, H., and Narabayashi, H. (1978). Disturbances of rhythm formation in patients with Parkinson's disease: part I. Characteristics of tapping response to the periodic signals. Percept. Mot. Skills *46*, 63–75.

Namboodiri, V.M.K., Huertas, M.A., Monk, K.J., Shouval, H.Z., and Hussain Shuler, M.G. (2015). Visually Cued Action Timing in the Primary Visual Cortex. Neuron *86*, 319–330.

Nikolić, D., Häusler, S., Singer, W., and Maass, W. (2009). Distributed Fading Memory for Stimulus Properties in the Primary Visual Cortex. PLOS Biol. *7*, e1000260.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsáki, G. (2008). Internally Generated Cell Assembly Sequences in the Rat Hippocampus. Science *321*, 1322–1327.

Penney, T.B., Meck, W.H., Roberts, S.A., Gibbon, J., and Erlenmeyer-Kimling, L. (2005). Interval-timing deficits in individuals at high risk for schizophrenia. Brain Cogn. *58*, 109–118.

Perrett, S.P., Ruiz, B.P., and Mauk, M.D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. J. Neurosci. *13*, 1708–1718.

Rajan, K., Harvey, C.D., and Tank, D.W. (2016). Recurrent Network Models of Sequence Generation and Memory. Neuron *90*, 128–142.

Scott, S.H. (2003). The role of primary motor cortex in goal-directed movements: insights from neurophysiological studies on non-human primates. Curr. Opin. Neurobiol. *13*, 671–677.

Scott, S.H. (2004). Optimal feedback control and the neural basis of volitional motor control. Nat. Rev. Neurosci. *5*, 532–546.

Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical Control of Arm Movements: A Dynamical Systems Perspective. Annu. Rev. Neurosci. *36*, 337–359.

Shuler, M.G., and Bear, M.F. (2006). Reward Timing in the Primary Visual Cortex. Science *311*, 1606–1609.

Song, S., Sjostrom, P.J., Reigl, M., Nelson, S., and Chklovskii, D.B. (2005). Highly nonrandom feature of synaptic connectivity in local cortical circuits. PLoS Biol *3*, e66.

Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K.V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. Nat. Neurosci. *18*, 1025–1033.

Tallal, P., and Piercy, M. (1973). Defects of Non-Verbal Auditory Perception in Children with Developmental Aphasia. Nature *241*, 468–469.

Wang, J., Narain, D., Hosseini, E., and Jazayeri, M. (2017). Flexible control of speed of cortical dynamics. BioRxiv 155390.

Zhou, X., Villers-Sidani, É. de, Panizzutti, R., and Merzenich, M.M. (2010). Successive-signal biasing for a learned sound sequence. Proc. Natl. Acad. Sci. *107*, 14839–14844.

Chapter 2: Encoding Time in Feedforward Trajectories of a Recurrent Neural Network Model.

2.1: Abstract

Brain activity evolves through time, creating trajectories of activity that underlie sensorimotor processing, behavior, and learning and memory. Therefore, understanding the temporal nature of neural dynamics is essential to understanding brain function and behavior. *In vivo* studies have demonstrated that sequential transient activation of neurons can encode time. However, it remains unclear whether these patterns emerge from feedforward network architectures or from recurrent networks, and, furthermore, what role network structure plays in timing. We address these issues using a recurrent neural network (RNN) model with distinct populations of excitatory and inhibitory units. Consistent with experimental data, a single RNN could autonomously produce multiple functionally feedforward trajectories, thus potentially encoding multiple timed motor patterns lasting up to several seconds. Importantly, the model accounted for Weber's law, a hallmark of timing behavior. Analysis of network connectivity revealed that efficiency—a measure of network interconnectedness—decreased as the number of stored trajectories increased. Additionally, the balance of excitation and inhibition shifted towards excitation during each unit's activation time, generating the prediction that observed sequential activity relies on dynamic control of the E/I balance. Our results establish for the first time that the same RNN can generate multiple functionally feed-forward patterns of activity as a result of dynamic shifts in the E/I balance imposed by the connectome of the RNN. We conclude that recurrent network architectures account for sequential neural activity, as well as for a fundamental signature of timing behavior: Weber's law.

2.2: Introduction

The ability to accurately tell time and generate appropriately timed motor responses is essential to most forms of sensory and motor processing. However, the neural processes used to encode time remain unknown (Mauk and Buonomano, 2004; Buhusi and Meck, 2005; Ivry and Schlerf, 2008; Merchant et al., 2013). While the brain tells time across many scales, ranging from microseconds to days, it is on the scale of tens of milliseconds to a few seconds that timing is most relevant to sensory-motor processing and behavior. Several neural mechanisms have been proposed to account for temporal processing in this range (for reviews see Hardy and Buonomano, 2016; Hass and Durstewitz, 2016), including pacemaker/counter internal clocks (Treisman, 1963; Gibbon et al., 1984), ramping firing rates (Durstewitz, 2003; Simen et al., 2011), the duration of firing rate increases (Gavornik et al., 2009; Namboodiri et al., 2015), models that rely on the inherent stochasticity of sensory signals and neural responses (Ahrens and Sahani, 2008; Ahrens and Sahani, 2011), and finally "population clocks", in which timing is encoded in the evolving patterns of activity within recurrent circuits (Buonomano and Mauk, 1994; Mauk and Donegan, 1997; Medina and Mauk, 1999; Buonomano and Laje, 2010).

The theory that time is encoded in the dynamics of large populations of neurons has received experimental support in several brain regions including the cortex (Crowe et al., 2010; Merchant et al., 2011; Harvey et al., 2012; Kim et al., 2013; Crowe et al., 2014; Bakhurin et al., 2017), basal ganglia (Jin et al., 2009; Gouvea et al., 2015; Mello et al., 2015), hippocampus, (Pastalkova et al., 2008; MacDonald et al., 2011; Kraus et al., 2013; Modi et al., 2014), and area HVC in songbirds (Hahnloser et al., 2002; Long et al., 2010). Some of these studies report relatively simple, apparently feedforward, sequential patterns of activity in brain regions containing significant recurrent connectivity. A fundamental question is whether these patterns of activity are generated by truly feed-forward circuits, or rather by recurrent circuits generating “functionally feedforward” patterns of activity (Banerjee et al., 2008; Goldman, 2009). Here we define

functionally feedforward trajectories as those generated by recurrent neural networks, and that are characterized by sequential patterns of activation (“moving bumps”) in which any given unit only fires once during a pattern.

Synfire chains are perhaps the simplest network-based model that could account for the reports of functionally feedforward patterns of activity. Typically, synfire chain models consist of many pools of neurons connected in a feedforward manner such that activation of one pool results in the sequential activation of each downstream pool (Abeles, 1991; Diesmann et al., 1999). However, cortical circuits, where functionally feedforward activity is often observed, are characterized by recurrent connections and local inhibition, features that standard synfire chain models generally lack (Harvey et al., 2012). Moreover, the capacity of these synfire networks is limited because any given neuron generally participates in only one pattern (Herrmann et al., 1995). To address these issues, we use a model of recurrent neural networks (RNNs) to examine how they might produce functionally feedforward patterns of activity that encode time.

Previous studies of timing using RNNs have not sought to simulate experimentally observed patterns of neural activity and have used RNNs that do not follow Dale’s law. We expand on previous work (Laje and Buonomano, 2013; Rajan et al., 2016) by training RNNs that follow Dale’s Law to emulate experimentally observed activity patterns. In addition, unlike standard RNN models, the networks in this study only have positive value firing rates. The networks are trained using the innate-training learning rule to autonomously produce stable activity for up to five seconds, two orders of magnitude greater than the time constant of the units (Laje and Buonomano, 2013; Rajan et al., 2016). Our results demonstrate that RNNs can robustly encode time by generating functionally feedforward patterns of activity. Importantly, these networks account for a characteristic of motor timing known as Weber’s Law (Gibbon, 1977; Gibbon et al., 1997), and can encode multiple feedforward patterns. Analysis of trained networks revealed

changes in the balance of excitation and inhibition that account for the production of this feedforward activity, thus generating an experimentally testable prediction.

2.3: Results

2.3.A: Recurrent neural networks produce functionally feedforward trajectories.

We first examined if recurrent neural networks can generate the sequential patterns of activity observed in the cortex and hippocampus (Pastalkova et al., 2008; Crowe et al., 2010; MacDonald et al., 2011; Crowe et al., 2014). Typically, in these areas any neuron participating in a sequence is active for periods of hundreds of milliseconds to a few seconds. We used a modified version of standard firing-rate RNNs in which units are sparsely and randomly connected to one another (Sompolinsky et al., 1988). Specifically, to more closely mimic neural physiology we incorporated separate populations of excitatory and inhibitory units (**Fig. 2.1A**). Furthermore, the firing rate of each unit was bounded between 0 and 1 (see Materials and Methods).

We used a supervised learning rule to adjust the recurrent weights and train the network to produce a functionally feedforward trajectory in response to a brief (50 ms) input. Specifically, the networks were trained to produce a 5 second long target sequence of feedforward activity such that each unit in the network was transiently activated without being driven by external input (**Fig. 2.1C**). This activity pattern can be thought of as a moving bump of neural activity. After training, the network was able to reproduce the 5 sec long neural trajectory in response to the brief input (**Fig. 2.1D**). Importantly, after the end of the trajectory the RNN returned to an inactive rest state—thus in contrast to RNNs in high-gain regimes these networks were silent at rest. Training dramatically altered the distribution of synaptic weights in the network: the weight of many synapses converged to 0 (in part as a consequence of the boundaries imposed by Dale's law) while others were strengthened resulting in long tails (**Fig. 2.1B**). These long-tails of the synaptic weight distribution is in line with experimentally observed distributions of synaptic weights (Song

et al., 2005) and observations in previous models of neural dynamics in RNNs (Laje and Buonomano, 2013; Rajan et al., 2016).

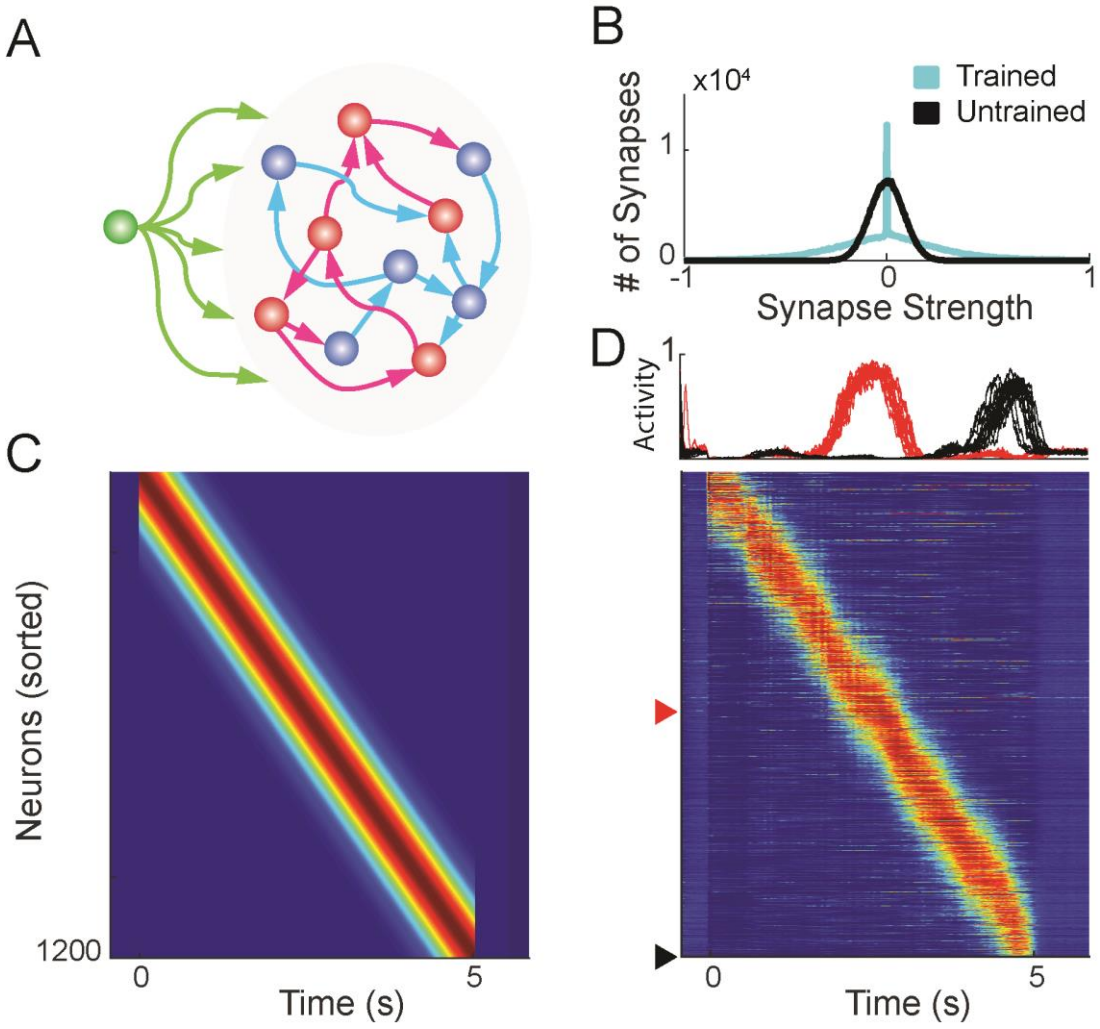


Figure 2.1. Generation of feedforward trajectories within a RNN.

A. Schematic of network architecture. The networks were composed of 600 excitatory (blue) and 600 inhibitory (red) firing rate units ($N^{Rec} = 1200$), with sparse recurrent connections. Neurons at the beginning of a sequence received input (green) during a 50 ms window to trigger the trajectory. **B.** Connection weights were initialized with a Gaussian distribution. After training to produce a single feedforward sequence, a large number of weights were pruned to zero, and some weights became stronger, resulting in a long tailed distribution. **C.** Example five second feedforward sequence target. **D.** After training, the RNN can produce a five second feedforward trajectory. Top: two units trained to activate in the middle and end of the trajectory, highlighted below. Each trace represents one of 15 trials. Bottom: Example network activity from a single trial.

2.3.B: RNNs can Encode Multiple Sequences

Many motor behaviors such as playing the piano or writing require the use of the same muscle groups activated in distinct temporal patterns. If the motor cortex is to drive these motor patterns, it must produce distinct well-timed trajectories of neural activity using the same sets of neurons activated in different orders. Traditional models of sequential neural activity (e.g. standard feedforward synfire chains) do not account for this because each unit generally participates in only a single sequence.

To examine the capacity of recurrent networks to encode multiple functionally feedforward trajectories, we trained RNNs to learn patterns in which all units participated in each trajectory. RNNs were trained to learn 1, 3, 5, 10, or 20 distinct sequences. Each sequence lasted one second, and was triggered by a distinct input. As shown in **Fig. 2.2A**, an RNN can generate multiple distinct patterns in response to distinct inputs. Importantly, each pattern recruits all the units in the network. To quantify the network capacity, we calculated the correlation between the evoked activity on each test trial and the corresponding target. We used the average correlation across targets as a measure of performance. Trained networks could reliably produce ten 1-second sequences with relatively little decrement in performance. However, when RNNs were trained on 20 patterns they showed a large decrease in performance (one-way ANOVA, $F_{4,45}= 193$, $p < 10^{-27}$, n = 10 networks; **Fig. 2.2B**) and increased failure rate (number of trials in which the input did not evoke a pattern or generated a partial sequence; one-way ANOVA, $F_{4,45}= 419$, $p < 10^{-35}$, n = 10 networks).

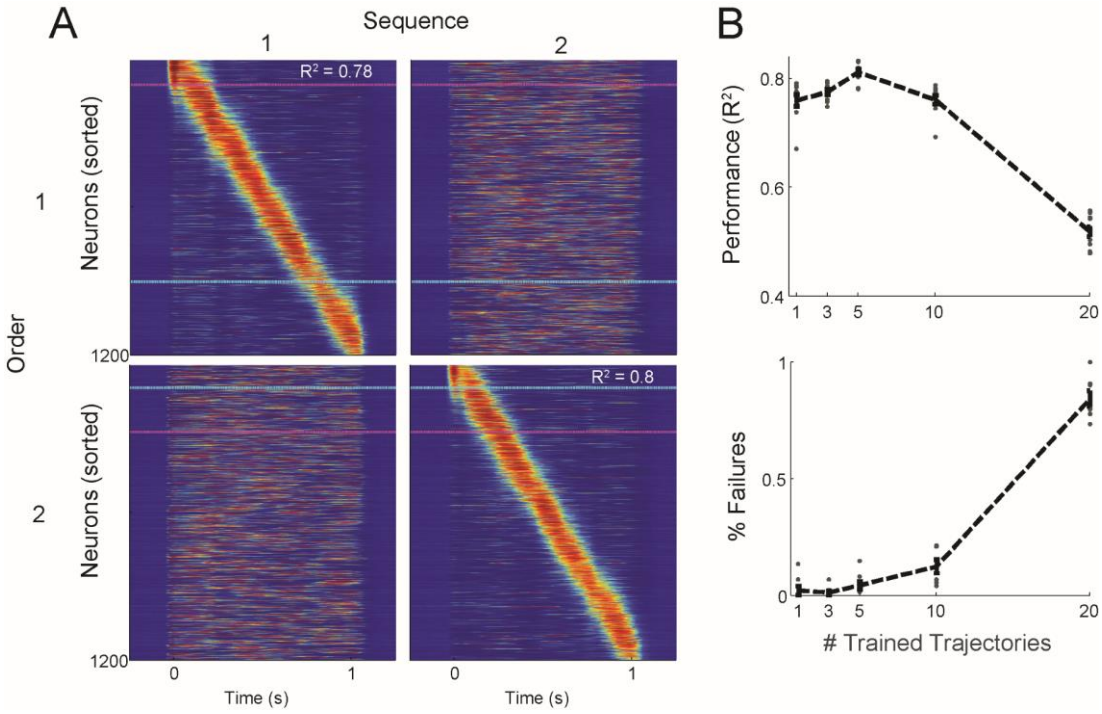


Figure 2.2. A single RNN can encode many different feedforward trajectories.

A. The units of the same trained RNN in response to two different inputs. Each column shows the spatiotemporal pattern of activity triggered by a single input. Each row shows the activity sorted according to feedforward trajectory #1 (top row) or #2 (bottom row). Blue and pink dashed lines highlight the same two units in all panels. **B.** Performance (top) and failure rate (bottom) for ten networks trained to produce up to 20 trajectories. Each dot represents the average across 15 trials per target for a single network. Each network can reliably produce up to ten feedforward sequences before performance decreases. Error bars show the mean ± SEM**.**

2.3.C: Functionally Feedforward Trajectories Account for Weber's Law

A defining feature of behavioral timing is that there is an approximately linear relationship between the standard deviation and mean of a timed response (Gibbon, 1977; Gibbon et al., 1997)—referred to as the scalar property or Weber's law. The ability to account for Weber's law is often taken as a benchmark for models of timing, and does not generally emerge spontaneously in many models (Ahrens and Sahani, 2008; Hass and Herrmann, 2012; Hass and Durstewitz, 2014). To examine whether the RNNs studied here obey Weber's law, we measured the temporal variability of each unit within a single feedforward pattern at different levels of noise. We fit the activity of each unit on every test trial with a Gaussian function, and calculated the standard deviation and mean of the peak time of each unit's fit across trials (see Materials and Methods). We used Weber's generalized law to fit the standard deviation as a function of time, and refer to the slope of this linear fit as the Weber coefficient (Ivry and Hazeltine, 1995; Merchant et al., 2008)—note that Weber's generalized law allows for a positive intercept. In each of ten trained networks, we found that the standard deviation of a unit's peak firing time across trials increased linearly with its mean activation time (**Fig. 2.3A**). This property was highly robust: while the Weber coefficient increased with the amount of noise injected into the network, the scalar property was preserved even at large noise amplitudes (**Fig. 2.3B**)—thus Weber's law was not limited to any specific noise parameter choice.

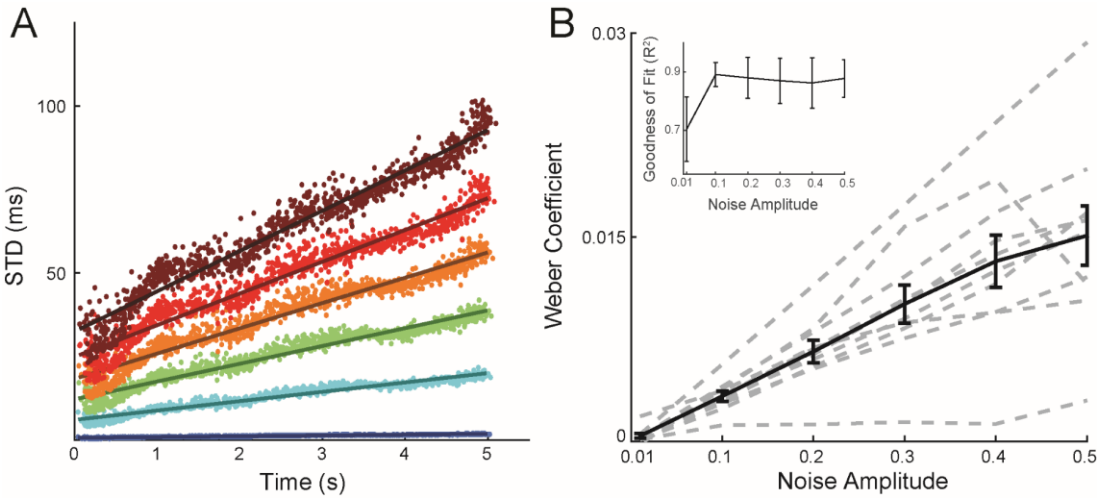


Figure 2.3. The timing generated by the RNN obeys Weber's law.

A. Temporal variability increases linearly with elapsed time. Each dot shows the standard deviation of the activation time of a unit of an example network plotted against its mean. The activation time of each unit on each trial was determined from the center of a Gaussian fit of that unit's activity. Each color represents one of six amplitudes of injected noise. Lines show the linear fit. **B.** The Weber law is robust to noise. The Weber Coefficient (slope of the linear fit shown in panel A) is shown across six noise amplitudes for ten trained networks. Each dotted line represent a single network, the solid black line represents the mean with error bars showing ± SEM. Inset: mean ± SEM of the goodness of fit (R^2) for the linear fits.

2.3.D: Sequences are Generated by Dynamic Shifts in the Balance of Excitation and Inhibition

Experimental studies have conclusively demonstrated that functionally feedforward patterns of activity occur *in vivo*, but these studies have not been able to explore the neural and network mechanisms underlying these patterns (Pastalkova et al., 2008; Harvey et al., 2012; MacDonald et al., 2013). Computational models allow us to address this question and generate experimental predictions. To determine how RNNs generate sequential activity, we first examined the balance of excitation and inhibition in the units during the trained patterns. This analysis parallels experimental studies which have examined the relative balance of excitatory and inhibitory currents (Shu et al., 2003; Froemke et al., 2007; Heiss et al., 2008). We examined the balance of excitation and inhibition by separately summing the total excitatory and inhibitory input onto each unit at all time points. **Fig. 2.4A** shows the relationship between the E/I balance and firing rate in a single example unit. Interestingly, when cells were not active they still received significant excitation and inhibition, and these inputs were usually approximately balanced (E/I $\cong$ 1), similar to a recent *in vitro* study of temporal processing (Goel and Buonomano, 2016). This balance shifted towards excitation (E/I > 1) primarily during a unit's target activity period. By plotting the E/I ratio of all the neurons in the network during a trajectory it is possible to visualize the progressive shift in the E/I balance during a feedforward sequence (**Fig. 2.4B**). These observations are consistent with experimental studies during up-state activity showing excitatory and inhibitory currents are balanced, and that changes in activity consist of subtle shifts towards excitation (Shu et al., 2003; Haider et al., 2006; Sun et al., 2010), as well as recent work showing timing-specific shifts in E/I balance *in vitro* (Goel and Buonomano, 2016).

To examine the origins of the dynamic shift in the E/I balance, we analyzed the synaptic connectivity matrix. When sorted according to activation order, connections from both inhibitory and excitatory units displayed a pattern of peak connection strength along the sequence of activation (**Fig. 2.4C**). To examine the structure of the connectome we shifted the sorted weight matrix to align the window of activity of each postsynaptic cell within the trajectory. Taking the mean across all cells of this shifted weight matrix revealed a peak of excitation pointing forward along the trajectory (that is, the excitatory weights are asymmetrically shifted to the right), bounded by peaks of inhibition (**Fig. 2.4D, upper panel**). Despite the rightward shift of the peak—and in contrast to feedforward networks—the excitatory units are clearly connected in the "forward" and "backward" directions. This anatomical feature allows for the local mutual excitation necessary to keep units active for durations of up to a second. This "Mexican hat" connectivity pattern, has been observed in other studies of sequence generation (Itskov et al., 2011; Rajan et al., 2016) and accounts for the moving bump of activation in feedforward RNNs. Interestingly, for a single pattern the shift of the E/I balance towards excitation was primarily driven by an increase in excitation (**Fig 2.4D upper panel**). However, when the same analysis is performed for networks that learned 10 patterns, the E/I shift driving activity forward was generated by both an increase in excitation and a decrease in inhibition (**Fig. 2.4D, lower panel**). Taken together, these results predict that recurrent networks in the cortex generate functionally feedforward sequences of activity using asymmetries in the connectivity patterns between neurons, and dynamic shifts in the E/I balance.

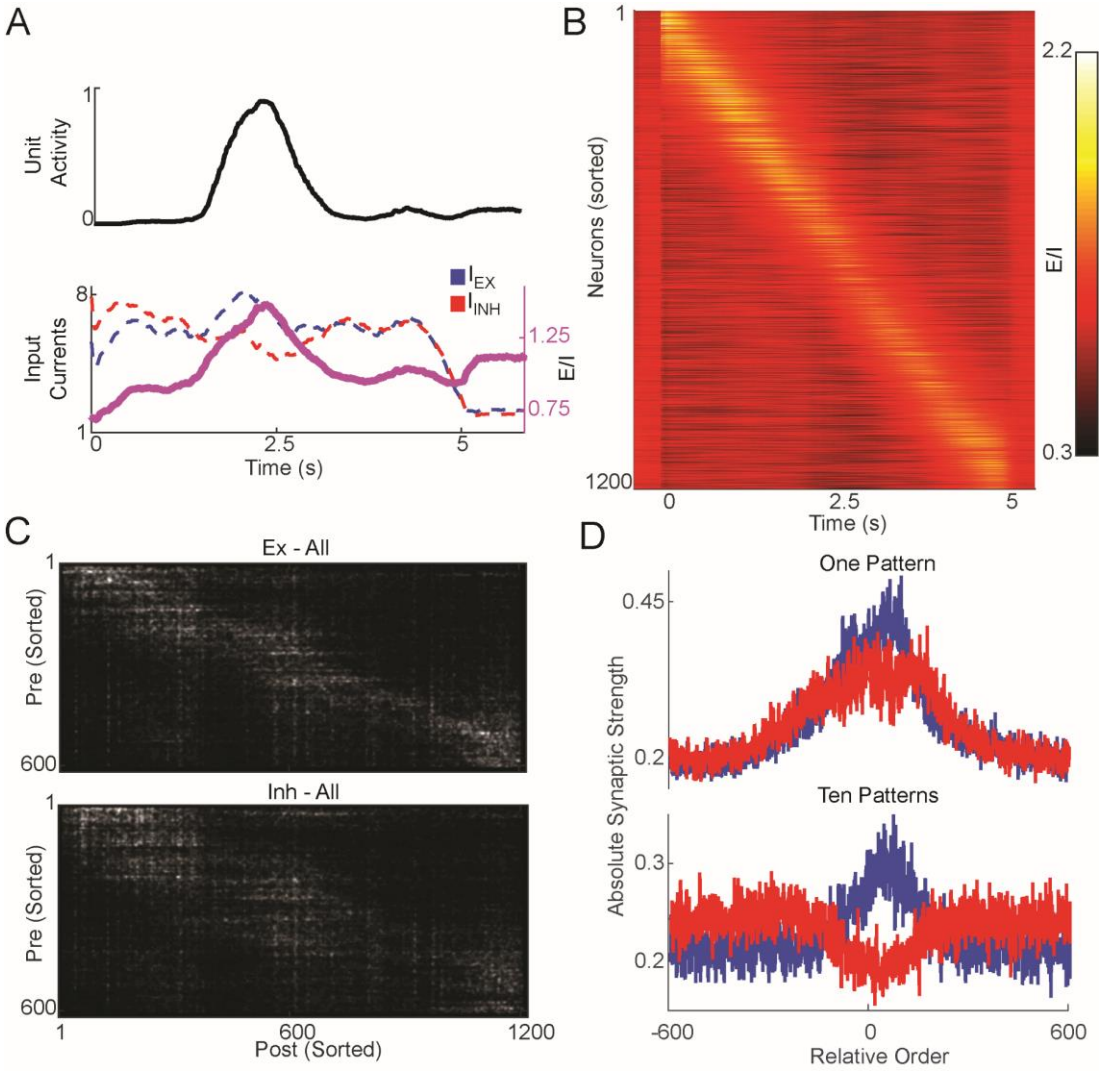


Figure 2.4. Feedforward trajectories are generated by dynamic shifts in the E/I balance.

A. The excitation/inhibition (E/I) ratio peaks around a unit's activation time. Top: the activity of an example unit which peaked in the middle of a five seconds trajectory. Bottom: Inhibitory (red) and excitatory (blue) inputs are high but usually balanced throughout the trajectory, but the balance shifts towards excitation when the unit is active. E/I ratio shown in pink. **B.** A heat-map of the E/I ratio of the same network showing a shifting peak of the E/I ratio along the trajectory (during one trial). **C.** Excitatory (top) and inhibitory (bottom) weight matrices of a network sorted according to the activation order. Note the peak in synaptic strength following the activation order along the diagonal. Weights are smoothed for visualization. **D.** Averaging along the feedforward trajectory reveals a peak in excitatory weights pointing forward, bounded by inhibition. The absolute value of the weight matrix was centered along the feedforward sequence and averaged according to excitatory and inhibitory units. Top: Weights of a network trained for one pattern. Bottom: The same initial network, but trained for ten feedforward patterns.

2.3.E: Connectivity of RNNs Reflects the Number of Encoded Trajectories

There is increasing emphasis on characterizing the microcircuit structure, or the connectome, of biological neural circuits. To further examine the relationship between the microcircuit structure in our model, and potentially generate experimental predictions, we calculated network efficiency—a standard measure from graph theory that captures the "interconnectedness" of the units in a graph (Boccalettii et al., 2006). Specifically, it measures the minimal weighted path length between units, such that a larger efficiency value corresponds to a shorter path (see Materials and Methods). Because we were interested in the relationship between structure and function we compared efficiency measures of coactive and non-coactive units. Furthermore, since the activity in any unit is the result of the interaction between excitatory and inhibitory inputs we separately calculated the net excitatory and inhibitory connection strengths between pairs of units—generating topological representations of recurrent excitation and inhibition. We then calculated the weighted efficiency index of these connections in trained untrained networks.

As expected, when averaged along a single trained sequence of feedforward activity, we observed that units that co-activate (i.e, active at neighboring points in time) had a higher than average efficiency value (**Fig. 2.5A**), similar to the observed "Mexican hat" architecture in Fig. 4D. However, when networks were trained for sustain more coactive units (**Fig. 2.5B**), or were trained to produce a larger number of targets (**Fig. 2.5C**), the connection efficiency between coactive and non-coactive units approached the network mean. Finding the average disynaptic efficiency between coactive and non-coactive units revealed that efficiency sharply increased when networks were trained for a single target,

with coactive efficiency exceeding non-coactive (**Fig. 2.5D**). As more trajectories were encoded, this difference decreased, indicating that efficiency became uniform with respect to unit pairs' active relationship. Moreover, networks with more coactive units (40% active) were initially more uniform than those with fewer coactive units (i.e, the efficiency between coactive and non-coactive units was more similar), consistent with the notion that higher local efficiency may be necessary to maintain temporally sparser trajectories in order to support more local positive feedback between coactive units.

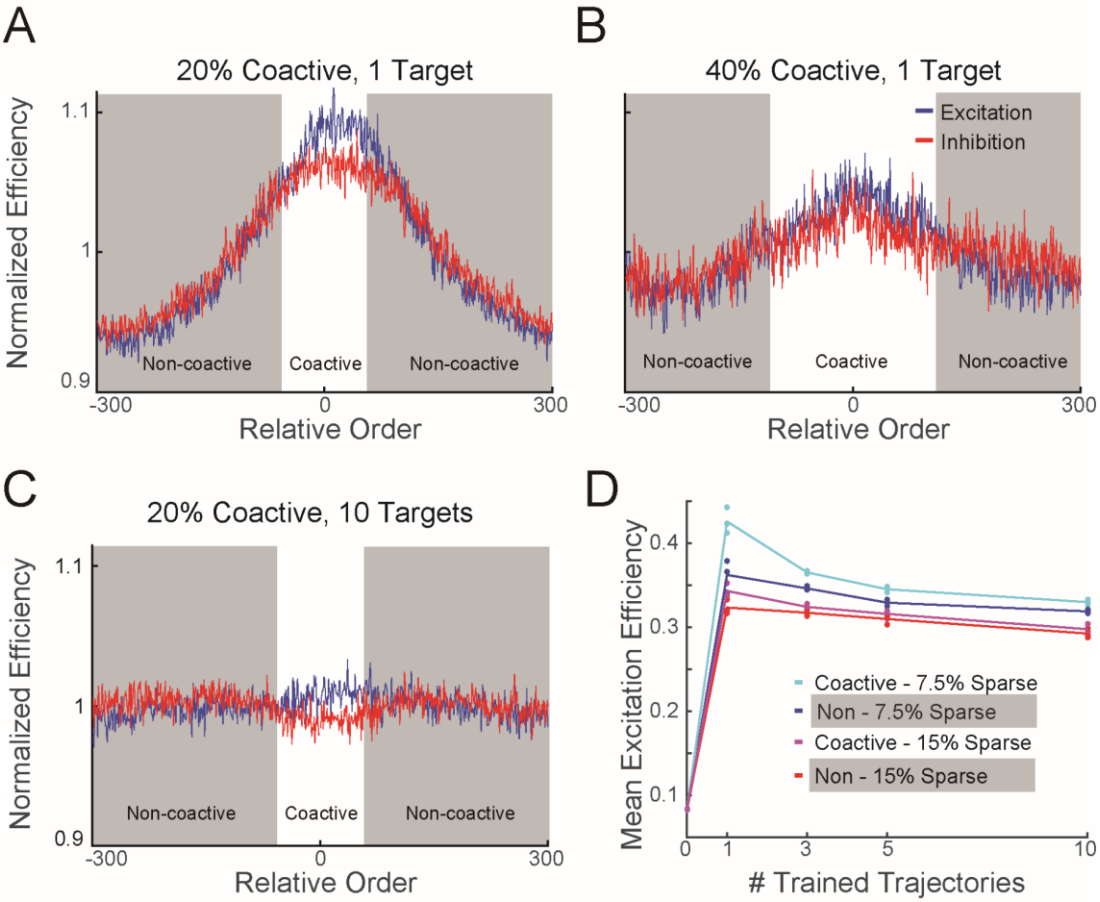


Figure 2.5. Multiple feedforward patterns are embedded via uniform path lengths between units.

A. Average disynaptic efficiency between coactive (unshaded) and non-coactive (shaded) unit pairs in an RNN trained for a single feedforward target. Units that are coactive during a trajectory have a higher mean efficiency compared to non-coactive pairs. Efficiency values were normalized to the mean to aid visualization. **B.** Same as **A** but showing the same initial network trained such that more units are coactive at a given time (lower temporal sparsity). The efficiency across the network is more uniform. **C.** Same as in **A** now trained for 10 targets. In **A-C**, postsynaptic efficiency values were aligned according to the activation order of the presynaptic unit within the same target order and averaged across units. **D.** Average efficiency of connections between coactive and non-coactive units according to the number of trained trajectories. Efficiency increases sharply from the naive weights. As more trajectories are encoded, efficiency becomes uniform across the network, i.e. the difference between coactive and non-coactive units decreases.

2.4: Discussion

Functionally feedforward patterns of activity have been observed in a wide range of different brain areas (Pastalkova et al., 2008; MacDonald et al., 2011; Kraus et al., 2013; Mello et al., 2015; Bakhurin et al., 2017). These patterns have been proposed to underlie a number of different behaviors, including memory, planning, and motor timing. Here we have focused primarily on the potential role of such patterns in timing, specifically in tasks which animals must learn to generate timed motor patterns or anticipate when external events will occur on the scale of hundreds of milliseconds to seconds. Our results show that even though many of these experimentally reported patterns of sequential activation are apparently accounted for by feedforward architectures, recurrent neural networks are more consistent with the data. Furthermore, recurrent architectures are computationally more powerful in that they can store many different trajectories in which each unit participates in each trajectory. We propose that networks with recurrent excitation underlie the functionally feedforward trajectories observed in cortical areas.

A number of models have proposed mechanisms for generating functionally feedforward patterns within recurrent networks (Buonomano, 2005; Liu and Buonomano, 2009; Fiete et al., 2010; Itskov et al., 2011; Rajan et al., 2016). These studies have used both spiking and firing rate models, and relied on a number of different mechanisms, but they have not explicitly addressed a standard benchmark for behavioral timing—Weber's law. One recent model developed by Rajan et al. (2016) also trained RNNs using the an RLS-based learning rule, and our results complement their findings: specifically that by tuning the recurrent weights of an initially randomly connected network it is possible to robustly encode multiple functionally feedforward patterns of activity. That study,

however, focused primarily on sequence generation and encoding memory-dependent motor behaviors, and did not encode time *per se* as the network was driven in part by time-varying inputs (that is, external information about the time from trial onset was present).

2.4.A: Weber's Law

Here we show that an RNN can encode time and account for Weber's law, more specifically Weber's generalized law, which states that the standard deviation of a timer increases linearly with elapsed time (Ivry and Hazeltine, 1995; Merchant et al., 2008; Laje et al., 2011). The origins of Weber's law in timing models is a longstanding and vexing problem, because according to the simplest model in which an accumulator integrates the pulses of a noisy oscillator, the standard deviation of the latency of a neuron, or of a motor output, should increase as a function of the square root of total time (Hass and Herrmann, 2012; Hass and Durstewitz, 2014; Hass and Durstewitz, 2016). In contrast, the current model naturally captures Weber's law (at least within the parameter regimes used here), even at high noise amplitudes. Major issues remain, such as the properties underlying Weber's law in recurrent networks and why the brain "settles" for the observed linear relationship (Hass and Herrmann, 2012). We hypothesize that these properties may be related: 1) recurrency may inherently amplify internal noise, producing long-lasting temporal correlations (Hass and Herrmann, 2012), and 2) evolutionarily speaking, the tradeoff was adaptive because it increased computational capacity.

The current model also establishes that RNNs can robustly store multiple patterns, in which each neuron participates in every pattern. This feature is consistent with experimental findings demonstrating that the same neuron can participate in multiple

patterns of network activity, firing within different windows in each (Pastalkova et al., 2008; MacDonald et al., 2013). Thus the experimental data and the current model are consistent with stimulus-specific timing, in which time codes are generated in relation to each stimulus or task condition as opposed to an absolute time code. The capacity of the RNNs described here appears to be fairly large. But as demonstrated in a previous study, the true capacity of RNNs is likely to be strongly dependent on model assumptions, most notably noise levels (Laje and Buonomano, 2013).

Within this population clock framework, the same RNN does not function as a single clock, but rather implements many event-specific timers. That is, the network does not encode absolute time but elapsed time from stimulus onset, and there is an entirely different time code for each stimulus. This computational strategy ensures that the activity vector at any given instant not only encodes elapsed time, but also provides a dynamic memory of the current stimulus.

2.4.B: RNN Connectome

In order for sequences to propagate in a defined trajectory through a network, activity must generate imbalances that simultaneously push the activity forward and prevent it from deviating from the proper activation order (Ben-Yishai et al., 1995; Fiete et al., 2010). Here we find that training an RNN composed of distinct excitatory and inhibitory populations produces synaptic connectivity resembling an asymmetric “Mexican hat” architecture, with excitation propagating and maintaining network activity and inhibition bounding this activity to prevent off-target activation. Importantly, the recurrency of the network enables multiple “Mexican hats” to be embedded in a single connectivity matrix, allowing multiple functional feedforward patterns to be produced by a single network.

An important characteristic of the connectome of a network is how efficiently individual units exchange information. Surprisingly, we found that the weighted efficiency of a feedforward RNN was negatively correlated with the number of sequences stored, and that this change was largely driven by reduced efficiency between coactive units. Indeed, multi-trajectory networks exhibited uniform path lengths between units, regardless of their relative activation order. This “flattening” of the efficiency within a network is likely necessary to allow units that are highly separated in one sequence to also co-activate in another. Thus a prediction that emerges from this study is that learning may induce an overall decrease in the efficiency of cortical circuits, as the networks embed more uniform connection structures, making individually learned patterns difficult to distinguish using connectomics.

Reports of sequential patterns of activity in multiple brain areas appear to be superficially consistent with feedforward synfire-like architectures. However, recurrent networks are likely responsible for generating the experimentally observed patterns for two reasons. First, although the patterns of activity comprise sequential activation of neurons, the duration of activity over which a neuron fires (in the range of hundreds of millisecond to a few seconds) likely relies on local positive feedback maintained by recurrent connections; and second, purely feedforward network architectures are unlikely to account for the ability of networks to generate multiple trajectories in which any given neuron can participate in many different patterns.

2.5: Materials and Methods

2.5.A: Network Structure and Dynamics

The network dynamics were governed by the standard firing rate equations (Abeles, 1982; Sompolinsky et al., 1988; Jaeger and Haas, 2004):

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N^{Rec}} W_{ij}^{Rec} r_j + \sum_{k=1}^{N^{In}} W_{ik}^{In} y_k + I_i^{Noise} \quad (1)$$

Where x_i represents the state of unit i. The sparse, $N^{Rec}\ x\ N^{Rec}$ matrix W^{Rec} describes the recurrent connectivity, with nonzero values initially drawn from a Gaussian distribution of $\mu = 0$ and $\sigma = g/\sqrt{p_c * N^{Rec}}$, where $g = 1.6$ is the synaptic scaling constant and $p_c = 0.3$ is the probability a given unit will connect to another unit in the network (autapses were eliminated). The firing rate, r_i, of unit i is given by the logistic function:

$$r_i = \frac{1}{1 + e^{-a \cdot x_i + b}} \quad (2)$$

where $a = 2$ and $b = 4$ correspond to the gain and "threshold" of the units, respectively. Compared to the traditional *tanh* function, this provides a more biologically plausible model in which activity is low at "rest" (i.e. without input), and rates are bounded between 0 and 1.

After initialization, the efferent synapses from a randomly selected half of the recurrent units were set to be positive and the other half set to be negative to create inhibitory and excitatory populations. The $N^{Rec}\ x\ N^{In}$ matrix W^{In} describes the input connections from input units y to the recurrent units, and is set to stimulate only those units active at the start of a functionally feedforward sequence with weight equal to each unit's target activity

at t = 0. The activity of the input units was set to 0 except during the 50 ms input window at the beginning of a trial when it was set to 3. $N^{Rec} = 1200$ is the network size, and $\tau = 25\ \mathrm{ms}$ is the time constant of the units. The random noise current (I^{Noise}) was drawn from a Gaussian distribution of μ = 0 and σ = 0.5 (noise amplitude) unless otherwise indicated. A single unit of the RNN contacted all other units and was tonically active, providing a "bias" to each unit, but containing no temporal information as its activity was set to 1 at all times.

2.5.B: Training the RNNs

The RNNs were trained to generate target patterns of sequential activity designed to mimic the functionally feedforward activity observed in neural circuits during temporal tasks (Hahnloser et al., 2002; Pastalkova et al., 2008; MacDonald et al., 2011; Harvey et al., 2012). These targets were generated by setting each unit to activate briefly in sequence so the entire population tiled the interval defined by t_{max} (**Fig. 2.1C**). The activation order was generated randomly for each target pattern and constrained so that the order of inhibitory and excitatory units was interleaved. The pattern of activation for each unit was set by a Gaussian with a μ equal to the unit's activation time. The temporal sparsity of the pattern, defined by $\frac{N^{Active}}{N^{Rec}}$, where N^{Active} represents the number of units in the target pattern that are active at any given point in time, was set to approximately 20% by making the σ of the Gaussian target function 7.5% of t_{max} (Rajan et al., 2016).

Training was performed using the innate-training learning rule which tunes the recurrent weights based on errors generated by each unit (Laje and Buonomano, 2013), similar to the learning rule used in Rajan et al. 2016. The error was determined by taking

the difference between the activity of the unit r_i and it's target activity at time t and used to update its weights using the recursive least squares (RLS) algorithm (see also, Haykin, 2002; Jaeger and Haas, 2004; Sussillo and Abbott, 2009; Mante et al., 2013; Carnevale et al., 2015; Rajan et al., 2016). As described previously (Laje and Buonomano, 2013) the weights onto a given unit were updated proportional to its error, the activity of its presynaptic units, and the inverse cross-correlation matrix of the network activity. To maintain Dale's law, efferent weights from all units were bounded so that they could not cross zero (i.e. negative weights were prevented from becoming positive and vice versa). The weights were bounded to a maximum value of $g + 5/\sqrt{N^{Rec}}$ of the appropriate sign to prevent overfitting. Training was conducted at a noise amplitude of 0.5 and all recurrent units were trained.

2.5.C: Weber Analysis

To determine if the timing of the network obeyed Weber's law, we tested each network 15 times at six different I^{Noise} amplitudes. For each trial, we fit the activity of each unit with a Gaussian curve and used the center of that curve as a measure of the unit's activation time. Because each unit activates only once, time can be measured directly from the activity state of the network (Abeles, 1982; Long et al., 2010). Only units whose Gaussian fit had an $R^2 > 0.9$ where used for the Weber analysis. For a given noise level and network, we calculated the standard deviation (std_i) and mean (t_i) of these activation times for each unit i and found a linear fit of these values. We then used the slope of the linear fit as the Weber coefficient, std/t (excluding units outside a 95% confidence interval of the linear fit; **Fig. 2.1**).

2.5.D: Performance

Network performance was measured by a performance index, calculated as the correlation (R^2) between network activity on a given trial and the corresponding target pattern. The overall performance of a network was calculated as the network's mean performance index across all trials for all trained patterns. Particularly at high noise levels and numbers of trained targets, networks sometimes failed to complete a feedforward sequence; thus we also used the percentage of these failures to quantify capacity (**Fig. 2.2B**).

2.5.E: Network Efficiency

In graph theory, network efficiency measures the shortest path between two nodes of a network, and can be thought of as a measure of interconnectedness of the units (Boccalettii et al., 2006). Efficiency was calculated by determining the minimum weighted disynaptic excitatory and inhibitory path length between pairs of excitatory units in the network, with weights normalized to the maximum weight. Disynaptic connection strengths were calculated by taking matrix products $W_{Ex\,Ex}W_{Ex\,Ex}$ for the excitatory path and $W_{Ex\,Inh}W_{Inh\,Ex}$ for inhibitory path, creating two $N^{Ex} \times N^{Ex}$ matrices. The path length between two units was determined by finding the series of edges that connected the units with the smallest summed inverse weight. For example, if unit A is connected to unit B with a strength of 0.2, the path length will be 5; if unit A is also connected to unit C with a strength 0.5 and unit C connects to B with a strength of 0.5, then the length from A to B is 4 (2 + 2). Thus the minimum weighted path between A and B would be through unit C. This path length was calculated for all possible pairs in the disynaptic matrices, inverted, and normalized by the total number of possible connections $(N^{Rec} * (N^{Rec} - 1))$ to

generate an efficiency value. Therefore, a network with 100% maximal connections would have an efficiency value of 1.

2.6: References Cited

Abeles M (1982) Local Cortical Circuits: An Electrophysiological Study: Springer, Berlin.

Abeles M (1991) Corticonics. Cambridge: Cambridge University Press.

Ahrens MB, Sahani M (2008) Inferring elapsed time from stochastic neural processes. In: Advances in Neural Information Processing Systems (JC P, D K, Y S, S R, eds), pp 1-8. Red Hook, NY Curran Associates.

Ahrens MB, Sahani M (2011) Observers Exploit Stochastic Models of Sensory Change to Help Judge the Passage of Time. Current Biology 21:200-206.

Bakhurin KI, Goudar V, Shobe JL, Claar LD, Buonomano DV, Masmanidis SC (2017) Differential Encoding of Time by Prefrontal and Striatal Network Dynamics. J Neurosci 37:854-870.

Banerjee A, Series P, Pouget A (2008) Dynamical constraints on using precise spike timing to compute in recurrent cortical networks. Neural Comput 20:974-993.

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. Proceedings of the National Academy of Sciences 92:3844-3848.

Boccalettii S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: structure and dynamics. Physics Reports 424:175-308.

Buhusi CV, Meck WH (2005) What makes us tick? Functional and neural mechanisms of interval timing. Nature Reviews Neuroscience 6:755-765.

Buonomano DV (2005) A learning rule for the emergence of stable dynamics and timing in recurrent networks. J Neurophysiol 94:2275-2283.

Buonomano DV, Mauk MD (1994) Neural network model of the cerebellum: temporal discrimination and the timing of motor responses. Neural Comput 6:38–55.

Buonomano DV, Laje R (2010) Population clocks: motor timing with neural dynamics. Trends in Cognitive Sciences 14:520-527.

Carnevale F, de Lafuente V, Romo R, Barak O, Parga N (2015) Dynamic Control of Response Criterion in Premotor Cortex during Perceptual Detection under Temporal Uncertainty. Neuron 86:1067-1077.

Crowe DA, Averbeck BB, Chafee MV (2010) Rapid Sequences of Population Activity Patterns Dynamically Encode Task-Critical Spatial Information in Parietal Cortex. The Journal of Neuroscience 30:11640-11653.

Crowe DA, Zarco W, Bartolo R, Merchant H (2014) Dynamic Representation of the Temporal and Sequential Structure of Rhythmic Movements in the Primate Medial Premotor Cortex. The Journal of Neuroscience 34:11972-11983.

Diesmann M, Gewaltig MO, Aertsen A (1999) Stable propagation of synchronous spiking in cortical neural networks. Nature 402:529-533.

Durstewitz D (2003) Self-Organizing Neural Integrator Predicts Interval Times through Climbing Activity. The Journal of Neuroscience 23:5342-5353.

Fiete IR, Senn W, Wang CZH, Hahnloser RHR (2010) Spike-Time-Dependent Plasticity and Heterosynaptic Competition Organize Networks to Produce Long Scale-Free Sequences of Neural Activity. Neuron 65:563-576.

Froemke RC, Merzenich MM, Schreiner CE (2007) A synaptic memory trace for cortical receptive field plasticity. Nature 450:425-429.

Gavornik JP, Shuler MGH, Loewenstein Y, Bear MF, Shouval HZ (2009) Learning reward timing in cortex through reward dependent expression of synaptic plasticity. Proceedings of the National Academy of Sciences 106:6826-6831.

Gibbon J (1977) Scalar expectancy theory and Weber's law in animal timing. Psychological Review 84:279-325.

Gibbon J, Church RM, Meck WH (1984) Scalar timing in memory. Ann N Y Acad Sci 423:52-77.

Gibbon J, Malapani C, Dale CL, Gallistel CR (1997) Toward a neurobiology of temporal cognition: advances and challenges. Curr Opinion Neurobiol 7:170-184.

Goel A, Buonomano Dean V (2016) Temporal Interval Learning in Cortical Cultures Is Encoded in Intrinsic Network Dynamics. Neuron 91:320-327.

Goldman MS (2009) Memory without Feedback in a Neural Network. Neuron 61:621-634.

Gouvea TS, Monteiro T, Motiwala A, Soares S, Machens C, Paton JJ (2015) Striatal dynamics explain duration judgments. Elife 4.

Hahnloser RHR, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequence in a songbird. Nature 419:65-70.

Haider B, Duque A, Hasenstaub AR, McCormick DA (2006) Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. J Neurosci 26:4535-4545.

Hardy NF, Buonomano DV (2016) Neurocomputational models of interval and pattern timing. Current Opinion in Behavioral Sciences 8:250-257.

Harvey CD, Coen P, Tank DW (2012) Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature 484:62-68.

Hass J, Herrmann JM (2012) The Neural Representation of Time: An Information-Theoretic Perspective. Neural Computation 24:1519-1552.

Hass J, Durstewitz D (2014) Neurocomputational Models of Time Perception. In: Neurobiology of Interval Timing (Merchant H, de Lafuente V, eds), pp 49-73: Springer New York.

Hass J, Durstewitz D (2016) Time at the center, or time at the side? Assessing current models of time perception. Current Opinion in Behavioral Sciences 8:238-244.

Haykin S (2002) Adaptive Filter Theory. Upper Saddle River: Prentice Hall.

Heiss JE, Katz Y, Ganmor E, Lampl I (2008) Shift in the Balance between Excitation and Inhibition during Sensory Adaptation of S1 Neurons. J Neurosci 28:13320-13330.

Herrmann M, Hertz JA, Prügel-Bennett (1995) Analysis of synfire chains. Network: Computation in Neural Systems 6:403-414.

Itskov V, Curto C, Pastalkova E, Buzsáki G (2011) Cell Assembly Sequences Arising from Spike Threshold Adaptation Keep Track of Time in the Hippocampus. The Journal of Neuroscience 31:2828-2834.

Ivry RB, Hazeltine RE (1995) Perception and Production of Temporal Intervals across a Range of Durations - Evidence for a Common Timing Mechanism. Journal of Experimental Psychology-Human Perception and Performance 21:3-18.

Ivry RB, Schlerf JE (2008) Dedicated and intrinsic models of time perception. Trends in Cognitive Sciences 12:273-280.

Jaeger H, Haas H (2004) Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. Science 304:78-80.

Jin DZ, Fujii N, Graybiel AM (2009) Neural representation of time in cortico-basal ganglia circuits. Proc Natl Acad Sci U S A 106:19156-19161.

Kim J, Ghim J-W, Lee JH, Jung MW (2013) Neural Correlates of Interval Timing in Rodent Prefrontal Cortex. The Journal of Neuroscience 33:13834-13847.

Kraus BJ, Robinson RJ, White JA, Eichenbaum H, Hasselmo ME (2013) Hippocampal “Time Cells”: Time versus Path Integration. Neuron 78:1090-1101.

Laje R, Buonomano DV (2013) Robust timing and motor patterns by taming chaos in recurrent neural networks. Nat Neurosci 16:925-933.

Laje R, Cheng K, Buonomano DV (2011) Learning of temporal motor patterns: an analysis of continuous versus reset timing. Frontiers in Integrative Neuroscience 5:61.

Liu JK, Buonomano DV (2009) Embedding Multiple Trajectories in Simulated Recurrent Neural Networks in a Self-Organizing Manner. The Journal of Neuroscience 29:13172-13181.

Long MA, Jin DZ, Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. Nature 468:394-399.

MacDonald CJ, Lepage KQ, Eden UT, Eichenbaum H (2011) Hippocampal "time cells" bridge the gap in memory for discontiguous events. Neuron 71:737-749.

MacDonald CJ, Carrow S, Place R, Eichenbaum H (2013) Distinct hippocampal time cell sequences represent odor memories in immobilized rats. J Neurosci 33:14607-14616.

Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503:78-84.

Mauk MD, Donegan NH (1997) A model of Pavlovian eyelid conditioning based on the synaptic organization of the cerebellum. Learn Mem 3:130–158.

Mauk MD, Buonomano DV (2004) The Neural Basis of Temporal Processing. Annual Review of Neuroscience 27:307-340.

Medina JF, Mauk MD (1999) Simulations of Cerebellar Motor Learning: Computational Analysis of Plasticity at the Mossy Fiber to Deep Nucleus Synapse. Journal of Neuroscience 19:7140-7151.

Mello GBM, Soares S, Paton JJ (2015) A scalable population code for time in the striatum. Curr Biol 9:1113-1122.

Merchant H, Zarco W, Prado L (2008) Do We Have a Common Mechanism for Measuring Time in the Hundreds of Millisecond Range? Evidence From Multiple-Interval Timing Tasks. J Neurophysiol 99:939-949.

Merchant H, Harrington DL, Meck WH (2013) Neural basis of the perception and estimation of time. Annual Review of Neuroscience 36:313-336.

Merchant H, Zarco W, Pérez O, Prado L, Bartolo R (2011) Measuring time with different neural chronometers during a synchronization-continuation task. Proceedings of the National Academy of Sciences 108:19784-19789.

Modi MN, Dhawale AK, Bhalla US (2014) CA1 cell activity sequences emerge after reorganization of network correlation structure during associative learning. eLife 3:e01982.

Namboodiri VMK, Huertas MA, Monk KJ, Shouval HZ, Hussain Shuler MG (2015) Visually Cued Action Timing in the Primary Visual Cortex. Neuron 86:319-330.

Pastalkova E, Itskov V, Amarasingham A, Buzsaki G (2008) Internally Generated Cell Assembly Sequences in the Rat Hippocampus. Science 321:1322-1327.

Rajan K, Harvey Christopher D, Tank David W (2016) Recurrent Network Models of Sequence Generation and Memory. Neuron 90:128-142.

Shu Y, Hasenstaub A, McCormick DA (2003) Turning on and off recurrent balanced cortical activity. Nature 423:288-293.

Simen P, Balci F, deSouza L, Cohen JD, Holmes P (2011) A Model of Interval Timing by Neural Integration. The Journal of Neuroscience 31:9238-9253.

Sompolinsky H, Crisanti A, Sommers HJ (1988) Chaos in random neural networks. Physical Review Letters 61:259-262.

Song S, Sjostrom PJ, Reigl M, Nelson, Chklovskii DB (2005) Highly nonrandom feature of synaptic connectivity in local cortical circuits. PLOS Biology 3:508-518.

Sun YJ, Wu GK, Liu B-h, Li P, Zhou M, Xiao Z, Tao HW, Zhang LI (2010) Fine-tuning of pre-balanced excitation and inhibition during auditory cortical development. Nature 465:927-931.

Sussillo D, Abbott LF (2009) Generating Coherent Patterns of Activity from Chaotic Neural Networks. Neuron 63:544-557.

Treisman M (1963) Temporal discrimination and the indifference interval: implications for a model of the 'internal clock'. Psychol Monogr 77:1–31.

Chapter 3: A Model of Temporal Scaling Correctly Predicts that Motor Timing Improves with Speed

3.1: Abstract

Timing is fundamental to complex motor behaviors: from tying a knot to playing the piano. A general feature of motor timing is temporal scaling: the ability to produce motor patterns at different speeds. Here we report that after learning to produce a Morse code pattern performance was inversely related to the change in speed. Temporal scaling was not an intrinsic property of a recurrent neural network (RNN) model of timing, however, training the RNN on two speeds produced robust temporal scaling. The model captured a signature of motor timing—Weber's law—but predicted that temporal precision improves at faster speeds. Human psychophysics experiments confirmed this prediction: the standard deviations of responses in absolute time were lower at faster speeds. These results establish that RNNs can account for temporal scaling, and suggest a novel psychophysical principle: the Weber-speed effect.

3.2: Introduction

It is increasingly clear that the brain uses different mechanisms and circuits to tell time across different tasks. For example, distinct brain areas are implicated in sensory (Bueti et al., 2012; Namboodiri et al., 2015) and motor (Carnevale et al., 2015; Kawai et al., 2015; Mello et al., 2015; Perrett et al., 1993) timing tasks on the scale of hundreds of milliseconds to a few seconds. This "multiple clock" strategy likely evolved because different tasks have distinct computational requirements. For example, judging the duration of a red traffic light relies on the judgement of absolute temporal durations, but tying your shoe and playing the piano rely on the relative timing and order of activation of similar sets of muscles. A general property of these complex forms of motor control is temporal scaling: well-trained motor behaviors can be executed at different speeds. Despite the importance of temporal scaling in the motor domain, basic psychophysical and computational questions remain unaddressed. For example, is temporal scaling intrinsic to motor timing? In other words, once a complex pattern is learned can it be accurately sped-up or down, like choosing the speed at which a movie is played?

The neural mechanisms underlying temporal scaling remain unknown in part because the neural mechanisms underlying motor timing itself continue to be debated. Converging evidence from theoretical (Itskov et al., 2011; Laje and Buonomano, 2013; Medina et al., 2000) and experimental studies suggests that motor timing is encoded in patterns of neural activity, i.e. population clocks (Bakhurin et al., 2017; Carnevale et al., 2015; Crowe et al., 2014; Jin et al., 2009; Lebedev et al., 2008; MacDonald et al., 2011; Mello et al., 2015; Pastalkova et al., 2008). Numerous computational models have been proposed to account for timing (Hardy and Buonomano, 2016; Hass and Durstewitz, 2016) but the problem of temporal scaling remains largely unaddressed.

Here we show that under the appropriate training conditions RNNs exhibit temporal scaling. The model also accounts for a signature feature of motor timing referred to as the scalar property or Weber's law: the standard deviation of timed responses increases linearly with time (Gibbon, 1977). However, the model predicts that the relationship between variance and time is not constant, but dependent on speed. A psychophysical study in which humans produce a complex pattern of taps confirmed this prediction: humans were more precise at the same absolute time while producing motor patterns at higher speeds.

3.3: Results

It is well established that humans can execute well-trained complex movements such as speaking or playing a musical instrument at different speeds. However, it is not clear how well complex temporal pattern can be automatically executed at different speeds. A few studies have examined temporal scaling in humans (Collier and Wright, 1995; Diedrichsen et al., 2007), however, to the best of our knowledge no studies have trained subjects to learn an aperiodic temporal pattern at a single speed, across days, and then examined subject's ability to reproduce that pattern at faster and slower speeds. We thus first addressed whether temporal scaling is an intrinsic property of motor timing by training subjects on a temporal pattern reproduction task (Online Methods). To ensure that subjects were learning a novel pattern—and thus that any temporal scaling was not the result of previous experience—subjects learned to tap out a Morse Code pattern (the word "time") at a speed of 10 words-per-minute (the duration of a "dot" was 120 ms). The target pattern was composed of six taps and lasted 2.76 s (**Fig. 3.1A**).

After training for four days, subjects were tested on their ability to produce the pattern at the original speed, twice as fast (50% duration), and at half speed (200% duration) under freeform conditions—i.e., they were not cued with any target pattern during this test phase. At the 1x speed subjects were able to reproduce the word "time" in Morse code with a performance score

(correlation between the produced and target patterns) of 0.66 ± 0.04. As expected in a freeform condition, few subjects reached the speeds of 2x and 0.5x. Thus we were able to measure how well subjects were able to temporally scale the trained pattern, and whether generalization performance was related to the magnitude of scaling (speed). We quantified temporal scaling using a scaling index based on the time normalized correlation (Online Methods) between the 1x and scaled patterns (**Fig. 3.1B**). For both the fast and slow patterns there was a significant correlation between the scaling index and overall pattern duration (r=0.75, p=0.008; and r=-0.63, p=0.038, respectively). Further analysis revealed a trend for the normalized RMSE (NRMSE) to be smaller for the trained 1x speed, and that most of the NRMSE was attributable to the standard deviation as opposed to the bias (i.e., the difference in the average response and target times; **Supplementary Fig. 3.1**) pattern. These results confirm that with moderate levels of training, humans are intrinsically able to speed up or slow down a learned motor pattern, but that performance degraded at untrained speeds.

3.3.A: RNN Model of Motor Timing

How can neural circuits generate similar temporal patterns at different speeds? To begin to examine the potential mechanisms of temporal scaling, we turned to a population clock model of timing that has previously been shown to robustly generate both simple and complex temporal patterns (Laje and Buonomano, 2013). The model consists of an RNN composed of randomly connected firing rate units whose initial weights are relativity strong, placing the network in a high-gain regime. In this regime RNNs exhibit complex (high-dimensional) patterns of activity. In theory, this activity can encode time while retaining long-term memory traces on scales much longer than the time constants of the units. In practice, however, this memory is limited by chaotic dynamics in the RNN (Sompolinsky et al., 1988). Chaotic behavior impairs the networks' computational capacity because their activity is not reproducible across trials. It is possible, however, to tune the recurrent weights to "tame" the chaos while maintaining their complexity

(Online Methods). The result is the formation of locally stable trajectories, i.e. dynamic attractors, that robustly encode temporal motor patterns. We first asked whether these RNNs can account for temporal scaling.

An intuitive mechanism for temporal scaling is that increased external drive onto a network increases the speed of its dynamics. Thus, to test whether this class of RNNs could account for temporal scaling, we examined the effects of altered input drive on speed. The RNNs received two independent inputs: one transient cue to start a trial and a second tonic speed input (y^{SI}), to modulate the speed of the dynamics. The recurrent units generate motor patterns through synapses onto a single output unit (**Fig. 3.2A**).

We trained the RNNs to reproduce an "innate" pattern of network activity while receiving a fixed amplitude speed input (defined as speed 1x, $y^{SI} = 0.15$), then trained the output to produce an aperiodic pattern composed of five "taps" after the cue offset (Online Methods). Unlike biological motor systems, RNNs in high-gain regimes are typically spontaneously active, i.e., their activity is self-perpetuating. To increase the model's congruence with cortical dynamics and motor behavior, we developed a method of training the recurrent units to enter a rest state when not engaged in a cued task. In this procedure, the recurrent units are trained to maintain a firing rate of approximately zero after the target pattern has terminated (Online Methods). This training produces a "gated dynamic attractor": in response to a cued input the network produces the trained dynamics and then returns to a rest state (**Fig. 3.2B**). In contrast, in response to an untrained input the network activity quickly decays to the rest state. Consistent with the lack of spontaneous activity the real eigenvalues of the trained weights are less than one (**Supplementary Fig. 3.2**).

After training, the network was able to reproduce the target output at the trained speed. However, when tested at novel speeds—by changing the amplitude of the tonic speed input—the

network did not exhibit significant temporal scaling (**Fig. 3.2C**). Notably, the network's ability to produce the trained pattern progressively degraded with more extreme speed inputs. These findings establish that simply changing the amplitude of a tonic input cannot account for the degree of temporal scaling naturally observed in humans.

3.3.B: RNN Model of Temporal Scaling Predicts a Weber -Speed Effect

The above result suggests that robust temporal scaling is not an intrinsic property of timing in RNNs, so we next examined whether temporal scaling could be learned. We trained RNNs to produce the same pattern of activity in the recurrent units at two different speeds (0.5x and 2x, Online Methods). After the recurrent network was trained, we trained the output to produce the same pattern as in **Fig. 3.2** but only at the 2x speed. When tested at different speed input levels, these networks robustly and linearly scaled their activity to novel speeds (**Fig. 3.3A-B**). Note that because the output was trained only at 2x speed, any change in the speed of the output reflects an underlying change in the speed of recurrent activity. Compared to RNNs trained on a single speed, those trained on two speeds accurately compressed or dilated the motor pattern between the trained 2x and 0.5x speeds—there was a small degree of intrinsic temporal scaling in the RNNs trained at one speed (black lines in **Fig. 3.3C**), however, the scaling was very limited (0.9x to 1.15x). In other words, when trained on two speeds RNNs accurately interpolated to untrained speed inputs. Even at speeds outside the trained range, there was some temporal scaling, although it progressively degraded (**Supplementary Fig. 3.3**).

As mentioned above, the standard deviation (SD) of timed motor responses increases linearly with time (Church et al., 1994; Gibbon, 1977). Because Weber's law is often held as a benchmark for timing models (Hass and Durstewitz, 2016), we examined whether the SD of the model's cross-trial tap times was linearly related to absolute time (the mean time of each tap in the sequence). There was a strong linear relationship between standard deviation and time, (**Fig.**

3.4B) which allowed us to calculate the Weber coefficient (defined as the slope of the variance versus t^2). In contrast to a number of other models of timing—simple drift-diffusion models for example (Hass and Durstewitz, 2014)—RNNs can inherently account for Weber's law. This is in part because the recurrent nature of these networks can amplify noise and impose long-lasting temporal noise correlations, leading to near linear relationships between SD and time (Hass and Herrmann, 2012) (**Supplementary Fig. 3.4**)

Across speeds there was a clear inverse relationship between the speed and coefficient of variation (CV or Weber fraction, **Fig. 3.4C**), as well as with the Weber coefficient (**Fig. 3.4D**). Specifically, the lower the speed the higher the Weber coefficient. Moreover, this Weber-speed effect was robust to changes in network size, noise amplitude, and if the RNNs were trained to slow down at higher input amplitudes (**Supplementary Fig. 3.5**). This is a counterintuitive observation, as it implies that at the same absolute time, temporal precision is significantly lower at slower speeds. To use a clock analogy: it would be as if the same clock was more precise at timing a two second interval when that interval was part of a short (high speed) pattern compared to a two second interval that was part of a long (slow) pattern. In other words, the model predicts that humans will be less precise tapping halfway through a four second pattern, compared to the last tap produced at the end of the same pattern produced over two seconds.

The learning rule used to train the RNNs provides a robust means to generate complex and highly stable spatiotemporal patterns (but is not meant to represent biologically realistic learning rule in recurrent neural networks). It is possible that the Weber-speed effect emerged from some property specific to the innate-training learning rule. Thus, we also examined timing in RNNs trained with Hessian-free backpropagation through time (Martens and Sutskever, 2011; Sussillo and Barak, 2013) and a standard echo-state network (Jaeger and Haas, 2004; Sussillo and Abbott, 2009). While these rules are not as well suited to learn complex long-lasting aperiodic

temporal patterns, they nevertheless did generate RNNs that exhibit a Weber-speed effect (**Supplementary Fig. 3.6-7**). Thus, our results suggest that the Weber-speed effect is a robust property of timing generated by the dynamics of recurrent neural networks.

3.3.C: Test of Prediction: Humans Exhibit the Weber x Speed Effect

To the best of our knowledge the notion that temporal precision is worse for complex temporal patterns produced at low speeds has never been predicted or experimentally tested. Thus, we used a temporal reproduction task in which subjects were required to reproduce an aperiodic pattern composed of six taps at five different speeds. The pattern and speeds were the same as those in the model above, with the first tone in this task acting as a self-initiated cue. Subjects (n=25) listened to an auditory pattern composed of six tones, and were asked to reproduce it using a keypad (**Fig. 3.5A**, Online Methods). In each block subjects heard the pattern at one of five temporally scaled speeds (0.5x, 0.66x, 1x, 1.5x, and 2x) and reproduced the pattern (**Fig. 3.5B**, single subject). Based on the mean and SD of the taps it is possible to calculate the CV for each tap at each speed, as well as the Weber coefficient (inset **Fig. 3.5B** right, SD versus *t* is shown for visualization purposes). Under Weber's law the CV of timed responses should be the same for taps at the same absolute time, and the Weber coefficients should be the same across speeds. However, across subjects (**Fig. 3.5C**) CVs were significantly different across speeds ($F_{4,96}$=8.40, $p<10^{-5}$, speed effect of a three-way repeated ANOVA), and the Weber coefficient decreased with higher speed ($F_{4,96}$=4.18, p<0.005, one-way repeated ANOVA).

The above data is potentially confounded with task difficulty or learning—that is, the difference in the Weber coefficients across speeds could potentially reflect some nonspecific effect in which slower patterns are harder to memorize and reproduce. We thus trained a subset of subjects (n=14) on the fastest and slowest speeds over an 8-day period. Again, at the same absolute time the CV was lower for the faster speed across training days (e.g., ≈0.7 s in **Fig. 3.5D**). The Weber

coefficient was significantly smaller for the faster speeds across training days (**Fig. 3.5D**, inset; $F_{1,13}$=16.58, p<0.002, speed effect two-way repeated ANOVA; pairwise posthoc test on each day, maximum p = 0.056, Tukey-Kramer)—even as subjects showed asymptotic learning, seen in the progressive decrease in the Weber coefficients of both speeds across days. To further confirm the existence of a Weber-speed effect and examine its potential dependence on the degree of training we performed a second study in which subjects (n=15) were trained on three speeds (0.5x, 1x, and 2x) across three days. Analysis of the Weber coefficient across speeds and days again revealed a robust effect of speed (**Supplementary Fig. 3.8**), $F_{2,28}$=11, p<0.0005, two-way repeated ANOVA) as well as an improvement across days ($F_{2,28}$= 7.1, p<0.005), and no significant interaction between speed and day of training. These results confirm that temporal precision is better at faster speeds.

3.3.D: Speed or Subdivision?

The Weber-speed effect is potentially related to the so-called subdivision effect. Specifically, it is well established that the timing of a given absolute interval can be improved by subdividing that interval into smaller subintervals—e.g., tapping your foot or counting—can improve the timing of a longer interval (Grondin and Killeen, 2009; Grondin et al., 2004). Subdivision cannot account for the Weber-speed effect in the model, because the internal dynamics is independent of what the output unit is trained to do, but it could explain the psychophysical results because the component subintervals of the pattern are shorter at higher speeds. To directly compare both the speed and subdivision hypothesis in the psychophysics experiments we trained subjects on a standard periodic subdivision task over five days. Subjects produced a series of taps with a total duration of 2400 ms, with four different inter-tap intervals (speeds; **Fig. 3.6A**). Similar to results from the aperiodic temporal pattern, subjects showed reduced variability at the same absolute time when the inter-tap-interval was shorter (**Fig. 3.6B**). Here, the subdivision and speed hypotheses are confounded, but they can be dissociated based on the standard explanation of

the subdivision effect. Subdivision is hypothesized to improve timing because the "clock" is reset at each tap (Fetterman and Killeen, 1990), whereas in our population clock model timing of a complex pattern relies on a timer that is running continuously. In the case of a single interval both views generate the same variance, but in the case of a pattern composed of a sequence of intervals (t_1, t_2, …, t_n) they generate different variance signatures (**Fig. 3.6C**). Thus we fit each subject's responses assuming either a speed ("continuous") or subdivision ("reset") interpretation of Weber's generalized law. While the both fits captured the data very well, the goodness-of-fit of the speed prediction was significantly better (**Fig. 3.6C-D,** fits for day 5 shown, $F_{1,10} = 12.8$, $p=0.005$, two-way repeated ANOVA on Fisher transformed r^2 values). These results suggest that the standard subdivision effect may be best interpreted not as produced by the reset of an internal timer but the increase in speed of the internal dynamics of a population clock.

3.3.E: Mechanisms of Temporal Scaling

Having established a model of temporal scaling that generated a correct prediction we next used the model to examine the potential network-level mechanisms underlying temporal scaling. At first glance the notion that an RNN can generate the same trajectory at different speeds is surprising, because it seems to imply that different tonic inputs can guide activity through the same points in neural phase space at different speeds. Furthermore, it is important to emphasize that the relationship between input amplitude and speed is arbitrary: the model exhibits temporal scaling whether the network is trained so that larger speed inputs increase *or* decrease trajectory speed (**Supplementary Fig. 3.5G**), implying that temporal scaling is an emergent phenomenon. Additionally, it is not the case that any RNN will generate temporal scaling in response to changes in the amplitude of a tonic input. In the regime used here, untrained RNNs are chaotic so any change in input produces dramatically different trajectories, and even when trained on one speed the network did not exhibit robust generalization (**Fig. 3.2**), furthermore when RNNs were trained

on two speeds with the Hessian-free backprop algorithm robust temporal scaling was not observed (**Supplementary Fig. 3.6**).

Because the network is trained to reproduce the exact same trajectory at two different speeds, the most straightforward way temporal scaling to novel speeds could emerge in the trained RNNs is via the formation of parallel neural trajectories. But such a mechanism could take two forms: nearby trajectories that are traversed at different speeds, or distant trajectories that are traversed at the same speed. As a first step toward examining the underlying mechanisms we first visualized the trajectories in PCA space, which revealed that the trajectories at different speeds are not overlapping, but follow offset paths through neural phase space that are traversed at different speeds (**Fig. 3.7A**). In other words, the trajectories are arranged according to speed in an apparently "parallel" manner. To quantify this observation, we calculated the Euclidean distance in neural space (N=1800) between the trajectory at each speed and the 0.5x speed (**Fig. 3.7B**). Finding the minimum distance between the comparison speed and the 0.5x speed (the "reference" trajectory) revealed that the trajectories maintain a relatively constant distance from each other (**Fig. 3.7C**). Examining the time points at which the trajectories were closest also provided an unbiased estimate of the relative speed of each trajectory. For example, if the test trajectory is moving four times as fast as the reference trajectory, it should be closest to the reference trajectory when it has been active for ¼ the elapsed time. In other words, plotting t_{min}^{2x} vs $t_{elapsed}^{0.5x}$ should form a line with slope 0.25, which is indeed what we observed. Moreover, this relationship generalized to novel interpolated speed levels (**Fig. 3.7D**).

The above observations confirm that temporal scaling is a result of the RNN generating nearby parallel trajectories, but do not address how changing the magnitude of a static input results in trajectory speeds that scale approximately linearly with the input magnitude. Understanding the underlying dynamics of complex nonlinear neural networks is a notoriously challenging problem

with only a few tools available (Sussillo and Barak, 2013). Here we introduce a method to dissect the internal forces driving a network. We first quantified the total drive to the network: the time-dependent change in the total synaptic input onto each neuron in the RNN. Measuring the magnitude (Euclidean norm) of the total drive showed that—in contrast to untrained networks or to networks trained at a single speed—the total drive scaled with the cued speed (**Fig. 3.8A**). To address how the total drive scales the neural dynamics, we used a novel network drive decomposition method (Goudar and Buonomano, 2017). This approach decomposes the total network drive into its three components: the recurrent synaptic drive, the synaptic decay (which is always driving the network towards the origin), and the external tonic (time-independent) speed input (**Fig. 3.8C**). While the speed input magnitude scaled with speed as defined by the experimental conditions, the recurrent and decay drive magnitudes did not, meaning that the recurrent and decay components in isolation cannot account for temporal scaling (**Fig. 3.8B**).

Analysis of the dynamics also revealed that, at each speed, the trajectories traversed directions that are independent of the speed input—i.e., the projection of each trajectory onto the speed input axis has low variance (explained variance was <1% at all speeds). There are two consequences to the observations that the time-varying dynamics are not driven by the input, and that the recurrent drive and decay magnitudes did not exhibit temporal scaling: 1) at each speed, some combination of these internal drive components counterbalance the speed input; and 2) they *collectively* underlie temporal scaling of the trajectory. In order to isolate the contribution of these interactions we studied the internal drive components in the subspace orthogonal to the speed input axis (**Fig. 3.8C**). Measurements showed that even in this subspace, changes of recurrent drive and decay magnitudes did not explain temporal scaling of the total drive (data not shown). Instead, the recurrent synaptic drive and decay opposed each other (the angle between them is obtuse) throughout the trajectory, and the extent of this opposition altered the trajectory's

speed (**Fig. 3.8D**). Specifically, the angle between the two components decreases as the speed input increases ($\theta_2 < \theta_1$), amplifying the net (or total) drive.

Projecting the trajectories onto the speed input axis revealed that speed is encoded in the trajectory's position rather than its direction (**Fig. 3.8E**). Moreover, by traversing phase space along directions that are independent of the speed input, the trajectory's position with respect to the speed input stayed relatively constant. This allows the speed of the trajectories to stay relatively constant. To confirm this, we asked if—as with biological motor patterns—a network could switch speeds mid-trajectory. Indeed, by decreasing the speed input in the middle of a fast (2x) trajectory we observed a rapid transition to the slow trajectory (**Fig. 3.8F**). Network drive decomposition showed that a change in the speed input caused an imbalance between it and the internal drive, altering the position of the trajectory along the speed input axis. In turn, this increased the angle between the recurrent and decay drives, slowing the trajectory down. It also rebalanced the speed input and the internal drive components such that trajectory speed stops changing when the balance between input and internal drive was restored (**Fig. 3.8E**). Together, these results demonstrate that temporal scaling is the outcome of speed input-dependent balance between the recurrent and decay drives.

3.4: Discussion

It is increasingly clear that on the scale of hundreds of milliseconds to seconds the brain represents time as dynamically changing patterns of neural activity ("population clocks") (Bakhurin et al., 2017; Carnevale et al., 2015; Crowe et al., 2014; Mello et al., 2015; Stokes et al., 2013). Timing on this scale exhibits a number of properties including: 1) the ability to execute the same motor patterns at different speeds, and 2) temporal variability increases linearly with time (Weber's law). Here we unify and extend these observations by presenting a recurrent neural network model that not only performs temporal scaling and accounts for Weber's law, but also

predicts that Weber's law is speed-dependent. Specifically, the Weber coefficient is speed-dependent. We tested this prediction using human psychophysics experiments, and confirmed that in absolute time the temporal precision of motor responses is dependent on speed.

Few studies have quantified temporal scaling of complex aperiodic motor patterns in humans (Collier and Wright, 1995; Diedrichsen et al., 2007). However, studies in the sensory and sensory-motor domain have clearly established that learning of simple sensory intervals does not generalize to different intervals (Bueti and Buonomano, 2014) but can generalize to different modalities, including motor (Meegan et al., 2000; Merchant et al., 2008; Planetta and Servos, 2008). In a manner of speaking, temporal scaling of motor patterns (e.g., **Fig. 3.1**) represents generalization to different intervals. However it is important to emphasize the difference between interval and pattern timing (Hardy and Buonomano, 2016). Simple intervals are defined by their absolute duration—i.e., the difference between a scaled interval and a different interval is ambiguous—whereas patterns can be defined by the relationship of the component subintervals. Thus, the apparent difference between generalization of learned intervals and patterns could be related to different underlying neural mechanisms.

3.4.A: Weber's Law

While Weber's law is a well-established property of timing in humans (Cicchini et al., 2012; Ivry and Hazeltine, 1995a; Jazayeri and Shadlen, 2010) the neural underpinnings of Weber's law have long been debated (Hass and Durstewitz, 2014). Early models of timing (internal clock models) consisted of an accumulator that integrated the number of pulses of a noisy oscillator. In their simplest form, however, these models did not account for Weber's law because the standard deviation of such a clock will increase as a function of $\sqrt{t}$ rather than *t.* Thus, early internal clock models postulated that Weber's law arises from a second clock-independent noise source, such as the memory of the interval being generated (Gibbon, 1977; Grondin et al., 2004). Other models

(Ahrens and Sahani, 2008; Balcı and Simen, 2016; Hass and Durstewitz, 2014), including those based on the variance between multiple timers, can intrinsically account for Weber's law, but the biological plausibility of such variance-based models remains an open question. Our results suggest that population clock models based on recurrent dynamics can also intrinsically account for Weber's law. Theoretical analyses have shown that Weber's law can arise from temporal noise correlations (Hass and Herrmann, 2012); and because RNNs can actively amplify noise through internal feedback, this mechanism likely contributes to the emergence of Weber's law.

As has been pointed out, Weber's law raises an important question: if independent noise sources cause the SD to increase as a function of $\sqrt{t}$, why does the nervous system settle for Weber's law (Hass and Herrmann, 2012)? First, it is possible that this reduced accuracy is an unavoidable consequence of the abundance of correlated noise on time scales at which the intervals being timed (Osborne et al., 2004). For example, in any neural circuit, slow fluctuations produced by sensory inputs or other brain areas will impose local temporal correlations. Second, the amplification of internal noise by recurrent neural networks may make Weber's law a necessary cost of the increased computational capacity provided by such architectures.

3.4.B: Weber-Speed Effect

The observed increase in temporal precision with motor speed raises the question of why a Weber-speed effect has not previously been reported. One reason is that the vast majority of timing studies have relied on interval or duration tasks rather than pattern timing. Thus, in most studies the Weber coefficient is calculated by fitting the variance of timed responses of distinct intervals collected across blocks. With this approach it is not possible to explicitly address questions pertaining to temporal scaling and the Weber coefficient. In contrast, by looking at the timing of complex motor patterns consisting of multiple taps (Laje et al., 2011), it is possible to estimate the Weber coefficient within each condition (speed), revealing a dependence of the

Weber coefficient on speed. As mentioned above this Weber-speed effect is confounded with the subdivision effect, in which subdividing a target interval into subintervals can improve temporal precision (Grondin and Killeen, 2009; Grondin et al., 2004). We suggest that the subdivision effect may be best reinterpreted as a speed effect. First, in the RNN model the improvement in precision is clearly an effect of speed because, as implemented here, timing is independent of the behavior of the output unit (e.g. the number of taps). Second, the subdivision hypothesis predicts a timer reset at each tap, yet a goodness-of-fit analysis revealed that the continuous version of Weber's generalized law provided better fits (**Fig. 3.6**). We thus hypothesize that subdivision effects may in part reflect the speed of the underlying neural trajectories. However, future studies will have to further examine the interrelation or independence of the Weber-speed and subdivision effects, as well as determine if the Weber speed effect represents a smooth linear transition or discrete steps reflecting different timing mechanisms.

As with Weber's law, the Weber-speed effect raises the question of why the nervous system would utilize a timing mechanism that is inherently better—more precise across trials—when engaged in a fast versus a slow motor pattern. Again, the answer may lie in part in the inherent properties of recurrent circuits. Our analysis of temporal noise correlations revealed larger and longer lasting noise covariance in the RRN during slow trajectories compared to fast trajectories (**Supplementary Fig. 3.9**). Additionally, there is an inherent inverse relationship between the rate-of-change of a dynamical system and the effects of noise. Consider a sinusoidal function at a fast (short period) and slow (long period) speed in the presence of additive noise. If we were to consider each peak in the amplitude of the function as a tic of a clock, additive noise will produce more temporal variance in the peaks of the slow curve because noise added to a slowly changing function is more likely to change the times of the peaks.

3.4.C: Predictions

The model of temporal scaling presented here makes a number of experimental predictions. The most important prediction, that movements executed at higher speeds are more temporally precise in absolute time, has been tested and confirmed. However, a number of important questions remain, including whether simple interval production tasks correspond to executing the same neural trajectories at different speeds. Two studies suggest that during a timing task it is indeed the case that different intervals are timed by similar neural patterns unfolding at different speeds (Mello et al., 2015; Wang et al., 2017). However, no electrophysiological studies have examined temporal scaling during the production of aperiodic temporal patterns similar to those studied here.

The model makes a number of additional neurophysiological predictions. First, electrophysiological recordings during temporal scaling to untrained speeds should produce neural trajectories whose positions on a manifold (a “surface”) in high-dimensional space reflect the speed of the motor pattern. Second, the model predicts that slower trajectories should exhibit larger temporal noise covariance. In other words, on a trial-by-trial basis, when the population clock reads early at the beginning of a trajectory, that deviation will persist longer if the trajectory is moving slowly.

While we propose that the model presented here captures general principles of how neural dynamics account for timing and temporal scaling, the learning rule used to generate the neural trajectories driving timing is not biologically plausible. Future research will have to determine whether such regimes can emerge in a self-organizing manner. However, because the Weber-speed effect was observed across learning rules we expect it to be a general property of timing with population clocks (**Supplementary Fig. 3.5-6**). Additionally, while the model is agnostic to

what parts of the brain generate such patterns, we hypothesize that similar regimes exist in neocortical circuits characterized by recurrent excitation.

Overall the current studies support the notion that many neural computations can be implemented not by converging to a point attractor (Hopfield, 1982; Wang, 2001), but as the voyage through neural phase space (Buonomano and Maass, 2009; Durstewitz and Deco, 2008; Rabinovich et al., 2008). And more specifically, that these trajectories represent dynamic attractors, which can encode motor movements and are robust to perturbation—that is, they can return to the trajectory after being bumped off (Laje and Buonomano, 2013). Here we show that recurrent neural networks can exhibit regimes with parallel families of neural trajectories that are close enough to drive the same motor pattern but that are traversed at different speeds—accounting for temporal scaling. These regimes predict that the temporal precision of motor responses in absolute time is dependent on speed of execution. This prediction was confirmed in human timing experiments, establishing a novel psychophysical Weber-speed effect.

3.5: Figures

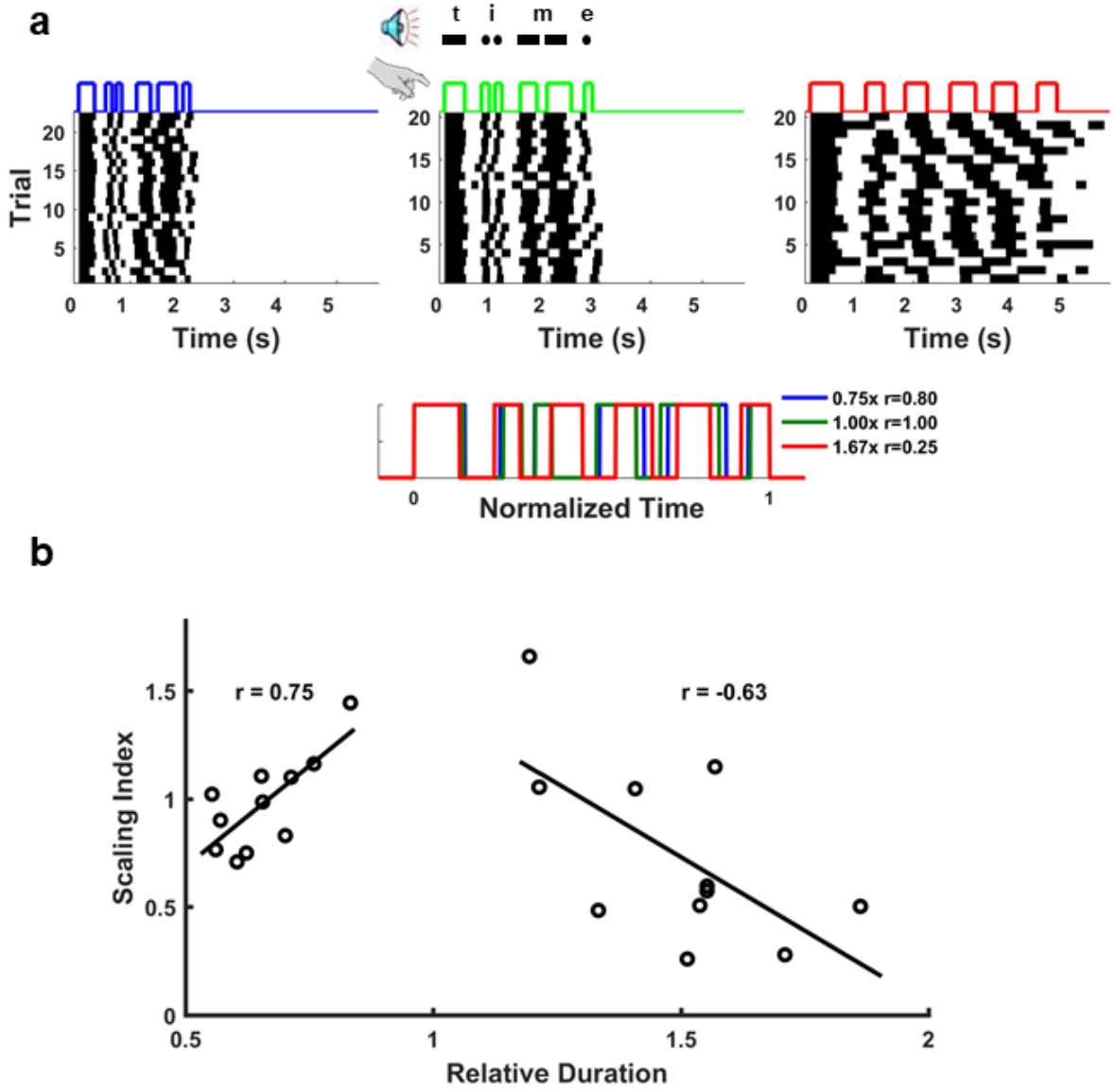


Figure 3.1. Limited temporal scaling of a learned Morse code pattern.

Subjects were trained to tap the Morse code for "time" at a speed of 1x (10 wpm) over four consecutive days (Online Methods). **A)** On the fifth day, subjects were asked to produce the pattern at three different speeds: twice as fast (2x), normal speed (1x), and twice as slow (0.5x) (data from a single subject). Bottom: Average of the responses above plotted in normalized time. The legend indicates the produced speed relative to the trained (1x) condition and the correlation of the mean response to the response at trained speed. **B)** The relationship between produced speed and temporal scaling accuracy for all eleven subjects. There was a significant correlation between speed and accuracy for both the fast (r=0.75, p=0.008) and slow (r=-0.63, p=0.038) patterns.

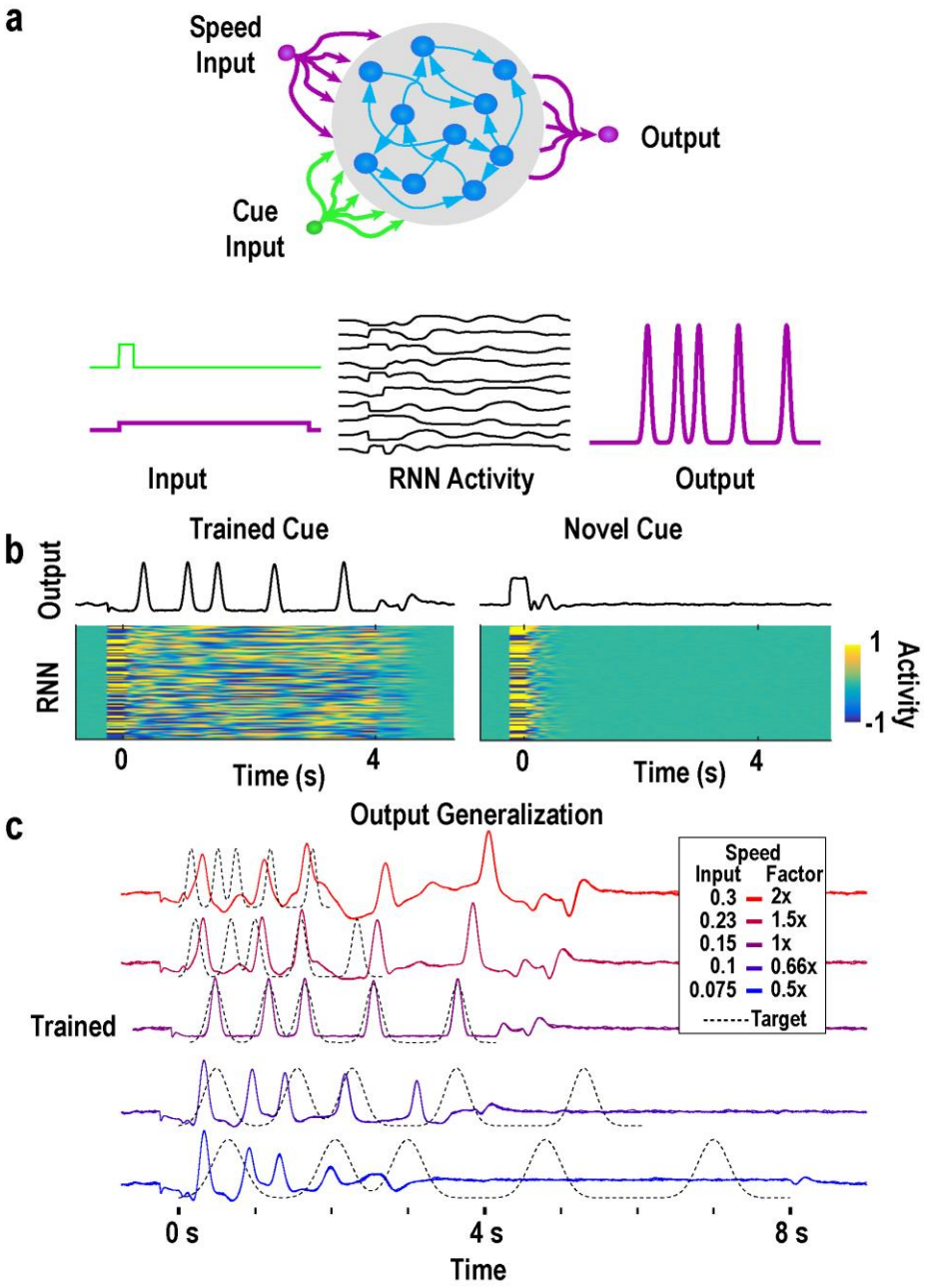


Figure 3.2. Robust temporal scaling is not produced by altered input drive of a RNN model.

A) The model was composed of recurrently connected firing rate units, which received two external inputs and connected to a single output. One input served as a start cue and was active briefly at the start of each trial between $t = [-250, 0]\ ms$. The second input delivered a constant low amplitude speed signal for the duration of a trial. **B)** The RNN was trained to autonomously produce a neural trajectory lasting four seconds at 1x speed ($y^{SI} = 0.15$). At the end of the trajectory, the recurrent network was trained to return to a "rest state" ($r = 0$), forming a "gated" attractor: networks only generate long-lasting stable dynamic activity in response to the trained cue. Following recurrent training, the output unit was trained to produce a series of

five “taps” at 325, 1025, 1500, 2400, and 3500 ms. In response to a novel cue input the RNN activity does not enter the trained dynamic attractor, and activity quickly returns to rest. **c)** Networks trained at one speed do not scale the speed of their dynamics in according to changing input drive. The speed signal was varied between $y^{SI} = [0.3, 0.23, 0.15, 0.1, 0.075]$ corresponding to 2x, 1.5x, 1x, 0.66x, and 0.5x speeds. Dotted lines illustrate the expected output pattern with ideal scaling.

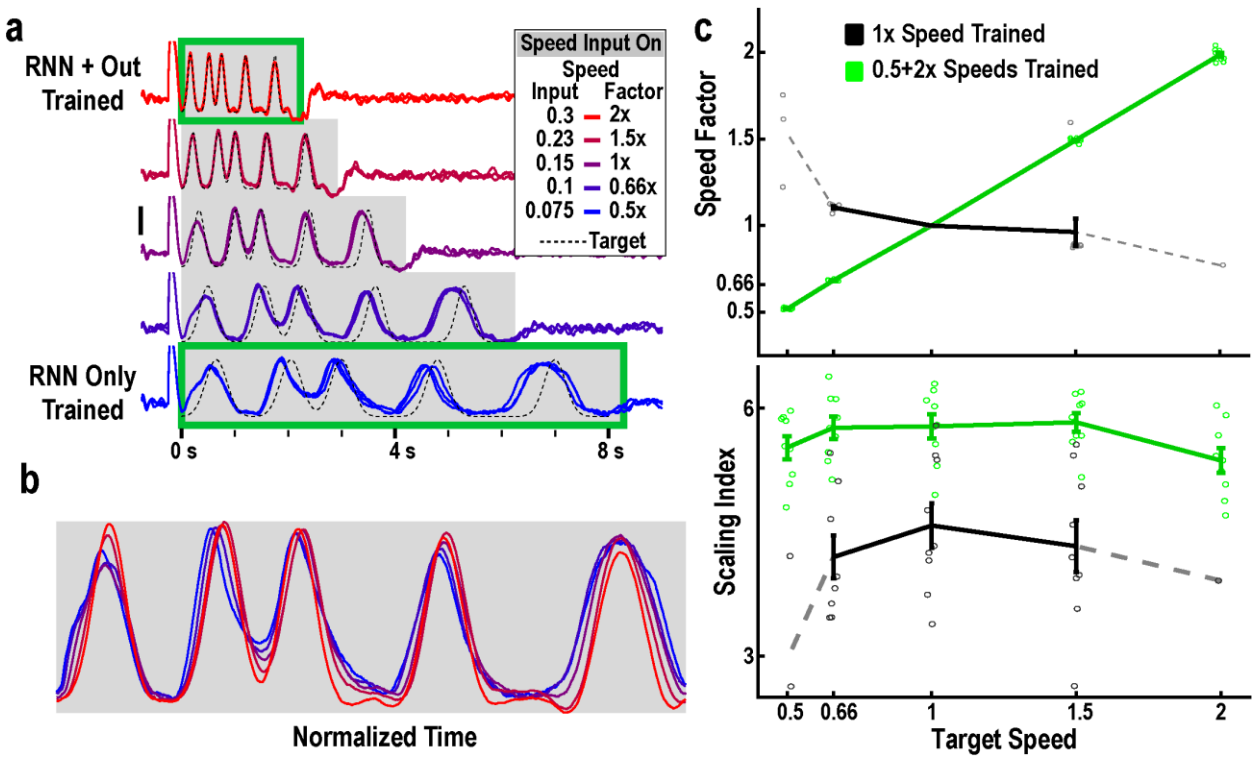


Figure 3.3. RNNs trained at multiple speeds exhibit robust temporal scaling.

A) Output activity of an RNN trained to produce the scaled patterns of recurrent activity at 0.5x ($y^{SI} = 0.075$) and 2x ($y^{SI} = 0.3$) speeds. The output was trained only at the 2x speed. After training (weight modifications stopped), the network was tested at different input speed levels ($y^{SI} = [0.075, 0.1, 0.15, 0.23, 0.3]$)—corresponding to speeds of 0.5, 0.66, 1, 1.5, and 2x. Three example test trials at each speed are overlaid. **B)** One trial from each test speed above shown with time normalized to the end of the active period. **C)** Networks (n=10) trained at two speeds generalize to untrained speed inputs. Top: The speed factor (the mean ratio of the final tap at each speed to the mean final tap time at 1x speed over twenty trials) of networks trained at two speeds matches the target speed (green), but the speed factor of networks trained at one speed does not (black). Bottom: The scaling index of networks trained on two speeds is higher than those trained on one speed. Error bars represent SEM, and circles show the value for each network. Because the activity of the one-speed networks degrades at more extreme speeds as shown in Fig. 3.1, many networks did not produce detectable "taps" (output peaks) at extreme speeds and we therefore could not calculate a scaling index or index for them. We show in grey the calculations for the networks that completed at least one trial at the extreme speeds.

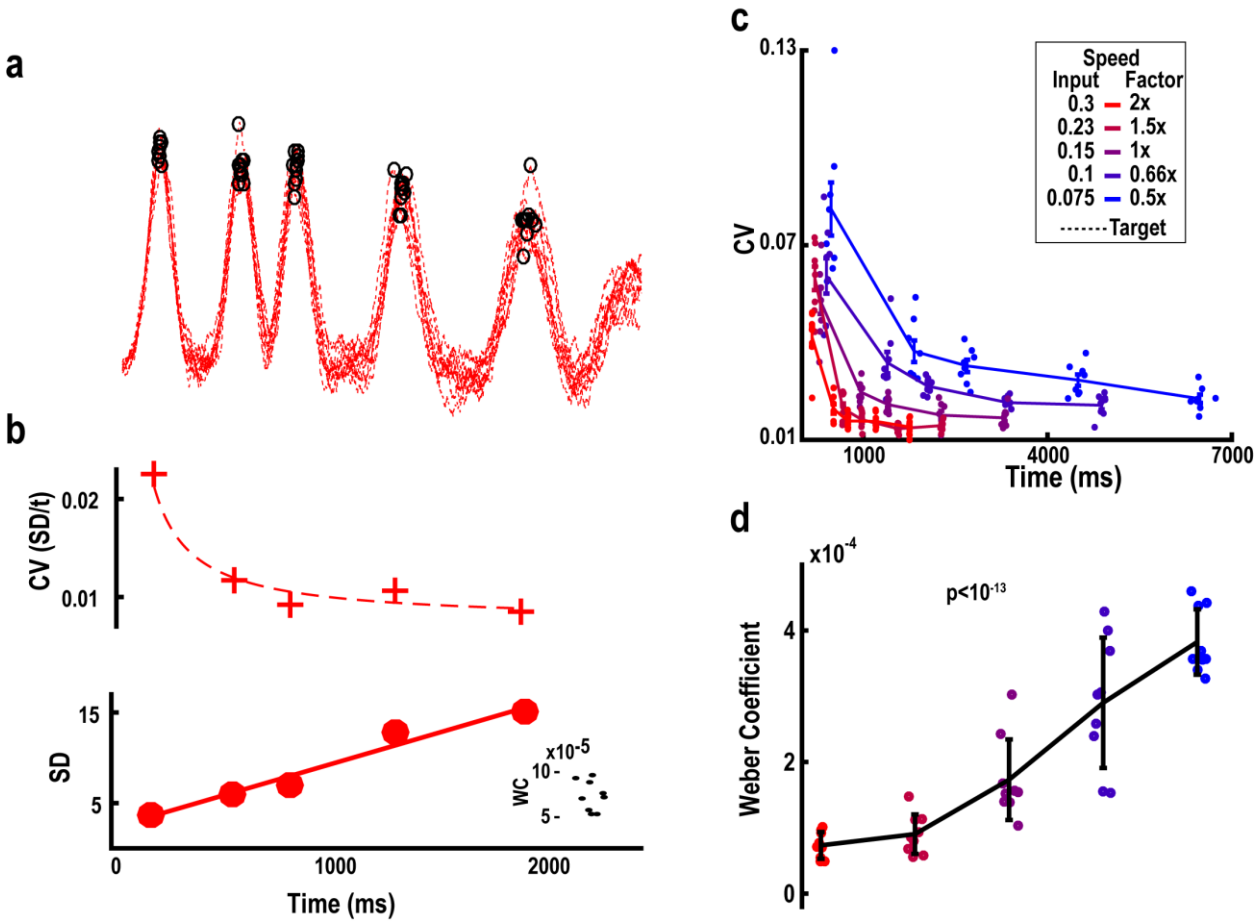


Figure 3.4. RNN models of temporal scaling predict a novel Weber-speed effect.

A) Ten trials of the output activity of one network at 0.5x speed with tap times indicated by black circles. **B)** Trained RNNs account for generalized Weber's Law, which predicts a linear relationship between the mean and standard deviation of timed intervals. Top: The coefficient of variation (CV, SD/t) at each of the five taps shown in **A**. The dotted line shows the CV calculated using the fit below. Bottom: standard deviation linearly increases with time. Line shows the linear fit (r^2=0.96). Inset shows the Weber Coefficient (the slope of variance vs. mean time) at 0.5x speed for all ten trained networks. **c)** The CV of ten networks calculated from twenty trials at each tested speed. Note that at the same absolute time across speeds, the CV is higher when speed is slower (the Weber-Speed effect). **D)** The Weber Coefficient increases at slower speeds (Repeated-measures one-way analysis of variance; F=54.4, $p<10^{-13}$). Networks (n=10) for this analysis were trained and tested at 0.25 noise amplitude.

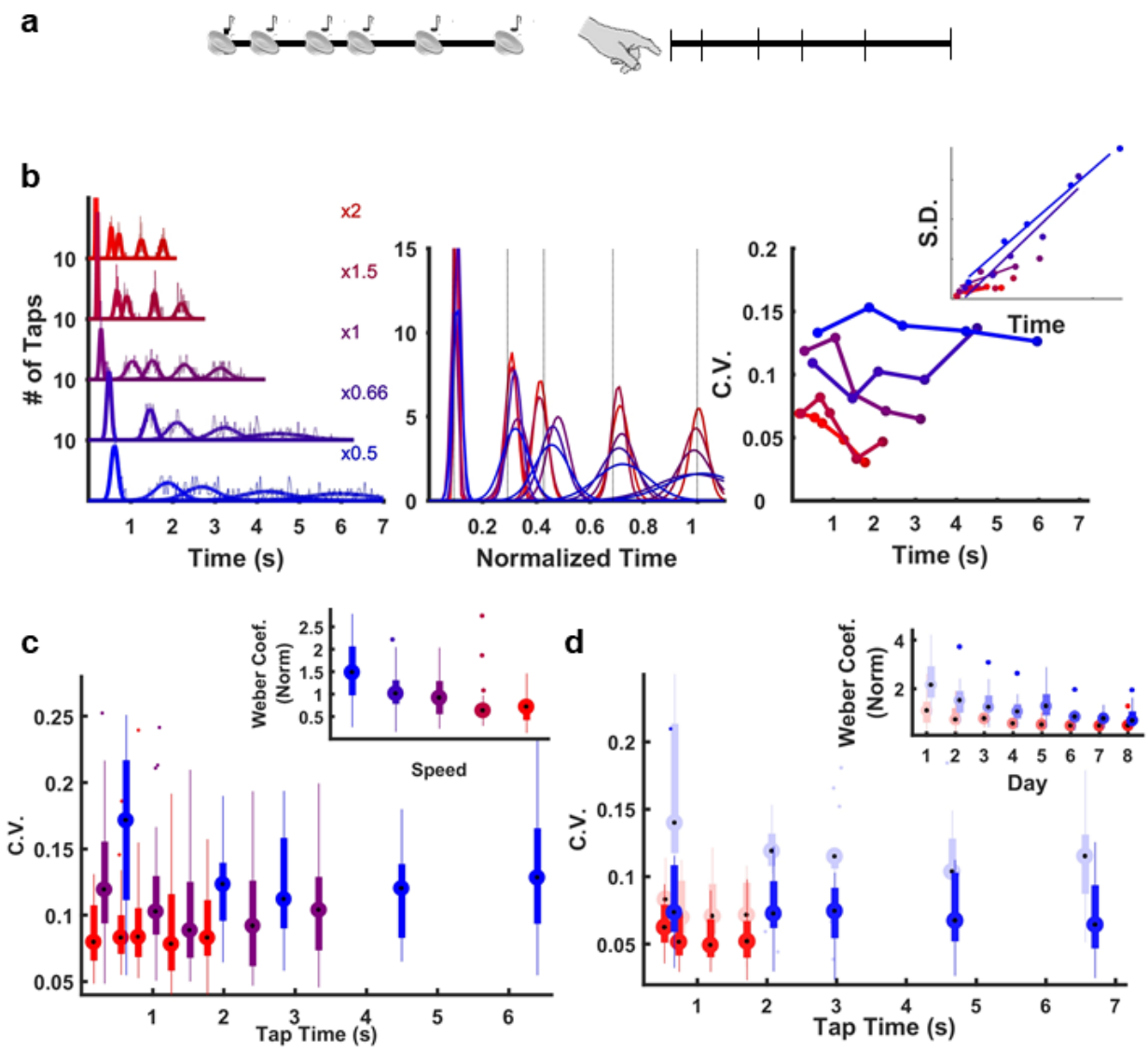


Figure 3.5. Test of the Weber-Speed effect prediction.

A) Subjects were trained on an auditory temporal pattern reproduction task, using the same aperiodic pattern and same five speeds used to test the RNNs. **B)** Left: Histogram (dashed lines) and Gaussian fits (solid lines) of the cued taps at all five speeds from a single subject (bin sizes scale with target duration). Middle: the fits shown with time normalized to the mean of the last tap (vertical lines represent target times)—note that the scaled fits do not overlap as expected by Weber's law. Right: CV of each tap at each speed, with SD vs mean time inset. The slope of the linear fit of the variance verses t^2 corresponds to the Weber coefficient (SD versus time is shown for visualization purposes). **C)** Whisker plots of the CV of all subjects (n=25) for three of the five speeds (0.5x, 1x, and 2x). Note that, as in the RNN model, the CV at the same absolute time is higher at slower speeds. Inset shows the Weber coefficient for all five speeds. **D)** The Weber-Speed effect is not due to inexperience with the task. A subset of fourteen subjects were trained to produce the 0.5x and 2x slow speeds over eight additional days. The Weber-Speed effect persists

over the course of training. CVs are shown for the first (light) and last (dark) day of training for both speeds. Inset: the Weber coefficients across all eight days of training. Whisker plots show the median, lower and upper quartile, 1.5x interquartile range, and outliers.

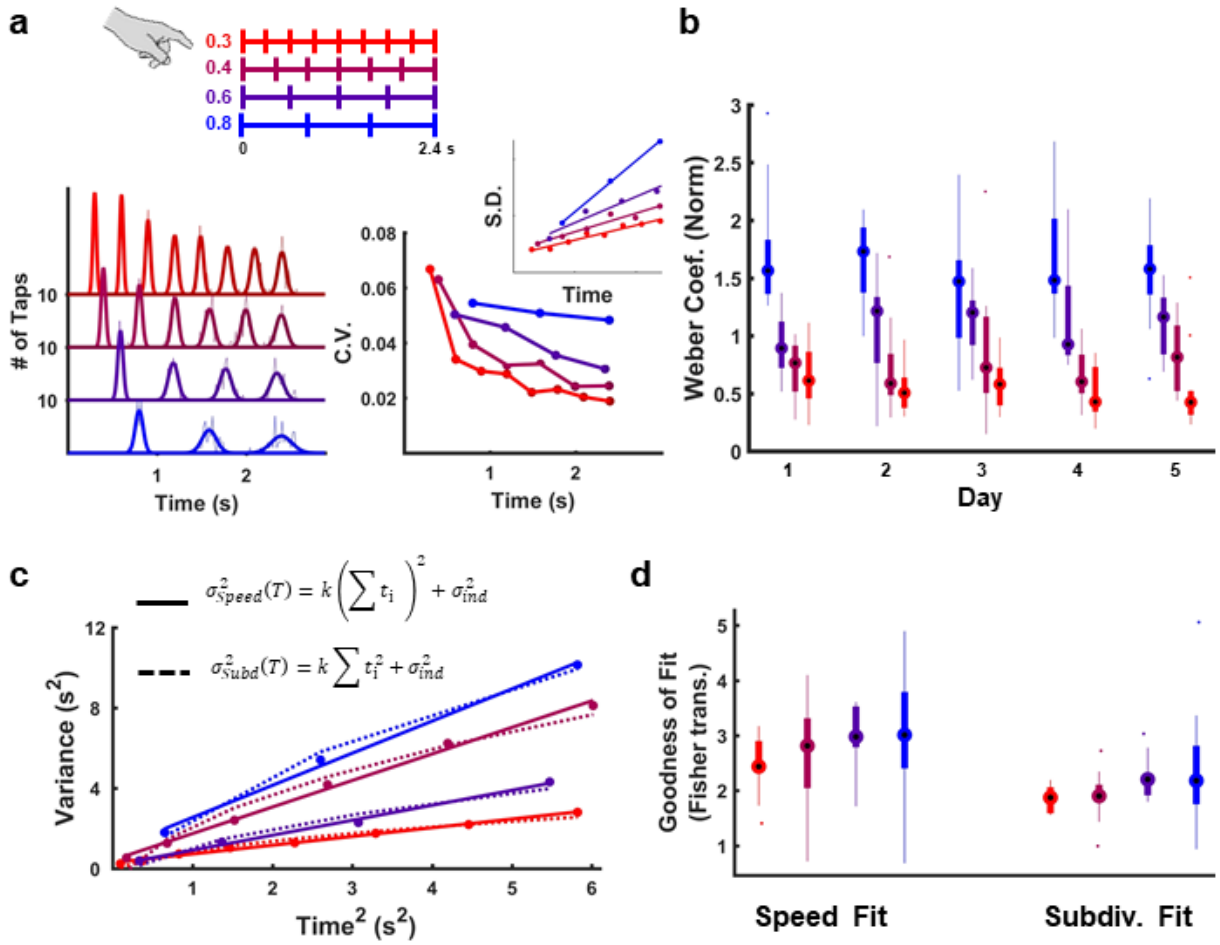


Figure 3.6. Comparison of the Weber-speed and subdivision hypotheses using a periodic task.

Subjects were trained on four periodic auditory temporal patterns all lasting 2.4 s (periods of 0.3, 0.4, 0.6, and 0.8 s) over five days. **A)** Left: Histogram (dashed lines) and Gaussian fits (solid lines) of the taps at all four speeds for a single subject. Right: CV of each tap at each speed, with SD vs mean time shown as the inset. **B)** Whisker plots of the Weber coefficient of all subjects (n=11) across the five days of training. **C)** Example fits of the variance at time *T* composed of n subintervals (t_1, t_2, ..., t_n) according to the speed (continuous, solid lines) and subdivision (reset, dashed lines) hypotheses (σ^2_{ind} represents the time independent source of variance). **D)** Goodness of fit values (Fisher transformed r^2) for both the speed and subdivision hypotheses for each speed across all subjects (data shown are from the last day of training).

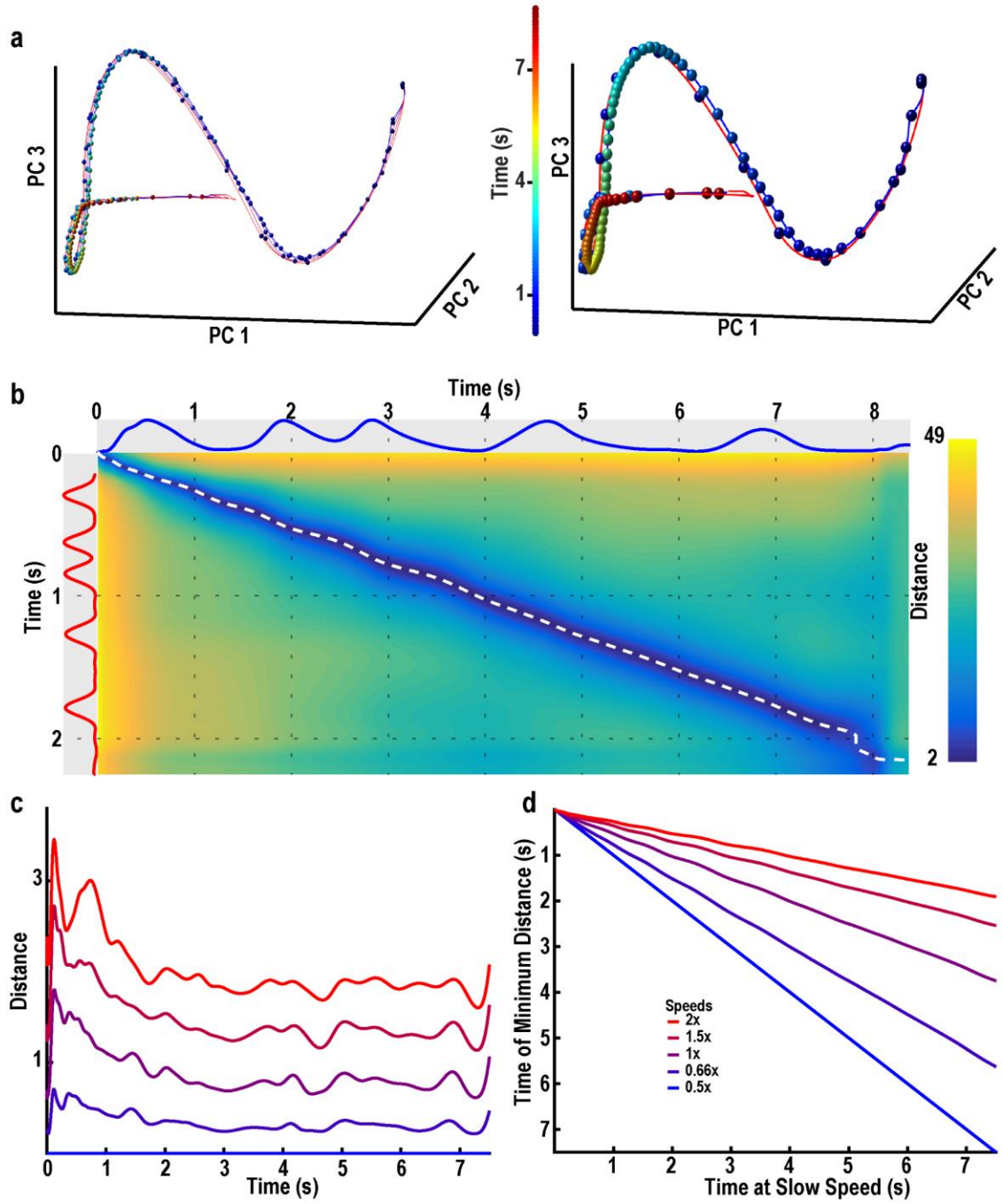


Figure 3.7. Temporal scaling relies on parallel neural trajectories at different speeds.

A) Trajectory of RNN activity at five speeds projected onto the first three principal components. Right: same data, but only the slowest (blue line) and fastest (red) speeds are plotted to highlight the difference in speed of the two trajectories. Colored spheres indicate absolute time in each trajectory (100 ms between spheres), and reveal the differences in the speeds of the trajectories in neural phase space. **B)** Euclidean distance matrix between the fast and slow trajectories in neural space at each point in time along each trajectory

(network size: N=1800). Blue and red traces along the axes show the output. White dotted line traces the minimum distance between the two trajectories, which never reaches zero. **C**) The minimum distance along the slowest trajectory from each other speed. **D**) The relative timing at which the minimum occurs in each trajectory. For example, at 4 s in the slowest speed (x-axis) the trajectory is closest to the 2x speed at 1 s.

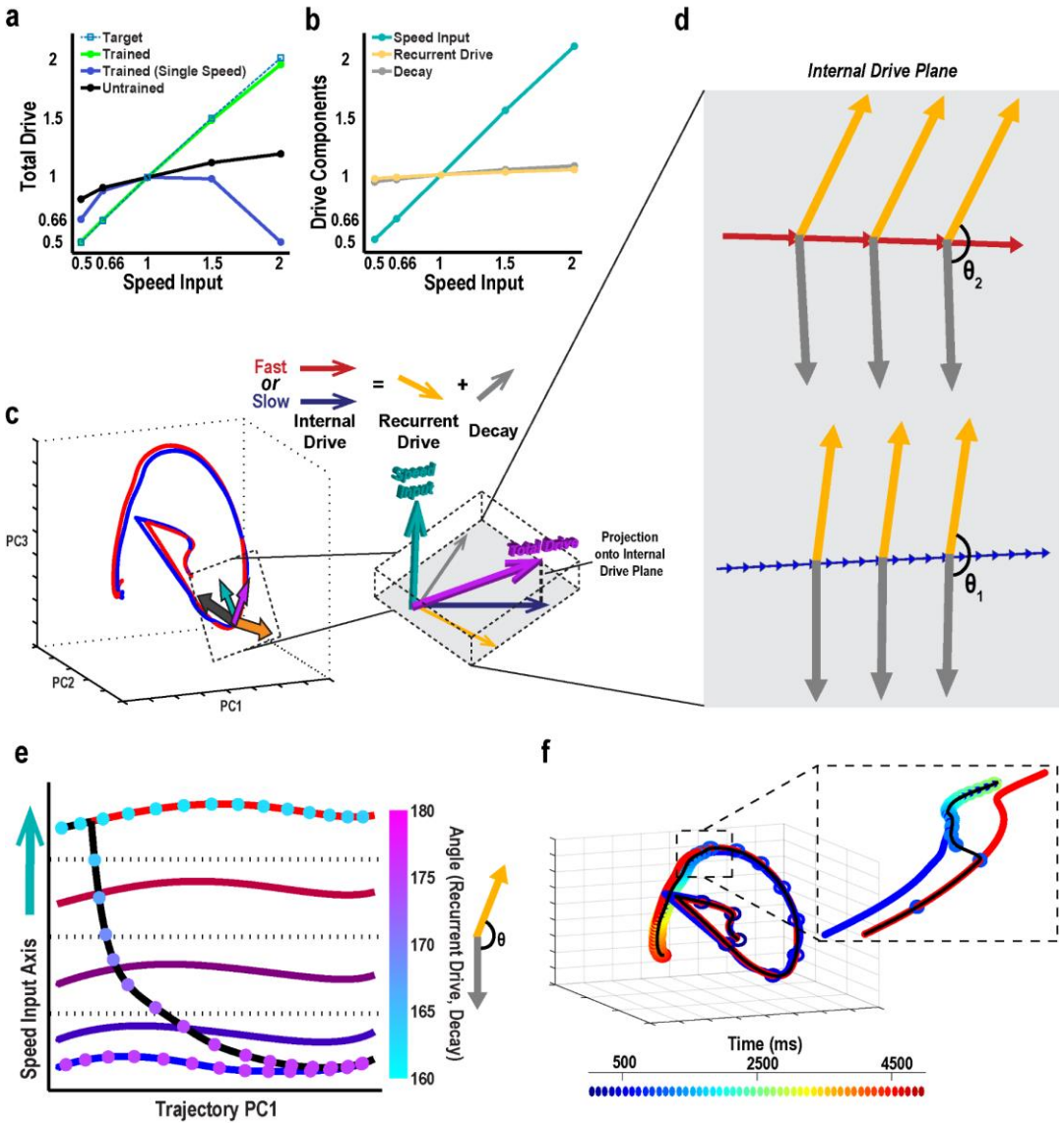
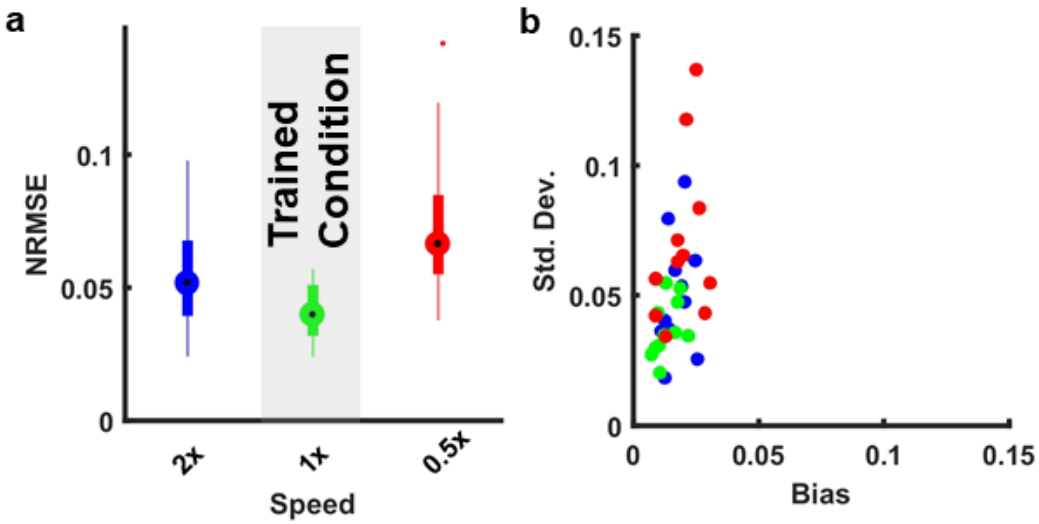


Figure 3.8. Mechanisms of temporal scaling in the RNN.

A) Magnitude of the instantaneous change in activity (trajectory speed) of the recurrent network (total drive) scales linearly with speed input value in networks trained at two speeds (green), but not in networks trained at one speed or untrained networks. Total drive is normalized to the 1x speed. **B)** Decomposing network drive into its three components (recurrent, decay, and input) revealed that the recurrent and decay components do not individually scale with speed input, thus neither of them in isolation can account for temporal scaling. **C)** To examine the relationship between the recurrent and decay components separate from the input drive, we projected them onto the internal drive plane, a subspace orthogonal to the speed input (Online Methods). **D)** This projection revealed that at faster speeds the angle between the recurrent and decay components decreases, creating a second-order effect that drives the network activity along the

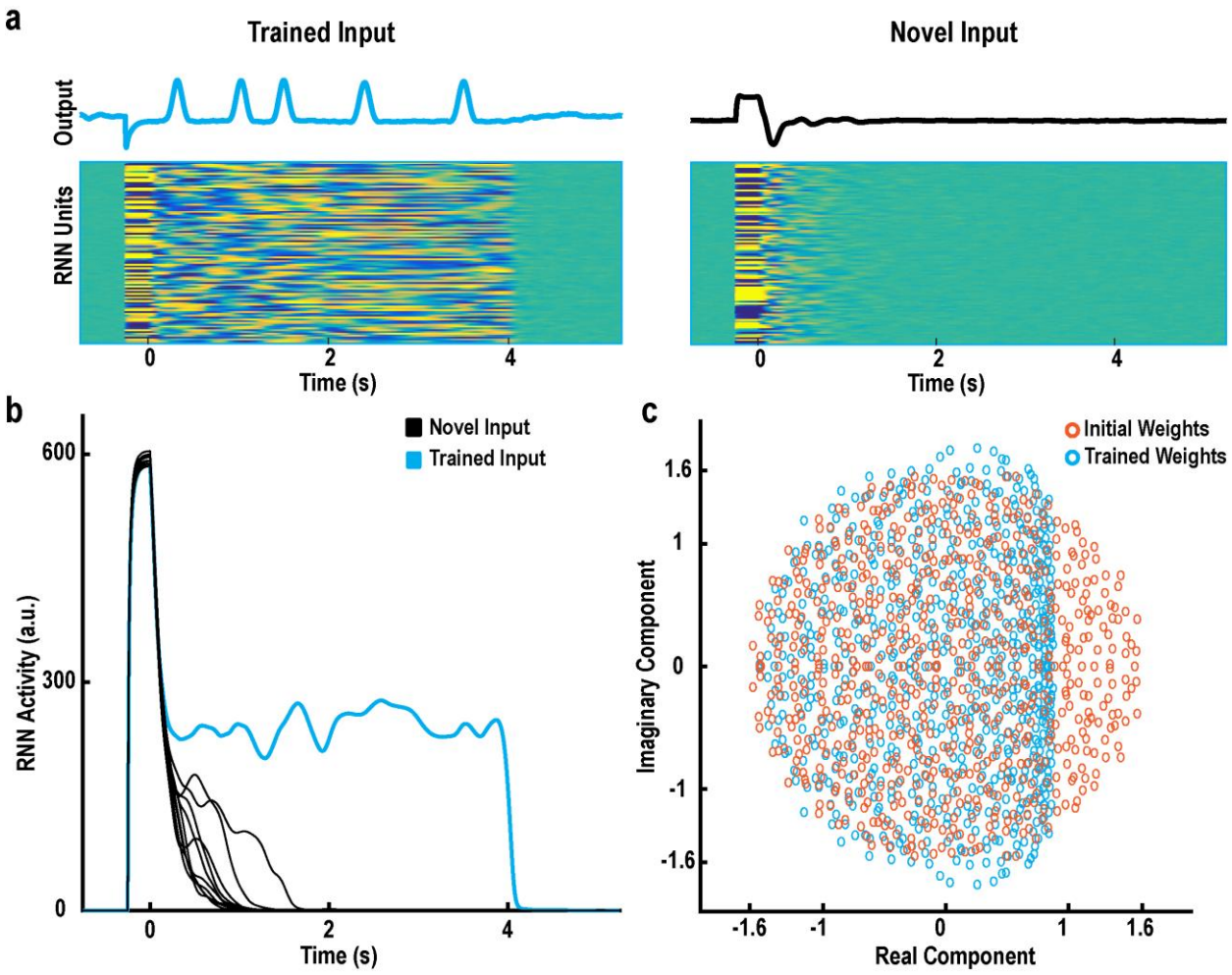
trajectory more quickly. **E)** Network activity projected onto the input axis and the first principal component of network activity (the dimension which accounts for the largest amount of variance). The colored markers indicate the angle between the recurrent and decay components. The position along the input axis does not change as a function of time, indicating that speed is encoded by the position along the input axis. When the speed input level is abruptly decreased partway through the trajectory (black line), the network switches from fast to slow speeds via an increase in the angle between the recurrent and decay components. **F)** Neural trajectories in the first three principal components during a mid-trajectory change in speeds. As the dynamics transition from fast to slow (inset), the trajectory (black line) moves along a hyperplane defined by the parallel trajectories shown in Fig. 3.6.

3.6: Supplementary Figures



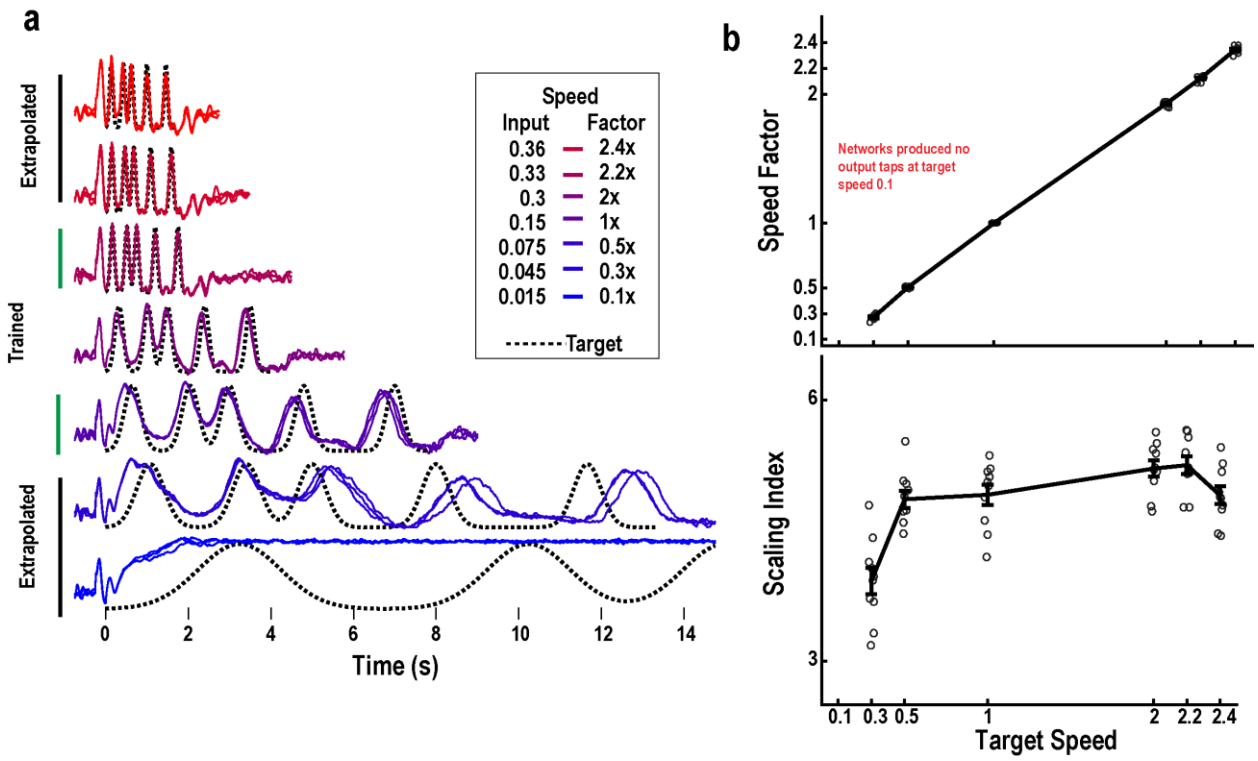
Supplementary Figure 3.1. Performance on the Morse code task based on normalized RMSE.

A) Normalized RMSE was calculated in relation to the target pattern and based on the normalized response times (normalized to the mean response time of the last response at each condition). As shown in **Fig. 3.1** subjects produced a range of different speeds—thus the actual speeds within a given group can vary significantly. There was a trend toward a significant difference between NRMSE between 2x and the trained 1x groups (t_{10}=2, p=0.073, paired t-test) and a significant difference between 0.5x and 1x (t_{10}=2, p=0.004, paired t-test). Note that the 2x and 0.5x labels reflect requested scaling factors, not actual produced speeds, thus this data must be interpreted with caution. **B)** Scatterplot of the mean SD (across responses) and mean biases (difference between the target response time and mean response time across all responses) for each subject on each condition.



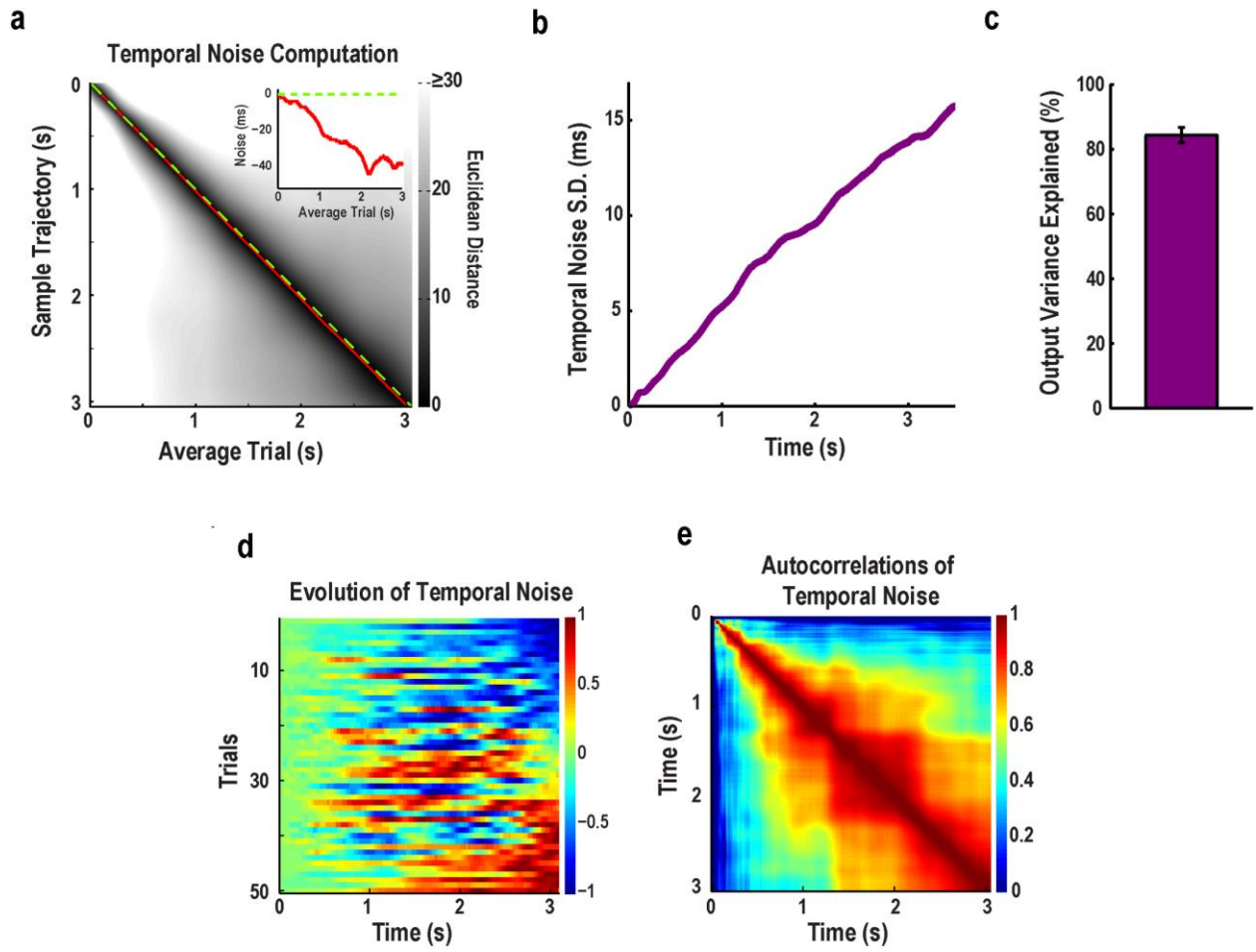
Supplementary Figure 3.2. Gated attractor networks suppress untrained activity.

A) RNNs were trained to suppress activity except in response to trained cue inputs. Left: Example activity after training in response to the trained cue input. Right: Response to an untrained cue. **B)** RNN activity (the norm of the firing rate (r) in response to the trained cue and ten untrained cue inputs. **C**) The eigenvalues of the recurrent weights before and after training. After training, the real components of the eigenvalues are less than one, meaning gated attractor networks are not spontaneously active.



Supplementary Figure 3.3. RNNs' temporal scaling degrades outside of the trained speed range.

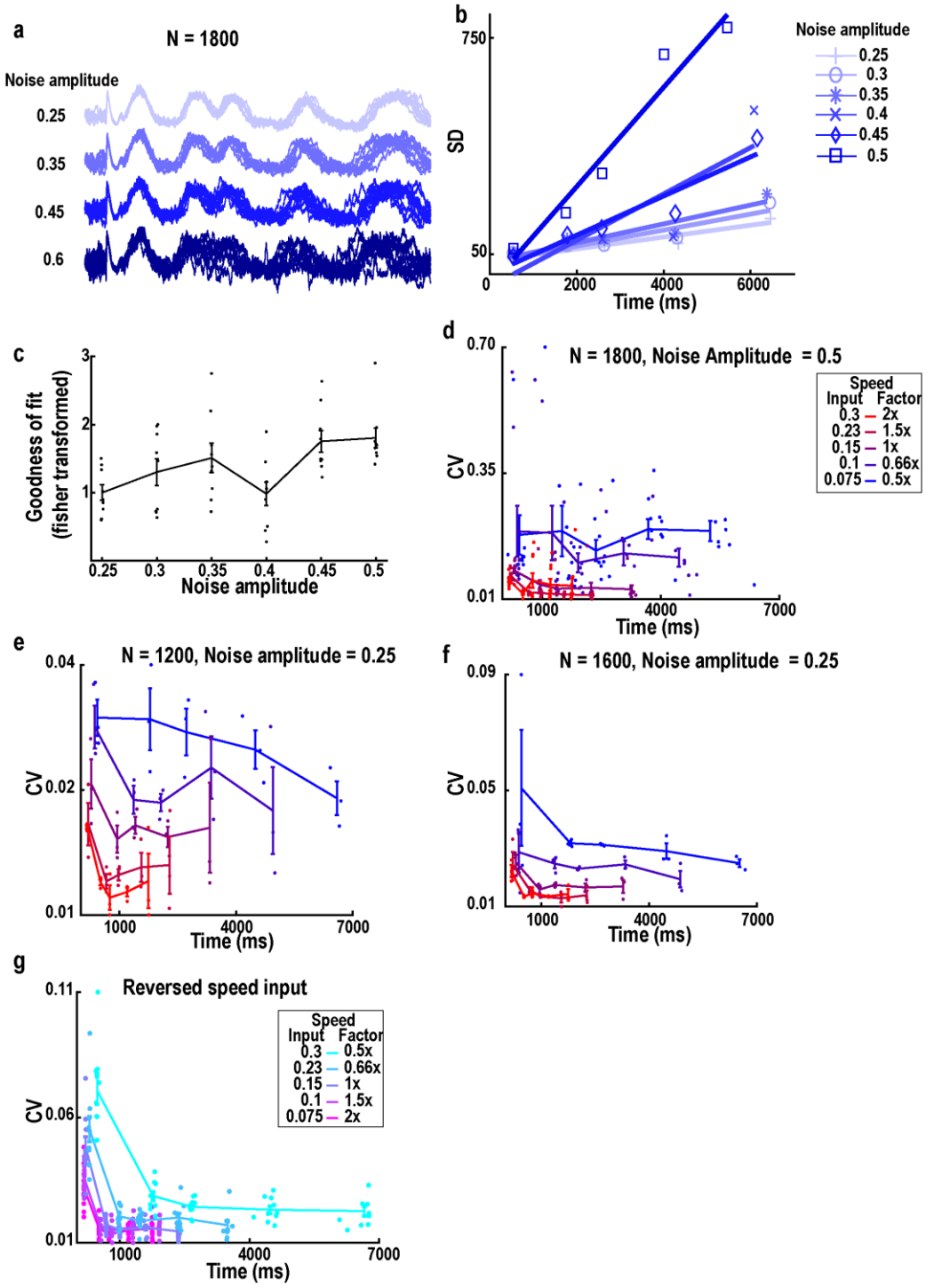
A) Output traces at interpolated and extrapolated speeds (outside of the trained speed range). **B)** Top: speed factor (top) and scaling index (bottom) of ten networks calculated from twenty trials at the speeds shown in **a**. Generalization degrades at slower speeds.



Supplementary Figure 3.4. RNNs produce long-lasting temporal noise correlations.

A) Euclidean distance matrix between the trial-averaged trajectory of a trained RNN and a single trial at speed 1x. The times at which the two trajectories are closest is represented by the red line. Matched time points between the trajectories is represented by the identity line (green dashed line). Over time the sample trajectory runs further ahead of the average trajectory (temporal noise), as evidence by the red line being below the green line. Inset: Temporal noise (red line) in the sample trajectory relative to the average trajectory. At the end of the average trajectory, the sample trajectory is ~40 ms ahead. **B)** The linear relationship of the SD of temporal noise in the trajectories and absolute time underlies Weber's law observed in the output unit. The SD of temporal noise is calculated over 50 trials, averaged across 10 networks. **C)** Output timing variability explained by temporal noise in the RNN (normalized mean squared error calculated between temporal noise in the trajectories and the output), averaged across 10 networks (error bars indicate SEM). **D)** Normalized temporal noise across 50 trials, sorted according to the noise at the end of the trajectory in the example network. **E)** Autocorrelation of temporal noise in one network. Each element in the matrix represents the correlation (across trials) of temporal noise at the corresponding pair

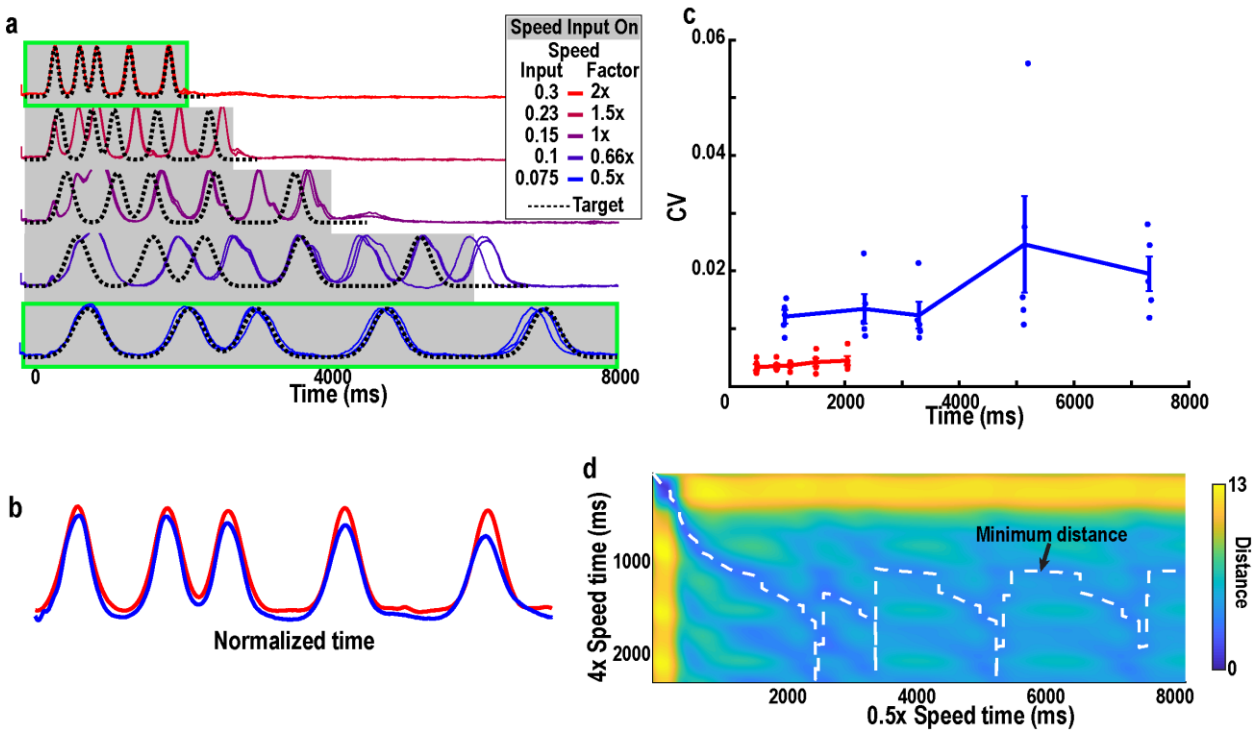
of time points (i.e., pair of columns in **D**). Deviations at early time points predict later deviations. Networks (n=10) trained and tested at noise amplitude of 0.25.



Supplementary Figure 3.5. Weber-speed effect in RNNs is observed across noise levels and network size.

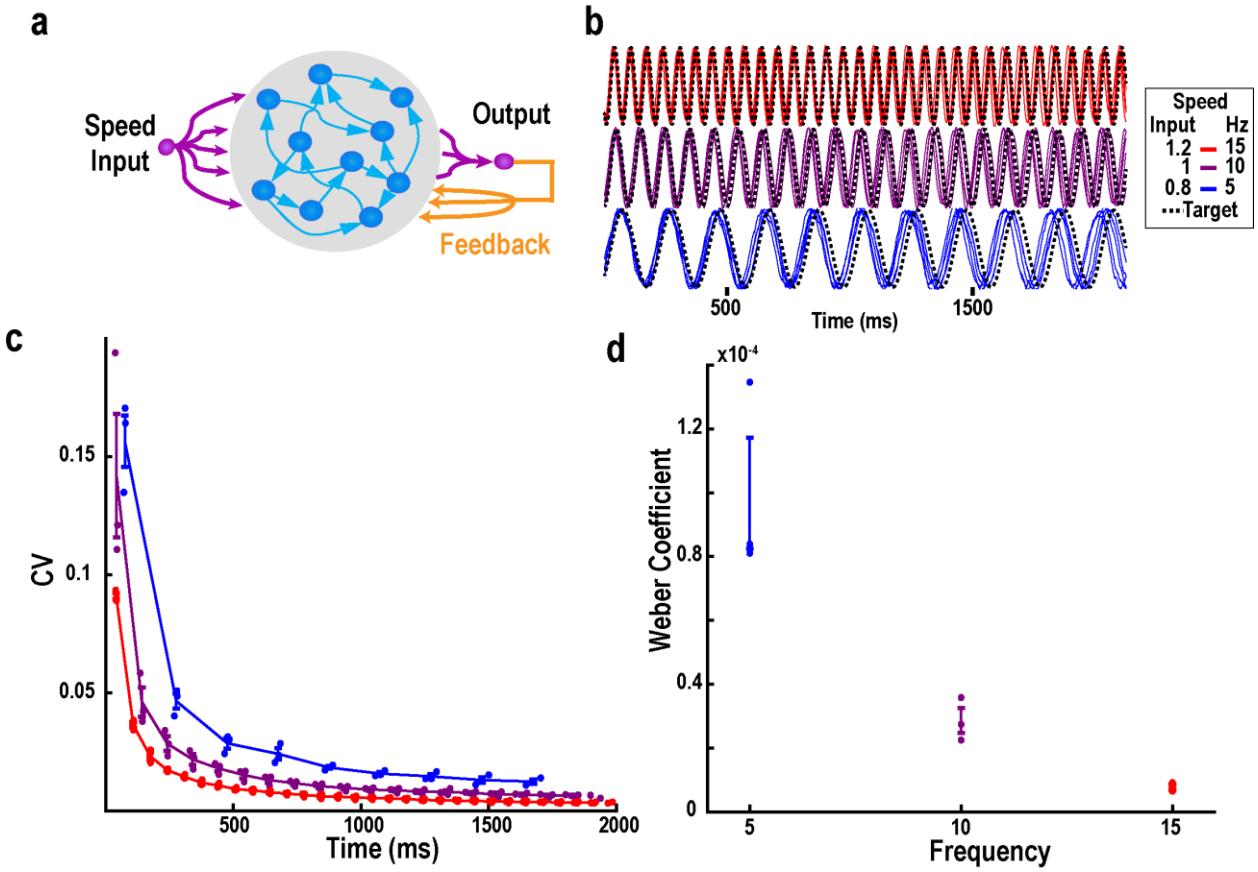
A) Ten example output traces from a single network at increasing levels of noise. The output pattern is discernable over a range of noise amplitudes (<0.5). **B)** SD vs t of hit times across noise amplitudes for the example network. Solid lines show the linear fits and symbols show the measured statistics. **C)** The Weber's

law (linearity of σ^2 vs t) is maintained within this stable range. Beyond a noise amplitude of 0.5 the output becomes too unsteady to reliably measure the hit time. **D-F)** The Weber- speed effect persists in high noise (panel d), and lower network size (n=3) (panels e, f). **G)** The Weber-speed effect is also observed when RNNs are trained with an inverse speed input amplitude vs. speed relationship.



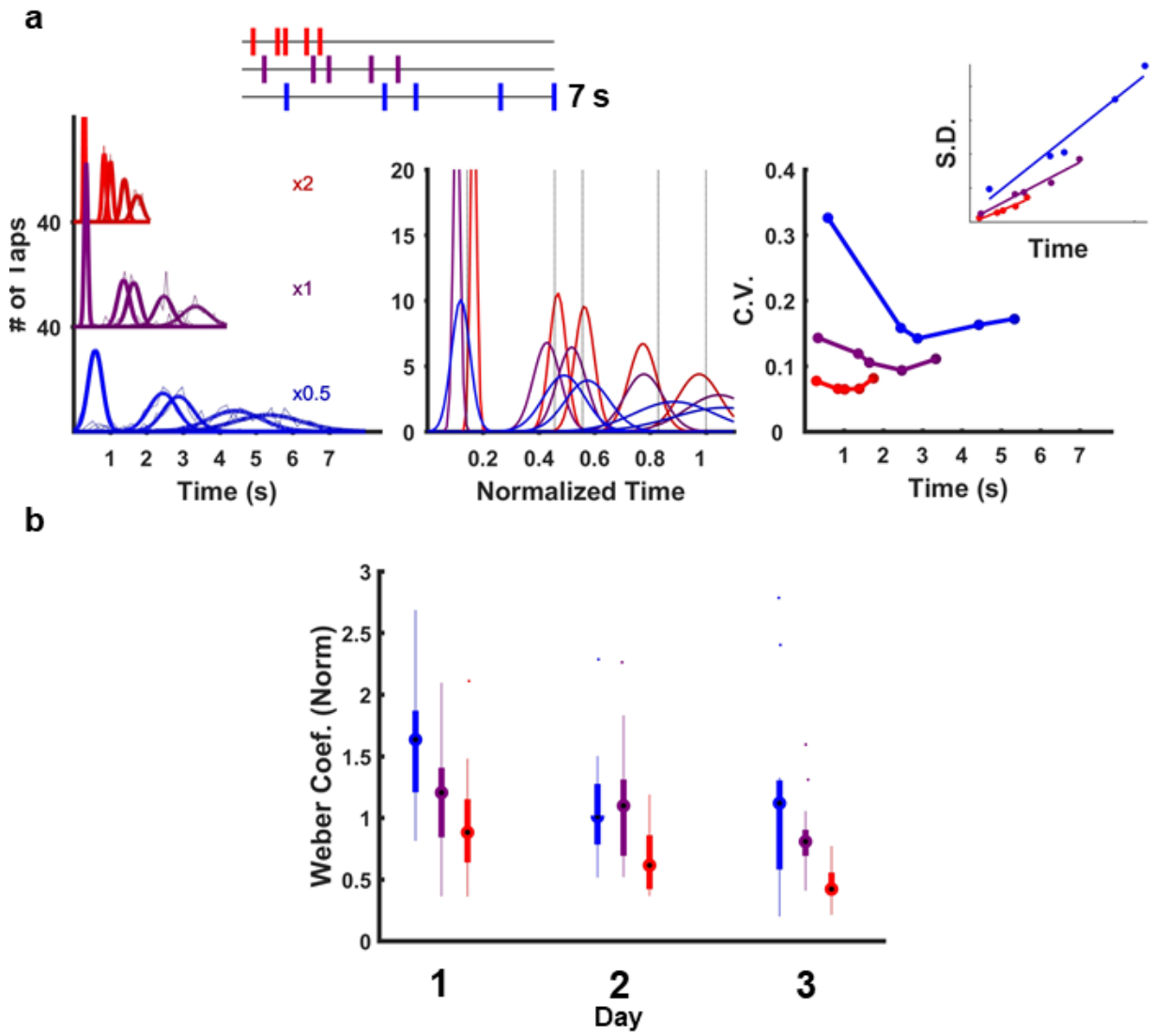
Supplementary Figure 3.6. RNN training based on output error does not result in robust temporal scaling.

Four control networks were trained at 0.5x and 2x speeds using the Hessian-free backprop algorithm, using the same speed-input relationship as the innate learning studies. **A)** Using the same training parameters used in the main text (i.e., training on only 2 speeds, and with the same noise levels) Hessian-free backprop did not result in temporal scaling to novel speeds (note the difference in the number of peaks). Three traces from an example network are shown at interpolated and trained speeds. **B)** Output at the trained speeds in normalized time. **C)** The Weber-speed effect is still observed at the two trained speeds, despite the absence of generalization to interpolated speeds. **D)** Pairwise Euclidean distance between the network trajectories at the trained speeds. The trajectories follow different paths (i.e. they are not parallel) as shown by the jagged trace of the minimum distance between the two speeds (white dashed line).



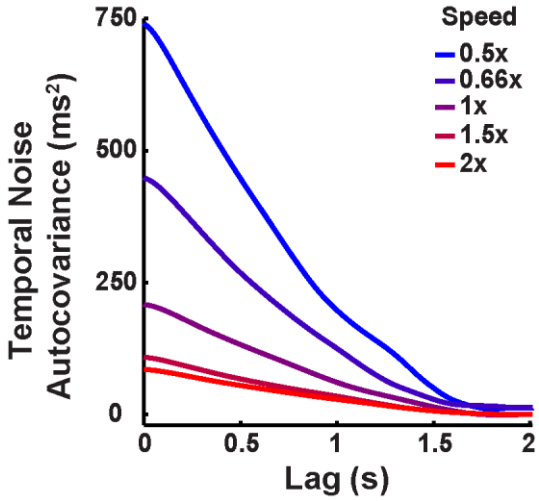
Supplementary Figure 3.7. Echo state network the produces sinusoids of different frequencies exhibits the Weber-speed effect.

A) Network schematic. The recurrent network receives a speed input, and generates a single output as in the other architectures. However, the output unit now provides feedback onto the recurrent units, and only the weights onto the output unit are trained to produce a sinusoid. **B)** Output targets and 5 example traces for each frequency. Networks were trained to produce sinusoidal output with frequencies 5, 10, and 15 Hz and input amplitudes 0.8, 1, and 1.2 by modifying the recurrent-to-output synapses (Online Methods). **C)** The coefficient of variation (Weber factor) and **D)** Weber Coefficient demonstrate reduced variability at higher frequencies (the Weber-speed effect).



Supplementary Figure 3.8. Weber-speed effect on an aperiodic task composed of three speeds in which subjects were trained across three days on all speeds.

A) Left: Histogram (dashed lines) and Gaussian fits (solid lines) of the taps at all three speeds (0.5x, 1x, and 2x) from a single subject. Middle: the fits shown with time normalized to the mean of the last tap (vertical lines represent target times). Right: CV of each tap at each speed, with the linear fig of the SD versus mean time plotted in the inset. **B)** Whisker plots of the CV of all subjects (n=15) for all three speeds across the three days of training.



Supplementary Figure 3.1. Slower speeds have longer lasting noise correlations.

Autocovariance of temporal noise at different speeds across 50 trials, averaged across 10 networks; note that the slowest speed has an elevated covariance even at a time lag of up to 1 s. Networks trained and tested at noise amplitude of 0.25.

3.7: Materials and Methods

3.7.A: Temporal Scaling of Motor Patterns in Humans

Human psychophysics experiments were performed using a temporal pattern reproduction task (Laje et al., 2011). During the experiments, the subjects sat in front of a computer monitor with a keyboard in a quiet room. On each trial, subjects heard a temporal pattern and then reproduced this pattern by pressing one key on a Cedrus Response Pad™. The target stimulus consisted of a series of brief tones (800 Hz). After the subjects reproduced the pattern, a visual representation of the target and of the subject's response appeared on the screen along with a score based on the correlation between the target and the reproduced pattern. Stimulus presentation and response acquisition were controlled by a personal computer using custom Matlab code and the Psychophysics Toolbox (Brainard, 1997). All experiments were run in accordance with the University of California Human Subjects Guidelines.

To test whether temporal scaling is an innate property of motor behavior, subjects we trained to produce the Morse code spelling of "time" at 10 words per minute (**Fig. 3.1**). Training occurred over four days, with 15 blocks of 15 trials per day. On the fifth day, subjects were asked to produce the trained pattern at 0.5x, 1.0x and 2x the speed under freeform conditions: subjects first completed 15 trials of the trained pattern, and then were asked to produce the same pattern at the same speed (1x), twice as fast (2x), and twice as slow (0.5x) in the absence of any additionally auditory stimuli. Subjects performed five blocks with five trials per speed in a random order for a total of fifteen trials per block. The subjects were 10 undergraduate students from the UCLA community who were between the ages of 18 and 21. Subjects were paid for their participation.

To test the Weber-speed prediction of the RNN model (**Fig. 3.5**), subjects performed a temporal reproduction task, wherein they heard a pattern of six tones (each lasting 25 ms) and were asked to reproduce the timing of the onset of each tone with a self-initiated start time

(representing the first tone). For the 1x speed the six tones were presented at 0, 325, 1025, 1500, 2400, and 3500 ms. This pattern was then scaled to five logarithmically distributed speeds: 0.5x, 0.6x, 1x, 1.5x, and 2.0x. Subjects completed four blocks of fifteen trials per speed in a random order. A pseudo-randomly chosen subset of the subjects were trained to produce the 0.5x and 2x speeds over eight additional days, consisting of ten blocks of fifteen trials per speed. The subjects for this study were 25 undergraduate students from the UCLA community between the ages of 18 and 21 and paid for their participation.

In the periodic/subdivision task (**Fig. 3.6)** subjects (n=11) were trained on a pattern reproduction tasks in which the four targets consisted of patterns lasting 2.4 seconds divided into subintervals of 300, 400, 600, or 800 ms. Subjects were trained for 5 days and performed four blocks of twelve trials on each condition per day. For the aperiodic timing task in **Supplementary Fig. 3.8**, subjects (n=15) reproduced a pattern of six tones presented at 0, 500, 1600, 1950, 2900, and 3500 ms. This pattern was then scaled to speeds 0.5x and 2.0x. Subjects were trained for three days with six blocks of fifteen trials per speed presented in a random order.

3.7.B: RNN Network Equations

The units of the RNNs used here were based on a standard firing rate model defined by the equations (Jaeger and Haas, 2004; Sompolinsky et al., 1988):

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} W_{ij}^{Rec} r_j + \sum_{j=1}^{I} W_{ij}^{In} y_j(t) + \varphi_i(t) \quad (1)$$

$$z = \sum_{j=1}^{N} W_j^{Out} r_j \quad (2)$$

where $r_i = \tanh(x_i)$ represents the output, or firing rate, of recurrent unit $i = [1, \ldots, N]$. The variable y represents the activity level of the input units, and z is the output. $N = 1800$ is the number of units in the recurrent network, and $\tau = 50\ ms$ is the unit time constant. The connectivity of the recurrent network was determined by the sparse $N\ x\ N$ matrix W^{Rec}, which initially had nonzero weights drawn from a normal distribution with zero mean and SD $g/\sqrt{Np_c}$. The variable $p_c = 0.2$ determined the probability of connections between units in the recurrent network which were drawn uniformly at random, and $g = 1.6$ represents the "gain" of the recurrent network (Rajan et al., 2010; Sompolinsky et al., 1988). The $N\ x\ I$ input weight matrix W^{In} was drawn from a normal distribution with zero mean and unit variance. For all figures, $I = 2$, except **Supplementary Fig. 3.2**, where additional input units were added to test the specificity of the network response to untrained cue inputs. One input served as cue to start a trial and its activity was set to zero except during the time window $-250 \leq t \leq 0$, when its activity was equal to 5.0. The second input unit served as a speed input and was set to a constant level during the time window $-250 \leq t \leq T$, where T represents the duration of the trial. Each unit in the recurrent network was injected with noise current $\varphi_i(t)$, drawn independently from a normal distribution with zero mean and SD 0.05, except for the Weber experiments where the SD was 0.25. The recurrent units were connected to the output unit z through the $Nx1$ vector W^{Out}, which was initially drawn from a normal distribution with zero mean and SD $1/\sqrt{N}$.

3.7.C: Recurrent Learning Rule

The networks in this study were trained using the Innate Learning Rule, which trains an initially chaotic recurrent network to autonomously yet reliably produce an arbitrary activity pattern in the presence of noise (Laje and Buonomano, 2013). It is based on the Recursive Least Squares (RLS) update rule (Haykin, 2002; Sussillo and Abbott, 2009). The recurrent weights onto unit i were updated every $\Delta t = 5\ ms$ as dictated by

$$W_{ij}^{Rec}(t) = W_{ij}^{Rec}(t - \Delta t) - e_i(t) \sum_{k \in \boldsymbol{B}(i)} P_{jk}^{i}(t) r_k(t) \tag{3}$$

where $\boldsymbol{B}(i)$ is the subset of recurrent units presynaptic to unit i. The error e_i of unit i is given by

$$e_i(t) = r_i(t) - R_i(t) \tag{4}$$

where r_i is the firing rate of unit i before the weight update, and R is the target activity of that recurrent unit. The square matrix $\boldsymbol{P}^i$ estimates the inverse correlation of the recurrent inputs onto unit i, updated by

$$\boldsymbol{P}^i(t) = \boldsymbol{P}^i(t - \Delta t) - \frac{\boldsymbol{P}^i(t - \Delta t) r(t) r'(t) \boldsymbol{P}^i(t - \Delta t)}{1 + r'(t) \boldsymbol{P}^i(t - \Delta t) r(t)} \tag{5}$$

3.7.D: Training Procedure

To train a network to perform the temporal scaling task, we first generated a target pattern of recurrent activity by stimulating the network with the cue input and capturing the dynamics generated according to equation (1) over 2000 ms in the presence of speed input level $y_{SI} = 0.3$ and zero noise (similar results are obtained if the target pattern is harvested in the presence of noise). We then produced a temporally dilated version of this target by linearly interpolating by a factor of four to produce a second scaled version of the target with a duration of 8000 ms. For **Fig. 3.3** and later, the recurrent network was then trained with random initial conditions and noise amplitude 0.05 according to the algorithm described in equations (3-5). The fast target (2x speed) was trained over the window $t \in [0,2000]$ with $y_{SI} = 0.3$ and the slow target (0.5x speed) over the window $t \in [0,8000]$ with $y_{SI} = 0.075$. Ten differently seeded networks were each trained for a total of 60 trials alternating between fast and slow targets. A similar procedure was used to train

networks at a single speed (**Fig. 3.2**). The initial target was captured with a duration of 4000 ms and $y_{SI} = 0.15$ and zero noise. The same initial networks used in the temporal scaling task were trained at this speed for 30 trials. To emulate a rest state all networks were trained to maintain zero r (firing rate) for 30 s following the end of each trained recurrent target. We dubbed networks trained in this manner "gated" attractor networks because they only entered the long-lasting dynamic attractor in response to a specific cued input (**Supplementary Fig. 3.2**).

After recurrent training was complete, the output unit was trained, only at the fastest trained speed, to produce a target function of a series of 5 Gaussian peaks ("taps") centered at 163, 513, 750, 1200, and 1750 ms (0.5x speed). The training algorithm for the output weights was similar to that used for recurrent training, as described previously (Laje and Buonomano, 2013).

3.7.E: Analysis of Temporal Scaling

To assess the ability of a network to generalize its activity to novel speeds, i.e. temporally scale, we tested the response of networks to a range of speed input levels after training was completed (weights were no longer modified). The network was set to a random initial state at $t = -750$ and given the trained cue input during $t \in [-250,0]$. The test speed input was delivered starting at $t = -250$ for a duration lasting 20% longer than when a perfectly timed last tap would occur. The timing of these peaks was used to measure the accuracy of the network's temporal scaling using a *speed factor* and *scaling index*. The speed factor was a coarse measure of temporal scaling calculated by dividing the final peak time of twenty test trials at each speed to the mean peak time at the 1x speed, and taking the mean across trials. The quality of temporal scaling (the scaling index) was calculated by taking the fisher-transformed correlation of the mean timing of the response time for each speed with the mean pattern of the 1x speed.

3.7.F: Weber Analysis

The Weber analysis was performed according to Weber's Generalized Law (Ivry and Hazeltine, 1995b; Merchant et al., 2008), which defines a relationship between mean and variance of perceived time as:

$$\sigma^2 = kT^2 + \sigma^2_{independent}$$

where σ^2 is the variance and T represent the mean of a given response time. We define the slope k as the Weber coefficient, and $\sigma^2_{independent}$ is the time independent source of variance—sometimes referred to as motor variance. We measured k independently at each speed, by performing a linear fit on the measured mean and variance of the five response times at that speed (peak times for the RNN in **Fig. 3.4** and button presses for psychophysics experiments). Note that for visualization purposes, in some plots we show the linear fit of the standard deviation by the mean time.

To test the subdivision hypothesis (**Fig. 3.6**), we additionally fit each subject's responses according to:

$$\sigma^2_{Subd}(T) = k\sum t_{\mathrm{i}}^2 + \sigma^2_{independent}$$

where t_i is the average interval between response i and the preceding response. We then calculated the goodness of fit for both the subdivision and continuous (speed-effect) fits by finding the Fisher-transformed coefficient of determination (r^2) between the predicted variance at each tap time and the measured variance.

3.7.G: RNN Trajectory Analysis

To analyze the position of the trajectories in relationship to one another, we tested the networks at each speed without noise. We then concatenated the active period of the trajectory at each speed, defined as the window between cue input offset and speed input offset, and performed principal components analysis (PCA) on these concatenated trajectories. We used the PCA coefficients to transform the individual trajectory at each speed for visualization in **Fig. 3.7A**. To measure the relationship between trajectories, we returned to full ($N = 1800$) neural phase space and measured the Euclidean distance between the slowest (0.5x speed) trajectory and the trajectories at each speed, at all pairs of points in time. This produced one t^{test} x $t^{0.5x}$ distance matrix per speed, as seen in **Fig. 3.7B** for test speed 2x. To confirm that the trajectories did not cross and followed a similar path, for each point on the slowest trajectory we found a corresponding point on the test trajectory that was closest to it. This produced a vector of approximately 8000 distance values (for each millisecond of the slowest trajectory) which we plotted in **Fig. 3.7C** for each of the five tested speeds. The distances were fairly constant for each test speed and never reached zero, indicating that the trajectories did not intersect. We also recorded the points t_{min}^{test} along the test trajectory where these minima occurred, allowing us to assess the relative speed of each trajectory along their entire length. For example, when the slowest trajectory is at its 400 ms mark, if a test trajectory is closest to it at the test trajectory's own 100ms mark, this would indicate that at that moment, the slowest trajectory was moving four times slower than the test trajectory. We plotted t_{min}^{test} for each of the five tested speeds in **Fig. 3.7D**.

3.7.H: Recurrent-Decay-Input Subspace Decomposition

In **Fig. 3.8**, the total drive $\left(\frac{\boldsymbol{dx(t)}}{\boldsymbol{dt}}, \text{Equation } 1\right)$ was decomposed into its three components: 1) synaptic decay $\left(\boldsymbol{DS}(t) = -\frac{1}{\tau}\boldsymbol{x}(t)\right)$; 2) recurrent synaptic drive $\left(\boldsymbol{RS}(t) = \frac{1}{\tau}\boldsymbol{W}^{Rec}\boldsymbol{r}(t)\right)$; and its

external component, the tonic speed input $\left(\boldsymbol{IS}(t) = \frac{1}{\tau}\boldsymbol{W}^{In}\boldsymbol{y}(t)\right)$. The magnitude of each of these components was calculated as the time-averaged L2-norm of the corresponding population vectors. **Fig. 3.8C** illustrates the generation of an orthonormal basis set $\{\boldsymbol{is}, \boldsymbol{ds}, \boldsymbol{rs}\}$ for the total drive at time *t*, which was computed by applying the Gram-Schmidt orthonormalization process as follows:

$$\boldsymbol{is} = \frac{\boldsymbol{IS}(t)}{\|\boldsymbol{IS}(t)\|}$$

$$\boldsymbol{ds} = \frac{\boldsymbol{DS}(t) - (\boldsymbol{DS}(t)'\boldsymbol{is})\boldsymbol{is}}{\|\boldsymbol{DS}(t) - (\boldsymbol{DS}(t)'\boldsymbol{is})\boldsymbol{is}\|}$$

$$\boldsymbol{rs} = \frac{\boldsymbol{RS}(\boldsymbol{t}) - (\boldsymbol{RS}(t)'\boldsymbol{is})\boldsymbol{is} - (\boldsymbol{RS}(t)'\boldsymbol{ds})\boldsymbol{ds}}{\|\boldsymbol{RS}(\boldsymbol{t}) - (\boldsymbol{RS}(t)'\boldsymbol{is})\boldsymbol{is} - (\boldsymbol{RS}(t)'\boldsymbol{ds})\boldsymbol{ds}\|}$$

Here, $\|.\|$ represents the L2-norm and the apostrophe represents the vector transpose operation. Collectively, these unit orthonormal vectors fully describe the total drive and its components at *t*, and therefore, form a basis set for these vectors. The plane described by the basis set $\{\boldsymbol{ds}, \boldsymbol{rs}\}$ is denoted the internal drive plane, with $\boldsymbol{DS}(t)$ projected onto this plane in grey, and $\boldsymbol{RS}(t)$ in yellow. In **Fig. 3.8D**, we visualize the relationship between these vector projections over a short sequence of time steps along the slow and fast trajectories, on a common internal drive plane. For this, we constructed a common orthonormal set by applying the Gram-Schmidt process to the sequence-averaged component vectors. While doing so precludes the orthonormal set from forming a basis for the vector sequences, restricting the length of these sequences to a small fraction of the network unit time constant (τ), renders the information loss negligible. Finally, in **Fig. 3.8E**, to show that the trajectories consistently encode their desired speeds, we plot the projection of the state variable $(\boldsymbol{x}(t))$ onto $\boldsymbol{is}$, against its projection onto the first principal

component in the subspace orthogonal to $\boldsymbol{is}$. That is, the x-axis represents the first principal component of $(\boldsymbol{x}(t) - (\boldsymbol{x}(t)'\boldsymbol{is})\boldsymbol{is})$.

3.7.I: Control Networks

We trained five control RNNs using Hessian-free optimization (Martens and Sutskever, 2011; Sussillo and Barak, 2013) to produce the same aperiodic output pattern as RNNs trained using the innate learning rule, at the 0.5x and 2x speeds. These networks were defined by:

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} W_{ij}^{Rec} r_j + \sum_{j=1}^{I} W_{ij}^{In} y_j(t) + b_i^x + \varphi_i(t)$$

$$z = \sum_{j=1}^{N} W_{ij}^{Out} r_j + b^z$$

where network size is $N = 300$ and $r_i = \tanh(x_i)$ is the firing rate of recurrent unit $i = [1, \dots, N]$. As in the innate learning RNNs, there was a cue and speed input, and Gaussian noise $\varphi_i(t)$ drawn from a normal distribution with SD 0.25. The Hessian-free learning algorithm adjusts the recurrent weights W^{Rec} by backpropagating the error in the output unit during a trial across W^{Rec}, defined as $e_i(t) = z(t) - Z(t)$, where Z is target output activity. Training resulted in the modification of bias terms b^x and b^z, and the weight matrices W^{In}, W^{Rec}, and W^{Out}. In this study, W^{Rec} was fully connected, unlike the sparsely connected RNNs used elsewhere. Other parameters were the same as in the innate learning studies. The code used to train these networks was based on Dr. David Sussillo's Hessian-free optimization implementation in MATLAB available at: https://github.com/sussillo/hfopt-matlab .

We also trained three Echo State Networks (Jaeger and Haas, 2004; Sussillo and Abbott, 2009) (ESNs) to produce a sinusoidal outputs (ESN's are not well-suited to produce long aperiodic patterns) at three different frequencies: 5, 10, and 15 Hz. ESNs have a similar architecture, except

there is feedback from the output unit to the recurrent units. These networks were governed by the equations:

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} W_{ij}^{Rec} r_j + W_i^{In} y(t) + W_i^{FB} z(t) + \varphi_i(t) \quad (1)$$

$$z = \sum_{j=1}^{N} W_{ij}^{Out} r_j \quad (2)$$

which are the same as those of the innate learning RNNs, but with the additional feedback term W_i^{FB} defining the weight of the feedback from output z onto recurrent unit x_i. The networks size was set to $N = 300$, and W^{FB} was drawn from a uniform distribution on the open interval $(-1,1)$ and delivered feedback to each unit in the recurrent network. As before, W^{In} and W^{Out} were drawn from a normal distribution with zero mean and unit variance for W^{In} and SD $1/\sqrt{N}$ for W^{Out}. Recurrent weights W^{Rec} were drawn from a normal distribution with zero mean and SD $g/\sqrt{Np_c}$, with gain $g = 1.2$ and connection probability $p_c = 0.2$. These networks were trained by modifying the weights onto the outputs units to match the output target Z based on the error term $e_i(t) = z(t) - Z(t)$ (using the FORCE algorithm (Sussillo and Abbott, 2009)). During training and testing, the networks received a single input I of amplitudes 1.2, 1, or 0.8 which determined the target output frequency, with higher amplitudes corresponding to higher frequency.

Data and code availability

Data and code used to generate the main simulation in this manuscript will be made available upon request, or code can be downloaded from: https://github.com/nhardy01/RNN

3.8: References Cited

Ahrens, M.B., and Sahani, M. (2008). Inferring elapsed time from stochastic neural processes. Paper presented at: NIPS 20 (Cambridge: MIT Press).

Bakhurin, K.I., Goudar, V., Shobe, J.L., Claar, L.D., Buonomano, D.V., and Masmanidis, S.C. (2017). Differential Encoding of Time by Prefrontal and Striatal Network Dynamics. The Journal of Neuroscience *37*, 854-870.

Balcı, F., and Simen, P. (2016). A decision model of timing. Current Opinion in Behavioral Sciences *8*, 94-101.

Brainard, D.H. (1997). The Psychophysics Toolbox. Spatial vision *10*, 433-436.

Bueti, D., and Buonomano, D.V. (2014). Temporal Perceptual Learning. Timing and Time Perception *2*, 261-289.

Bueti, D., Lasaponara, S., Cercignani, M., and Macaluso, E. (2012). Learning about Time: Plastic Changes and Interindividual Brain Differences. Neuron *75*, 725-737.

Buonomano, D.V., and Maass, W. (2009). State-dependent Computations: Spatiotemporal Processing in Cortical Networks. Nat Rev Neurosci *10*, 113-125.

Carnevale, F., de Lafuente, V., Romo, R., Barak, O., and Parga, N. (2015). Dynamic Control of Response Criterion in Premotor Cortex during Perceptual Detection under Temporal Uncertainty. Neuron *86*, 1067-1077.

Church, R.M., Meck, W.H., and Gibbon, J. (1994). Application of scalar timing theory to individual trials. J Exp Psychol Anim Behav Process *20*, 135-155.

Cicchini, G.M., Arrighi, R., Cecchetti, L., Giusti, M., and Burr, D.C. (2012). Optimal Encoding of Interval Timing in Expert Percussionists. The Journal of Neuroscience *32*, 1056-1060.

Collier, G.L., and Wright, C.E. (1995). Temporal rescaling of sample and complex ratios in rhythmic tapping. J Exp Psychol Hum Percept Perform *21*, 602-627.

Crowe, D.A., Zarco, W., Bartolo, R., and Merchant, H. (2014). Dynamic Representation of the Temporal and Sequential Structure of Rhythmic Movements in the Primate Medial Premotor Cortex. The Journal of Neuroscience *34*, 11972-11983.

Diedrichsen, J., Criscimagna-Hemminger, S.E., and Shadmehr, R. (2007). Dissociating Timing and Coordination as Functions of the Cerebellum. J Neurosci *27*, 6291-6301.

Durstewitz, D., and Deco, G. (2008). Computational significance of transient dynamics in cortical networks. European Journal of Neuroscience *27*, 217-227.

Fetterman, J.G., and Killeen, P.R. (1990). A componential analysis of pacemaker-counter timing systems. J Exp Psychol Hum Percept Perform *16*, 766-780.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. Psychological Review *84*, 279-325.

Goudar, V., and Buonomano, D.V. (2017). Encoding Sensory and Motor Patterns as Time-Invariant Trajectories in Recurrent Neural Networks. arXiv:170100838.

Grondin, S., and Killeen, P.R. (2009). Tracking time with song and count: different Weber functions for musicians and nonmusicians. Atten Percept Psychophys *71*, 1649-1654.

Grondin, S., Ouellet, B., and Roussel, M.-E. (2004). Benefits and Limits of Explicit Counting for Discriminating Temporal Intervals. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale *58*, 1-12.

Hardy, N.F., and Buonomano, D.V. (2016). Neurocomputational models of interval and pattern timing. Current Opinion in Behavioral Sciences *8*, 250-257.

Hass, J., and Durstewitz, D. (2014). Neurocomputational Models of Time Perception. In Neurobiology of Interval Timing, H. Merchant, and V. de Lafuente, eds. (Springer New York), pp. 49-73.

Hass, J., and Durstewitz, D. (2016). Time at the center, or time at the side? Assessing current models of time perception. Current Opinion in Behavioral Sciences *8*, 238-244.

Hass, J., and Herrmann, J.M. (2012). The Neural Representation of Time: An Information-Theoretic Perspective. Neural Computation *24*, 1519-1552.

Haykin, S. (2002). Adaptive Filter Theory.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A *79*, 2554-2558.

Itskov, V., Curto, C., Pastalkova, E., and Buzsáki, G. (2011). Cell Assembly Sequences Arising from Spike Threshold Adaptation Keep Track of Time in the Hippocampus. The Journal of Neuroscience *31*, 2828-2834.

Ivry, R.B., and Hazeltine, R.E. (1995a). Perception and Production of Temporal Intervals across a Range of Durations - Evidence for a Common Timing Mechanism. J Exp Psychol-Hum Percept Perform *21*, 3-18.

Ivry, R.B., and Hazeltine, R.E. (1995b). Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. Journal of Experimental Psychology Human Perception and Performance *21*, 3-18.

Jaeger, H., and Haas, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science *304*, 78-80.

Jazayeri, M., and Shadlen, M.N. (2010). Temporal context calibrates interval timing. Nat Neurosci *13*, 1020-1026.

Jin, D.Z., Fujii, N., and Graybiel, A.M. (2009). Neural representation of time in cortico-basal ganglia circuits. Proc Natl Acad Sci U S A *106*, 19156-19161.

Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, Antoniu L., Dhawale, Ashesh K., Kampff, Adam R., and Ölveczky, Bence P. (2015). Motor Cortex Is Required for Learning but Not for Executing a Motor Skill. Neuron *86*, 800-812.

Laje, R., and Buonomano, D.V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. Nat Neurosci *16*, 925-933.

Laje, R., Cheng, K., and Buonomano, D.V. (2011). Learning of temporal motor patterns: An analysis of continuous vs. reset timing. Frontiers in Integrative Neuroscience *5*.

Lebedev, M.A., O'Doherty, J.E., and Nicolelis, M.A.L. (2008). Decoding of Temporal Intervals From Cortical Ensemble Activity. J Neurophysiol *99*, 166-186.

MacDonald, Christopher J., Lepage, Kyle Q., Eden, Uri T., and Eichenbaum, H. (2011). Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events. Neuron *71*, 737-749.

Martens, J., and Sutskever, I. (2011). Learning recurrent neural networks with Hessian-free optimization. Proc 28th Int Conf Machine Learn.

Medina, J.F., Garcia, K.S., Nores, W.L., Taylor, N.M., and Mauk, M.D. (2000). Timing mechanisms in the cerebellum: testing predictions of a large-scale computer simulation. J Neurosci *20*, 5516-5525.

Meegan, D.V., Aslin, R.N., and Jacobs, R.A. (2000). Motor timing learned without motor training. Nat Neurosci *3*, 860–862.

Mello, G.B.M., Soares, S., and Paton, J.J. (2015). A scalable population code for time in the striatum. Curr Biol *9*, 1113-1122.

Merchant, H., Zarco, W., and Prado, L. (2008). Do We Have a Common Mechanism for Measuring Time in the Hundreds of Millisecond Range? Evidence From Multiple-Interval Timing Tasks. J Neurophysiol *99*, 939-949.

Namboodiri, Vijay Mohan K., Huertas, Marco A., Monk, Kevin J., Shouval, Harel Z., and Hussain Shuler, Marshall G. (2015). Visually Cued Action Timing in the Primary Visual Cortex. Neuron *86*, 319-330.

Osborne, L.C., Bialek, W., and Lisberger, S.G. (2004). Time Course of Information about Motion Direction in Visual Area MT of Macaque Monkeys. Journal of Neuroscience *24*, 3210-3222.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsaki, G. (2008). Internally Generated Cell Assembly Sequences in the Rat Hippocampus. Science *321*, 1322-1327.

Perrett, S.P., Ruiz, B.P., and Mauk, M.D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. J Neurosci *13*, 1708–1718.

Planetta, P.J., and Servos, P. (2008). Somatosensory temporal discrimination learning generalizes to motor interval production. Brain Research *1233*, 51-57.

Rabinovich, M., Huerta, R., and Laurent, G. (2008). Transient Dynamics for Neural Processing. Science *321*, 48-50.

Rajan, K., Abbott, L.F., and Sompolinsky, H. (2010). Stimulus-dependent suppression of chaos in recurrent neural networks. Phys Rev E Stat Nonlin Soft Matter Phys *82*, 011903.

Sompolinsky, H., Crisanti, A., and Sommers, H.J. (1988). Chaos in random neural networks. Phys Rev Lett *61*, 259-262.

Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic Coding for Cognitive Control in Prefrontal Cortex. Neuron *78*, 364-375.

Sussillo, D., and Abbott, L.F. (2009). Generating Coherent Patterns of Activity from Chaotic Neural Networks. Neuron *63*, 544-557.

Sussillo, D., and Barak, O. (2013). Opening the Black Box: Low-Dimensional Dynamics in High-Dimensional Recurrent Neural Networks. Neural Computation *25*, 626-649.

Wang, J., Narain, D., Hosseini, E., and Jazayeri, M. (2017). Flexible control of speed of cortical dynamics. bioRxiv.

Wang, X.J. (2001). Synaptic reverberation underlying mnemonic persistent activity. Trends Neurosci *24*, 455–463.

Chapter 4: State dependent encoding of temporal information in cortical neural networks

4.1: Background and motivation

To process dynamic sensory input, the cortex must be able to encode temporal structure. For example, the words “lady” and “delay” have distinct temporal, but similar spectral, content. However, the mechanisms of embedding temporal information in cortical activity remain unknown. Reservoir computing, a prominent theory of cortical function, provides a potential framework. This theory states that input is processed by recurrent neural networks in a state-dependent manner, such that successive stimuli generate increasingly divergent population activity that encodes the spatiotemporal structure of the input (Buonomano and Maass, 2009; Buonomano and Merzenich, 1995). Several experimental studies have shown evidence for state-dependent processing, including *in vitro* in the hippocampus (Buonomano et al., 1997; Hyde and Strowbridge, 2012), in networks of dissociated cortical neurons (Dranias et al., 2013; Ju et al., 2015) and *in vivo* in cat primary visual cortex (Nikolić et al., 2009). However, these studies often examined input patterns with changing spatial (i.e. different stimuli), as well as temporal, components. Thus, it is unclear how the cortex encodes a pure temporal pattern in which the same stimulus components are presented repeatedly with different temporal relations.

In addition to processing temporal information, the brain must also generate a short-term, persistent representation of sensory information. Neural activity related to short-term working memory about the identity of a stimulus has been be observed in the cortex, often as spiking activity that persists after the stimulus is removed (Funahashi et al., 1989; Fuster and Jervey, 1981). This activity has been modeled in neural networks as locally stable activity (point attractors) that is largely static in time, or as activity that represents the level of a stimulus by shifting its state along a line attractor (Compte et al., 2000; Miller et al., 2003; Wang, 2001). On the other hand,

studies of temporal working memory (i.e. memory for elapsed time) have observed time-varying activity in the cortex that can encode temporal information (Namboodiri et al., 2015; Wang et al., 2018). Models of this type working memory have shown that semi-stable reverberating activity can encode representations of input stimuli for up to several seconds (Laje and Buonomano, 2013; Rajan et al., 2016). However, how naïve neural networks (i.e. networks without prior training) encode temporally dynamic input remains unknown.

Here, I present evidence that the cortex can autonomously process temporal information and maintain a representation of that information in a state-dependent manner. Using organotypic slice cultures expressing both the calcium indicator GCaMP6f and the red-shifted channelrhodopsin Chrimson, I stimulated cortical networks using two purely temporal patterns while recording high-dimensional network activity. Analyzing the population activity revealed that each stimulus pattern generated a distinct spatiotemporal population response. Moreover, I show that these responses encode the temporal structure of the input on a single trial basis at times well beyond the last input. Together, the results of this study support theories of state-dependent processing, demonstrating a mechanism for encoding sensory information in dynamic population activity over extended periods of time.

4.2: Results

4.2.A: Optogenetics and two-photon calcium imaging in vitro

To understand its role in processing sensory input, it is important to examine to what extent the cortex performs computations autonomously (i.e. without input from other brain regions). To do so, I examined the *in vitro* activity of individual neurons evoked by broadly stimulating organotypic cortical slice cultures. I virally expressed the genetically encoded calcium indicator GCaMP6f under the synapsin promoter, inducing expression in a subset of neurons. Chrimson was also expressed under the synapsin promoter, allowing me to stimulate cortical networks by

exposing them to red light. This preparation allowed me to record the activity of 40-120 neurons using 2-Photon calcium imaging while stimulating the slice optogenetically (**Fig. 4.1A**; Materials and Methods). In addition, using a stimulus duration of 25 ms (less than the duration of an imaging frame), made the stimulus artifact easy to remove from the recorded activity (**Fig. 4.1B, inset**). Calcium imaging is a powerful method for measuring neural activity, but the slow decay time of GCaMP6f (> 1 s) limits its temporal resolution (Chen et al., 2013). To mitigate these effects, many studies use event detection algorithms to generate discrete representations of neural activity. I compared two methods for detecting calcium events: 1) constrained sparse nonnegative calcium deconvolution (CaImAn package; Pnevmatikakis et al., 2016), and 2) a rectified high pass filter (Materials and Methods). In practice, both methods provided similar results (**Fig. 4.1B**). However, I found the simple high pass filter to be preferable because it had fewer parameters to adjust and was easy to apply across neurons with different GCaMP6f expression levels. Thus, the data in the remainder of this chapter were analyzed using the event magnitudes extracted using the rectified high-pass filter.

4.2.B: Temporal pattern stimulation

Theories of state-dependent processing posit that sensory input to the cortex is encoded differentially depending on the current state of cortical networks. The state of a network can include, for example, the current firing rate of its neurons and the effects of recent activity on short-term plasticity. Thus, these theories predict that repeatedly presenting the same stimulus (e.g. a musical note or an oriented bar) within a sufficiently short time window will evoke distinct population responses upon each successive stimulation. In contrast, computations that are completely insensitive to the networks' state would generate stationary responses over every presentation of the stimulus (**Fig. 4.2A**). Put another way, if the cortex operates in a state-dependent framework, recent stimulus history will be probabilistically encoded in the current population activity of cortical networks.

To examine state-dependent computations in cortical networks, I stimulated cortical slices using two temporal patterns composed of three 25 ms red light pulses each: Pattern A (short-long), with stimuli at 0, 300, and 1000 ms, and pattern B (long-short), with stimuli at 0, 700, and 1000 ms (**Fig. 4.2B**). Importantly, the stimulus patterns were balanced, i.e. each provided the same amount of total stimulation, and differed only in the timing of the second stimulus. Thus, if state-dependent theories are correct, the temporal context (i.e. the difference between the two stimulus patterns) will be encoded in the network's response to the final stimulus. Each slice was stimulated with sixty repetitions of each pattern, with a 30 s inter-trial interval. To prevent spontaneous network activity from contaminating evoked responses, trials with network activity occurring one second preceding trial start were excluded from all analyses (Materials and Methods).

Neural activity is stochastic, making it difficult to extract stimulus information from single trials. Indeed, the responses of individual neurons in this experiment were highly variable across presentations of both stimuli, despite qualitative differences on average (**Fig. 4.2C**). At the population level, network activity appeared to be differentially shaped by each stimulus pattern, indicating that the trajectories of population activity encoded some amount of information about the stimuli (**Fig. 4.2D**). However, it was unclear whether this average population signal was strong enough to overcome the inter-trial noise.

4.2.C: Analysis of network trajectories

The correlation of population responses under the same conditions, or noise correlation, can affect the amount of information encoded by a neural population (Averbeck et al., 2006). To examine how the two stimulation patterns altered noise correlations in network activity I measured the average pairwise correlation of neural activity across trials induced by the stimulation (**Fig. 4.3A**). Each pattern induced significant noise correlations in the network activity. In addition, there

was a significant interaction between stimulation pattern and time, due to the different structure of each pattern (n = 15; F^{Time} (47, 658) = 21.6, $p < 10^{-4}$; $F^{Pattern:Time}$ (47, 658) = 6.07, $p < 10^{-4}$).

The wide-field stimulation used in this experiment is likely to evoke activity in a broad set of neurons. To determine whether each pattern evoked differing amounts of overall network activity, I compared the average magnitude of network events (the L-2 norm of the vector of event magnitudes for each neuron) evoked by each pattern over time (**Fig. 4.3B**). Unsurprisingly, because each pattern delivers stimulation at different times, the network event magnitudes evoked by the patterns were distinct over time, i.e. there was a significant interaction between stimulation pattern and time (2-way ANOVA with repeated measures, $F^{Pattern:Time}$ (47, 658) = 8.9, $p < 10^{-4}$). Interestingly, however, event magnitudes were significantly different between patterns well after the final stimulus (post-hoc Tukey-Kramer test for time; each time point from 160-1650 ms was significant with $p < 0.05$).

To more thoroughly examine whether the two temporal patterns differentially altered each network's state, I next examined the trajectories of network activity evoked by each pattern. As a first step, principal component analysis (PCA) was used to generate a low dimensional representation of each trajectory. PCA calculates dimensions which maintain the statistical structure of a dataset and sorts them by the amount of variance they capture. Projecting network activity onto its first three principle components revealed that the average trajectories evoked by each pattern appeared to be distinct (**Fig. 4.3C**, example recording). To determine whether the stimulation patterns generated separate population responses, I measured the within-pattern and across-pattern Euclidean distance between the trajectories evoked by each presentation of the stimuli (**Fig. 4.3D**). This analysis revealed a significant interaction in the within-pattern and between-pattern trajectory distances over time ($F^{Pattern:Time}$ (26,364) = 10.3, $p < 10^{-4}$). In addition, the trajectories were more separated across patterns than they were within up to 350 ms (6

frames) after the final stimulus (post-hoc Tukey-Kramer test for time, patterns were significantly different at each time point from 350-1350 ms with significance level $p < 0.05$). Together, these analyses confirmed that the response to the final stimulus was modulated by the recent stimulus history, or state, of the network.

4.2.D: Single trial decoding of population activity

Humans and rodents can interpret sensory input after a single presentation. To examine whether cortical slices can perform similarly I trained multiclass linear support vector machines (SVM) to decode the two stimuli using the evoked network dynamics on a trial-by-trial basis. Each SVM was trained to classify population activity in each frame as either pattern A or B. The training set consisted of 90% of trials selected randomly, holding out 10% of trials for testing (Materials and Methods). This training/testing procedure was repeated ten times. After training, the SVMs were able to classify the held-out trials with an average accuracy significantly above chance across slices (**Fig. 4.4**). Specifically, classification accuracy changed significantly over time, with an average performance of 89% at the time of the final stimulus (ANOVA with repeated measures, F^{Time} (45, 630) = 35.8, $p < 0.0001$). In addition, post-hoc tests revealed that accuracy was signifigantly higher than baseline from 300-1300 ms after the start of stimulation, (post-hoc Tukey-Kramer test, $p < 0.05$ for all frames from 300-1300 ms after trial start). These results confirm that the network dynamics encode information about the stimulus patterns well after the last light pulse.

4.3: Discussion

State dependent computations have been proposed to underlie timing and sensory processing. In this work, I examine this issue using a reduced preparation in order to determine whether the cortex can perform these computations autonomously, or if input from other brain regions is required. I show that two balanced, purely temporal stimulation patterns induce distinct network

states. In addition, these trajectories can be used to decode the two stimulus patterns on a single trial basis. Though these findings demonstrate that the cortex can act as a reservoir computing device *in vitro* (Buonomano and Maass, 2009), it is clearly important to validate and extend these findings using *in vivo* studies of state-dependent processing.

4.3.A: Encoding temporal information

Previous work has examined state-dependent processing in the hippocampus (Buonomano et al., 1997; Hyde and Strowbridge, 2012) and in networks of dissociated cortical neurons (Dranias et al., 2013; Ju et al., 2015). I extend these findings by examining the ability of cortical slices to process purely temporal patterns, with important implications. First, input that varies in structure (i.e. input patterns that use multiple stimuli) over time can itself be a temporal code. Thus, experiments designed to study how the brain generates an internal representation of time must avoid injecting external temporal information. This work does so by using the same stimulation within and across each input pattern, meaning that information the cortex encodes about each pattern represents the cortex's tracking of time. Further, structurally dynamic stimuli directly alter network state over time, making it difficult to determine the contribution of intrinsic network dynamics on state-dependent encoding of information. Here, the differential encoding of the two input patterns relies entirely on changes in the "hidden" states of cortical networks, such as short-term plasticity. While real-world sensory input is spatially dynamic, it is important to understand the role of intrinsic activity in sensory processing.

4.3.B: Future directions

The work I present here examines temporal processing in naïve slices, demonstrating that the ability to encode temporal information is an intrinsic property of cortical neural networks. However, it is likely that experience plays a significant role in timing. Indeed, previous work using a similar preparation has shown that slices can be trained to encode intervals by repeatedly stimulating

them with paired electrical or optical pulses (Goel and Buonomano, 2016; Johnson et al., 2010). Future work should examine whether training slices can improve their ability to distinguish temporal input patterns.

In addition, neuromodulators may modulate state-dependent encoding of temporal information. Neuromodulators such as acetylcholine and dopamine are known to have a significant impact on temporal processing in humans and rodents (Coull et al., 2011). In fact, cholinergic signaling has been show *in vitro* and *in vivo* to play a role in encoding trained temporal intervals (Chubykin et al., 2013; Namboodiri et al., 2015). However, whether these findings will extend to the framework presented here is unclear.

Lastly, though this work demonstrates that cortical networks are capable of encoding temporal input patterns via state dependent computations, the specific network mechanisms are unclear. One potentially interesting mechanism is short term plasticity (STP). Theoretical studies have shown that STP can generate neural responses at specific intervals (Buonomano, 2000). In addition, dynamic modulations of the balance of excitation and inhibition are known to play a role in tuning the temporal profile of neural dynamics *in vitro* (Goel and Buonomano, 2016). Future work must be done to examine the role of STP in the state-dependent processing of temporal information.

4.4: Figures

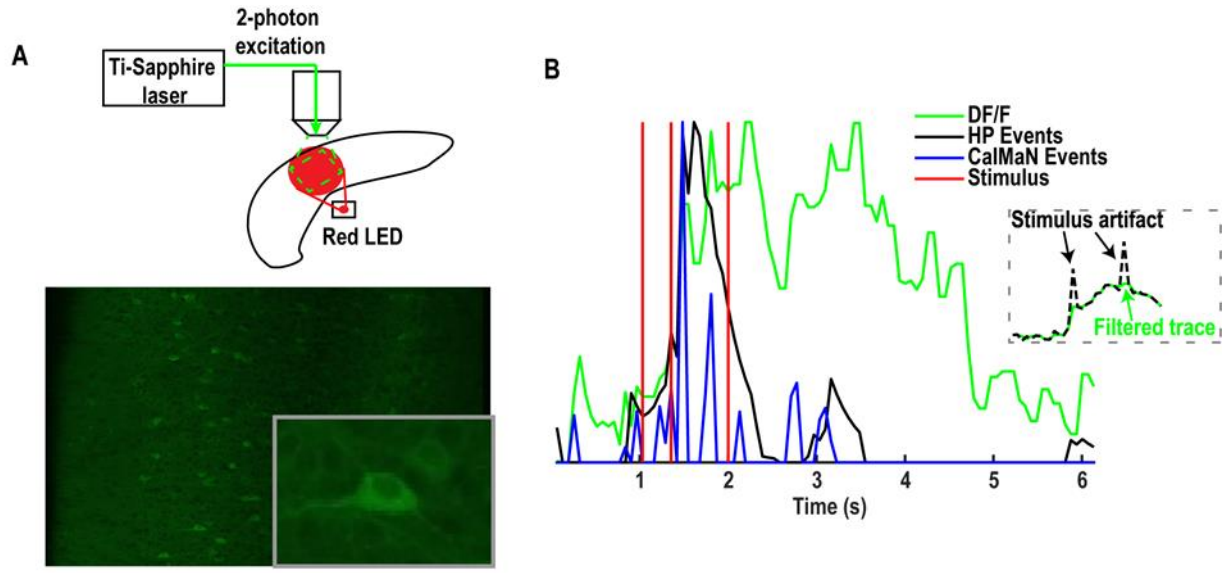


Figure 4.1. Recording setup and data processing.

A) Top: Calcium activity in organotypic slice cultures expressing GCaMP6f was imaged using a 2-photon microscope at 15.5 Hz (Materials and Methods). Stimulation was delivered using a 625 nm red LED positioned under the slice. Bottom: Example slice image. The imaging field captured the activity of 40-120 neurons per slice. The inset shows an expanded image of an example neuron. **B**) After imaging, the change in fluorescence over time was calculated for each slice. Activity in each neuron was discretized using two methods: constrained sparse nonnegative calcium deconvolution (CaImAn) and a rectified high pass filter (Materials and Methods). Inset: Filtering of the stimulus artifact. Each stimulus was 25 ms long (less than the duration of an imaging frame), allowing the stimulus artifact to be filtered out using a 3-frame median filter.

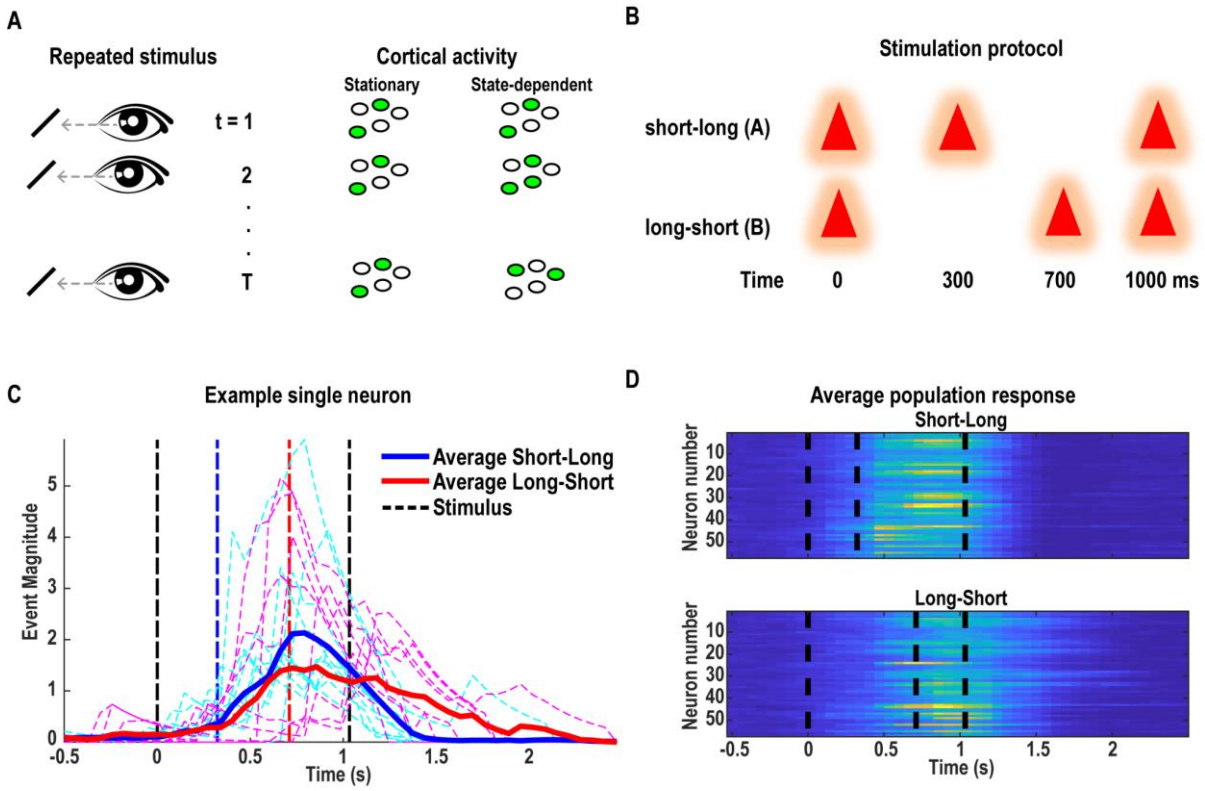


Figure 4.2. Experimental protocol and example activity.

A) Illustration of state-dependent computations. In this schematic a simple stimulus is presented repeatedly over time, generating a response in a population of cortical neurons. In the stationary case, each stimulus presentation results in the same population response. In the state-dependent case, each presentation cumulatively modifies the network state, generating a different population response at each presentation. **B**) Stimulation protocol. Slices were stimulated with 2 purely temporal patterns of three 25 ms red light pulses. The stimulation times for each pattern were: pattern A (short-long) – 0, 300, and 1000 ms and pattern B (long-short) – 0, 700, and 1000 ms. **C**) Example responses of a single neuron. Stimulation generated variable neural responses across trials (dashed lines). However, the average response to each pattern (solid lines) appeared qualitatively different. **D**) The average population response to each pattern in the same example slice shown in **C**. Again, the network activity was qualitatively different across patterns.

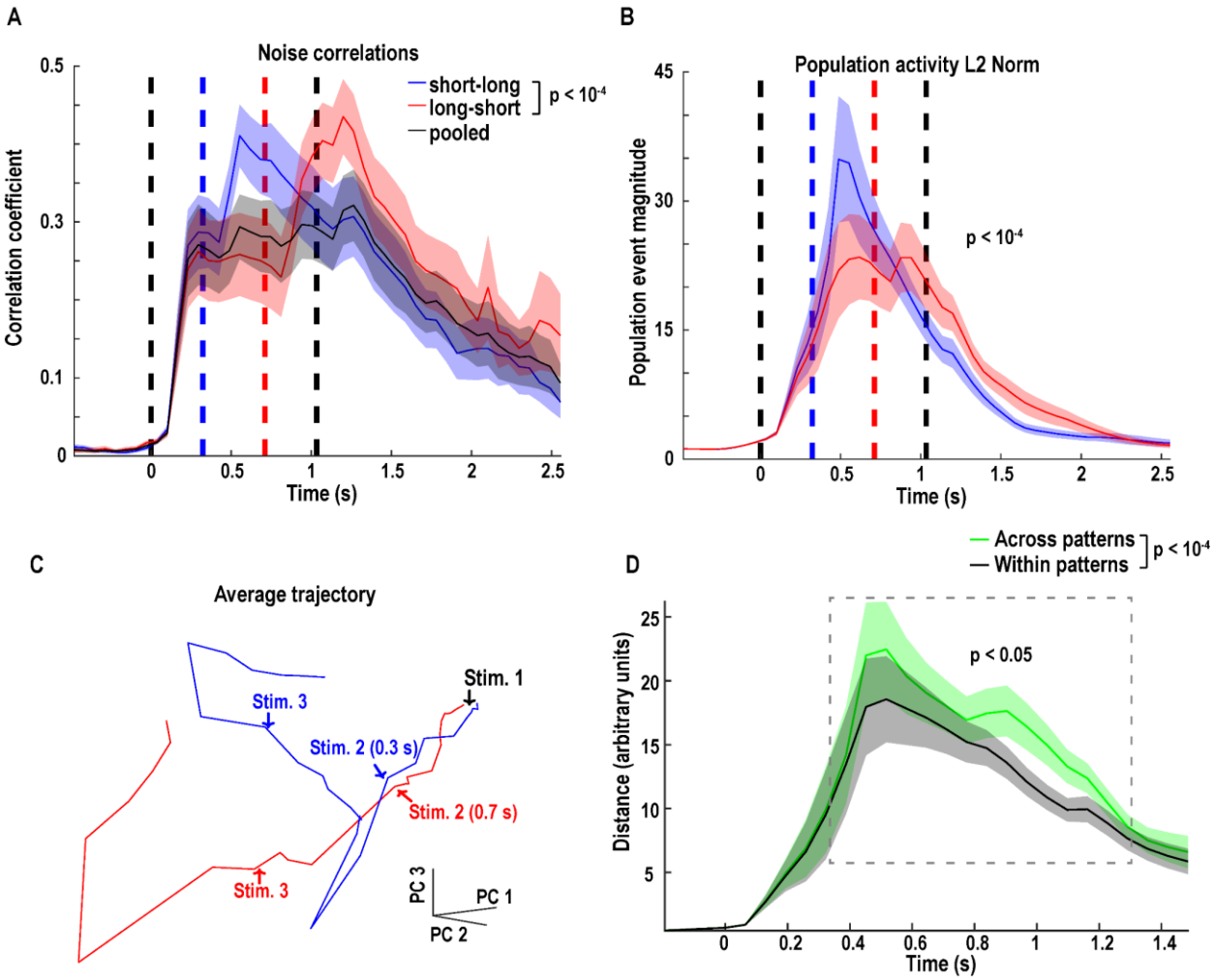


Figure 4.3. Temporal pattern stimulation evoked distinct population trajectories.

A) Noise correlations evoked by each stimulus pattern. Noise correlations were calculated by taking the pairwise Pearson's correlation coefficient for each frame across trials. Stimulation generated a significant change in the noise correlation over time, and the correlations generated by each pattern were significantly different ($p < 10^{-4}$). Shaded regions in this and later panels represent the standard error of the mean, and solid lines show the average across slices. **B**) Population event magnitude generated by each temporal pattern. Population event magnitude was calculated as the L-2 norm of the vector composed of each neuron's calcium events in each frame. The event magnitudes evoked by each pattern were significantly different over time ($p < 10^{-4}$). In addition, event magnitudes were significantly different up to from 150-1650 ms ($p < 0.05$). **C**) Average trajectories of neural activity evoked by each pattern for an example slice. Trajectories are shown plotted in the space defined by their first three principle components. **D**) The average pairwise Euclidean distance between the first 3 principle components of single trial network trajectories. The distance between trajectories evoked by the same pattern (within) and the distance between trajectories evoked by different patterns (across) showed a significant interaction with time ($p < 10^{-4}$). In

addition, the across pattern distance was significantly greater than the within pattern distance from 350-1350 ms, up to 6 frames after the final stimulus (boxed area, $p < 0.05$).

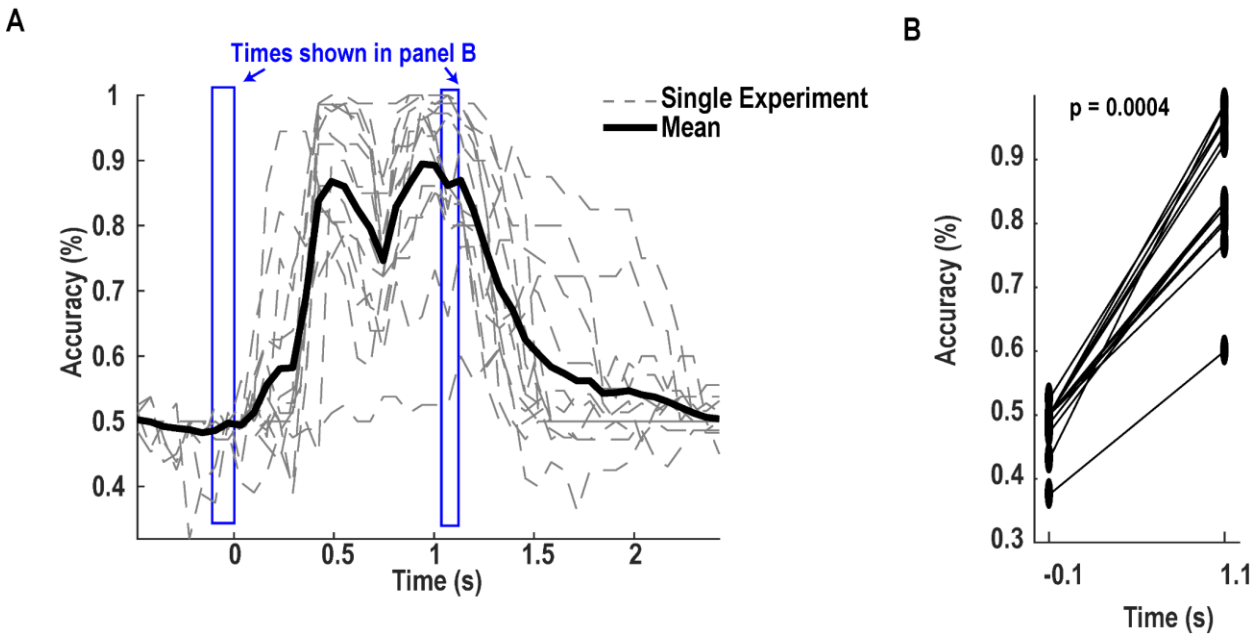


Figure 4.4. Single trial decoding of temporal patterns

A) Single trial classification accuracy of evoked network activity. A multiclass leaner SVM was trained to classify each frame of network activity as either pattern A or B (Materials and Methods). Accuracy was calculated as the percent of held-out trials correctly classified. Grey dashed lines show the classification results for each slice, and the black line shows the average of all slices. **B**) Classification accuracy was significantly higher than baseline after the final stimulus ($p = 0.0004$). Each line represents a single slice. Accuracy values were taken from the frame 100 ms before trial start and 100 ms after the final stimulus. These time points were chosen to emphasize that network activity contained information about the stimulus patterns even after the end of stimulation. Classification accuracy was significantly above baseline from 300-1300 ms ($p < 0.05$).

4.5: Materials and Methods

4.5.A: Organotypic slice preparation

All animal procedures followed the National Institutes of Health (NIH) guidelines and were approved by the UCLA Institutional Animal Care and Use Committee (IACUC). Postnatal day 6-7 FVB mice were anesthetized with isoflurane and decapitated. The brain was removed and placed in chilled cutting media. 400 µm thick coronal slices containing primary somatosensory or auditory cortex were cut using a vibratome. Culture media was changed 1 and 24 hrs after cutting and every 2-3 days thereafter. Cutting media consisted of EMEM (MediaTech cat. #15-010) (final concentration in mM): MgCl2, 3; glucose, 10; Hepes, 25; and Trisbase, 10. Culture media consisted of EMEM plus (final concentration in mM): glutamine, 1; CaCl2, 2.6; MgSO4, 2.6; glucose, 30; Hepes, 30; ascorbic acid, 0.5; 20% horse serum, 10 units/L penicillin, and 10 µg/L streptomycin. Slices were incubated in 5% CO2 at 35°C.

4.5.B: Viral transfection

At 7 days *in vitro* (DIV), slices were transduced with adeno-associated viruses (AAV) encoding GCaMP6f under the control of the synapsin promoter (Penn Vector Core, AAV9.Syn.GCaMP6f.WPRE.SV40) and Chrimson under the control of the Synapsin promoter (Penn Vector Core, AAV9.Syn.ChrimsonR-tdTomato.WPRE.bGH). The pan-neuronal synapsin promoter was selected to allow stimulation and imaging of calcium activity in a subset of all neurons (Chen et al., 2013; Klapoetke et al., 2014). Viral injection was performed using glass micropipettes containing 0.8 µL of each virus. Slices were removed from the incubator and the pipette was carefully lowered onto the coronal surface of the slice using a micromanipulator. The viral solution was ejected by applying positive pressure gently by hand and then slices were immediately returned to the incubator. Slices were kept in the incubator for 14-21 days post transduction to allow for sufficient expression of both proteins.

4.5.C: Simultaneous 2-photon calcium imaging and optogenetic stimulation

After allowing GCaMP6f and Chrimson to reach sufficient expression levels (21-28 DIV), slices were removed from the incubator and placed in the recording chamber. The chamber was perfused with oxygenated artificial cerebrospinal fluid (aCSF) formulated to match the culture media with final concentrations (mM): NaCl, 125; KCl, 5.1; MgSO4, 2.6; NaHCO3, 26.1; NaH2PO4, 1; glucose, 25; and CaCl2, 2.6. Temperature was maintained at 32-33°C, and the aCSF perfusion rate set to 4-8 ml/min.

Calcium imaging was performed using a Neurolabware 2-Photon microscope (Neurolabware, Los Angeles, CA; http://neurolabware.com/) controlled by Scanbox acquisition software (Scanbox, Los Angeles, CA). The light source was a Coherent Chameleon Ultra II laser (Coherent Inc, Santa Clara, CA) running at 920 nm. Laser power was set individually for each slice to optimize recording conditions under variable GCaMP6f expression levels. The objective was a 16x water immersion lens (Nikon, 0.8 NA, 3 mm working distance) and the imaging rate was 15.5 Hz (512 lines with an 8 kHz resonant mirror). GCaMP6f photon emission was measured using a PMT and the gain was adjusted to prevent saturation during imaging.

Optogenetic stimulation was delivered by activating a red LED positioned under the slice (Thorlabs, 625 nm) at 150 mA output power. The light was collected and focused onto the surface of the slice using a collimating lens near the LED, followed by a condenser lens near the slice. The size of the stimulated region was controlled using an iris diaphragm mounted just below the condenser lens. To prevent PMT saturation during stimulation, excess green light was filtered out using 610 nm longpass filters (Thorlabs). The LED was triggered using 25 ms TTL pulses controlled by a National Instruments data acquisition board delivered to both the LED controller and imaging hardware to record the stimulation frames.

4.5.D: Data processing

The recorded images were preprocessed as described by (Garcia-Junco-Clemente et al., 2017). In brief, frames were aligned to correct for slice motion using a two-step process. First, pairs of images were recursively rigidly aligned using the peak of their cross-correlation to correct for slow drift in the imaging plane. Next, the images were nonrigidly aligned to a reference mean image to correct for fast in-plane movement caused by vibrations in the recording camber. Because there was little movement during slice recording, this alignment did not result in significant changes in the images, but it was nonetheless performed to ensure maximal alignment.

Following motion stabilization, individual cells were segmented using the Scanbox graphical user interface segmentation tool. This tool computes an image from the average correlation of a pixel and its eight neighbors across time to identify candidate cell bodies. This image was used to manually select candidate seed pixels within an image patch, and then automatically flood-fill a region of interest (ROI) according to the thresholded correlation of the selected pixel and every other pixel in the image.

Following segmentation, the fluorescence within each ROI was extracted for every frame of the recorded movie. Stimulation artifacts were removed using a 3rd-order median filter, while long term linear trends in the signal were removed using a 0.5 second sliding median filter. The change in fluorescence over time ($\Delta F/F$, dFF) was calculated by normalizing the change in fluorescence at each frame by the minimum fluorescence value of the ten preceding seconds. Calcium events were then calculated using both the constrained sparse nonnegative calcium deconvolution (CaImAn) algorithm (Pnevmatikakis et al., 2016) and a rectified high-pass 2 Hz filter.

4.5.E: Temporal pattern stimulation

Slices were stimulated using the red LED described above. At the start of each trial, the LED was activated for three 25 ms periods in one of two temporal patterns: A) 0, 300, and 1000 ms; and B) 0, 700, and 1000 ms. The patterns were presented alternately, with 60 repetitions per pattern and a 30 s inter-trial-interval, resulting in 120 total trials and 65 minutes total time including a 5-minute baseline period. Calcium activity was recorded continuously for the entire session. To minimize the effects of spontaneous network activity on future analyses, trials with network events in the one second preceding a trial were excluded. Network events were defined as a period of more than 64 ms (one frame) in which average network activity exceeded 2% of the maximum. Trials without network events in the two seconds following the start of a trial or with spontaneous events one second preceding a trial were excluded.

4.5.F: Decoding stimulus patterns from network activity

The stimulus pattern presented in each trial was decoded by requiring a classifier to label a vector of population event rates as either stimulus A or B. Population vectors were formed from the total calcium event magnitude of each segmented neuron in each frame. The classification task was performed by a multiclass support vector machine (SVM) with a linear kernel, using the LIBSVM library (Chang and Lin, 2011). In this approach, the SMV takes a population vector as input and generates an output for each of the two classes. The input vector is classified as A or B by determining which output had the highest activation. A separate classifier was trained for each time bin, using 90% of accepted trials for each pattern, and the remaining trials were reserved as a test-set. This training and testing procedure was repeated ten times. Classification accuracy was calculated as the average over all training and testing sets.

4.6: References Cited

Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nat. Rev. Neurosci. *7*, 358–366.

Buonomano, D.V. (2000). Decoding temporal information: A model based on short-term synaptic plasticity. J. Neurosci. Off. J. Soc. Neurosci. *20*, 1129–1141.

Buonomano, D.V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. Nat. Rev. Neurosci. *10*, 113–125.

Buonomano, D.V., and Merzenich, M.M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. Science *267*, 1028–1030.

Buonomano, D.V., Hickmott, P.W., and Merzenich, M.M. (1997). Context-sensitive synaptic plasticity and temporal-to-spatial transformations in hippocampal slices. Proc. Natl. Acad. Sci. *94*, 10403–10408.

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. ACM Trans Intell Syst Technol *2*, 27:1–27:27.

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

Chubykin, A.A., Roach, E.B., Bear, M.F., and Shuler, M.G.H. (2013). A cholinergic mechanism for reward timing within primary visual cortex. Neuron *77*, 723–735.

Compte, A., Brunel, N., Goldman-Rakic, P.S., and Wang, X.-J. (2000). Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. Cereb. Cortex *10*, 910–923.

Coull, J.T., Cheng, R.-K., and Meck, W.H. (2011). Neuroanatomical and Neurochemical Substrates of Timing. Neuropsychopharmacology *36*, 3–25.

Dranias, M.R., Ju, H., Rajaram, E., and VanDongen, A.M.J. (2013). Short-Term Memory in Networks of Dissociated Cortical Neurons. J. Neurosci. *33*, 1940–1953.

Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. *61*, 331–349.

Fuster, J.M., and Jervey, J.P. (1981). Inferotemporal Neurons Distinguish and Retain Behaviorally Relevant Features of Visual Stimuli. Science *212*, 952–955.

Garcia-Junco-Clemente, P., Ikrar, T., Tring, E., Xu, X., Ringach, D.L., and Trachtenberg, J.T. (2017). An inhibitory pull–push circuit in frontal cortex. Nat. Neurosci. *20*, 389–392.

Goel, A., and Buonomano, D.V. (2016). Temporal Interval Learning in Cortical Cultures Is Encoded in Intrinsic Network Dynamics. Neuron *91*, 320–327.

Hyde, R.A., and Strowbridge, B.W. (2012). Mnemonic representations of transient stimuli and temporal sequences in the rodent hippocampus in vitro. Nat. Neurosci. *15*, 1430–1438.

Johnson, H.A., Goel, A., and Buonomano, D.V. (2010). Neural dynamics of in vitro cortical networks reflects experienced temporal patterns. Nat. Neurosci. *13*, 917–919.

Ju, H., Dranias, M.R., Banumurthy, G., and VanDongen, A.M.J. (2015). Spatiotemporal Memory Is an Intrinsic Property of Networks of Dissociated Cortical Neurons. J. Neurosci. *35*, 4040–4051.

Klapoetke, N.C., Murata, Y., Kim, S.S., Pulver, S.R., Birdsey-Benson, A., Cho, Y.K., Morimoto, T.K., Chuong, A.S., Carpenter, E.J., Tian, Z., et al. (2014). Independent optical excitation of distinct neural populations. Nat. Methods *11*, 338–346.

Laje, R., and Buonomano, D.V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. Nat. Neurosci. *16*, 925–933.

Miller, P., Brody, C.D., Romo, R., and Wang, X.-J. (2003). A Recurrent Network Model of Somatosensory Parametric Working Memory in the Prefrontal Cortex. Cereb. Cortex *13*, 1208–1218.

Namboodiri, V.M.K., Huertas, M.A., Monk, K.J., Shouval, H.Z., and Hussain Shuler, M.G. (2015). Visually Cued Action Timing in the Primary Visual Cortex. Neuron *86*, 319–330.

Nikolić, D., Häusler, S., Singer, W., and Maass, W. (2009). Distributed Fading Memory for Stimulus Properties in the Primary Visual Cortex. PLOS Biol. *7*, e1000260.

Pnevmatikakis, E.A., Soudry, D., Gao, Y., Machado, T.A., Merel, J., Pfau, D., Reardon, T., Mu, Y., Lacefield, C., Yang, W., et al. (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron *89*, 285–299.

Rajan, K., Harvey, C.D., and Tank, D.W. (2016). Recurrent Network Models of Sequence Generation and Memory. Neuron *90*, 128–142.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. Trends Neurosci. *24*, 455–463.

Wang, J., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. Nat. Neurosci. *21*, 102–110.

Chapter 5: Conclusions and future directions

The findings presented in this dissertation support state dependent theories of temporal processing in the cortex. Three major findings are discussed. First, Chapter 2 examines the mechanisms of embedding functionally feedforward neural activity in recurrent networks. Next, Chapter 3 develops a model of temporally invariant motor behavior and describes a novel feature of timing behavior: the Weber-speed effect. Finally, Chapter 4 studies state dependent temporal processing *in vitro* and reveals that cortical neural networks can embed time information on a single trial basis. In this chapter I attempt to integrate these findings while discussing potential shortcomings and future directions for research.

5.1: The role of theory in neuroscience

Studying the dynamics of neural networks is a difficult problem. Some of this difficulty arises from technological limitations that have historically restricted the ability of researchers to record the activity of large numbers of neurons simultaneously. Some of the most seminal experiments in neuroscience examined the activity of single neurons, and laid the groundwork for decades of future work (Hodgkin A. L. and Huxley A. F., 1952a; Hubel D. H. and Wiesel T. N., 1959). The techniques used in these early days of neurophysiology are commonly used to this day, and continue to provide valuable information about neural systems. As technology has developed, neuroscientists are now able to record from hundreds of neurons while animals perform behavioral tasks (Schwarz et al., 2014). However, as the scale of recorded data has increased, so has the difficulty in interpreting that data. To aid in interpreting large amounts of neural data, neuroscientists have long used mathematical models to provide an intuition for how neural systems work.

Even in the earliest days of neurophysiology, theoretical models have been used to describe neural activity and generate testable predictions for future experiments (Ben-Yishai et al., 1995;

Hodgkin A. L. and Huxley A. F., 1952b). As data has become more complex, so have the models used to describe that data. At the scale of neural populations, theoreticians now focus on simulating neural networks that can perform complex tasks. These simulations have been useful in describing the computational mechanisms of the brain, often by mimicking recorded network activity (Carnevale et al., 2015; Mante et al., 2013). Thus, the interaction of theoretical and experimental neuroscience has provided valuable insights into the workings of the brain, a pattern that will likely increase in significance (Churchland and Abbott, 2016).

In the previous chapters, I employed a combination of experimental and theoretical techniques to analyze the activity of neural networks. Chapter 2 exclusively uses mathematical models to examine and predict the mechanisms underlying previously observed experimental data. Chapter 3 goes a step further by combining theoretical and experimental work to generate and test predictions about the role of network dynamics in temporal scaling. Lastly, Chapter 4 tests some of the predictions of previous theoretical work on how the brain processes temporal information. Currently, it is relatively unusual for computational and experimental work to be combined in this way. However, this is likely to become increasingly common in future work.

5.2: Limitations and future directions

Theoretical neuroscience is currently limited by the lack of biologically inspired unsupervised learning rules. Due to this limitation, the research in Chapters 2 and 3 relies on supervised learning to train simulated neural networks perform a task. Recent research on reward-based learning in neural networks is a significant advance in overcoming this limitation (Song et al., 2017). However, this work still relies on backpropagation to modify synaptic connections, which does not yet have a clear biological correlate (Scellier and Bengio, 2016). As experiments uncover new learning rules in the brain, theoreticians should incorporate those these findings into future models of neural networks (Bittner et al., 2017).

Another important limitation of the theoretical studies in this dissertation is their reliance on simplified neural models. The networks used in both Chapters 2 and 3 are composed of analog, or firing rate, units which can intuitively be described as representing the firing rate of a large population of neurons (DePasquale et al., 2018). Unlike biological neurons that generate output using discrete spikes, firing rate units generate output using saturating nonlinear functions. Networks of firing rate units are useful because they can spontaneously generate complex dynamics that resemble experimentally recorded activity. However, generating similar activity patterns in networks of spiking units is still challenging. Recent advances have begun to address this challenge (Nicola and Clopath, 2017; Ostojic, 2014). However, these studies still relied supervised learning algorithms or hand-tuned network architectures, leaving open the question of how networks in the brain generate such activity.

An important limitation of the research in Chapter 4 is the use of *in vitro* cortical networks. As mentioned in the Discussion section of that chapter, using a reduced preparation allowed me to examine whether cortical networks can perform state dependent computations autonomously. However, it is clear that this work must be verified *in vivo*. With current advances in recording techniques, future work will be able to examine state dependent computations using minimally invasive procedures to record and manipulate the activity of many neurons simultaneously *in vivo* (Garcia-Junco-Clemente et al., 2017; Hochbaum et al., 2014).

5.3: Conclusions

In the current work, I examine the role of state dependent neural network dynamics in temporal processing. The theory that neural networks represent time using emergent dynamics is increasingly accepted by the timing field. Though the mechanisms the brain uses to process time are still unclear, this work advances our understanding of temporal processing and neural network dynamics and provides inspiration for future work.

As a closing remark, I wish to emphasize the value of embracing complexity in studying neuroscience. The brain is often referred to as a “black box” which receives sensory input, performs some unknown computations on that input, and produces a behavioral output. How the brain performs those computations is still largely unclear, and neuroscientists have necessarily relied on distilling neural activity into a simpler form to make it more readily interpretable. Though this provides us with a stronger intuition of what the brain does, it carries the risk of oversimplification. As the field advances, new methods must be developed which provide an intuitive representation of neural activity while not imposing overly simplified interpretations on data. To accomplish this task, basic research on the functions of neural circuits will be essential.

One especially exciting area of future neuroscience research is brain machine interface (BMI). The fundamental goal of BMI research is to develop technology to translate brain activity (colloquially, *thoughts*) into computer commands that control, for example, a computer cursor or robotic device. Developing functional BMI technology has profound clinical applications, such as allowing patients who cannot move (e.g. those with ALS or locked-in syndrome) to volitionally interact with their environments by generating speech or moving robotic limbs via a computer connected to their brain. In addition, BMI technology will allow researchers to study basic neuroscience questions that were previously impossible, such as the flexibility of the brain’s ability to regulate its own activity. It is well established that if a young patient loses a portion of their cortex through a disease (e.g. a tumor) or accident, the remaining intact cortex is able to compensate for the loss by performing the functions normally executed by the missing tissue. In this way, a patient whose auditory cortex, for example, is damaged at a sufficiently young age (usually before the critical period closes) will often recover close to normal hearing because nearby cortical regions develop to become functionally similar to the missing auditory cortex. However, recent BMI research has shown that the brain can flexibly control neural activity in the absence of severe brain damage. In one study, researchers showed that rats are able to operantly

control a computer-generated auditory tone by modulating the activity of randomly selected neurons in the somatosensory cortex, in order to receive a reward (Clancy et al., 2014). This study demonstrated that the brain can quickly learn to control neurons which don't normally process auditory information or generate motor output to perform an auditory input/output processing task, challenging canonical assumptions about the functional organization of the cortex.

Inspired by such BMI research and similar findings in other fields, including my own, I anticipate that as BMI and other neuroscientific technologies advance, our current understanding of the relationship between brain structure and function will be revised significantly. In the most extreme case, it is possible that neuroscientists will no longer be able to demarcate the cortex into static functional areas like visual, motor, and auditory regions because we will discover that all parts of the brain play some role nearly every task we perform. If this prediction proves true, we will have to examine the functions of the entire cortex in any given task, regardless of the specific stimuli or motor outputs used in an experiment. Such a revision in our understanding of brain function would be monumental, and require rejecting many strongly held simplifying assumptions about neural function. Though those assumptions have enabled significant progress in neuroscience research, scientific advances require that assumptions be challenged and revised as new data becomes available. In the next stage of my career, I hope to contribute to these advances by developing BMI technology, using it both to aid the disabled and to increase our understanding of basic neural function. It is an exciting time to be a neuroscientist, and I look forward to the many new discoveries the field will make in the near future.

5.4: References cited

Ben-Yishai, R., Bar-Or, R.L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. *92*, 3844–3848.

Bittner, K.C., Milstein, A.D., Grienberger, C., Romani, S., and Magee, J.C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. Science *357*, 1033–1036.

Carnevale, F., de Lafuente, V., Romo, R., Barak, O., and Parga, N. (2015). Dynamic Control of Response Criterion in Premotor Cortex during Perceptual Detection under Temporal Uncertainty. Neuron *86*, 1067–1077.

Churchland, A.K., and Abbott, L.F. (2016). Conceptual and technical advances define a key moment for theoretical neuroscience.

Clancy, K.B., Koralek, A.C., Costa, R.M., Feldman, D.E., and Carmena, J.M. (2014). Volitional modulation of optically recorded calcium signals during neuroprosthetic learning. Nat. Neurosci. *17*, 807–809.

DePasquale, B., Cueva, C.J., Rajan, K., Escola, G.S., and Abbott, L.F. (2018). full-FORCE: A target-based method for training recurrent networks. PLOS ONE *13*, e0191527.

Garcia-Junco-Clemente, P., Ikrar, T., Tring, E., Xu, X., Ringach, D.L., and Trachtenberg, J.T. (2017). An inhibitory pull–push circuit in frontal cortex. Nat. Neurosci. *20*, 389–392.

Hochbaum, D.R., Zhao, Y., Farhi, S.L., Klapoetke, N., Werley, C.A., Kapoor, V., Zou, P., Kralj, J.M., Maclaurin, D., Smedemark-Margulies, N., et al. (2014). All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. Nat. Methods *11*, 825–833.

Hodgkin A. L., and Huxley A. F. (1952a). The dual effect of membrane potential on sodium conductance in the giant axon of Loligo. J. Physiol. *116*, 497–506.

Hodgkin A. L., and Huxley A. F. (1952b). A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. *117*, 500–544.

Hubel D. H., and Wiesel T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. *148*, 574–591.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Nicola, W., and Clopath, C. (2017). Supervised learning in spiking neural networks with FORCE training. Nat. Commun. *8*, 2208.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. Nat. Neurosci. *17*, 594–600.

Scellier, B., and Bengio, Y. (2016). Equilibrium Propagation: Bridging the Gap Between Energy-Based Models and Backpropagation. ArXiv160205179 Cs.

Schwarz, D.A., Lebedev, M.A., Hanson, T.L., Dimitrov, D.F., Lehew, G., Meloy, J., Rajangam, S., Subramanian, V., Ifft, P.J., Li, Z., et al. (2014). Chronic, Wireless Recordings of Large Scale Brain Activity in Freely Moving Rhesus Monkeys. Nat. Methods *11*, 670–676.

Song, H.F., Yang, G.R., and Wang, X.-J. (2017). Reward-based training of recurrent neural networks for cognitive and value-based tasks. ELife *6*, e21492.